Lung Cancer Detection from X-ray Image

by

Exam Roll: Curzon Hall-209
Registration No: 2014-416-633
Session: 2014-15

Exam Roll: Curzon Hall-210
Registration No: 2014-216-635
Session: 2014-15

A project submitted in partial fulfilment of the requirements for the degree of
Bachelor of Science in Computer Science and Engineering

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
UNIVERSITY OF DHAKA

January 5, 2019

Abstract

Lung cancer is the deadliest disease in current times in all over the world. Mortality rate of both men and women affected by lung cancer is higher than any other types of cancer. As increasing step refers more spread cancer, survival rate of a certain years is significantly dropped. So early detection of lung cancer is the most promising way to prevent patient from early death. Computed Tomography, Mechanical resonance are the mostly used method to detect lung cancer. As most of these common methods are cost and time inefficient we proposed a new method to detect lung cancer. Our proposed method detects lung cancer from X-ray image using template matching technique. According to our method, some templates of lung cancer, collected from JSRT and other trusted source, are matched with a portion of X-ray image and calculate the distance. Image portion with minimum distance is detected as cancer region. As, X-ray is very popular method in medical science and also very cost efficient, our method will help doctor to detect lung cancer more accurately.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cancer is not a specific term. It means the group of disease that result from uncontrolled division of cell. Each cell in human body is under a tissue and has a specific task with a life circle. After completing the life circle the cell is noticed to die. If cell is divided without control continuously then it results in difficulties or disease. This called cancer.

The affected tissue with uncontrolled dividing cell is called mass tissue. Most of the mass tissues are injurious for human body. Some of them are visible and these are called tumor. Many mass tissues are invisible too.

Cancer may occur at any part of human body. Though breast cancer, lung cancer, prostate cancer and colon cancer are most common, lung cancer is the leading cause of cancer-related death.

Number of people affected by cancer in recent years is growing rapidly. As the treatment of cancer is highly expensive and time consuming, the survival rate

is not very poor. Current condition of cancer is divided into four stages. Many cancer organizations show that survival rate of the people whose cancer is detected in first or second stage is significantly high compared to the survival rate of other. So, early detection of cancer is the main concern of cancer treatment.

We want to propose X-ray based lung cancer detection method that will contribute in cancer treatment by detecting cancer in early stage. As the method is cost and time efficient having greater accuracy, this will bring a good result in cancer treatment.

1.1 Problem Definition

Cancer is a term that is used for diseases in which abnormal cells divide without control and spread nearby tissue. Lung cancer is the deadliest among other type of cancers. Most of the people having lung cancer are detected in last stage of cancer. As the survival rate of the people whose cancer is detected in very last stage is very poor so most of them have to accept premature death.

Modern technology is used to diagnosis lung cancer including computed tomography (CT), magnetic resonance imaging (RMI) and some highly efficient technology. Most of the people are not interested to examine of these medical tests because of their highly cost. So, most of the patient are not detected in early period that leads them to early death.

We propose an efficient approach that can detect lung cancer very efficiently. This approach is neither time consuming nor costly. This approach will detect lung cancer from X-ray image of lung. Using machine learning, this approach is

learned by hundreds of lung image including nodule and non-nodule images and will predict whether cancer tissue is appearing in lung or not including exact position.

1.2 Motivation

Early detection of cancer is the most promising way to improve the patients chance for survival from lung cancer. A mass tissue or nodule in lung can be a tumor, cancer tissue and infection. To detect a chest mass tissue, a region of interest needs to be identified. Medical imaging methods such as magnetic resonance imaging (MRI) and computed tomography (CT) have been developed and almost all parts of a living body can be investigated. But these methods are very complex, time consuming and cost inefficient. One the other hand X-ray imaging method are simple, time saving and cost efficient and used to find out abnormalities in heart, rib, especially lung.

Many of the people of our country cant afford the cost of the tests like MRI, CT because of their highly cost. As X-ray is lower cost compare with them and a very popular to all this method have a strong possibility to be popular and save many of lives.

1.3 Challenges that remain

Our proposed method has a good precision and recall percentage as performance measure. But for female both precision and recall are slightly lower than male patient with a lower margin. For a defined threshold both precision and recall are not excellent.For lower threshold recall is very high but precision is poor. For

higher threshold recall is poor but precision is very good. As precision refers exactness and recall refers completeness for a given threshold both of them are not excellent. So for overall case we have to pay attention on both and that gives a moderate result. To make both precision and recall higher is the main challenge for us in future.

1.4 Project objective

We extract cancer region in lung in our project. Basic objectives are listed below

1. To investigate Image enhancement techniques to improves the qualities of an X-ray image.
2. To propose a new frame work for X-ray images enhancement and detect lung nodule.
3. To provide noise reduction capabilities, with considerably less blurring by using effective filter which is median filter.
4. To propose method which will increase the sharpness of X-ray images to visualize cancerous region
5. To design lung cancer detection system based on proposed methods for better diagnosis.

1.5 Organization of the Report

The project report is organized as follows:
Chapter 2 talks about the preliminary concepts that are relevant to the topic and discusses some of the state-of-the-art works that directly influence this study. In **Chapter 3**, the proposed method is developed and an example is worked out. Performance of the proposed method is discussed in **Chapter 4**, along with other

result variations in JSRT dataset. In **Chapter 5**, the summary is provided as conclusion, and directions for future work is discussed.

Chapter 2

Background Study and Related Work

2.1 Introduction

Cancer is a condition where cells are divided without control. Uncontrolled dividing cells never die though each cell has a life cycle including death after a certain period. Majorly lung cancer cell are two in kinds. These are small lung cancer cell and non-small lung cancer cell. Almost 85% of lung cancers are in non-small type.

2.2 Preliminary Concepts

We need to know about some terms those are very important to understand thesis paper related to our topic.

2.2.1 Lung Cancer

Lung cancer is a common disease around the world. Both male and female are subjected to premature death for lung cancer. Mortality rate by cancer is greatly

raised by death by lung cancer. However, lung cancer is curable if it diagnosed at early stage.

2.2.2 Types of Lung Cancer

Lung cancer occurs for uncontrolled growth of lung cells and form tumor. Tumor can be cancerous or non-cancerous. If the tumor contains large number of cancerous cell that have the ability to spread another part of the body then it called lung cancer. The rate of mortality by lung cancer depends on different types of lung cancer. These cancers grow differently and diagnosis of them are also different.

There are two major types of lung cancer defined by the size of cells. Their type of growth and the way of treatment are different.

2.2.2.1 Small Cell Lung Cancer (SCLC)

About 10-15% of lung cancers are in small cell lung cancer. This type of lung cancer grows faster and spread other organ more quickly than non-Small Cell Lung Cancer (NSCLC). The number of deaths by SCLC is higher than NSCLC.

2.2.2.2 Non-small Cell Lung Cancer (NSCLC)

About 85-90% of lung cancers are in Non-Small Cell Lung Cancer. This type of lung cancer grow slower than small type of lung cancer and cured by surgery if cancer is detected in early stage.

2.2.3 Stages of Lung Cancer

The stage of a cancer is determined by the location of the cancer cells, the size of the cancer cells, whether they have spread from whether they initially began. The stage of a cancer plays a significant role to determine the appropriate method of treatment by doctor to cure the diseases and prolong the patients life.

2.2.3.1 Stages of Small Cell Lung Cancer

Small cell lung cancer is divided into mainly two stages.

Limited Stage: Cancer cells are found in lung area only. They havent spread.

Extensive Stage: Cancer cells have spread other parts of the body.

2.2.3.2 Stages of Non-small Cell Lung Cancer

Non-small cell lung cancer is divided into four stages.

Stage 1: Cancer cells are found only lung area. They are not found in lymph nodes and have not spread other organ of human body.

Stage 2: This stage is divided into two sub stages.

- **2A:** In this stage, cancer cells are relatively small and exist in lymph nodes.
- **2B:** In this stage, cancer cells are larger than primary stage and seen on lymph nodes. Cancer cells have spread over the affected lung cells.

Stage 3: Stage 3 is also divided into two sub stages like stage 2.

- **3A:** Cancer cells have extended to the lymph node far away from the invaded lung. Cancer cells are seen in lymph node and other portion of the affected lung. They are also seen other organs besides the lung area.
- **3B:** Cancer cells have spread outside of the lymph. Infected cells are seen in collarbone, chest or there is more than one tumor. Cancer cells also hold fluid around the lung.

Stage 4: In this stage lung cancer is seen liver, brain and bone.

2.2.4 Morphological Transformation

Morphological transformations are some simple operation based on image shape. Normally it is performed on binary image. Morphological transformation takes

two inputs. One is original binary image and another is kernel that decides the characteristics of the operation. Simply erosion and dilation are the morphological transformation. Some other operation also performed based on erosion and dilation. We will see here different morphological transformation of an image.

FIGURE 2.1: Original Image

2.2.4.1 Erosion

Erosion is like as soil erosion. In erosion it takes a kernel and scan whole the image by substituting the pixel based on kernel. If whole pixel values are upper than kernel then the value of considering pixel is 1 otherwise 0. So, in border line the whole pixel values are not upper than the kernel value. For this the value of the pixel in a border line is substituted by 0. This shrinks the object and looks like the erosion of the object.

Erosion is used to remove white noise from an object. A noise is easily removed by erosion as noise cant keep the pixel value under kernel higher than kernel value. Normally 4x4 or 5x5 kernels are used with full of 1s

FIGURE 2.2: Erosion

2.2.4.2 Dilation

Dilation is just the opposite of erosion. It also scans the whole image. But in this time, the pixels under the kernel is calculated and substituted considering pixel by on 1 if at least one pixel is 1. So, it increases the white region of the image. Normally, in case of noise removal erosion is followed by dilation because erosion shrinks the object and dilation reconstructs the object. Since noise is gone for erosion they will not come back but the object get its own shape. Dilation is also useful to join a broken part of an object.

FIGURE 2.3: Dilation

2.2.4.3 Opening

Opening is the combined operation of erosion and dilation. Here, erosion is followed by dilation. It uses to remove noise from image.

FIGURE 2.4: Opening

2.2.4.4 Closing

Closing is just the opposite of opening. Here dilation is followed by erosion. Closing is used to remove black spot from image.

FIGURE 2.5: Closing

2.2.5 Precision and Recall

To calculate precision and recall, true positive (TP), false positive (FP), true negative (TN) and false negative (FN) are calculated before. True positive refers the number of events those are truly predicted as positive occurrence. True negative refers the number of events those are truly predicted as negative occurrence. False positive refers the number of events those are falsely predicted as positive occurrence. False negative refers the number of events those are falsely predicted as negative occurrence. So, in these contexts a cell detected as cancerous cell will count as true positive if the cell was actually cancerous and false positive if the cell was not cancerous. Likewise a cell detected as a non-cancerous cell will count as true negative if the cell was non-cancerous cell and false negative if the cell was actually cancerous. So, only true positive detect truly a cancerous cell as cancerous and true negative detect a cell as non-cancerous truly. Any false positive and false negative decreases the performance of a system.

All those are shown in a table called confusion matrix.

Table 2.1: Confusion Matrix

	Predicted class			
		yes	no	Total
Actual class	yes	TP	FN	P
	no	FP	TN	N
	Total	P'	N'	$P + N$

Based on these, performance of a system can be measured by the term ***Precision*** and ***Recall*** Accuracy and precision are measured by following formulas

$$Precision = \frac{TP}{TP + FP} \tag{2.1}$$

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{P} \tag{2.2}$$

From equation (2.1) we see that to calculate precision false positive is a parameter that reduce the precision. So, precision refers how the method can detect a TP exactly. For this precision is called exactness. Equation (2.2) says that recall is percentage of true positive with respect to total positive. How many result is taken is not an issue here. How complete the true positive is the main issue. For this reason, recall is called completeness.

2.3 Related Works

Here we want to discuss about some papers that are related with our topic. We fully discussed three papers. The problems addressed by those papers, their positive and negative aspects, described methods and challenges are mainly discussed in this section.

2.3.1 Problems Those Papers Address

In [1] local contralateral subtraction method is proposed. They uses computer aided diagnosis (CAD) scheme to detect lung nodules in chest X-ray.

In [2] here three different methods of nodule detection from posterior and anterior chest Radiographs are described. In first method, nodule region is extracted from chest radiograph and find its geometrical features. Then some morphological transformations are happened with predefined threshold to extract the nodule. In second method, multiscale blob detection method is used to detect nodule region and then support vector machine (SVM) classifier [3] is used to reduce false positive. Third method gives more attention to select candidate nodule including

subtle region. Then Restricted Boltzmann machine [4] and SVM classifier are used to detect actual nodule region.

2.3.2 Methods and Positive Aspects of Those Papers

In [1] the total works in divided into some parts. First is preprocessing. X-ray image is converted into 512 x 512 pixels image and then contrast stretching is implemented to visible the image more clearly. The image is converted into binary image then using Otsu method [5] with predefined threshold value. The binary image contains noise. To eliminate noise a morphological method, closing, is performed. Morphological closing method is simply a dilation followed by erosion.

The second step is region of interest (ROI) extraction. They apply mass tissue classifier only left and right lung image. Any other part of the X-ray image is not examined. So, only left and right lung is extracted as ROI extraction.

In third or last step, mass tissue detection classifier is applied. Some classified mass tissue is collected from reliable source and most of them are in circular shape. The whole image is masked with 8x8, 16x16 and 32x32 and so on. Then each mask is transformed into discrete cosine transform (DCT). The DCT image is used to classify mass tissue by calculating Euclidian distance from top pixel of DCT image with a predefined threshold value.

Paper [1] shows that the average accuracy and precision are respectively 76.12% and 76.16%. By using X-ray technology the performance is pretty good.

In [2] three different methods are described. Two of them are discussed here.First method extract the nodule using threshold and then performing some morphological transformations. To separate nodule region based segmentation technique Region Growing [6-7] is used. Then some geometrical features are calculated from extracted nodule. Here diameter, perimeter and area is calculated. These are used to calculate irregularity index. As the growth of tumor is circular in nature according to literature survey [8], Irregularity index is calculated using this simple formula:

$$I = \frac{4 * pi * area}{perimeter^2} \tag{2.3}$$

Irregularity index is also known as circularity index or roundness value. This roundness value is 1 for a circular shape object and less than 1 for any other shape. So a object is marked as a nodule if roundness value is close to 1

In second method, for image enhancement two methods are applied. First one is contrast limited adaptive histogram equalization and second one is local contrast enhancement using Gaussian low pass filter.
To normalize the contrast within the image and and accross all the images local enhancement is done using this equation,

$$y(m,n) = \frac{x(m,n) - \mu(m,n)}{\alpha(m,n)} \tag{2.4}$$

Here,

x(m,n) is input image

y(m,n) is local contrast enhanced image

μ(m,n) is local mean estimate

α(m,n) is local standard deviation estimate

μ(m,n) is calculated by

$$\mu(m,n) = x(m,n) * h(m,n) \tag{2.5}$$

h(m,n) is the Gaussain low pass filter response with standard deviation 16. The local standard deviation is calculated by

$$\alpha(m,n) = \sqrt{X^2(m,n) * h(m,n) * \mu^2(m,n)} \tag{2.6}$$

Then lung area is segmented. After segmentation blob detection method is used to detect nodule in lung. Blob detection method scans the whole image and identify difference of intensity, darkness of any circular shape. Then Laplacian of Gaussian filter is used to detect blob on enhanced image.

After that SVM classifier is to classify a nodule as mass tissue. SVM classifier classify a nodule based on some features.Here, five features are calculated for each candidate nodule. Mean, standard deviation, skewness, kurtosis and gradient magnitude are those five features.SVM classifier classify a candidate nodule based on these features.

In [9] convolutional neural networks(CNN) [10] is ensemble to reduce false positive. CNN has a great advantage to detect nodule directly with the help of

great amount of data. Here chest X-rays are used. At first CNNs are architectured and then they are ensembled. Five fold cross validation and free-response receiver operating characteristic (FROC) [11] were used to analyze the performance of nodule detection. This method gives 94% sensitivity which is the other name of recall with 4.6 false positive per image and 80% sensitivity with 2 FPs/images. In both [12] and [13] the sensitivity was 80% with 5.4 and 5 FPs/image. So [9] shown a great improvement to reduce false detection.

2.3.3 Challenges that Remains

In first paper, though accuracy and precision is very good enough with respect to X-ray image some challenges also remain. One of the critical task is segment the lung area. In sometimes areas outside of lung are entered into the region of interest and detect some nodule that are not actually in lung. DCT image values are used to calculate Euclidian distance that is compared to a threshold value. Sometimes Euclidian distances higher than the threshold value though the region was not a mass tissue. Calculating the threshold value is needed some labeled images that are actually classified by doctor or collected from trusted source. Other methods also have high false positive rate.

2.4 Summary

Some methods are described here. All of them detect lung cancer form X-ray image in differnt approach. Circularity of cancer tissue, difference intensity of cancerous and non cancerous region, ensemble of CNN, multi-scale blob detection technique, SVM classifier are commonly used in those paper. Most of the method are sensitive to false positive.

Chapter 3

The Proposed Approach

3.1 Introduction

In Chapter 2 we discussed the preliminary concepts and some existing methods of detecting lung abnormalities. In this chapter we merge those concepts to propose a new method for detection of lung cancer from X-ray image.

3.2 Data Collection Method

At first we search some papers those are in related topic. We addressed the authors of those papers to collect data but we have seen most of them used JSRT (Japanese Society of Radiological Technology) database for their work. Then we decided to collect data from same source. So, all data are collected from JSRT database. JSRT has a free database of digital image. Both nodule and non-nodule images are in this database. All of these are labeled data. JSRT database has limited access for potential user. We collected whole JSRT database for our project.

3.3 Data Description

JSRT database contain both nodule and non-nodule data. Some characteristics of JSRT data are listed below:

- Useful for ROC analysis (154 nodule and 93 non-nodule images)
- High resolution (2048 x 2048 matrix size, 0.175mm pixel size)
- Wide density range (12bit, 4096 gray scale)
- Universal image format (no header, big-endian raw data)
- Useful for diagnostic training and testing
- Images are in IMG format
- Some additional information is also provided
  - Patient age
  - Gender
  - Diagnosis (malignant or benign)
  - X and Y coordinates of nodule
  - Simple diagram of nodule location
  - Degree of subtlety in visual detection of nodules

3.4 Data Preprocessing

As all the images of JSRT database are in IMG format and image in IMG format are not easy to open so we have convert all the images into BMB format according to user guide provided by JSRT database. We converted a 2048 x 2048 image into two 1024 x 1024 image so that we can segment left lung and right lung. We detect spinal cord using histogram and divide the image.Images contain different kinds of noise. We try to eliminate noise by applying morphological transformation.

Then lung region is extracted to find lung cancer within only lung area. All the preprocessing steps are discussed below for an original X-ray image

FIGURE 3.1: Original X-ray image

3.4.1 Contrast Stretching

Without proper image enhancement we will not get expected output from X-ray image. Contrast Stretching is such a technique that removes black and white shadow from input image. We use the flowing equation for contrast stretching.

$$Pout = (Pinc)\left(\frac{b-a}{d-c}\right) + a \tag{3.1}$$

Where, Pin and Pout are respectively the intensity level of input pixel and output pixel. Likewise, a and b are the upper and lower intensity limit of an image. And, c and d are highest and lowest pixel value of the image

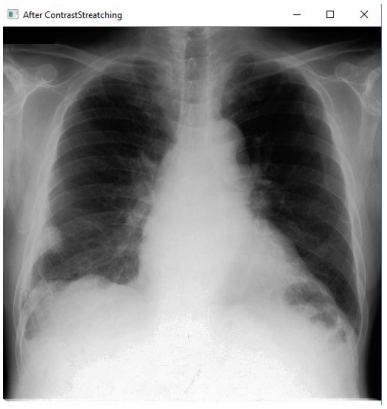


FIGURE 3.2: Contrast stretched image

3.4.2 Binary Imaging

After contrast stretching, binary image is created by taking a threshold value (T) selected by Otsu's method using the equation that is given below

$$g(x,y) = \begin{cases} 1, & f(x,y) > T \\ 0, & f(x,y) <= T \end{cases} \tag{3.2}$$

Where, g(x,y) is the binary image and f(x,y) is the image after contrast stretching.

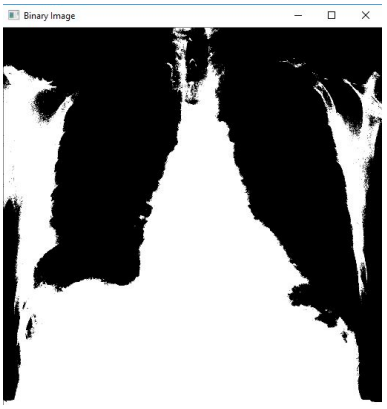


FIGURE 3.3: Binary Image

3.4.3 Closing

Closing is simply a morphological dilation operation followed by an erosion operation.This is also discussed in chapter 2 also. However, the equation of closing is:

$$f \bullet s = (f \oplus s) \ominus s \tag{3.3}$$

Where, s is a 3x3 structuring element or kernel and $\oplus and \ominus$ are respectively dilation and erosion operation.

FIGURE 3.4: X-ray Image After Closing

3.4.4 ROI Extraction

Region of interest extraction is a segmentation process. Using ROI extraction, we extract the portion of lung that is only needed for us. We only detect mass tissue from this region further. To extract region of interest we use the algorithm for both left and right lung:

1. Detect the spinal cord.

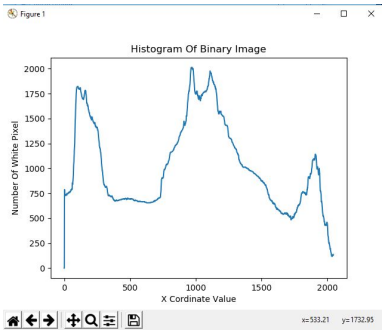


FIGURE 3.5: Histogram

From histogram the region of spinal cord is detected

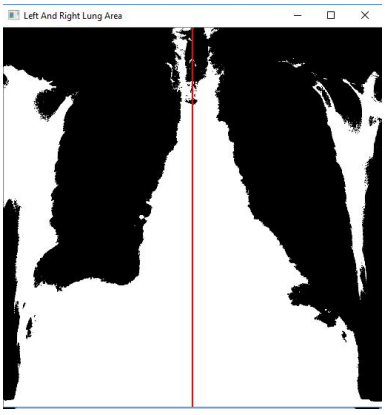


FIGURE 3.6: Spinal Cord

2. Save the contour image co-ordinate using

$$B(A) = A - (A \oplus B) \tag{3.4}$$

where, A is the original image and B is a 3x3 structuring element. *tmp(x,y)* represents the co-ordinate of contor image of B.

FIGURE 3.7: Cantor Image_1

A lot of cantor image is provided by JSRT database. All of them are used to extract region of interest

FIGURE 3.8: Cantor Image_2

3. Save the original image co-ordinate as $org(x, y)$

4. Compare the corresponding $temp(x, y)$ and $org(x, y)$ and calculate total Hit and $Miss$. If $temp(x, y)$ and $org(x, y)$ both are equal with 1 then it called $Hit$and otherwise called $Miss$.

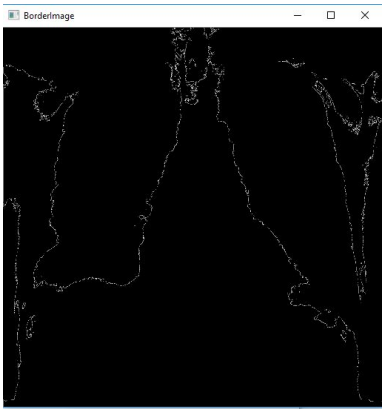


FIGURE 3.9: Selected Lung Border

5. Calculate Hit ratio using

$$HitRatio = \frac{TotalHit}{TotalHit + TotalMiss} \tag{3.5}$$

6. Find the highest Hit ratio in contour image and plot the boundary in original binary image.

FIGURE 3.10: Extracted ROI

This is the segmented lung region.

3.5 Description of Proposed Method

We detect nodule based on template matching. From the observation of JSRT database we understood that the average intensity of a cancerous area and normal area has a great difference. Normally intensity of normal area of a binary image is much lower than a cancerous region. Bone region also shows greater intensity but lower than cancerous region.From this observation we proposed a method that match a region with template that is declared as cancer by doctor.Our method is described below:

- Abnormalities in left and right lung are searched in differently since lungs are divided into two separate images.
- Collect some extracted circular shape which are declared as cancer region by doctor or manually extract some cancer region from X-ray image collected from trusted data set.
- Convert the extracted template 8x8, 16x16, 32x32, 64x64 and 128x128 and search the whole ROI.
- Mark the region where the template matches for a given threshold.
- Continue the procedure for every template image.
- If a region is marked multiple time declare the region as cancer.

3.6 Summary

This method detects the cancer region correctly but some other regions also detected as cancer. This percentage is not so high. In other word this method has a moderate false positive rate. Most of the false positive region are located in the border of the lung. If we want to minimize the false positive rate some cancerous region may be unidentified. ROI is not properly extracted for some image and for those image greater candidate cancerous region are classified as cancer.

Chapter 4

Experiment and Result

4.1 Introduction

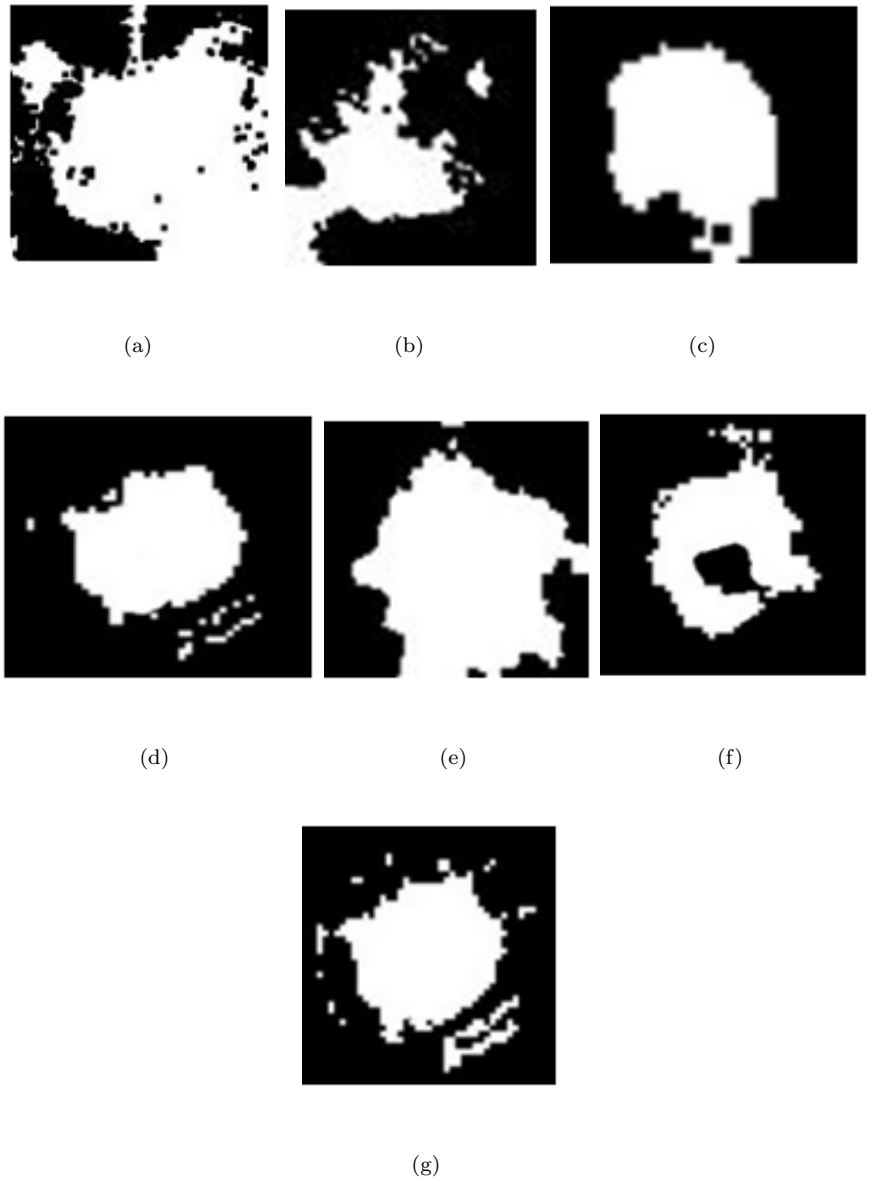
In this chapter we discussed the performance of the proposed method. Proposed method is applied on 154 images collected from JSRT dataset. ***Precision*** and ***Recall*** are used to evaluate performance of this method.Different threshold value is applied to detect cancer in lung X-ray image.

4.2 Environmental Setup for Experiment

We use JSRT dataset in our experiment.All the X-ray image in JSRT are used in our experiment. We Extracet 39 lung cancer template from collected image those are detected as cancer by doctor. All the experiments are catagorised by male and female patient.Here experimental results are shown for cancer stage 1-2 in together and cancer stage 3-4 in together.Here 22 images are used to find cancer of stage 1-2 and 15 images are examined to find cancer of stage 3-4. For each case performance is measured using identified measure.

Our method is implemented in Python and experiments were performed in Windows environment (Windows 10), on a core-i3 intel processor which operates at 2.2GHz with 4 GB of memory.

We use 39 templates and all of them are extracted from JSRT dataset. All the templates are detected as abnormalities by doctor. Only few of them are given here,



(a) (b) (c)

(d) (e) (f)

(g)

FIGURE 4.1: Extracted Template Images

Before going to performance analysis we will show some experimental output of our proposed method. Some X-ray images are given below as output.

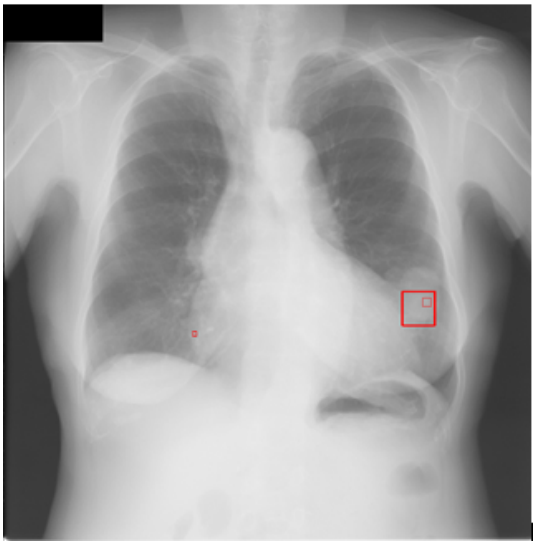


FIGURE 4.2: Output_1

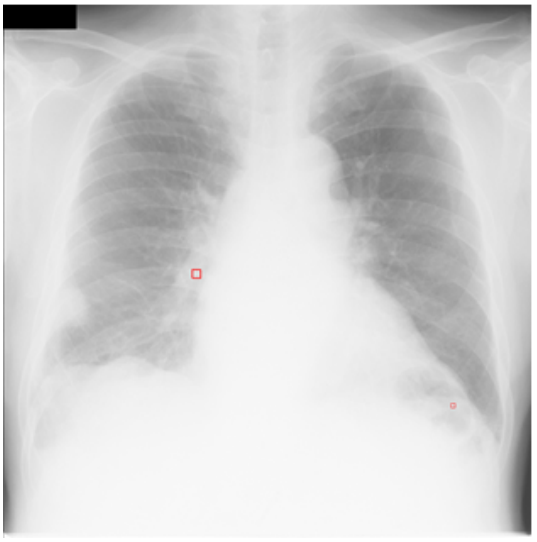


FIGURE 4.3: Output_2

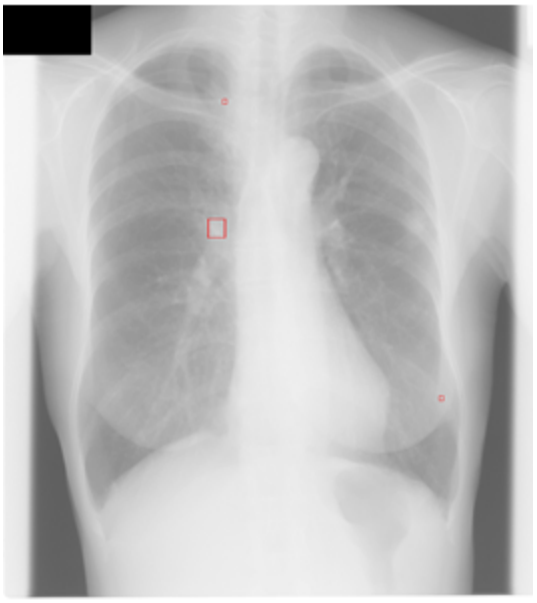


FIGURE 4.4: Output_3

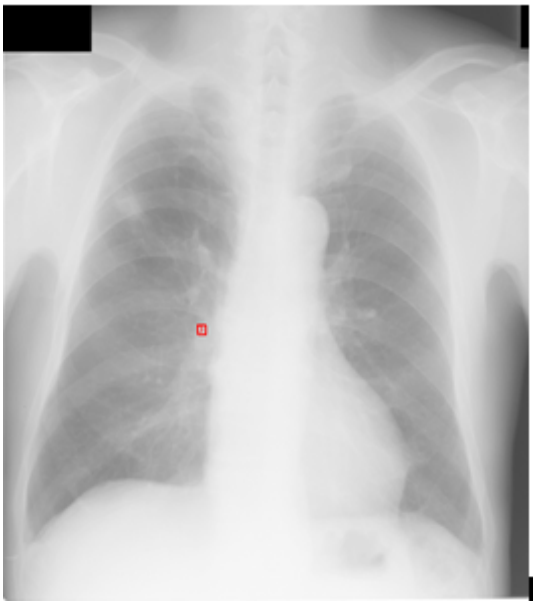


FIGURE 4.5: Output_4

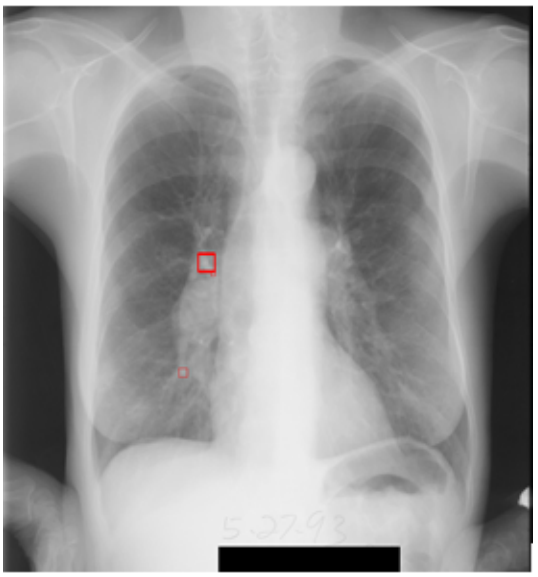


FIGURE 4.6: Output_5

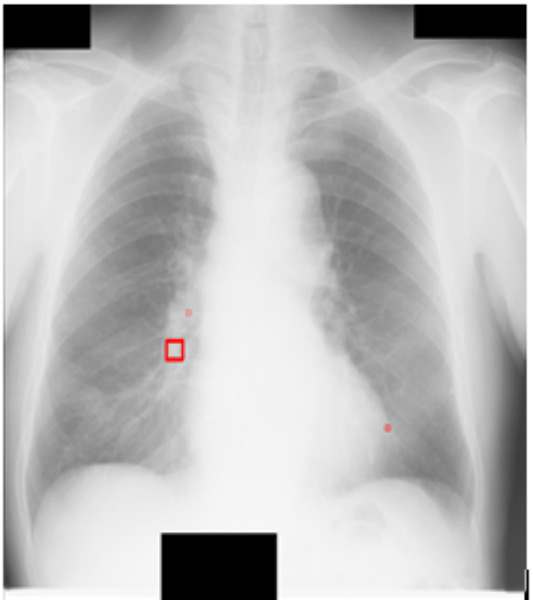


FIGURE 4.7: Output_6

4.3 Performance Analysis

We examined performance in some categories.We used different threshold values to measure the number of correctly detected and falsely detected cancer tissue. Hence we calculated Accuracy and Recall. We measured all of these parameters for both male and female patient.

4.3.1 Threshold

Threshold is a predefined value that is used to classify a region abnormal or not. Some features are calculate for a region and a template and then difference of them is compared with threshold. Threshold value is like a binary classifier. If the difference is lower than threshold then the region is detected as cancer and otherwise the region is classified as non-cancerous region.

As first the tables made from result of experiment is given below. All the graphs given below are constructed form these tables.

Table 4.1: Threshold 0.75

	Male (Total 13)			Female (Total 9)			Total (Total 22)		
Stage 1-2 (Total 22 images)	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected
	49	21	28	34	15	19	83	36	47
	Male (Total 6)			Female (Total 9)			Total (Total 15)		
Stage 3-4 (Total 15 images)	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected
	17	7	10	36	10	26	53	17	36

Table 4.2: Threshold 0.80

	Male (Total 13)			Female (Total 9)			Total (Total 22)		
Stage 1-2 (Total 22 images)	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected
	9	7	2	13	7	6	22	14	8
	Male (Total 6)			Female (Total 9)			Total (Total 15)		
Stage 3-4 (Total 15 images)	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected
	30	17	13	17	11	6	47	28	19

Table 4.3: Threshold 0.85

	Male (Total 13)			Female (Total 9)			Total (Total 22)		
Stage 1-2 (Total 22 images)	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected
	15	13	2	8	7	1	23	20	3
	Male (Total 6)			Female (Total 9)			Total (Total 15)		
Stage 3-4 (Total 15 images)	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected	Total number of abnormalities	correctly detected	Falsely detected
	3	3	0	4	3	1	7	6	1

Table 4.4: Precision vs Recall

Thr	Male						Female							
esho ld	Recall in %			Precision in %			Recall in %			Precision in %			Recall in	
	Stage 3-4	Stage 1-2	Total	Stage 3-4	Stage 1-2	Total	Stage 3-4	Stage 1-2	Total	Stage 3-4	Stage 1-2	Total	Stage 3-4	Stage 1-2
0.75	87.5	91.3	89.4	41.1	42.8	42.0	90.9	88.2	89.5	38.4	44.1	41	89.4	90
0.80	87.5	73.9	80.7	77.7	56.6	67.2	63.6	64.7	4.1	53.8	64.7	59	73.6	70
0.85	37.5	56.5	47.0	100	76.4	88.2	27.2	41.1	34.2	75	87.5	81	31.5	50

Some graph is given below depends on different threshold and different stage of cancer.0.75, 0.80 and 0.85 are chosen for our experiment with respect to stage 1-2 and stage 3-4.

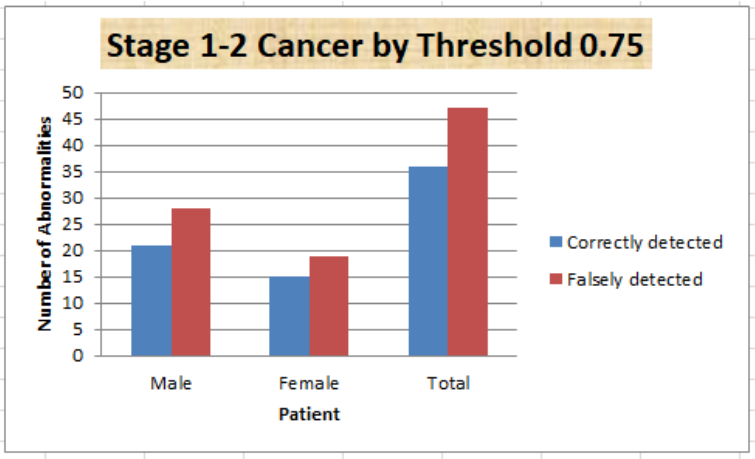


FIGURE 4.8: Stage 1-2 Cancer by Threshold 0.75

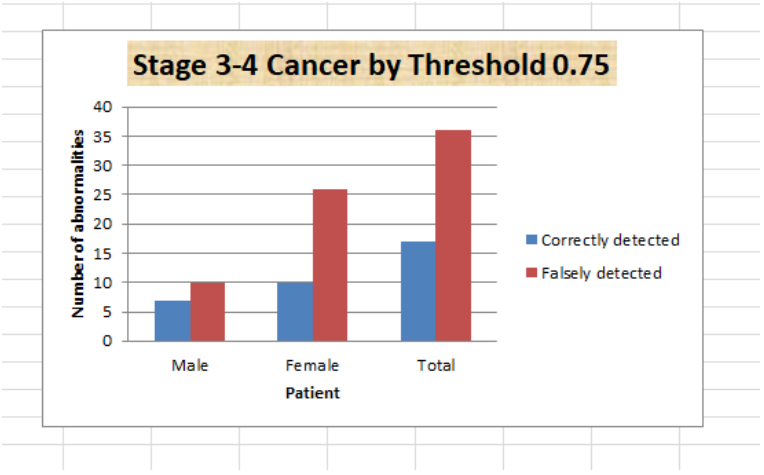


FIGURE 4.9: Stage 3-4 Cancer by Threshold 0.75

These are the graphs of abnormality detection constructed from our methodś resulted by threshold 0.75 where the actual abnormalities are shown in below.

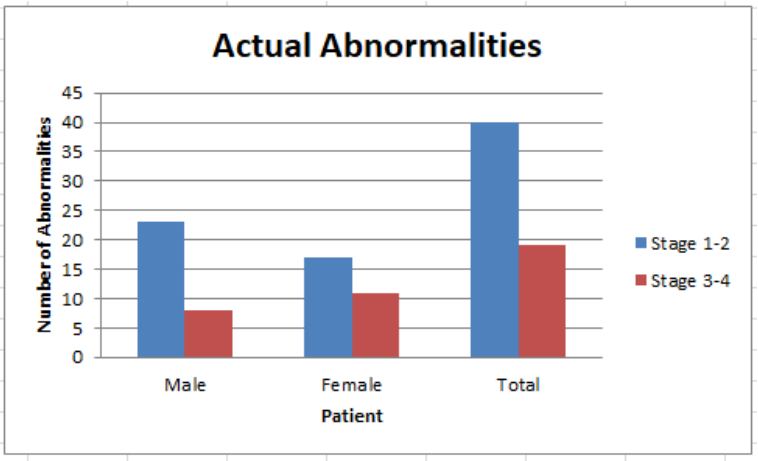


FIGURE 4.10: Actual Abnormalities

By threshold 0.75 most of the abnormalities are detected with slightly higher false positive rate. In stage 1-2, total number of actual abnormalities is 19 and 17 of them are correctly detected. In stage 3-4, 36 abnormalities are correctly detected from 40 actual abnormalities. As most of the abnormalities are detected by threshold 0.75, this shows a greater ***Recall*** as ***Recall*** refers the completeness. Graph of ***Recall*** by threshold 0.75 is attached here.

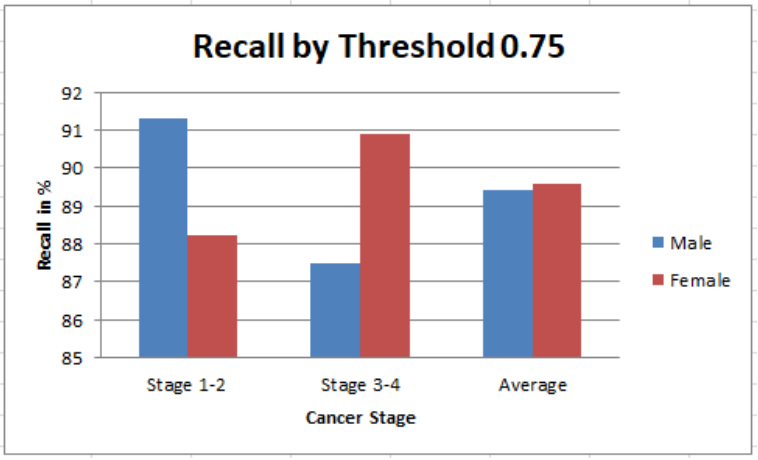


FIGURE 4.11: Recall by Threshold 0.75

Graph shows that average 89% ***Recall*** for both male and female patient. This is a great result for completeness. If any abnormality is actually have our method will give the guarantee to detect it almost 90% cases if the experiment is measured by threshold o.75.

As ***Recall*** and ***Precision*** are mostly used measures for performance evolution of any proposed method, we also measured precision of our proposed method. As precision refers the exactness it is very sensitive of false positive. As threshold o.75 gives much attention on completeness it produces a good number of false positive. So precision by threshold 0.75 is not good enough. Graph of Precision by threshold 0.75 is included here.

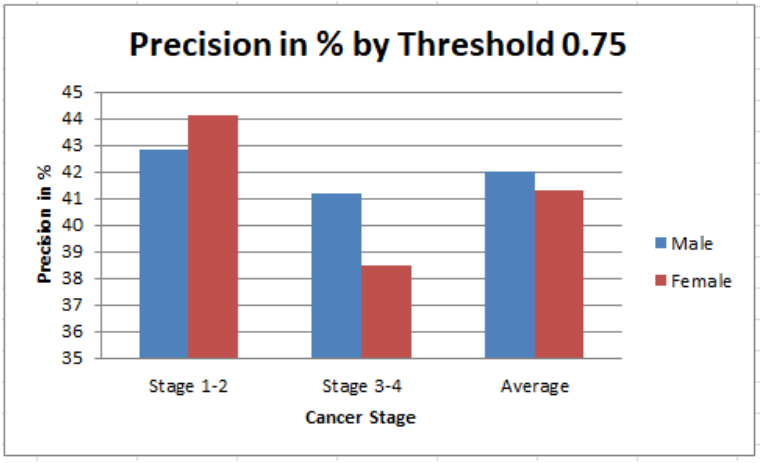


FIGURE 4.12: Precision by Threshold 0.75

Graph shows that average precision is about 42% that means 42% of abnormalities are exact abnormalities from candidate solution. Precision is more important than recall when an abnormality should have been correctly detected.

Similar graphs are also included here by threshold 0.80 and analysed also.

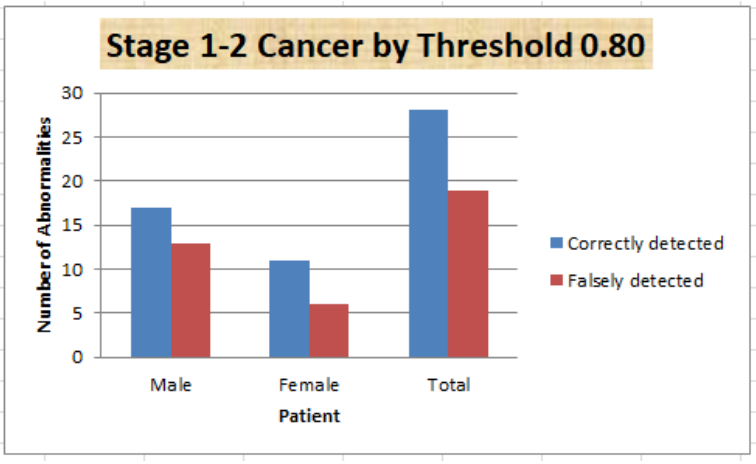


FIGURE 4.13: Stage 1-2 Cancer by Threshold 0.80

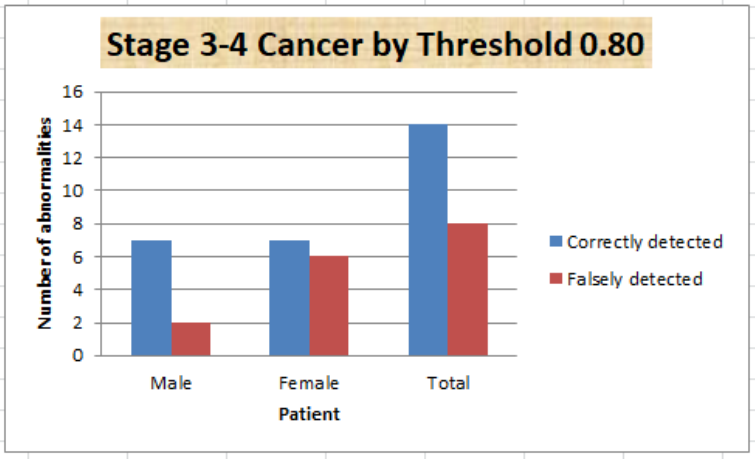


FIGURE 4.14: Stage 3-4 Cancer by Threshold 0.80

From above figures we can see that these figures are different from figures by threshold 0.75 in case of correctly detected rate. In both figure number of abnormalities those are correctly detected in both stage 1-2 and stage 3-4 are higher than falsely detected in those stage. This indicates that precision is increased. But Number of correctly detected abnormalities is lower than previous experiment by

threshold 0.75. This indicates that recall is decreased. Both graph of recall and precision is given below.

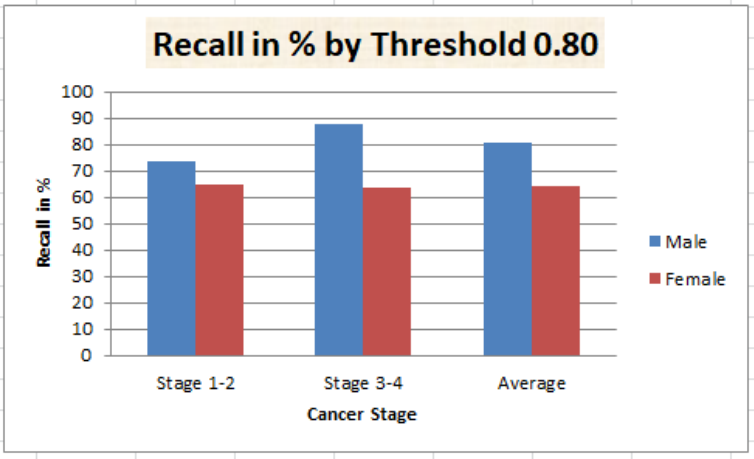


FIGURE 4.15: Recall by Threshold 0.80

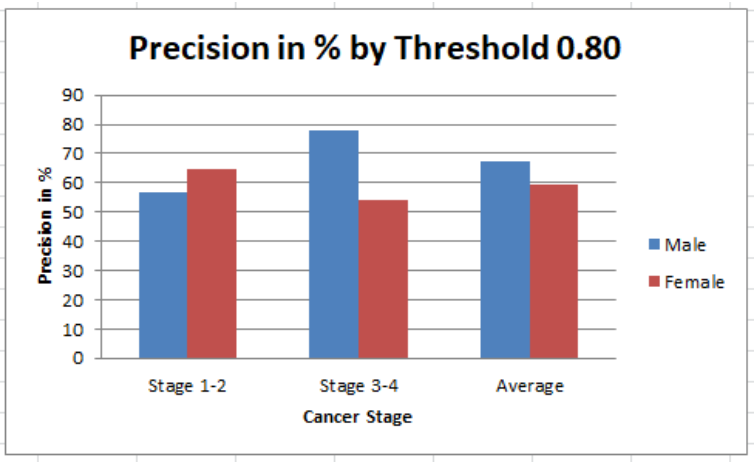


FIGURE 4.16: Precision by Threshold 0.80

From above two figures we can see the recall and precision of experimental result of our proposed method by threshold 0.80. In both case male patients have greater result. Recall of male patient is 80% which is pretty good. For male patient Precision is also excellent and is close to 68%.

For female those figures show almost 65% recall and 60% precision. So for threshold .80 male patients have excellent recall and precision but female patient has slightly lower recall and precision.

However, threshold .80 shows pretty good result with respect to precision and recall. As both of them are needed by good percentage, threshold 0.80 balanced them both.

Experiment by threshold 0.85 produced slightly different output from threshold o.75 and 0.80

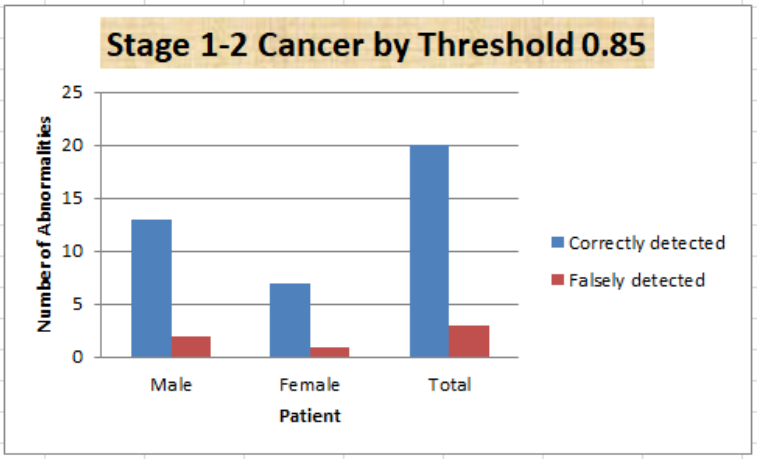


FIGURE 4.17: Stage 1-2 Cancer by Threshold 0.85

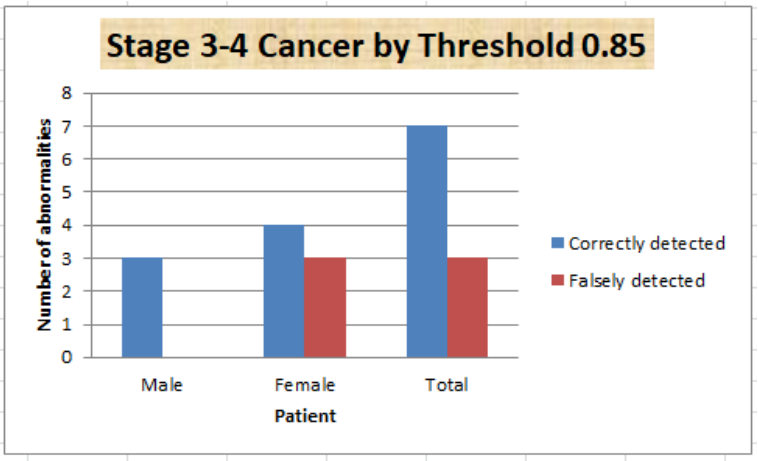


FIGURE 4.18: Stage 3-4 Cancer by Threshold 0.85

From above figure, we can see that number of abnormalities correctly detected by our proposed method is more higher than falsely detected. In both cases stage 1-2 and stage 3-4 shows a great difference between correctly detected and falsely abnormalities.

In stage 1-2, for male patient 13 abnormalities are correctly detected where only 2 abnormalities are falsely detected. For female patient only one abnormality is detected as falsely where 7 abnormalities is correctly detected.

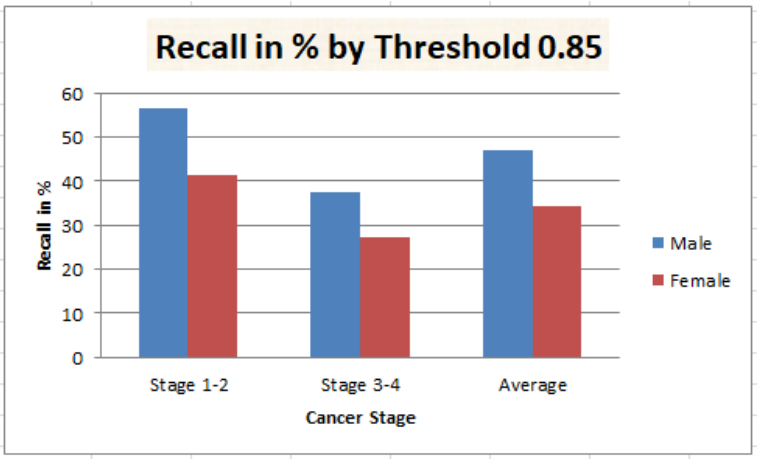


FIGURE 4.19: Recall by Threshold 0.85

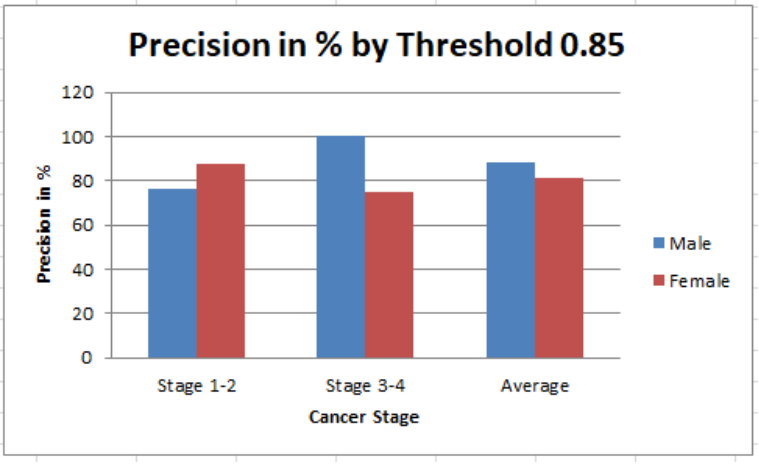


FIGURE 4.20: Precision by Threshold 0.85

In stage 3-4, 3 abnormalities are detected without any false positive for male. For female, false positive is only one where 3 abnormalities is correctly detected.

A great news is that most of the detected region is cancerous but by 0.85 threshold our method detected only few abnormalities where the number of actual abnormality is very high.

So performance by recall and precision is opposite of one another. For threshold 0.75 recall is excellent but precision is not so good and for threshold 0.85 precision is very good but recall is poor. Threshold 0.80 gives moderate precision and recall. Performance graphs by recall and precision for both male and female are included below.

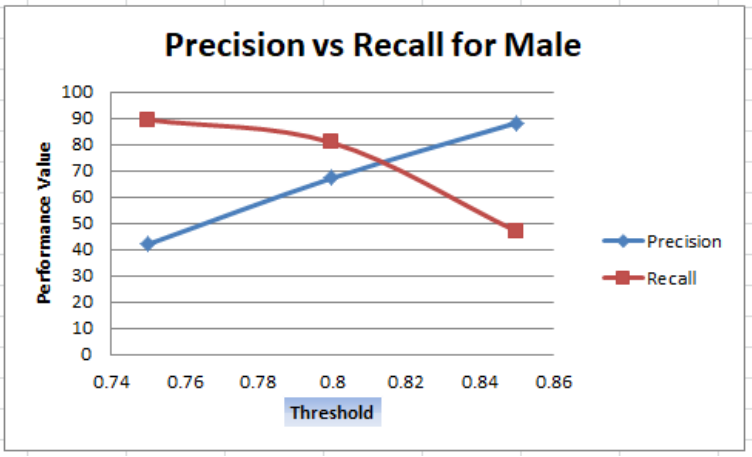


FIGURE 4.21: Precision vs Recall for Male

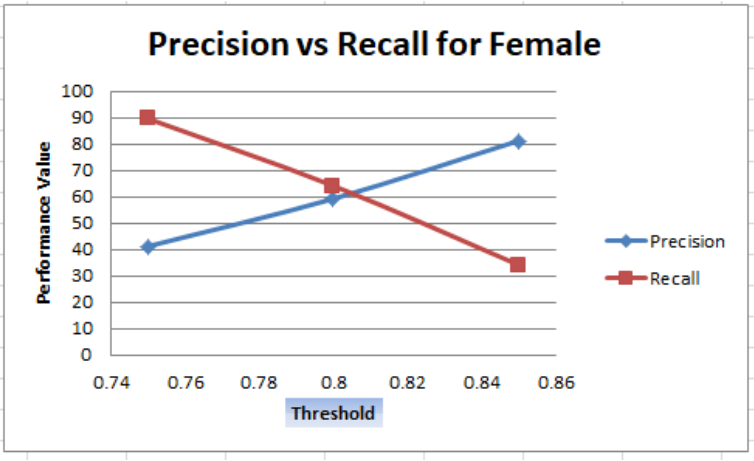


FIGURE 4.22: Precision vs Recall for Female

Both of the graphs show that similar characteristics. Small difference is shown here. In both case recalls by threshold 0.75 is 90% but by threshold 0.80 recall for male and female are nearly 80% and 65% respectively. Recall for female is also lower than recall for male by threshold 0.85 also. Here recall for male and female are nearly 50% and 35%. So, Recall for male is relatively better than recall for female.

In case of precision both graphs have great similarity. Both gives 40% precision for threshold 0.75 but in case of threshold 0.85 precision for male is nearly 90% but precision for female is close to 80%. Both are pretty good. For threshold 0.80 dissimilarity is also here in 10% margin. Here precision for male is close to 70% but precision for female is 60%. So, precision for male is relatively better from precision for female like recall.

Now we will see the overall precision vs recall graph and summarize the experimental result.

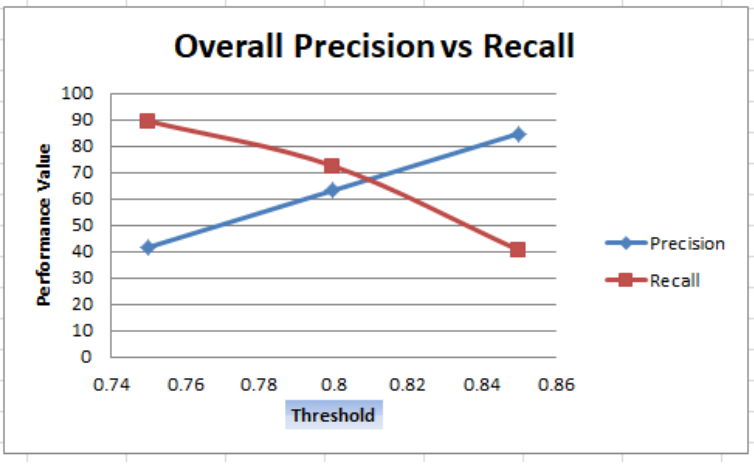


FIGURE 4.23: Overall Precision vs Recall

FIGURE 4.16 shows that the overall precision and recall of our proposed system. For threshold 0.75, 0.80 and 0.85 precision and recalls are respectively 37.73%, 61.61%,86.34 and 89.73%, 71.84%, 40.79%.This graph also shows that recall is higher in lower threshold and lower in higher threshold. For precision the relationship is exactly opposite. Precision is lower in lower threshold and higher in higher threshold.

Our initial goal was the reduction of false positive rate. We achieved recall 90% with 2.24 FPs/image and 74% with only 0.73 FPs/image. These are shown in graph.

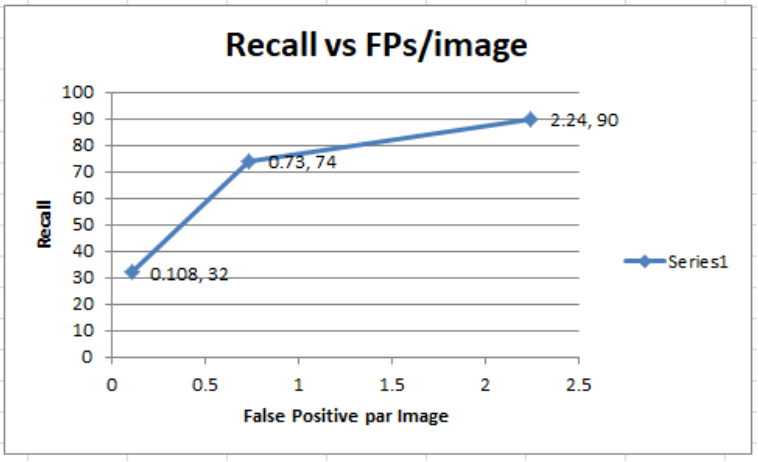


FIGURE 4.24: False Positive w.r.t. Recall

A comparison table is given below False positive rate w.r.t recall.

TABLE 4.5: Performance Comparisons of CAD Schemes in the JSRT Database

Methods	Recall	**FPs/image**	**Database**
Wei et al. (2000) [12]	80%	5.4(1333/247)	JSRT
Coppini et al. (2003) [14]	60%	4.3(1333/247)	JSRT ,Nodule
Schilham et al. [15]	51%	2.0(308/154)	JSRT, Nodule
	67%	4.0(616/154)	
Hardy et al. [13]	80%	5.0(700/140)	JSRT, Nodule
	63%	2.0(280/140)	
Chen et al. [16-17]	71.4%	5.0(1235/247)	JSRT, Nodule
	85%	5.0(1165/233)	Non-nodule
Li et al. [9]	94%	4.6(1131/247)	JSRT
	84%	2.0(480/247)	
Proposed method	90%	2.24(83/37)	JSRT, Nodule
	74%	0.73(27/37)	Non-nodule

The table reflects an excellent performance of our proposed method. 90% recall with 2.24 FPs/image is pretty good.

4.4 Summary

More than 150 images from JSRT dataset is used for experiment with 39 extracted cancerous template. Three different thresholds are used for each image. Lower the threshold higher the recall and higher the threshold lower the recall is shown. For precision proportional relationship is found. Lower the threshold lower the precision and higher the threshold higher the precision. For a given threshold both recall and precision are not excellent. So completeness and exactness are not found for same threshold. Moderate completeness and exactness is the result of threshold 0.80.

Chapter 5

Conclusions

5.1 Summary of Research

Lung cancer is detected from cancer by our proposed method and all the images are collected from JSRT dataset. 39 cancerous regions from collected images are extracted as template. Experimental result is pretty good. Different threshold may be applied to diagnosis cancer. Lower threshold gives more number of abnormalities and so most of the cancer regions are detected that higher the recall but lower the precision. For higher threshold less number of abnormalities are detected but almost all of them are correctly detected as cancer. So precision is very high for higher threshold but recall is not so high because many abnormalities are not detected. For moderate precision and recall our proposed method is applied with 0.80 threshold.

5.2 Future Work

We want to upgrade our work for better result in future. Our main goal was to reduce the false positive rate. For this we have to apply our method with greater threshold that reduces recall. We want excellent recall and precision both. This is the main challenge in future.

Bibliography

[1] E. M. Dey and H. M. Muktadir, "Chest X-Ray Analysis to Detect Mass Tissue in Lung", 3rd International Conference on Informatics, Electronics Vision 2014,Dhaka

[2] T. Satya Savithri and S. K. Chaya Devi; "Nodule Detection from Posterior and Anterior Chest Radiographs with Different Methods"; 2016 Future Technologies Conference (FTC),6-7 Dec;San Francisco, CA, USA;

[3] Xiaowu Sun, Lizhen Liu, Hanshi Wang and Wei Song & Jingli Lu, "Image classification via support vector machine"; 2015 4th International Conference on Computer Science and Network Technology (ICCSNT); DOI: 10.1109/ICCSNT.2015.7490795

[4] A. Fischer and C. Igel, "An Introduction to Restricted Boltzmann Machines" L. Alvarez et al. (Eds.): CIARP 2012, LNCS 7441, pp. 1436, 2012.

[5] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms." Automatica 11.285-296(1975):23-27.

[6] Matei Mancas, Matei Mancas, Bernard Gosselin, Bernard Gosselin, Benoit Macq and Benoit Macq, "Segmentation using a region-growing thresholding", Proc. SPIE 5672, Image Processing: Algorithms and Systems IV, (1 March 2005); DOI: 10.1117/12.587995

[7] S. A. Hojjatoleslami and J. Kittler, "Region Growing: A New Approach" IEEE Transactions on Image Processing (Volume: 7 , Issue: 7) ,pp 1079 - 1084, Jul 1998; DOI:10.1109/83.701170

[8] Blob Detection by Scale-Space Filtering. by James Fishbaugh

[9] C. Li , G. Zhu , X. Wu and Y. Wang, "False-Positive Reduction on Lung Nodules Detection" in Chest Radiographs by Ensemble of Convolutional Neural Networks; IEEE Access (Volume: 6); pp 16060 - 16067;19 March 2018 ; DOI: 10.1109/ACCESS.2018.2817023

[10] T. Guo , J. Dong , H. Li and Y. GaoSimple, "convolutional neural network on image classification" IEEE 2nd International Conference on Big Data Analysis (ICBDA),Beijing, China;10-12 March 2017; DOI: 10.1109/ICBDA.2017.8078730

[11] J. P. Egan, G. Z. Greenberg and A. I. Schulman, "Operating characteristics signal detectability and the method of free response", J. Acoust. Soc. Amer., vol. 33, no. 8, pp. 993-1007, 1961.

[12] J. Wei et al., "Optimal image feature set for detecting lung nodules on chest X-ray images" in Computer Assisted Radiology and Surgery, Berlin, Germany:Springer, pp. 706-711, 2002.

[13] R. C. Hardie, S. K. Rogers, T. Wilson and A. Rogers, "Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs", Med. Image Anal., vol. 12, no. 3, pp. 240-258, 2008.

[14] G. Coppini, S. Diciotti, M. Falchini, N. Villari and G. Valli, "Neural networks for computer-aided diagnosis: Detection of lung nodules in chest radiograms", IEEE Trans. Inf. Technol. Biomed., vol. 7, no. 4, pp. 344-357, Dec. 2003

[15] A. M. R. Schilham, B. van Ginneken and M. Loog, "A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database", Med. Image Anal., vol. 10, no. 2, pp. 247-258, 2006.

[16] S. Chen, K. Suzuki and H. MacMahon, "Development and evaluation of a computer-aided diagnostic scheme for lung nodule detection in chest radiographs by means of two-stage nodule enhancement with support vector classification", Med. Phys., vol. 38, no. 4, pp. 1844-1858, 2011.

[17] S. Chen and K. Suzuki, "Computerized detection of lung nodules by means of virtual dual-energy radiography", IEEE Trans. Biomed. Eng., vol. 60, no. 2, pp. 369-378, Feb. 2013.

List of Notations

TotalHitt - Number of total hit

TotalMiss - Number of total miss