National Lung cancer detection using region properties of CT scan images

Term Project
Of
Digital Image Processing and Pattern Recognition

(EC-714)

MASTER OF TECHNOLOGY
in
ELECTRONICS AND COMMUNICATION ENGINEERING

(Communication System & Networks)

Submitted by

SURAJ KUMAR **19M428**

Submitted to

Dr. ABHIJIT BHATTACHARYYA

Department of Electronics & Communication Engineering

NIT Hamirpur (H.P.)

ABSTRACT

CANCER is one of the most serious health problems in the world field. 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. The earlier the detection is, the higher the chances of successful treatment are. An estimated 85% of lung Cancer cases in males and 75% in females are caused by cigarette smoking.

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 methods 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 binarization thresholding ,erosion morphological process and region extraction have been discussed in detail.

Keywords: binarization thresholding ,erosion morphological process ,region extraction

Contents

1. 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. 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. 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.

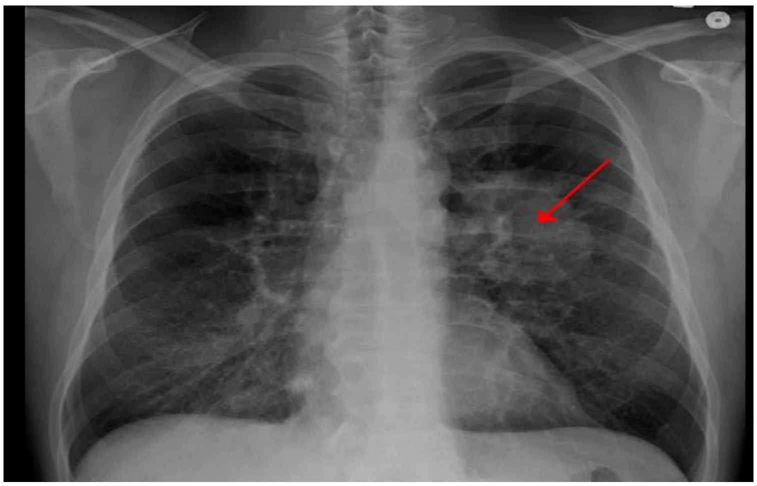


Fig1 A tumor in the lung (marked by arrow)

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. first stage starts with taking a collection of CT images (normal and abnormal) from the available Database from IMBA Home (VIA-ELCAP Public Access). The second stage applies several techniques of image enhancement, to get best level of quality and clearness. The third stage applies image segmentation algorithms which play an effective rule in image processing stages, and the fourth stage obtains the general features from enhanced segmented image which gives indicators of normality or abnormality of images. At present, only 15% of all diagnosed lung cancers are detected at an early stage, which causes a five-year survival rate of only 16%. The purpose of screening is to detect cancer at an early stage when curative treatment options are better.

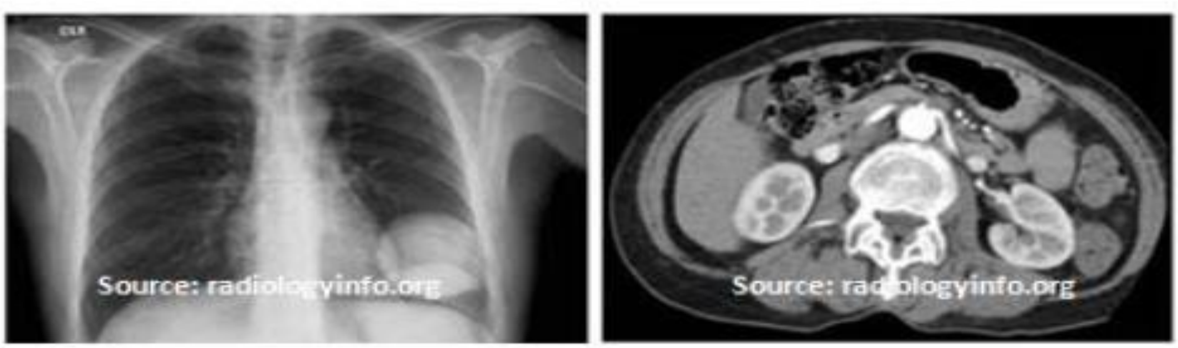
1.1 Motivation

The process of early detection plays an important role to prevent cancer cells from spreading. Although computed tomography (CT) scan imaging is one of the best imaging techniques for medical uses, it is hard to interpret and identify just from CT scan images. Hence, image processing is used to enhance the detection process. For this project, we plan to use our knowledge we learn from the FIT3081 class and the research papers we read up to come out an effective program that could help to contribute in this field. Our system can be adapted and used as a tool to benefit the doctors in the following ways:

- Provides a rough idea whether the CT scan image likely belong to normal people or lung cancer patient
- Provides an overview of the characteristics features of the lung nodules to judge the possible existence of tumour in the lungs.

There are several diagnostic imaging techniques such as MRI, PET, Chest X-ray and CT scan. MRI scans are less popular due to low availability and costly (Aggarwal, Furqan, & Kalra, 2015). On the other hand, CT scans are more sensitive and provides a higher contrast than chest X-ray. A low-dose computed tomography (LDCT) scan can detect early stage diseases 6-10 times more frequent than X-ray (Hollings & Shaw, 2002). However, there are some limitations to LDCT scans. Peikert et. al. (2018) has raised that high false positive rates from CT leads to unnecessary invasive procedures and patient anxiety. Hollings and Shaw (2012) highlighted the difficulty in differentiating benign and malignant nodules.

According to Dolejsi (2007), image processing and visualization methods would help to reduce the workload of radiologist sifting through the huge amount of data per examination, which complicates the interpretation due to the tedious process. This implies that image processing techniques help reduce false-negative rates that are caused by the radiologist's overlook. Another motivation would be to increase the sensitivity of detection rate, especially for smaller nodules of 1-7mm in diameter that are only detected 63% of the time (Dolejsi, 2007). Our methodology would point out the smaller nodules often left undetected by radiologist, hence contributing to our cause of early detection.



Traditional X-ray   CT Scans showing internal body

Fig2 X-ray vs CT scan

2. LITERATURE SURVEY

Ginneken [1] has classify 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 [2-3], 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. On the other hand, there is another approach used in lung regions extraction process based on pixel classifications, where each pixel in the CT image is classified into an anatomical class (usually lung or background, but in some cases more classes such as heart, mediastinum, and diaphragm). Classifiers are various types of neural networks, or markov random field modeling, trained with a variety of local features including intensity, location, and texture measures [1].

CADs can be divided into two groups [4]: density-based and model-based approaches. Considering the fact that lung nodules have relatively higher densities than those of lung parenchyma, density-based detection methods employ techniques such as multiple thresholding, region-growing, locally adaptive thresholding in combination with region growing, opening and closing, using the histogram, the top 20% gray values considered as initial cancerous candidate regions, using the histogram the normal tissues are removed, then elliptical-shaped regions, which is in general represent abnormalities, are detected, and fuzzy clustering used to identify nodule candidates in the lungs. False-positive results can then be reduced from the detected nodule candidates by employing a priori knowledge of small lung nodules. For the model-based detection approaches, the relatively compact shape of a small lung nodule is taken into account while establishing the models to identify nodules in the lungs. Techniques such as Morphological filter and the anatomy based generic model have been proposed to identify sphere shaped small nodules in the lung. Nodule candidates are detected using template matching or a modified Hough transform in which edge pixels vote for circles that could cause these edges.

After getting the segmentation results, different features should be extracted to be used in the diagnosis phase where sets of rules are formulated to distinguish between true and false cancerous candidates.

Different features were extracted in different papers depending on the methods used by the authors in the diagnosis phase. In some approaches uniformity, connectivity, and position features were extracted [5]. In [3] the features such as size, circularity, and mean brightness of region of interests (ROIs) were extracted. Area, thickness, circularity, intensity, variance, localization, and distance from the lung wall are the extracted features in [2]. The underlying idea of developing a CAD system is not to delegate the diagnosis to a machine, but rather that a machine algorithm acts as a support to the radiologist and points out locations of suspicious objects, so that the overall sensitivity is raised.

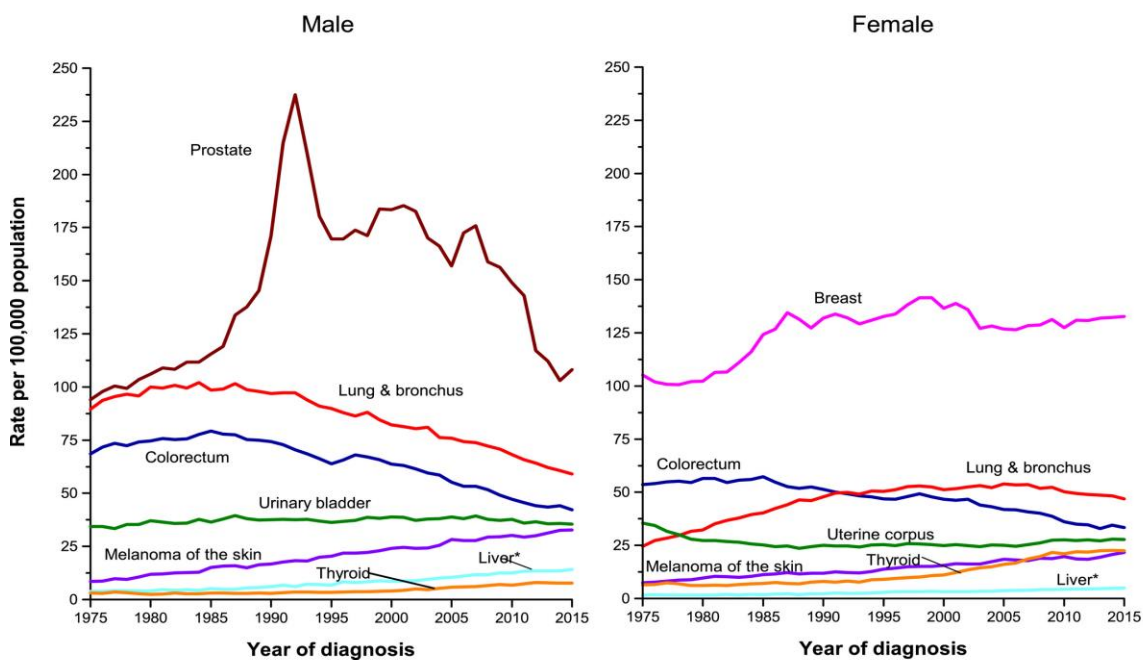


Fig3. lung cancer graph for Male vs Female

CAD systems meet four main objectives, which are improving the quality and accuracy of diagnosis, increasing therapy success by early detection of cancer, avoiding unnecessary biopsies and reducing radiologist interpretation time [6,7].While many image segmentation algorithms exist, when objects of the same predefined class are near 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.

3. RESEARCH METHODOLOGY

3.1 Overview

General procedure can be seen in the figure 1.

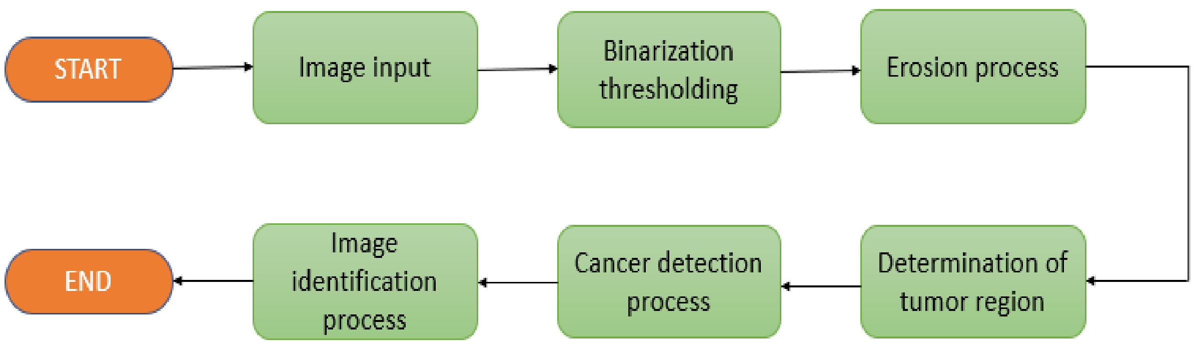


Fig 4. Flowchart of process

The first process begins by giving input to the CT scan image that will be identified then resized so that the size is smaller so that the next process will run faster. After that do pre-processing image by performing image processes Binarization thresholding which aims to simplify the process of morphology that will be done next. Then do the erosion morphology process by using the structuring element " diamond " which aims to separate the cancer objects with lung objects. Then do the process of determining the region properties in the image that has been eroded which will be determined by the region properties that will be removed if there are image properties that exceed the predetermined threshold.
After obtaining a tumor image, the identification process of the lungs can be carried out, if there is residual image after the process of removal then the lungs are considered abnormal while if not it will be considered as normal.

3.2 Data Sources

The sample used in this study was taken randomly from the Google Images site by carrying out the process of searching for a CT scan of the lung in detail with an abnormal lung image of 14 and a normal lung of 4, bringing the total image used in the experiment this amounts to 18 images.

4. DESCRIPTION

4.1 Lung Cancer

Lung cancer or lung carcinoma is a malignant lung tumor that is characterized by cell growth that is not known in the lung tissue . The highest rates are in North America, Europe and East Asia, with more than one third of new cases in China. Levels in Africa and South Africa are much lower . Worldwide in 2012, lung cancer occurred in 1.8 million people and caused 1.6 million deaths . This makes it the most common cause of cancer-related death in men and second. most common in women after breast cancer . These cases are often caused by a combination of factors and exposure to random gas, asbestos, passive smoking, or other forms of air pollution. The most common clinical manifestations are cough (including coughing up blood), weight loss, shortness of breath and chest pain . The diagnosis is mainly by chest radiography and *computed tomography* (CT) scan. The diagnosis is confirmed by biopsy by bronchoscopy or CT guidance . Common treatments include surgery, chemotherapy, and radiography.

A chest radiograph is one of the first steps in an investigation if someone reports symptoms that might indicate lung cancer. This may reveal a clear mass, mediastinal dilation (suggestive of spread to the lymph nodes there), atelectasis (collapse), consolidation (pneumonia) or pleural effusion . CT imaging is usually used to provide more information about the type and extent of the disease. CT-guided bronchoscopy or biopsy is often used to sample tumors for histopathology.

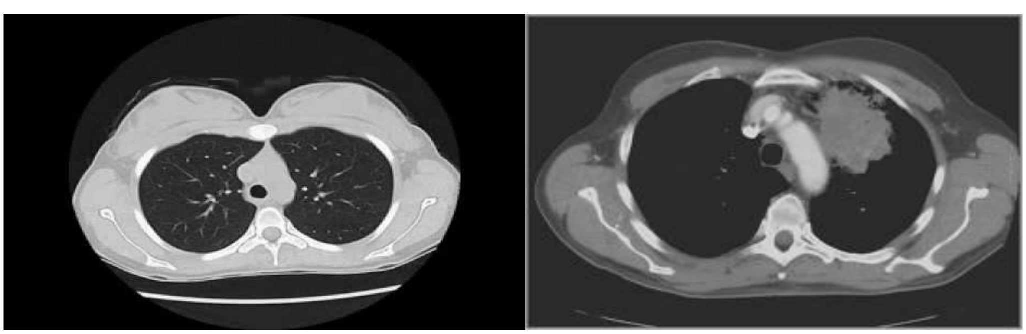


Fig 5. Image ... Normal and abnormal image

Lung cancer often appears as a solitary pulmonary nodule on chest radiography. Many other diseases can also give this appearance, including metastatic cancer, hamartomas, and granuloma infections such as tuberculosis, histoplasmosis, and coccidioidomycosis . Lung cancer can also be an incidental finding as a solitary pulmonary nodule on chest radiography or CT scan done. for unrelated reasons .

4.2 Data Acquisition

Data acquisition is the process of converting data from analog to RGB images with the help of a digital camera and the image will be saved in .jpg format which will then be processed to the *preprocessing* stage.
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.

4.3 Binary

Image binarization is the process of separating pixel values into two groups, white as background pixels and black as foreground pixels or objects. Thresholding plays a major role in image binarization. The thresholding method is the most commonly used method for binarizing images using global thresholds. Image binarization process is an important and effective for image processing and pattern recognition. The thresholding method involves iteration through all possible threshold values and calculating the spread size for pixel levels that fall in the foreground or background. The goal is to find the optimal threshold value where the number of foreground and background spreads to a minimum. This is important in image processing for selecting a threshold gray sufficient to extract objects from backgrounds.

Covert into Binary image. Let f(x, y) is an input image. T is the threshold value and g(x, y) is the output image of thresholding process then the mathematical equation of this conversion is: $g(x, y) = I$, if $f(x, y) \geq 2$ T otherwise 0

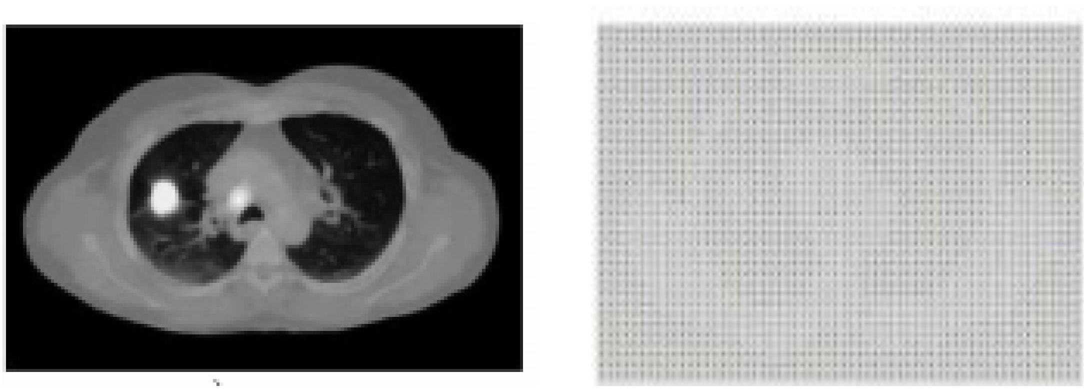


Fig6. (a) gray scale image (b) binarized image

4.3 Binary Thresholding

Thresholding is the simplest segmentation method where pixels are partitioned depending on the intensity value. Binary thresholding is a simple but effective global automatic thresholding method for binarizing grayscale images such as foregrounds and backgrounds. In the method of binary thresholding, iteration done for all possible available threshold values provided from the image. Then the size of the spread for the pixel level of each side of the threshold is calculated. So it can be found whether the pixel belongs to foreground or background. Finally a threshold value is found which gives a minimum value as the number of foreground and background spreads.

4.6 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.

4.5 Erosion

Feature extraction is done using the process of erosion morphology using structuring elements (SE) and ordinary morphological processes. Picture shows a diamond structuring element .

The structuring elements (SE) determines the exact detail of the operator effect on the image . SE consists of patterns, which are determined as coordinates of a number of discrete points relative to some starting points. The circular area in the SE is represented as the starting point (origin) . Changes in the size or shape of the structuring elements give a reflection on the output image.

0	0	0	1	0	0	0
0	0	1	1	1	0	0
0	1	1	1	1	1	0
1	1	1	1	1	1	1
0	1	1	1	1	1	0
0	0	1	1	1	0	0
0	0	0	1	0	0	0

Fig 7. Diamond structure element

Erosion is an operation used to shrink binary image objects. Mathematically, Erosion combines two sets using the vector reduction set element . Erosion

X and B is a set of points from all possible reduction of element pair vectors, one from each set of X and B

where, X is an image with size M × N and B is the arrangement element or kernel. The basic effect of the operator on binary images is to erode the boundaries of the foreground pixel area . Thus, the foreground pixel area shrinks in size, and the hole in the area becomes larger.

4.6 Region Properties

Nearby areas are also called objects, connected components, or lumps. The label matrix that contains contiguous regions might look like this:

1 1 0 2 2 0 3 3
1 1 0 2 2 0 3 3

Element L equals 1 belonging to the first contiguous region or connected component; element L equals 2 belonging to the second connected component; etc. Unclear regions are areas that may contain several connected components to it . A label matrix containing irregular regions might look like this:

1 1 0 1 1 0 2 2
1 1 0 1 1 0 2 2

Element L is equal to 1 belonging to the first region, which is unclear and contains two connected components. Element L equals 2 belonging to the second region, which is a single connected component.

5. EXPERIMENTS AND RESULTS

The experiments in this study used the Matlab programming language which was already in the form of a program using GUI (Graphical User Interface) picture shows the main view of the program

The first process is inputting the image to be identified by pressing the "Browse Image" button, the syntax for running the command is in the Program Code

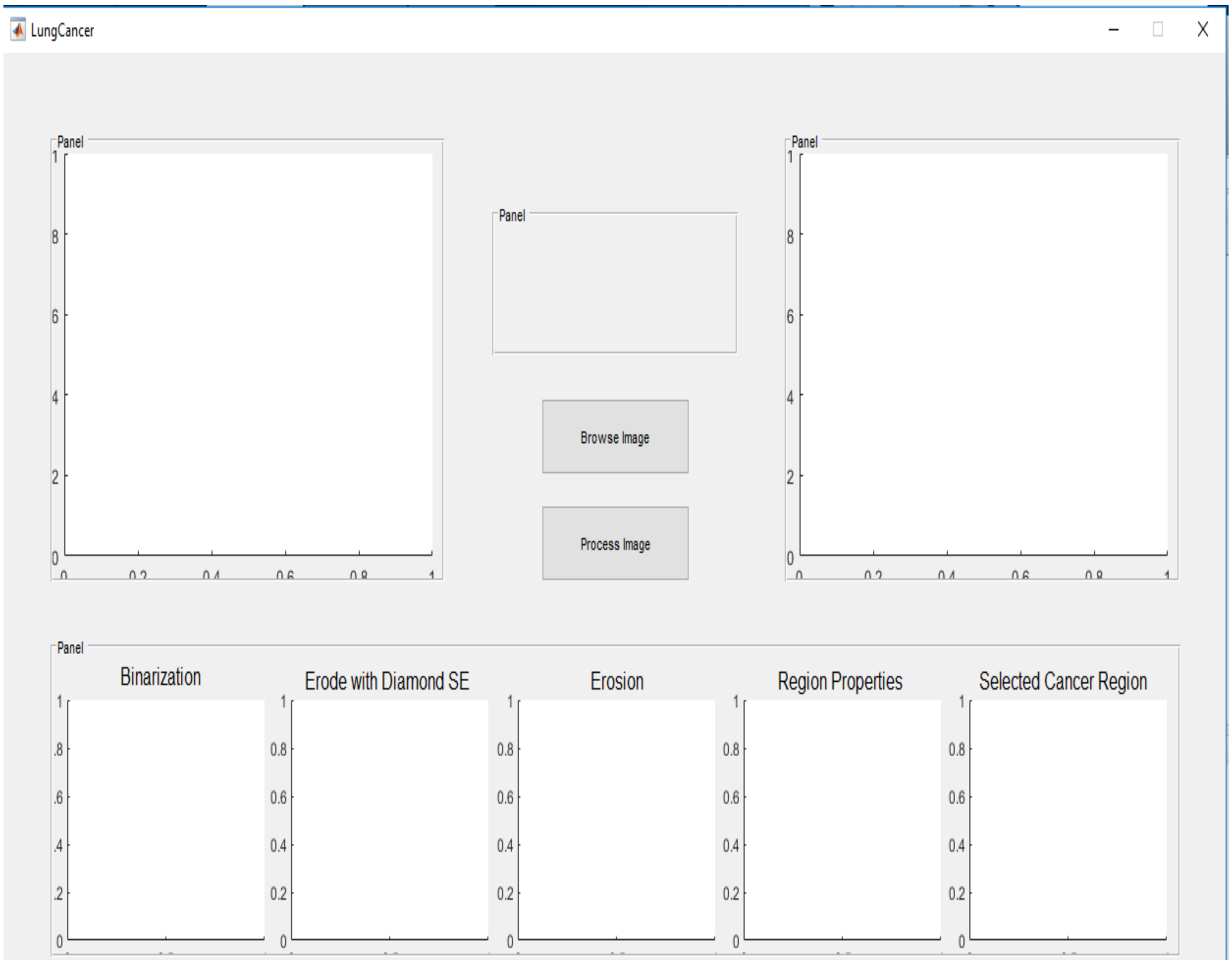


Fig 8. main view of the program

```
global img;
[nama_file, nama_path] =
uigetfile('*.png;*.jpg;*.bmp;*.gif;*.tif','Select
Image');
if ~isequal (nama_file,0)
    img = imread(fullfile(nama_path,nama_file));
    guidata(hObject,handles);
    axes(handles.axes1);
    imshow(img);

    cla(handles.axes3);
    cla(handles.axes4);
    cla(handles.axes5);
    cla(handles.axes6);
    cla(handles.axes7);
    cla(handles.axes8);
else
    return;
end
```

Program Code ... is the code to run the command to open the image after pressing the "Browse Image" button. If before the image is already open, then when opening a new image of that of the axes - axes particular will be reset . Then press the "Process Image" button, it will run a series of processes which will be explained as follows.

The first process is to carry out a binarization process. The syntax for doing this can be seen in the Program Code

```
global img;
level = graythresh(img);
img_bin = rgb2gray(img);
img_bin = imbinarize(img_bin, level);
axes(handles.axes4);
imshow(img_bin);
```

Binary process uses thresholding so that the thresholding process can be adaptive in accordance with the conditions of the image, then the thresholding will be carried out binarization to facilitate the next process. The image of the results of the process can be seen in Figure ...

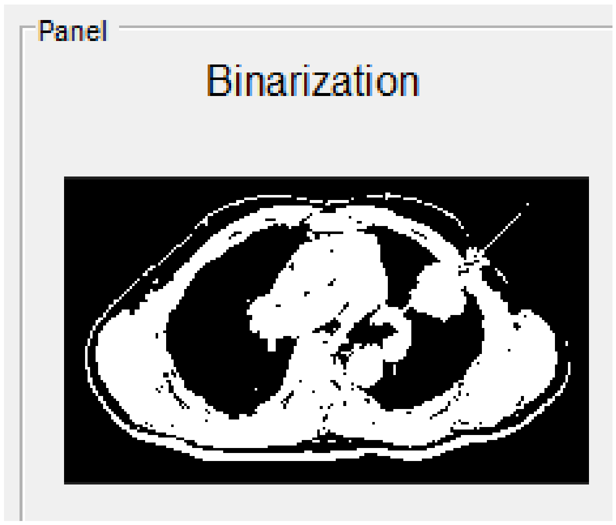


Fig 9 Binarization

After binarization, the erosion process will then be carried out. The syntax for doing this can be seen in the Program Code ...

```
se = se = strel('diamond',3);
img_er = imerode(img_bin,se);
axes(handles.axes5);
imshow(img_er);

img_er = bwmorph(img_er,'erode',1.5);
axes(handles.axes6);
imshow(img_er);
```

The erosion morphology process is carried out twice, first by using the " diamond " structuring element and the second by erosion morphology with a threshold . This process is done so that the tumor object can be separated from the object of the lungs so that later it will facilitate the next process. The erosion process image can be seen in Figure ...

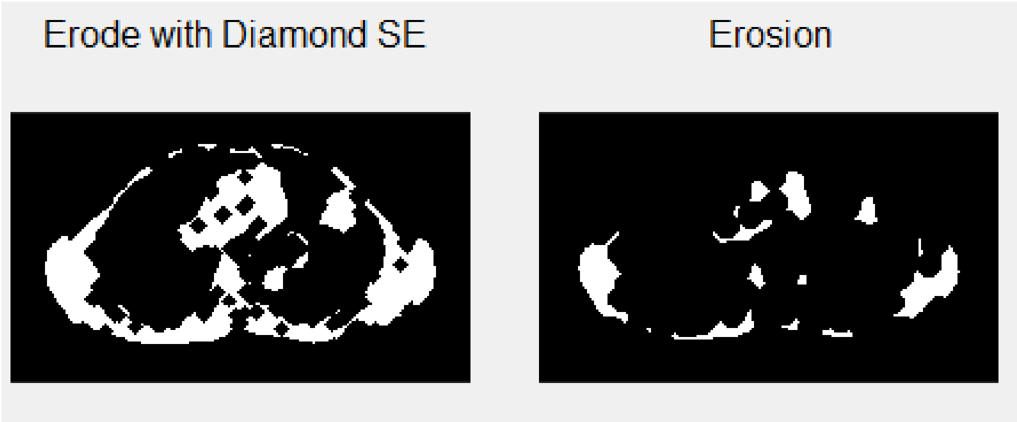


Fig 10 Erosion

After the erosion process, it will proceed with determining the region properties in the image that has been eroded. The syntax of the process can be seen in the Program Code ...

```
BWr = regionprops(img_er, 'BoundingBox', 'Area',
'Image');
axes(handles.axes7);
imshow(img_er);
hold on;
pixelH = [];
pixelW = [];
for i=1:size(BWr,1)
    rectangle('Position',
BWr(i).BoundingBox,'edgecolor','red');
    pixelH(1,i) = BWr(i).Area;
end

cc = bwconncomp(img_er);
num = cc.NumObjects;

arr = sort(pixelH, 'descend');
arrMaxP = arr(1,num);
[c_max, idx_max] = max(arrMaxP);
```

```
T = 1500;
removeMask = [BWr.Area]>T;

BWremove2 = img_er;
BWremove2(cat(1, cc.PixelIdxList{removeMask})) = false;
BWremove2 = bwmorph(BWremove2, 'open', Inf);
axes(handles.axes8);
imshow(BWremove2);
```

The process of determining region properties is to find the whole properties of each object detected in the image. The aim is to remove the detected tumor object that has not been removed from the lung . The process of region properties, will first determine the region in the form of detected objects. Then set a threshold of 1500 which later if there are objects that result from region properties that have a value of more than 1500, the object will be removed , so that objects that are less than that are considered as tumors. The object that has been detected as a tumor can be used as a tumor detection on the CT scan image . The results of this process can be seen in Figure ...

Fig 11 Region detection

Because there are objects that are not removed in the previous process , the object is detected as a tumor. So the result of identification of the image is an abnormal lung condition.

Fig 12. Tumor detection

6.OUTPUT

Below is the output of program :

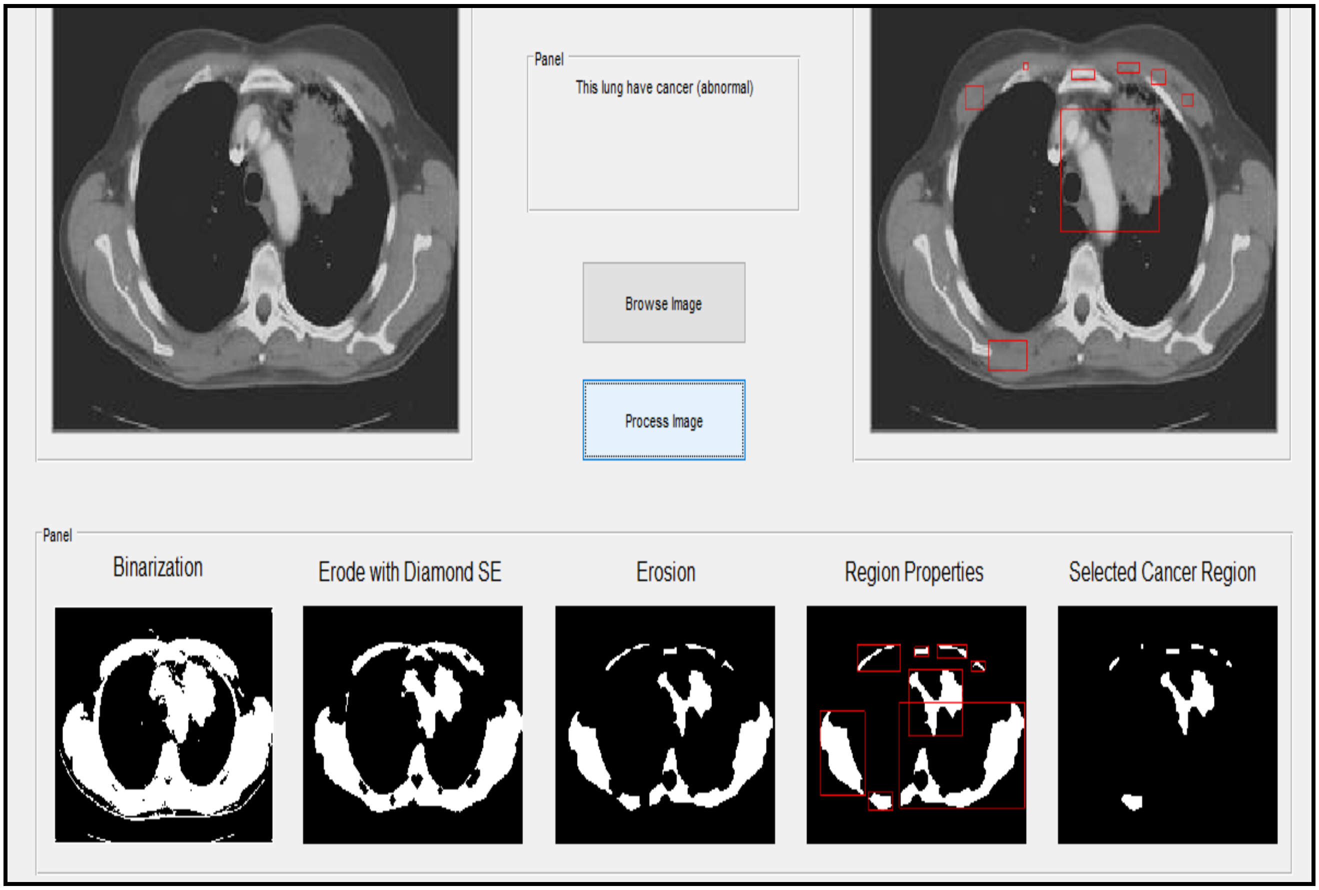


Fig 13 Tumor detection in lung cancer

7. CONCLUSION

An image improvement technique is 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 is the core factors of this research, image quality assessment as well as enhancement stage where were adopted on low pre-processing techniques. The proposed technique is efficient for segmentation principles to be a region of interest foundation for feature extraction obtaining. The proposed technique gives very promising results comparing with other used techniques. Relying on general features, a normality comparison is made. The main detected features for accurate images comparison are pixels and tumor region with high accuracy and robust operation.

8. REFERENCE

[1] 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, (2001).*

[2] K. Kanazawa, Y. Kawata, N. Niki, H. Satoh, H. Ohmatsu, R. Kakinuma, M. Kaneko, N. Moriyama and K. Eguchi, "Computer-aided diagnosis for pulmonary nodules based on helicalCT images", *Compute. Med. Image Graph, vol. 22, no. 2(1998),pp. 157-167.*

[3] D. Lin and C. Yan, "Lung nodules identification rules extraction with neural fuzzy network", *IEEE, Neural Information Processing, vol. 4,(2002).*

[4] 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, (2013).*

[5] A. El-Baz, A. A. Farag, PH.D., R. Falk, M.D. and R. L. Rocco,M.D.,"detection, visualization, and identification of lung abnormalities in chest spiral CT scans: phase I", *Information Conference on Biomedical Engineering, Egypt (2012).*

[6] Yamamoto. T, Ukai. Y, Kubo. M, Niki. N, Satou. H and Eguchi. K, "Computer aided diagnosis system with functions to assist comparative reading for lung cancer based on helical CT image", *Image Processing, 2016 International Conference on Proceedings, pp: 180 – 183, vol.1, 2016.*

[7] Cheran. S. C, Gargano. G, "Computer aided diagnosis for lung CT using artificial life models", *Seventh International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2005, 2005.*