CSE4019-Image Processing

Lung Cancer Detection Using Segmentation

Submitted by : - Aditya Khandelwal , Aditya Chaudhary , Saksham Agarwal

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 pre-processing, segmentation and feature extraction have been discussed in detail. We are aiming to get the more accurate results by using various enhancement and segmentation techniques.

Keywords: Image Segmentation, Image Extraction, Image Enhancement

Introduction:

Lung cancer is a disease of abnormal cells multiplying and growing into a tumor. The mortality rate of lung cancer is the highest among all other types of cancer. Lung cancer is one of the most serious cancers in the world, with the smallest survival rate after the diagnosis, with a gradual increase in the number of deaths every year. Survival from lung cancer is directly related to its growth at its detection time. But people do have a higher chance of survival if the cancer can be detected in the early stages [1]. Cancer cells can be carried away from the lungs in blood, or lymph fluid that surrounds lung tissue. Lymph flows through lymphatic vessels, which drain into lymph nodes located in the lungs and in the centre of the chest. Lung cancer often spreads toward the centre of the chest because the natural flow of lymph out of the lungs is toward the centre of the chest. Lung cancer can be divided into two main groups, non-small cell lung cancer and small cell lung cancer. These assigned of the lung cancer types are depends on their cellular characteristics. As for the stages, in general there are four stages of lung cancer; I through IV. Staging is based on tumor size and tumor and lymph node location. Presently, CT are said to be more effective than plain chest x-ray in detecting and diagnosing the lung cancer. An estimated 85 percent of lung cancer cases in males and 75 percent in females are caused by cigarette smoking. Objective of this study is to detect lung cancer using image processing techniques. CT scanned lung images of cancer patients are acquired from various hospitals. Using image processing techniques like pre-processing and feature extraction, area of interest is separated. Developing the algorithm, features like area, perimeter and eccentricity are extracted from all the images. The parameter values obtained from these features are compared with the normal values suggested by a physician. From the comparison result, cancer stage is detected. A graphical user interface is developed to scan all the images and display the features and cancer stage. This system can help in early detection of lung cancer more accurately.

The rank order of cancers for both males and females among Jordanians in 2008 indicated that there were 356 cases of lung cancer accounting for (7.7 %) of all newly diagnosed cancer cases in 2008. Lung cancer affected 297 (13.1 %) males and 59 (2.5%) females with a male to female ratio of 5:1 which Lung cancer ranked second among males and 10th among females. Figure 1 shows a general description of lung cancer detection system that contains four basic stages. The 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.

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. 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.

Methodology:

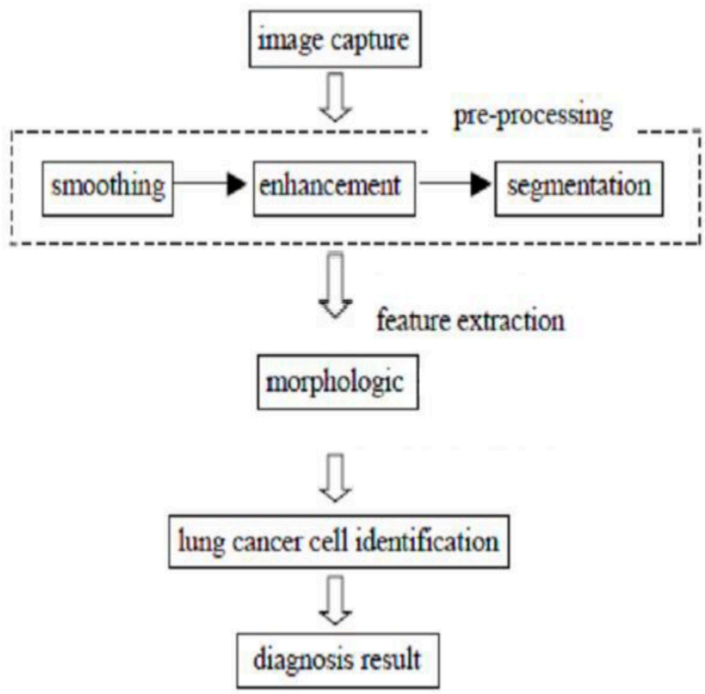


Image Acquisition:

First step is to acquire the CT scan image of lung cancer patient. The lung CT images are having low noise when compared to X-ray and MRI images; hence they are considered for developing the technique. The main advantage of using computed tomography images is that, it gives better clarity and less distortion. For research work, the CT images are acquired from NIH/NCI Lung Image Database Consortium (LIDC) dataset. DICOM (Digital Imaging and Communications in Medicine) has become a standard for medical Imaging. Figure 3.2 shows a typical CT image of lung cancer patient used for analysis. The acquired images are in raw form. In the acquired images lot of noise is ob-served. To improve the contrast, clarity, separate the background noise, it is required to pre-process the images. Hence, various techniques like smoothing, enhancement are applied to get image in required form.

Binarization:

Binarization 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.

Erosion and Dilation:

Image segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify or change the representation of an image into something that is more meaningful and easier to analyze. Segmentation divides the image into its constituent regions or objects. The result of image segmentation is a set of segments that collectively cover the entire image or a set of contours extracted from the image. Marker-controlled watershed segmentation follows this basic procedure. Compute a segmentation function. This is an image whose dark regions are the objects you are trying to segment. 2) Compute foreground markers. These are connected blobs of pixels within each of the objects. 3) Compute background markers. These are pixels that are not part of any object. 4) Modify the segmentation function so that it only has minima at the foreground and background marker locations. 5) Compute the watershed transform of the modified segmentation function.

Feature Extraction:

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

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

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

Perimeter = P =(Pi,j , X edge[P] = i, Y edge[P] = j) Where, X edge [ ] and Y edge[ ] are vectors represent the co-ordinate of the ith and jth pixel forming the curve, respectively.

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

Eccentricity = Length of Major Axis /Length of Minor Axis

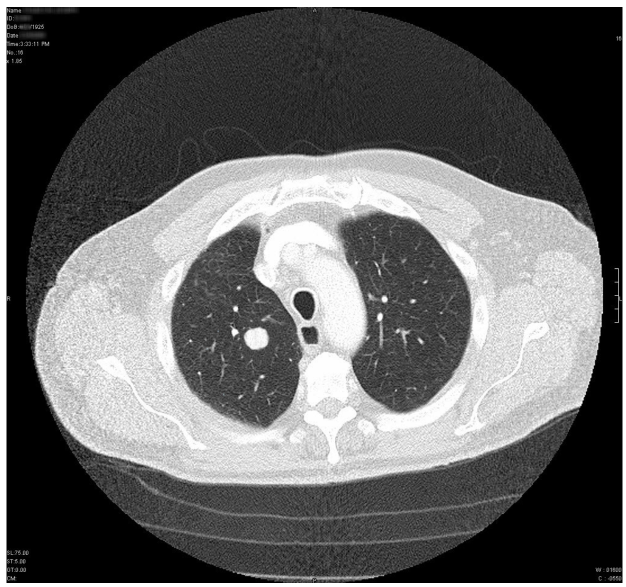
Results and Analysis

```
clc;
clearvars;
close all;
file = 'ok2.jpg';
imtool close all;
FontSize = 12;
initImage = imread(file);
[rows, columns] = size(initImage);
initImage = rgb2gray(initImage);
initImage = medfilt2(initImage,[10 10]);
[B, A] = imhist(initImage)
C=A.*B;
D=A.*A;
E=B.*D;
```

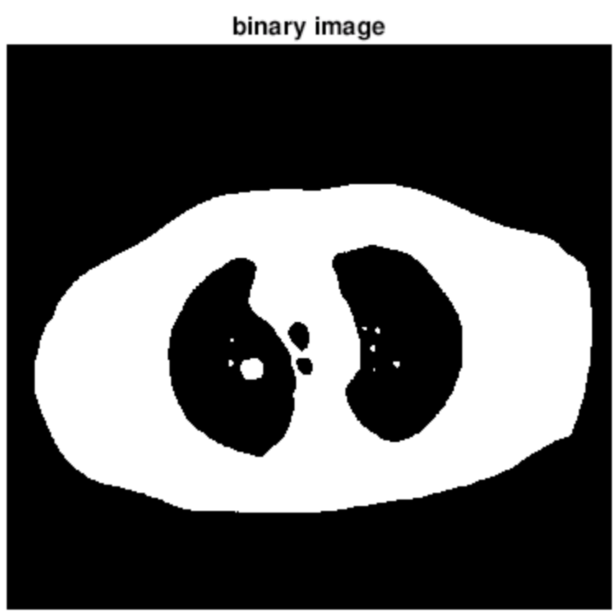
```
n=sum(B);
Mean=sum(C)/sum(B);
var=sum(E)/sum(B)-Mean*Mean;
std= (var)^0.5;
thresholdValue = Mean+0.5*std;
bwImage = initImage > thresholdValue;
figure
imshow(bwImage)
title('binary image');
img_dil = imdilate(bwImage , strel('arbitrary', 20));
figure
imshow(img_dil);
title('dilated image');
bwImage = imerode(img_dil , strel('arbitrary', 20 ));
figure
imshow(bwImage);
title('eroded image');
bigMask = bwareaopen(bwImage, 2000);
finalImage = bwImage;
finalImage(bigMask) = false;
bwImage=bwareaopen(finalImage,55);
figure
imshow(bwImage)
labeledImage = bwlabel(bwImage, 8);
RegionMeasurements = regionprops(labeledImage, initImage, 'all');

Ecc = [RegionMeasurements.Eccentricity];
RegionNo = size(RegionMeasurements, 1);
allowableEccIndexes = (Ecc< 0.98);
keeperIndexes = find(allowableEccIndexes);
RegionImage = ismember(labeledImage, keeperIndexes);
bwImage=RegionImage;
figure
imshow(RegionImage)
%%%%%
clear labeledImage;
clear RegionMeasurements;
clear RegionNo;
```

```
labeledImage = bwlabel(bwImage, 8);
RegionMeasurements = regionprops(labeledImage, initImage, 'all');
figure
imshow(initImage);
title('Outlines', 'FontSize', FontSize);
axis image;
hold on;
boundaries = bwboundaries(bwImage);
numberOfBoundaries = size(boundaries, 1);
for k = 1 : numberOfBoundaries
thisBoundary = boundaries{k};
plot(thisBoundary(:,2), thisBoundary(:,1), 'r', 'LineWidth', 3);
end
hold off;
RegionMeas = regionprops(labeledImage, initImage,
'all'); RegionNo = size(RegionMeas, 1); textFontSize
= 14;

labelShiftX = -7;
RegionECD = zeros(1, RegionNo);
fprintf(1,'Region number Area Perimeter Centroid
Diameter\n'); for k = 1 : RegionNo
RegionArea = RegionMeas(k).Area;
RegionPerimeter =
RegionMeas(k).Perimeter;
RegionCentroid =
RegionMeas(k).Centroid; RegionECD(k) =
sqrt(4 * RegionArea / pi);
fprintf(1,'#%2d %11.1f %8.1f %8.1f %8.1f % 8.1f\n', k,
RegionArea, RegionPerimeter, RegionCentroid,
RegionECD(k));

text(RegionCentroid(1) + labelShiftX, RegionCentroid(2), num2str(k),
'FontSize',
textFontSize, 'FontWeight', 'Bold');
end
```

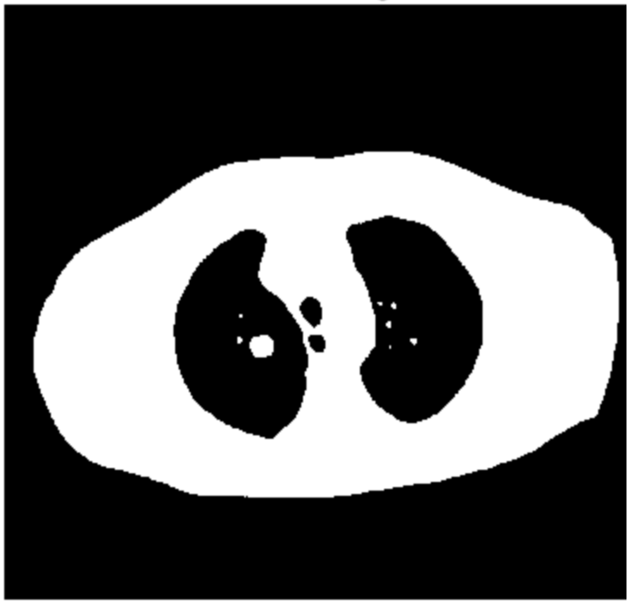
Original Image:



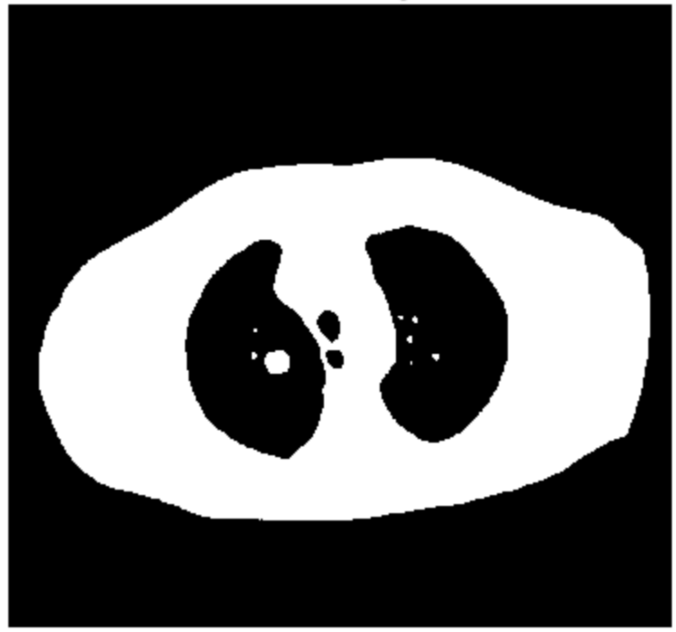
Processes images :

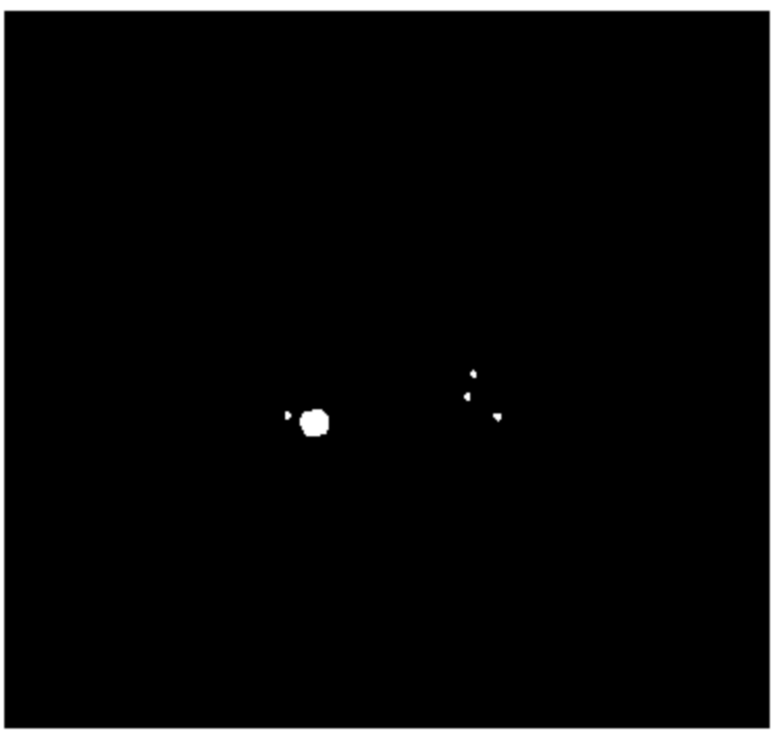


dilated image



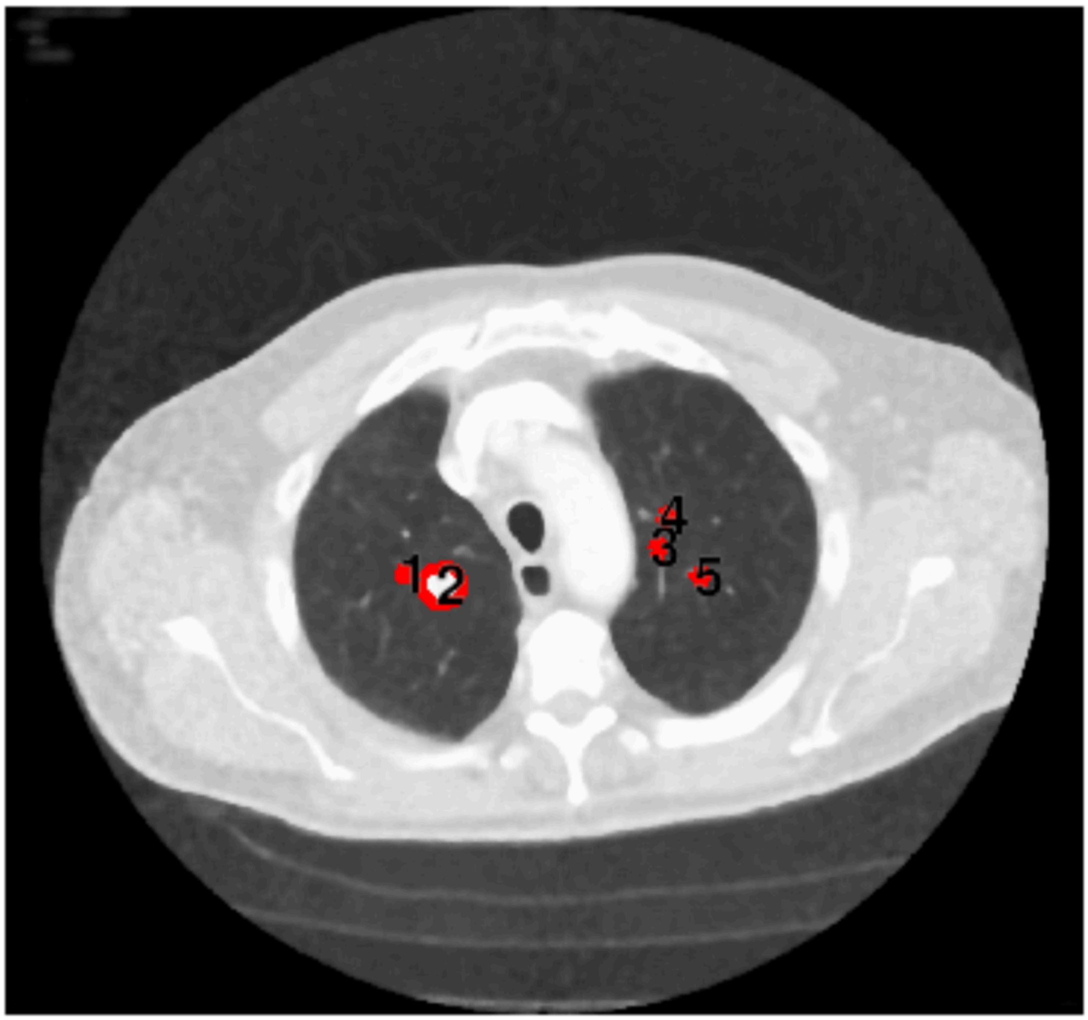
eroded image





Final Image :

Outlines



The Module Outputs are as expected.

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 based on Gabor filter within Gaussian rules. 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' percentage and mask-labelling with high accuracy and robust operation.

References:

[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.