ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTING

Gray level Co-occurrence Matrix for Miliary Tuberculosis Pathologies

A Thesis Submitted in the Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Systems

By

Yosef Taddesse

January 2017

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTING

Gray level Co-occurrence Matrix for Miliary Tuberculosis Pathologies

A Thesis Submitted in the Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Systems

By

Yosef Taddesse

Advisor

Prof. Dr.-Ing. Achim Ibenthal

Co_ Advisor

MSc. Gedamu Alemu

Declaration

I, the undersigned, declare that this thesis is my original work and has not been presented for a degree in any other work and that all source of materials used for the thesis has been duly acknowledged.

Declared by:

Name: Yosef Taddesse

Signature: _____

Date: _____

This thesis has been submitted for examination with my approval as university advisor

Confirmed by advisor:

Name: Prof. Dr.-Ing. Achim Ibenthal

Signature: _____

Date: _____

Co-advisor:

Name: Mr. Gedamu Alemu (MSc)

Signature: _____

Date: _____

Place and date of submission: Adama, Jan 2017

Name and signature of Members of the Examining Board

Name	Title	Signature	Date
_____	Advisor	_____	_____
_____	Co- Advisor	_____	_____
_____	External Examiner	_____	_____
_____	Internal Examiner	_____	_____
_____	Chairperson	_____	_____
_____	Department Head	_____	_____
_____	School Dean	_____	_____

Acknowledgement

First of all, I would like to thank Almighty GOD who makes things possible and helps me in every aspect throughout my life.

Next, I wish to express my deepest gratitude to my advisor Prof. Dr.-Ing. Achim Ibenthal, co-advisor Mr. Gedamu Alemu (MSC) for their continued support, guidance and encouragement, without them would not have been possible to succeed in my research. The financial support from the Adama science and Technology University (ASTU) and the school of electrical engineering and computing in the form of a graduate/research assistantship are gratefully acknowledged. Special thanks to all my friends for their enthusiastic and generous support through my research study.

Above all, I wish to express my deep appreciation to my family, and friends who with their love and encouragement have made all this possible.

Table of Contents

List of Figures

List of Tables

Abbreviations

Term	Explanation
AAM	Active Appearance Model
AuC	Area under RoC Curve
BAC	Breast Arterial Calcifications
CT	Computer Tomography
CPM	Competition Performance Metric
ConvNet	Convolutional Network
CXR	Chest X-Ray
DCXR	Digital Chest X-Ray
DIP	Digital Image Processing
DNA	Deoxyribonucleic Acid
DSP	Digital Signal Processing
EMR	Electronic Medical Record
ETB	Extra Pulmonary TB
HIV	Human Immunodeficiency Virus
HoG	Histogram of Oriented Gradients
KLD	Kullback–Leibler Divergence
kNN	k-Nearest Neighbor

LBP	Local Binary Pattern
LDC	Linear Discriminant Classifier
LoG	Laplacian of Gaussian
MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
SVM	Support Vector Machine
PTB	Pulmonary TB
ROC	Receiver Operating Characteristic
TB	Tuberculosis
TN	True Negative
TP	True Positive
RoI	Region of Interest
WHO	World Health Organization

Medical Terms

Bronchioles — branches of the bronchi (→*bronchus*).

Bronchus — (pl. bronchi) major air passages of the lung.

Clinical Correlation — diagnosis by multiple clinical means.

Coccidioidomycosis — fungus disease affecting the lung and other tissue.

Differential Diagnosis —differentiating pathologies of similar →*manifestation*.

Fibrous Tissue — tissue composed of bundles of white fibers of which are rows of connective tissue cells.

Florid —highly developed.

Granuloma — a collection of immune cells forming tissue around substances that cannot be eliminated by the body's immune system.

Granulomatous —characterized by →*granulomas*.

Hematogenous — distributed via blood flow.

Hila — contains major bronchi (→*bronchus*) and the →*pulmonary* arteries and veins of the lung.

Hilarandmediastinal Lymph Nodes — lymph nodes close to the →*hila* and the root of the lung.

Histoplasmosis — fungus disease from bird or bat excrements.

Lesion — region in an organ or tissue that has been damaged by injury or disease, such as a wound, ulcer, abscess or tumor.

Malignant — very virulent or infectious; in case of tumor: tending to invade normal tissue or to recur after removal;

Manifestation — here: pattern or manifestation of a given →*pathology* on a diagnosis, e.g. on an x-ray image.

Mediastinum — median partition of the thoracic (→*thorax*) cavity.

Melioidosis — bacterial →*pulmonary* disease transmitted by rodents.

Miliary — nodule-like → *lesions* with the appearance of millet seeds.

Necrosis — cell death in an organ or tissue.

Parenchymal — referring to the functional elements of an organ or tissue rather than the structural parts.

Pathology — here: combined characteristics and patterns of a distinct disease.

Pulmonary — related to the lung.

Silicosis — disease caused by inhalation of crystal silica by e.g. miners.

Tubercle — nodule caused by →*tuberculosis*.

Tubercular, **Tuberculous** — relating to, or affected by →*tuberculosis*.

Tuberculosis — infectious bacterial disease characterized by nodules or cavities in tissues, predominantly but not exclusively in the lungs.

Thorax — chest.

Technical Terms

Appearance — combination of →*shape*- and →*texture* information.

Computer Vision — a computer or DSP application emulating human vision, including machine learning and logical deduction based on visual stimuli.

Descriptor — a set of numbers encoding interesting information about an image object, an area or the whole image, also called a feature vector. Ideally, the information is invariant against certain types of geometrical and/or gray-level transformation.

Digital Image Processing — any form of signal processing which' input and output are digital images and which may retrieve additional features including the recognition of individual objects.

Feature — here: interesting part of an image such as corner, blob, edge or lines.

Feature Vector →*descriptor*.

Parametric Model — use different parameters to describe an abstraction of reality, where the used parameters are derived from the object. The model quality is determined by the similarity between the model and the real object. Modifying the model's parameters causes the model to update and change its appearance according to the modified parameters.

Principal Component Analysis — (abbrev. PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a (smaller) number of uncorrelated variables called principal components.

Shape — here: a set of characteristics of an object. These characteristics are hereafter referred to as landmark points. A particular trait all Landmark points must have in common is that they are invariant with respect to translation, rotation, and scaling.

Texture — here: pixel intensities of the object comprised in the boundaries of the landmark points.

Abstract

Tuberculosis (TB) is currently one of the most dangerous diseases worldwide as announced by the World Health Organization. Sputum tests or thorax X-rays are used to identify infected patients. In some regions of Africa, the infection rate is up to 90% while the density of radiologists is too low to screen patients' area-wide while diagnosing tuberculosis still remains a challenge. When left undiagnosed and thus untreated, mortality rates of patients with tuberculosis are high. In an effort to reduce the burden of the disease, this paper presents our automated approach for detecting miliary tuberculosis in conventional poster anterior chest radiographs. We first preprocess the lung image using different preprocessing techniques and segment the image using a thresholding algorithm in order to clearly identify the texture feature of miliary tuberculosis. For this lung region, we compute a set of statistical texture features using Gray Label Co-occurrence Matrix (GLCM) algorithms, which enable the X-rays to be diagnosis as normal or miliary TB using four statistical features. We measure the performance of our system datasets: the set collected by download from a various website like https://ceb.nlm.nih.gov/repositories/tuberculosischest- x-ray-image-data-sets/ and Japanese Society of Radiological Technology (JSRT) built a database. The proposed computer-aided diagnostic system for miliary TB screening achieves a performance that approaches the performance of human experts. We achieve a contrast under the confusion matrix of 20 image dataset from these 10 miliary TB chest x-ray image and 10 normal chest x-ray image, from this data set we get accuracy 95% and sensitivity 100% and specificity 100% according to this result this work reduce false negative result and increase true negative result and also decrease false positive result and increase true positive result.

Keywords: Gray level Co-occurrence matrix, miliary tuberculosis, a texture feature, confusion matrix.

CHAPTER ONE

1. Introduction

Tuberculosis is an infectious disease caused by bacillus mycobacterium that can spread through the lymph nodes and bloodstream to any organ in the body [1]. World health organization collect the data from 205 countries and territories, which account for more than 99% of the world's population, this global TB report documents advances in prevention, diagnosis, and treatment of the disease. It also identifies areas where efforts can be strengthened. According to this data in 2014, TB killed 1.5 million people (1.1 million HIV-negative and 0.4 million HIV-positive). The toll comprised 890 000 men, 480 000 women and 140 000 children [2].

The Bacillus Mycobacterium tuberculosis is the cause of TB among humans. According to the global tuberculosis report of the WHO approximately 1.3 million people died, out of 8.6 million reported with TB, in 2012. Most of TB deaths can be prevented if they are detected at the early stage. According to the WHO, 40% of TB cases are not detected globally. Global Tuberculosis Report 2013 reports that "we have to find new measures to eradicate TB to meet the millennium development goals". But case detection is one of the difficulties faced in eradicating this disease [3].

1.1. Tuberculosis in Ethiopia

In [4] the following are stated as "Ethiopia ranks ninth among the world's 22 high-burden countries with TB and one of the top five in Africa, with regard to the prevalence of TB. According to WHO 2010 Global TB Report, the country had an estimated 44,398 TB cases in 2009, with an estimated incidence and prevalence rate of 300 and 470 cases per 100,000 populations respectively". But According to 2011 WHO global TB report, the incidence, and prevalence of TB for the year of 2010 burden decreases to 261 and 394 cases per 100,000 populations respectively. Similarly, case detection was 50% for all forms of TB. Among all new TB cases 30% were smeared positive, 35% smear negative, 34% extrapulmonary and,<1% smear unknown cases. In addition among re-treatment cases 2,259 (64%) were relapse case, treatment after failure 381(11%), treatment after default 478(13%) and 56,040 had both TB and HIV co-infection. According to the Ministry of

Health hospital statistics data, TB is one of the leading cause of morbidity and the fourth cause of hospital admission and the second cause of hospital death in Ethiopia [4].

1.2. Morphology of Tuberculosis

As [5] describe the morphology of tuberculosis on a radiograph a nodule can produce a tiny dense shadow at the edge of the lung field, often accompanied by a shadow representing lymph nodes in the center of the field. These two shadows represent the primary complex of tuberculosis and such an infection is common in children, who often do not suffer from illness and may develop immunity to TB. In adults, TB usually takes a different form A blood vessel can be eroded and then the patient coughs up blood. This spreading type of the disease was popularly known as consumption, a name that vividly describes TB's destructive progress.

If the patient is treated with drugs or builds up resistance, the bacilli may be sealed up in the lungs, which by then contain much scar tissue. Consequently, the disease can, in its various stages, give rise to a large variety of signs on an X-ray. Acute lung injury has been defined as a syndrome of inflammation and increased pulmonary permeability associated with a constellation of clinical, radiological and physiological abnormalities which cannot be explained by left atrial or pulmonary capillary hypertension [6].

Accounts of tuberculosis can be found as far back as the writings of the ancient Egyptians. TB was the leading cause of death for all age groups in the western world from that period until the early 20th century. Finding an effective cure for tuberculosis represents one of the great advances in 20^{th}-century medicine. Fifty years after Robert Koch, in 1882, discovered the tubercle bacillus; several antimicrobial drugs were discovered that can cure the disease [7]. Sample X-ray image of miliary TB as shown in figure1.1.

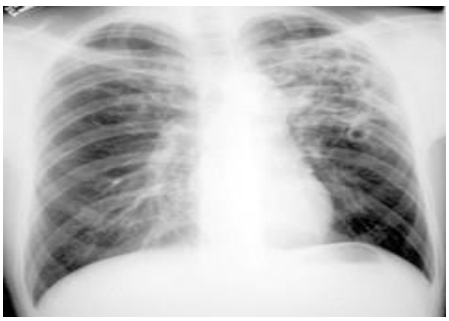


Figure 1.1: Miliary tuberculosis, image source from [5]

1.3. Burden of the Problem

According to [8] state as "TB treatment prevented 49 million deaths globally between 2000 and 2015, but important diagnostic and treatment gaps continue." Even if effective therapy being available Mortality from this disease has remained. For a long time, miliary TB has been considered to be a childhood disease. However, during the last three decades, it is increasingly being recognized in adults as well. Most Schoharie' has several reasons are thought to be responsible for this changing epidemiological trend. These include human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS), ever increasing the list of causes of immunosuppression, such as the use of biological and immunosuppressive drugs for the treatment of various medical disorders, increasing occurrence of organ transplantation, chronic hemodialysis programmer, among others [9]. In [10] the following is stated: "Miliary tuberculosis (TB), is a fatal form of disseminated TB characterized by tiny tubercles evident on gross pathology similar to uncountable millet seeds in size and appearance."

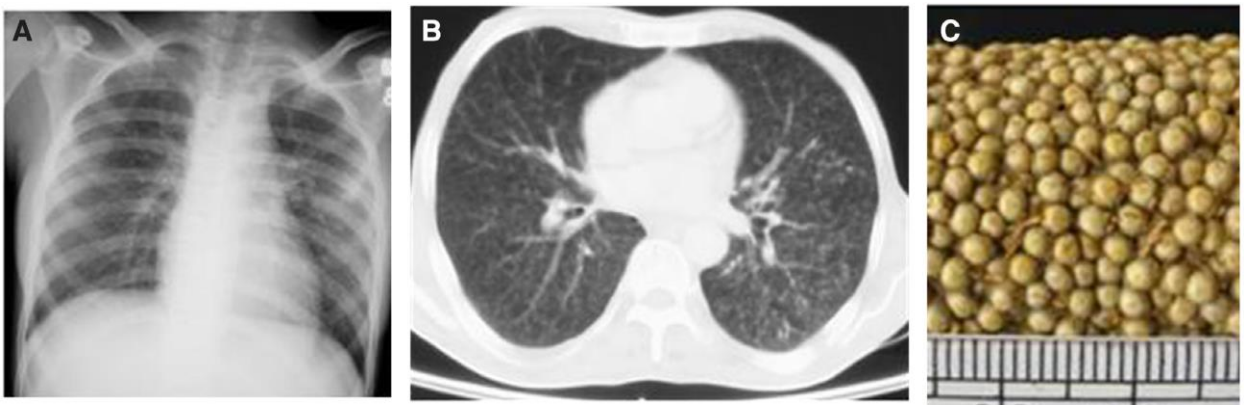


Figure 1.2: Chest radiograph (posteroanterior view) (A) and chest CT (lung window) (B) showing a classical miliary pattern. The nodules (< 2 mm) evident in miliary tuberculosis resemble the grains of pearl millet (Pennisetum typhoid, Baja) (C). Source form [11].

Finally, as we have seen in Figure1.2 disseminated tuberculosis are distribute all over the lung fairly and equally due to this reason the bacterium may spread increasingly. Therefore, treating the patient early will save the life of the patient for this solution identifying the texture future of the miliary Tuberculosis is effective techniques because the texture is a significant feature of an image that has been widely used in medical image analysis.

2. Statement of the Problem

Since miliary tuberculosis is a serious death disease in the world especially in sub-Saharan Africa country like Ethiopia. Survive such type of disease is impossible due to the distribution of mycobacterium all over the lungs, its main site of infection is the lungs because of the bacillus preference for high oxygen environments as we disused in introduction part. While diagnosing tuberculosis still remains a challenge. The pathogenesis of TB is complex, and some of its features are not fully understood yet.

Ethiopia is the second largest country in Africa, with a population of 100 million people from this 85% of the people lives in the rural area. In Ethiopia, a number of professional existing at health care is less and number of patients is rising every year in all regions. The physician's residents are located at the center, therefore, people who have traveled hundreds of kilometers away from their home seeking treatment like Black Lion hospital. When the patient travel to the hospital have no sufficient transportation access so they use another alternative like mule or footpath. These patients don't have the means to take care of themselves in terms of accommodation, food, and support during their stay in Addis. Even the patient face lot of challenge and reach to hospital getting treatment is also difficult because of a shortage of physicians other problem. The diagnosis process of chest x-ray/CT tuberculosis image is done by just looking at the X-rays/CT images by the Doctors/technicians. But in rural areas are difficult to find such a doctor or technicians, therefore, patients will lead to death due to the luck Doctors/technicians. Due to this reason, the automatic detection texture feature by Gray level co-occurrence matrix for miliary tuberculosis from an x-ray is helpful in the rural area and where an expert Radiologist is not always available.

The following question should be answered at the end of this study:

RQ1: what is the current state-of-the-art in machine-based miliary tuberculosis?

RQ2: which texture classification algorithm is best for miliary tuberculosis pathologies?

RQ3: how to decrease false positive (FP) and false negative (FN)?

RQ4: is the GLCM suitable and sufficient for the diagnosis of miliary TB features.

RQ5: what evaluation matrix can be used for the detection algorithm?

3. Objective

The general and specific objectives of this study are described hereunder.

3.1. General Objective

The general objective of this study is identifying the texture feature of miliary tuberculosis pathology using Gray level Co-occurrence Matrix.

3.2. Specific Objective

- Literature Study related to GLCM-based texture feature analysis.
- Collect the test image containing miliary TB and normal chest x-ray image.
- Select and implement the most interesting texture analyses methods.
- Calculate texture feature of each chest x-ray image from the dataset
- Compare the result of the test image and the reference image of the dataset.
- Increase true positive and true negative findings and reduce false positive and false negative results.
- Evaluate the performance based on the result
- Documentation.

4. Methodology

To achieve our objective we have selected appropriate methods and materials. Overall, there are four main processes used throughout the thesis; Pre-processing, segmentation, feature extraction and finally the diagnosis process. MATLAB® R2012b is used in every process made throughout the experiment.

4.1. Literature Review

In order to have a deep understanding of the problem, it is vital to review several kinds of literature that have been conducted in the field so far. For this reason, related literature such as books, articles, proceeding papers and some other sources that are retrieved from the internet have been consulted so as to understand about miliary tuberculosis, image analysis and methods that are important for developing texture classification algorithm.

4.2. Image Analysis

We used different types of image analysis techniques to identifying the texture of miliary tuberculosis. In this work, we identified three analyzing techniques image enhancement, segmentation and feature extraction.

4.2.1. Image Enhancement

Image enhancement involves operations that improve the appearance to a human viewer, or operations to convert an image to a format better suited to machine processing. Image enhancement processes consist of a collection of techniques that seek to improve the visual appearance of an image or to convert the image to a form better suited for analysis by a human or a machine. In an image enhancement system, there is no conscious effort to improve the fidelity of a reproduced image with regard to some ideal form of the image, as is done in image restoration [12]. For enhancing the image we used rolling ball algorithm and image equalization techniques.

4.2.2. Image Segmentation

Image segmentation divides the input image into multiple segments or regions, which show objects or meaningful parts of objects. It segments the image into homogenous regions thus making it easier to analyze images. The success of image analysis depends on the reliability of segmentation [13]. In our work, we used optimal thresholding techniques to segment the chest x-ray images.

4.2.3. Feature Extraction

In this case, it uses the process of acquiring higher level information of an image, such as texture feature. Features contain the relevant information of an image and will be used to identify characteristics of any image. It aims to classify data (patterns) based either on the occurrences of miliary tuberculosis. Much attention will be focused on methods that reduce the number of features by selecting the most interesting ones either through supervised feature selection or unsupervised dimensionality reduction approaches by applying the GLCM algorithm. In this work GLCM used as feature extraction techniques.

4.3. Material and Tools

Here mentioned what material and tools used throughout the research work, Toshiba laptop with RAM 2 and Pentium(R)dual processor 2.3GHz each and 300 hard disk computer. Microsoft word 2013 use for documentation purpose and excel 2013 for good format tables and Microsoft power point2013 for good preparation of the presentation, and we used Edraw Max2016 software for drawing an attractive diagram. Chest x-ray dataset used for the experiment.

The whole system is designed on the MATLAB platform. MATLAB, which stands for matrix laboratory, is a very powerful technical language for mathematical programming. It has a very extensive library of predefined programs or functions designed to help engineers and scientists to

solve their problems in a faster and less painful way. In MATLAB having over a number of toolboxes has made it easy for different subjects of study. A toolbox of a particular subject contains mainly the functions or programs required to solve problems related to the subject. The present day professional version of MATLAB is having graphical and GUI features. Writing programs in MATLAB is much easier compared to other programming languages like FORTRAN, C, C++ or Java. This is because when writing a program in MATLAB, There is no worry about the declaration of variables, types, sizes and memory requirements, which are the main sources of troubleshooters in other programming languages [14], [23].

4.4. Evaluation Methods

An important aspect of our work is the confusion matrix because confusion matrix is the convenient way of summarizing the errors and forms the base for many evaluation measures. The confusion is labeled as 2 by 2 relation that is positive and negative. For two class labels of being "positive" or "negative" for some property. For example, a patient can have miliary TB (he or she is "positive") but have a negative classification result. The cells of the confusion matrix then get a specific meaning: on the diagonal are the correctly classified true positive (TP) and true negative (TN), the other two cells count the false positive (FP) and false negative (FN).

5. Scope and Limitation

The scope of the study is limited to Gray level Co-occurrence Matrix (GLCM) for miliary tuberculosis pathologies. This work concentrate on understanding image preprocessing, segmentation, and future extraction. The study did not include other types of tuberculosis like Primary Tuberculosis, Pneumonia, Tuberculosis Pleurisy, Laryngeal TB, and Extrapulmonary Tuberculosis. And our work is not focused on the treatment of the patient. In the experiment process, we concentrated on chest x-ray image only other image like CT scan and MRI images will not include through our experiment. The system has no graphical user interface.

6. Significance of the Study

Identifying problems and attempting to offer solutions is a significant step in any sort of investigation. The study will create awareness for researchers a person who wants to do in this area as we know this area is new for Ethiopian student. Using the recommended solutions, hospitals and clinic at remote areas will also be initiated to use computer aid miliary tuberculosis

identification. The findings of this study will benefit health care persons can easily identify and diagnosis of miliary tuberculosis diseases.

Image analysis can be used to enhance problem-solving capacity in a flexible manner. Such systems can also document knowledge for future use and training. This leads to having increased quality in the problem-solving process. It is also advantageous for remote and rural areas that have a scarcity of medical experts and solves the problems of the burden of experts. In general, the benefits and beneficiaries of the research will be the following:

6.1. Benefits of the research

The system will help miliary tuberculosis diseases patients and nurses provide useful information about the disease of statistical texture feature. The system will help non-experts in miliary tuberculosis diseases identifier and diagnosis worker. Non-experts can improve their knowledge and be effective in Miliary Tuberculosis diseases diagnosis.

The beneficiaries of this work will be the following:

- Researchers use as reference who works in this area
- Patients, spatially in a rural area which is an insufficient radiologist.
- Hospital professionals /nurses
- Governmental and Non-governmental institutions, health institutions, etc.

7. Thesis Organization

The remainder of the document is organized as follows. Chapter 2 describes how GLCM is working and how to extract texture features using GLCM algorithms. Chapter 3 discusses the literature review. Chapter 4 discusses our proposed solution. Chapter 5 show the result and discussion briefly. Chapter 6 evaluate the performance of the result. Finally, Chapter 7 gives the conclusions of the work done and future work.

CHAPTER TWO

2. Theory

2.1. Gray Level Co-occurrence Matrix

In 1973, Haralick introduced the co-occurrence matrix and texture features which are the most popular second order statistical features. Haralick proposed two steps for texture feature extraction. The first step is computing the co-occurrence matrix and the second step is calculating texture feature based on the co-occurrence matrix [15]. This technique is useful in wide range of image analysis applications from biomedical and remote sensing techniques. In statistical texture analysis, texture features are computed from the statistical distribution of observed combinations of intensities at specified positions relative to each other in the image [16].

According to the number of intensity points (pixels) in each combination, statistics are classified into first-order, second-order and higher-order statistics. The Gray Level Co-occurrence Matrix (GLCM) method is a way of extracting second order statistical texture features [17] [18]. The approach has been used in a number of applications. A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G, in the image. The matrix element $p(i,j|\ \Delta x,\Delta y)$is the relative frequency with which two pixels, separated by a pixel distance $\Delta x,\Delta y$, occur within a given neighborhood, one with intensity i and the other with intensity$j$. One may also say that the matrix element $p(i,j\mid d,\theta)$contains the second order [19]

The four matrices are used separately for classification, then the final decision is formed by combining the four decisions. As these matrices are symmetric, it is more convenient to use the upper or lower diagonal matrix coefficients in forming the vectors [20].

Haralick also introduced 14 statistical features. These features are generated by calculating the features for each one of the co-occurrence matrices obtained by using the directions 0◦, 45◦, 90◦, and 135◦, then averaging these four values [21]. The symbol Δ, representing the distance parameter, can be selected as one or higher.

In general, the Δ value is set to 1 as the distance parameter. A vector of these 14 statistical features is used for characterizing the co-occurrence matrix contents. These features can be calculated by using the following equations.

Haralick Features are defined by the following fourteen functions [22], [23], [24], [25].

$$f_1 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)^2 \tag{2.1}$$

$$f_2 = \sum_{n=0}^{Ng-1} n^2 \left\{\sum_{i=1}^{Ng} \sum_{j=1}^{Ng}{}_{|i-j|=n} p(i,j)\right\} \tag{2.2}$$

$$f_3 = \sum_{i=1}^{Ng}\sum_{j=1}^{Ng} \frac{(i-\mu x)\,(i-\mu y)\;p(i,j)}{\delta x \delta y} \tag{2.3}$$

$$f_4 = \sum_{i=1}^{Ng}\sum_{j=1}^{Ng} (i-\mu)^2\; p(i,j) \tag{2.4}$$

$$f_5 = \sum_{i=1}^{Ng}\sum_{j=1}^{Ng} \frac{1}{1+(i-j)2}\; p(i,j) \tag{2.5}$$

$$f_6 = \sum_{k=2}^{2Ng} kpx + y^{(k)} \tag{2.6}$$

$$f_7 = \sum_{k=2}^{2Ng} (k - f_6)^2\; px + y^{(k)} \tag{2.7}$$

$$f_8 = -\sum_{k=2}^{2Ng} px + y^{(k)}\, log\{px + y^{(k)}\} \tag{2.8}$$

$$f_9 = -\sum_{i=1}^{Ng}\sum_{j=i}^{Ng} p(i,j)\, log\{p(i,j)\} \tag{2.9}$$

$$f_{10} = \sum_{k=0}^{Ng-1}\left[k - \sum_{l=0}^{Ng-1} lpx - y^{(l)}\right]^2 px - y^{(k)} \tag{2.10}$$

$$f_{11} = -\sum_{i=1}^{Ng} px - y^{(k)}\, log\{px - y^{k}\} \tag{2.11}$$

$$f_{12} = \frac{f_{9-HXY1}}{max(HX, HY)} \tag{2.12}$$

$$f_{13} = (1 - exp[-2(HXY2 - f_9)])^{1/2} \tag{2.13}$$

$$f_{14} = (second\ largest\ eignvalue\ of Q)^{1/2} \tag{2.14}$$

Even if Harlick introduces 14 statistical features, but in our work, we are concentrating on the second step because second order is calculating texture feature based on the co-occurrence matrix. The Gray Level Co-occurrence Matrix (GLCM) method is a way of extracting second order statistical texture features.

2.2. How GLCM Working

GLCM is the method of computing the frequency of pixel pairs having the same gray level in the image. The relationship between the reference pixel and the neighboring pixels is calculated to determine the textural features of the image [26]. In Figure 2.1, the resolution cell 1 and 5 are 0 degrees nearest neighbors to cell 0. Similarly, resolution cells 2 and 6, 3 and 7, 4 and 8 are 135 degrees, 90 degrees and 45 degrees nearest neighbors to cell 0. The co-occurrence matrix is formed by evaluating the count of pixel pairs with gray level occurring adjacent to the pixel with gray level 6. The relative frequencies of gray level pixel pairs separated by a distance *d* in a particular direction forms the displacement vector(d, θ).

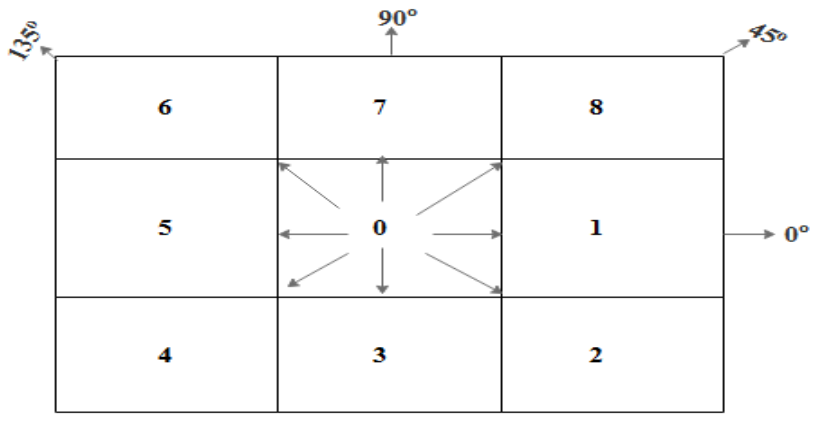


Figure 2.1: Resolution cells and their relationship with neighboring pixels at various angles [28].

Basic of GLCM texture considers the relation between two neighboring pixels in one offset, as the second order texture. The gray value relationships in a target are transformed into the co-occurrence matrix space by a given kernel mask such as 3 3, 5 5, 7 7 and so forth [27].

In the transformation from the image space into the co-occurrence matrix space, the neighboring pixels in one or some of the eight defined directions can be used; normally, four directions such as 0°, 45°, 90°, and 135° is initially regarded, and its reverse direction (negative direction) can be also counted into account [1]. It contains information about the positions of the pixels having similar gray level values. Each element (i, j) in GLCM specifies the number of times that the pixel with value i occurred horizontally adjacent to a pixel with value$j$. In Figure 2.2, computation has been made in the manner where, element (1, 1) in the GLCM contains the value 1 because there is only one instance in the image where two, horizontally adjacent pixels have the values 1 and 1. Element (1, 2) in the GLCM contains the value 2 because there are two instances in the image where two, horizontally adjacent pixels have the values 1 and 2.

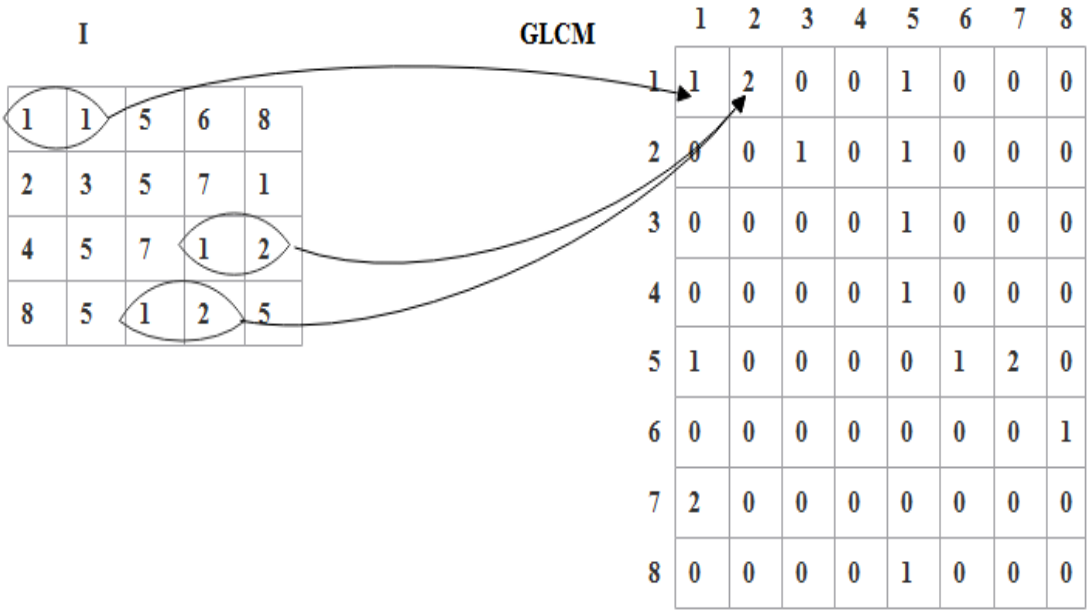


Figure 2.2: Creation of GLCM from image matrix. Source from [32]

Element (1, 2) in the GLCM contains the value 2 because there are two instances in the image where two, horizontally adjacent pixels have the values 1 and 2. The GLCM matrix has been extracted from input dataset imagery. Once after the GLCM is computed, texture features of the image are being extracted successively.

The distance between the pixel of interest and its neighbor, specified as a *p*-by-2 array of integers. Each row in the array is a two-element vector, [row_offset, col_offset], that specifies the relationship or *offset*, of a pair of pixels. row_offset is the number of rows between the pixel-of-interest and its neighbor. Col_offset is the number of columns between the pixel-of-interest and its neighbor. Because the offset is often expressed as an angle, the following table lists the offset values that specify common angles, given the pixel distance D.

As illustrates in Figure 2.3 the details of the process to generate the four co-occurrence matrices using *Ng* =5 levels for the offsets *{*[0 1], [-1 1], [-1 0], [-1 -1]*}* that are defined as one neighboring pixel in the possible four directions.

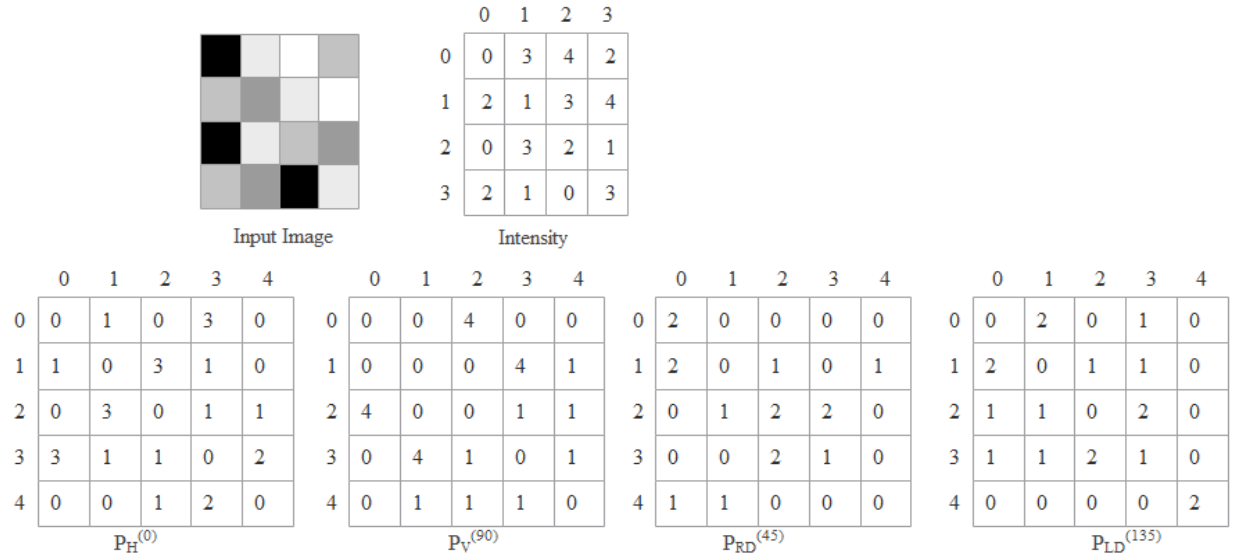


Figure 2.3: Co-occurrence matrix generation for Ng=5 levels and four different offsets: PH (0◦), PV (90◦), PRD (45◦), and PLD (135◦). Image from [27]

We can see that two neighboring pixels (2, 1) of the input image is reflected in P*H* concurrence matrix as 3 because there are 3 occurrences of the pixel intensity value 2 and pixel intensity value 1 adjacent to each other in the input image. The neighboring pixels (1, 2) will occur again 3 times in P*H*, which makes these matrices symmetric. In the same manner, the other three matrices P*V*, P*LD*, P*RD* are calculated.

2.3. Texture

In [28] [19] describe texture is a significant feature of an image that has been widely used in medical image analysis, image classification, automatic visual inspection and remote sensing. Generally speaking, textures are complex visual patterns composed of entities or sub-patterns that have characteristic brightness, color, slope, size, etc. A basic stage to collect such features through texture analysis process is texture feature extraction. Texture features can be extracted in several methods, using statistical, structural, model-based and transform information, in which a well-known method is using a Gray Level Co-occurrence Matrix (GLCM). GLCM contains the second-order statistical information of spatial relationship of pixels of an image [29].

GLCM is a useful tool in texture analysis, however, the overall calculations for the computation of GLCM and texture features are computationally intensive and time-consuming. Tahir described an example of calculations of GLCM and texture features in a medical application. With an image of size 5000 ^{x} 5000 pixels with 16 bands, the time required is approximately 350 seconds using Pentium 4 machine running at 2400 MHz. 75% of the total time spent is for the calculation of GLCM, 5% for the normalization, 19% for the calculation of texture features while 1% is for the classification using classical discrimination method. Due to such heavy computation, there were many types of research on accelerating the process of calculation GLCM and texture features [30].

2.3.1. Texture Feature Extraction

The texture features are extracted using GLCM algorithm. The GLCM represents second order statistics based on neighboring pixels [19]. The GLCM is a two-dimensional array which takes into account the specific position of a pixel relative to other pixels. The GLCM is a tabulation of how often a different combination of pixel brightness values occur in an image [31].

Feature extraction is used to denote a piece of information which is relevant for solving the computational task related to the certain application system. There are two types of texture feature measures. They are given as first order and second order measures. In the first order, texture measures are statistics, calculated from an individual pixel and do not consider pixel neighbor relationships. The intensity histogram and intensity features are first order calculations. In the second order, measures consider the relationship between neighbors The GLCM is a second order texture calculation [32]. In this research, we used second order texture calculation because we are measuring the relationship between neighbors of the pixel.

A co-occurrence matrix is one of the most traditional techniques for encoding texture information. The texture is one of the most important defining characteristics of an image. It describes spatial relationships among gray levels in an image. A cell defined by the position(i, j) in this matrix registers the probability at which two pixels of grey levels i and j occur in two relative positions [27] [26].

Table2: The table illustrates the array: offset = [0 1; -1 1; -1 0; -1 -1].

Angle	**Offset**
0	[0 D]
45	[-D D]
90	[-D 0]
135	[-D -D]

Each element of the GLCM is the number of times that two pixels with gray tone i and j are neighbourhood in distance d and direction θ. For 00 co-occurrence matrix, there are 2 occurrences of the pixel intensity value 1 and pixel intensity value 3 adjacent to each other in the input image [28].

Based on this theory, we studied the related publication to understand the current state-of-art and investigating algorithmic potential related to miliary tuberculosis.

CHAPTER THREE

3. Literature Study

This chapter presents a review of the literature on CAD systems for lung diseases and addresses some of the most relevant techniques, algorithms and methods for preprocessing, segmentation, region of interest (ROI) extraction, feature extraction and classification of chest abnormalities. The major algorithms and techniques used in this research work are also discussed. Finally, the novelty of the research work proposed in this thesis is presented.

[33] In their work have developed a CAD system for Automatic Tuberculosis Screening Using Chest Radiographs, which implemented methods for lung segmentation, feature computation, and classification.

They use different processing steps; First system segments the lung of the input CXR using a graph cut optimization method in combination with a lung model. For the segmented lung field, their system then computes a set of features as input to a pre-trained binary classifier. Segmentation in medical images has to cope with poor contrast, acquisition noise due to hardware constraints, and anatomical shape variations. Lung segmentation is no exception in this regard. They select these masks according to their shape similarity as follows: We first linearly align all training masks to a given input CXR. Then, they compute the vertical
and horizontal intensity projections of the histogram equalized images.

To measure the similarity between projections of the input CXR and the training CXRs, they use the Bhattacharyya coefficient. They then use the average mask computed on a subset of the most similar training masks as an approximate lung model for the input CXR. In particular, they use a subset containing the five most similar training masks to compute the lung model. As training masks, they use the publicly available JSRT set for which ground truth lung masks are available. The pixel intensities of the lung model are the probabilities of the pixels being part of the lung field. In a second step, they employ a graph cut approach and model the lung boundary detection with an objective function.

To formulate the objective function, they define three requirements a lung region has to satisfy: a) the lung region should be consistent with typical CXR intensities expected in a lung region, b) neighboring pixels should have consistent labels, and c) the lung region needs to be similar to the

lung model we computed. Mathematically, they described the resulting optimization problem as follows : Let f = {f1, f2, ..., fp, ..., fN} be a binary vector whose components fp correspond to foreground (lung region) and background label assignments to pixel p ∈ P, where P is the set of pixels in the CXR, and N is the number of pixels. E(f) = Ed(f) + Es(f) + Em(f), where Ed, Es and Em represent the region, boundary, and lung model properties of the CXR, respectively. Using the three energy terms given above, they minimize the objective function with a fast implementation of mincut/max-flow algorithm.

The minimum-cut is then the optimal foreground/background configuration of for the input CXR. To describe normal and abnormal patterns in the segmented lung field, they experimented with two different feature sets. 1) Object Detection Inspired Features here use the following shape and texture descriptors. Intensity histograms (IH), Gradient Magnitude Histograms (GM), Shape descriptor histograms (SD), Curvature descriptor histograms (CD), Histogram of oriented gradients (HOG), Local binary patterns (LBP). 2) CBIR-based Image Features - Set B: For their second feature set, Set B, they use a group of low-level features motivated by content-based image retrieval (CBIR). This feature collection includes intensity, edge, texture and shape moment features, which are typically used by CBIR systems. Classification to detect abnormal CXRs with TB, we use a support vector machine (SVM), which classifies the computed feature vectors into either normal or abnormal. They have been measured the performance of the system on two datasets: a set collected by the tuberculosis control program in the United States The MC set contains 138 posters anterior CXRs, among which 80 CXRs are normal and 58 CXRs are abnormal with manifestations of TB. And a set collected from Shenzhen Hospital, China using a Philips DR Digital Diagnostic system. The set contains 340 normal CXRs and 275 abnormal CXRs with TB.

Finally, they achieve an area under the ROC curve (AUC) of 87% (78.3% accuracy) for the first set and an AUC of 90% (84% accuracy) for the second set. For the first set, when they compare the system performance with the performance of radiologists. When trying not to miss any positive cases, radiologists achieve an accuracy of about 82% on this set, and their false positive rate is about half of our system's rate. Using decision rules and thresholds, the classifier outputs its confidence in classifying the input CXR as a TB positive case.

[34] In their work have developed a CAD system for Wavelet Analysis for Identification of Lung Abnormalities using Artificial Neural Network. They use these methods Image Processing:

Segmentation process is performed by using ‘mask’ in Matlab to obtain the lung region. Feature extraction performed after segmentation process using Daubechies wavelet. Training: Algorithm used in training process is ‘trainscg’ function.

They use these data set the data was taken in the Radiology Emergency Installation section of Wahid in Sudirohusodo Hospital with the amounts of 90 images. The images consist of 52 normal lung samples, 23 pleural effusion samples, and 15 pulmonary tuberculosis samples. These sample images are subdivided into two, samples for training data and samples for test data. A number of samples for the training data are 43 normal lung samples, 20 pleural effusion samples, and 12 pulmonary tuberculosis samples. The remaining 15 samples for the test data (9 normal lungs, 3 pleural effusion, and pulmonary tuberculosis 3) will be used to test accuracy rate of the system identification. Another 20 samples from other sources were added as secondary test data. They do an experiment the experiment using 75 training data samples showed that the system was able to identify 100% accuracy of abnormalities when testing with training data samples and 91.65% accuracy when testing with new data samples.

[35] In their work have developed a CAD system for Multiple-instance learning for computer-aided detection of tuberculosis. They data’s total of 917 CXRs from two sites in Zambia and South Africa were used in our experiments.

They were acquired by an Odelca-DR system with a slot-scan detector (Delft Imaging Systems, The Netherlands) and subsampled to a width of 1024 pixels. This database was divided into two independent sets of 461 and 456 CXRs that were respectively used for training and evaluation. The training set included 184 normal and 277 abnormal CXRs, while the test set was constituted by 208 normal and 248 abnormal CXRs. The reference standard was set by means of radiological assessment. This CAD system has been developed by the Diagnostic Image Analysis Group, Nijmegen, The Netherlands, and comprises four main stages: CXR normalization, lung field segmentation, texture feature extraction, and pixel classification. CXR normalization aims at minimizing the variability observed on images that, for example, come from different sources. This is achieved by dividing the image into frequency bands and scaling the pixel values in each band according to a set of reference CXRs. Afterward, lung segmentation is carried out to limit subsequent analysis to the region inside the lung fields. The utilized method is based on pixel classification after computing a multiscale local jet of second order. An independent set of 309

images is used for classifier training. The resulting lung likelihood map is post processed (blurring and morphology operations) to obtain a binary segmentation.

In the next stage, features consisting of the first four moments of the intensity distributions obtained after applying the above multiscale jet are extracted. The moments are computed on a grid with a spacing of 8 pixels considering circular patches with a radius of 32 pixels.3 In addition, several spatial features, such as the normalized horizontal and vertical position and the distance to the lung wall and to the center of gravity of both lungs, are computed.

In the last stage, classification of each of the previously obtained feature vectors is carried out. Although the original system utilizes a k-nearest neighbor (k-NN) approach, in this study, we have experimented with additional state-of-the-art classifiers: gentle ad boost with regression stumps as weak learners (GAB-RS), random forest (RF)5 and radial basis function support vector machine (RBF-SVM).6 These classifiers were trained by following a standard supervised setting, that is, they used the class labels assigned to each feature vector in the training set according to the available lesion annotations. After obtaining an abnormality score for each image in the set, those scores were pooled together and receiver operating characteristic (ROC) analysis was carried out. The area under the curve (AUC) was utilized as a performance measure. To assess statistically significant differences between the performances of pairs of CAD systems, bootstrapping9 with 5000 repetitions was applied. Differences in AUC were considered significant if $p < 0:05$. The results among the conventional approaches, only the system with RBF-SVM attained a significantly better performance than the miSVM-based system.

[25] In their work have analyzed the performance of several machine classifiers that were used to differentiate obstructive lung diseases in HRCT images. The dataset consisted of 67 normal lung images, 70 images of bronchiolitis obliterans, 65 images of centrilobular emphysema and 63 images of pan lobular emphysema from the Department of Radiology, As a Medical Centre, Korea. One ROI was sampled from each lung. Initially, automatic lung segmentation was performed. The blood vessels and the chest wall were removed by thresholding all regions in the image with densities corresponding to -400 to -1024 Hounsfield Units (HU). After segmenting the lung, different sizes of ROIs were considered and texture analysis was performed. For each CT image, radiologists labeled three sizes of ROIs: 16 x 16, 32 x 32 and 64 x 64 corresponding to the three diseases or normal lung. Four descriptors were used in this work and each ROI was represented

using a 13-dimensional feature vector. Naïve Bayesian, Bayesian, ANN and SVM classifiers were compared experimentally based on the training time, testing time, accuracy, sensitivity, and specificity. A five-fold cross-validation scheme was used to verify the cross-validity of the input data. The SVM classifier outperformed all the other classifiers as it was reasonably fast in training and its sensitivity and specificity were small. The overall accuracy of the SVM classifier was 64.1% for 16 x 16 ROI, 75.7% for the 32 x 32 ROI and 83.1% for the 64 x 64 ROI which indicated that as the ROI size increased the overall accuracy also showed improvement.

[36] In their work they have proposed Statistical Feature-based Neural Network Approach for the Detection of Lung Cancer in Chest X-Ray Images. The proposed algorithm consists of two stages.

The first stage involves a pipeline of image processing routines to separate the lung from other structures of original chest X-rays and identifying the region that is suspected of being a nodule. At this stage, the system extracts 65x65 square areas considering the suspicious point as the center. Because it involves a pixel-based technique, all the pixels in the square region are considered as the inputs to the system. The intensity values of the pixels that fall within this region are extracted and are stored in a database which is used to train the system at the next stage. The database is then divided into several subcategories, and the data available in these sub-categories will be used for the training as well as for testing the results.

In the second stage, a neural network is trained based on two types of inputs: pixel-based inputs in which we consider the intensity levels of pixels within the suspected region and statistical feature-based inputs where we take into account first and second order statistical features.

In the preprocessing stages, median filtering is required to remove the effect of poor contrast due to glare, noise, and effects caused by poor lighting conditions during image capture. The image was then converted to a binary image by using a thresholding technique. It computes a global threshold that can be used to convert an intensity image to a binary image. It utilizes Otsu's method, which chooses the threshold to minimize the intra-class variance of black and white pixels.in the Lung region segmentation tapes, Lung masks were prepared using active shape models that are available with the JSRT database. These images can be used to segment the lung region while the user can identify the scope when selecting suspicious points.

These images consist of 154 lung nodules (100 malignant cases, 54 benign cases), and 93 non-nodules. The chest radiographs were digitized with a 0.175mm pixel size, a matrix size of 2048 x 2048 and a 12-bit grayscale level. The criteria for inclusion of radiographs in the database were: (1) absence of nodules larger than 35 mm, (2) absence of suspicious nodules that were not confirmed by CT examination, (3) no more than one nodule per patient, and (4) absence of nodules with margins that could not be confirmed by radiologists.

They used connected components labeling to scan the binary thresholded image and group its pixels into components based on pixel connectivity. A 49x49 mask was selected within a suspicious region via an interactive user interface that we have developed for this purpose. After selecting the suspicious region during the first stage of our algorithm, we train the neural network until convergence is achieved. Two types of inputs were considered as inputs to the neural network.

They use GLCM as a feature extraction based on texture statistical analysis and neural network as a classification. A neural network with one hidden layer of 1000 neurons and an input layer of 10 neurons to accommodate the ten first and second order textures was used at the training stage based on these the proposed approach detected nodules in the diseased area of the lung with an accuracy of 96% using the pixel-based technique while the feature-based technique produced an accuracy of 88%.

[37] In their work they have proposed Classification of medical x-ray images for automated annotation In this work, an effort has been made to automatically search and classify the X-ray images into six classes namely chest, spine, foot palm, head, and neck. Each class consists of 30 images and it is collected from IRMA image dataset representing different ages, genders, views, positions and pathologists, wherein the image quality varies significantly.

The medical X-ray images are pre-processed so as to avoid noises by using a median filter and they are segmented by applying CCl. The shape and texture features are extracted by ZM, GLCM, PCA and DCT. The extracted features are classified by SVM classifier.

The pre-processing Image enhancement technique is applied to improve the image quality it provides better gray intensities distribution. In this work, a median filter is applied to enhance the image, the wherein segmentation process is done to find the region of interest (ROI). The ROI is found by segmenting the biggest region in the image. Connected Component Labelling (CCL) is

applied for this purpose. CCL scans an image and groups its pixels into components based on pixel connectivity.

Zernike moments have been reported to be a good descriptor for shape description. Therefore, this study uses the Zernike moments for describing the medical X-ray images because of the following two properties: The Zernike basis function satisfies the orthogonal property, implying that the contribution of each moment coefficient to the underlying image is unique and independent, i.e. no redundant information overlap between the moments;

Calculation of the Zernike moments does not require knowledge of the precise boundary of an object. This makes Zernike moments suitable for representing X-ray images with obscure boundaries.

The shape features of the images are as follows:

Normal

$$\varphi 1 = \mu 20 + \mu 02 \quad (3.1)$$

Rotation

$$\varphi 2 = (\mu 20 + \mu 02)^2 + (4\mu 11)^2 \quad (3.2)$$

Inversion

$$\varphi 3 = (\mu 30 + 3\mu 12)^2 + (3\mu 31 - \mu 03) \quad (3.3)$$

Texture Feature Extraction Co-occurrence matrix is one of the most traditional techniques for encoding texture information. The texture is one of the most important defining characteristics of an image. It describes spatial relationships among gray levels in an image. A cell defined by the position (i, j) in this matrix registers the probability at which two pixels of gray levels I and j occur in two relative positions. A set of co-occurrence probabilities (such as energy, entropy, contrast) has been proposed to characterize textured regions.

The authors sue this steps and achieve the result, initially the X-ray images are pre-processed and segmented using CCL. Then the features are extracted by applying GLCM, ZM, a combination of DCT & PCA and GLCM & ZM. The extracted features are fed into the SVM classifier which

gives an accuracy of 96.56%. The preprocessed and segmented X-ray image and the results of the classified images using SVM classifier.

[38] In their work they have proposed An Efficient Detection of Tuberculosis from Chest X-rays. The method for the detection of tuberculosis includes the main steps like segmentation, feature extraction, and classification.

In the localized region based segmentation have two main categories exist for active contours: edge-based and region-based. Edge-based active contour models utilize image gradients in order to identify object boundaries. This type of highly localized image information is adequate in some situations but has been found to be very sensitive to image noise. One benefit of this type of flow is the fact that no global constraints are placed on the image. Thus, the foreground and background can be heterogeneous and a correct segmentation can still be achieved in certain cases. These approaches model the foreground and background regions statistically and find an energy optimum where the model best fits the image. Some of the most well-known and widely used region-based active contour models assume the various image regions to be of constant intensity. Mean separation Energy another important global region-based energy that uses mean Intensities is the one proposed by Yezzi which Refer to as means separation energy.

Feature and classification to describe normal and abnormal patterns in the segmented lung field, following features are extracted. Intensity histograms (IH).This histogram is a graph showing the number of pixels in an image at each different intensity value found in that image. For an 8-bit greyscale image there are 256 different possible intensities, and so the histogram will graphically display 256 numbers showing the distribution of pixels amongst those greyscale values. Gradient Magnitude Histograms (GM), Shape descriptor histograms (SD). Where $\lambda 1$ and $\lambda 2$ are the eigenvalues of the Hessian matrix, with $\lambda 1 \leq \lambda$. Histogram of oriented gradients (HOG) is a descriptor for gradient orientations weighted according to gradient magnitude. The image is divided into small connected regions, and for each region, a histogram of gradient directions or edge orientations for the pixels within the region is computed. The combination of these histograms represents the descriptor.

Local binary patterns (LBP) is a texture descriptor that codes the intensity differences between neighboring Pixels by a histogram of binary patterns, LBP is thus a histogram method in itself.

The Binary patterns are generated by thresholding the Relative intensity between the central pixel and its neighboring pixels. Because of its computation simplicity and efficiency, LBP is successfully used in various computer vision applications often in combination with HOG.

To detect abnormal CXRs with TB, here use a support vector machine (SVM), which classifies the computed feature vectors into normal and abnormal class. An SVM is a supervised non-probabilistic classifier that generates hyperplanes to separate samples from two different classes in a space with possibly infinite dimension. The important characteristic of an SVM is that it classifying the data by computing the hyperplane with the largest margin. It uses the hyperplane with the largest distance to the nearest training data point of any class. The feature vectors of abnormal CXRs will have a positive distance to the separating hyperplane, and feature vectors of normal CXRs will have a negative distance.

In the experimental results, they have implemented on MATLAB 2015 which takes nearly 18 sec to complete the detection process. Her 1800 iteration used. When the iteration increases better result will obtain.

Finally, they conclude like this TB can be detected from Chest x-ray images by using image processing methods like segmentation, Feature Extraction, and classification. Existing diagnostics method such as sputum staining has become less reliable in high population so this method will be helpful in rural areas. When increasing the features selected and when using another segmentation method it may get more accurate result.

[27] In their work have proposed a CAD system for selecting a significant slice from a set of slices of a CT scan to analyze the nodules that correspond to lung cancer. A set of chest CT images, each consisting of multiple slices were selected. Each slice was initially threshold using Otsu's algorithm which was adopted for operation on Digital Imaging and Communications in Medicine (DICOM) images. The Canny operator was used to detecting the edges of the binary image. Morphological and connectivity operations were carried out on the edge detected image to obtain the thresholded lung. A greedy snake algorithm was then applied to reconstruct the lung borders and get the segmented lung parenchyma. The tumor regions were extracted by using a region growing approach and all regions that were non-candidate nodules were removed using a morphological erosion filter. The regions that were not present at the same location in the previous

and next slice were pruned and regions that were not pruned constituted the ROIs. The shape and texture features of the ROIs were extracted to form the feature vectors. The feature vectors were used to train a Radial Basis Function Neural Network (RBFNN). ROIs that were greater than 9 pixels and existed in at least three consecutive slices were considered as nodules. The slice containing the ROIs with the largest area was selected as the significant slice. The accuracy, recall, specificity and precision of the system were 94.44%, 92.31%, 94.92% and 80% respectively.

[18] In their work they have proposed Automatic Detection of Abnormalities in Chest Radiographs Using Local Texture Analysis in this approach consists of TB database and TB id The TB database contains 326 abnormal and 290 normal posterior– anterior (PA) chest radiographs collected from a tuberculosis screening program for people seeking political asylum in The Netherlands.

The abnormal images are representative of abnormal images encountered in mass chest screening: they consist of all abnormally as far as they could be traced from a consecutive period of 18 months of screening, during which approximately 25.000 examinations were made.

All images were read by two radiologists. If one of them or both considers the image to be abnormal, the patient is contacted for further examination. The chest units used in this screening program are mobile Electrode ca units (Oldelft BV, Delft, The Netherlands). The tube voltage was 117 kV and the images were printed on 10 by the 10-centimeter film. These films were digitized with a Lumisys 100 scanner (Lumisys, Inc., Sunnyvale, CA) to 932 by 932 pixels (pixel size: 0.1 mm; a small area at the sides of the film was cropped since it was not exposed) with 12-bit intensity. In order to be able to train classifiers to distinguish between normal and abnormal areas, two radiologists have outlined the areas containing abnormalities in each abnormal image (each radiologist did this for half of the total set of abnormal images). During this procedure, several abnormal images were re-classified by the radiologist as normal. These images have been excluded from the set, leaving a total of 279 abnormal images. The ground truth that we are trying to reproduce by a computer for this database is the judgment of the radiologists who read the films. In a large number of cases (106 out of the 279 abnormal cases), one of the radiologists judged the image to be normal, while the other favored contacting the patient for further examination. We consider these images as abnormal. TB is known to often affect the upper lung, gives an idea of the spatial distribution of the abnormal areas within the chest radiographs in the TB database.

For this study, all normal cases and only those abnormal cases with textural abnormalities were used. Furthermore, all images with large clothing artifacts (people are photographed fully dressed in this screening) were removed. This includes necklaces and buttons in the lung fields. Images containing artifacts outside the lung fields and small artifacts such as brassiere clips were not removed. The resulting database contained 147 abnormal cases and 241 normal cases.

The ID Database contains 100 normal PA chest radiographs and 100 abnormal PA chest radiographs with interstitial disease and is referred to as the ID database. The images were obtained at the University of Chicago Hospitals from daily clinical practice. All normal cases were selected based on the consensus of an independent review on each radiograph by four experienced chest radiologists. The abnormal cases, which ranged from subtle to severe, were selected based on radiological findings, CT, clinical data, and/or follow-up radiographs, by consensus of the same radiologists. Some images contain artifacts due to clothing or catheters. The radiographs were digitized to 2000 by 2000 pixels with 0.175 mm pixel size and 10-bits intensity. A chest radiologist classified each upper, middle and lower lung field of the abnormal cases as normal, possibly abnormal and definitively abnormal. For this study, the possibly and definitively abnormal parts were used to obtain abnormal areas for training classifiers. The researcher uses a different methodology like segmentation such as Active Shape Model (ASM) Segmentation: The purpose of the segmentation is to find corresponding regions within the lung fields. And Texture Analysis they use a multiscale filter bank to extract texture features. Filters are Gaussian derivatives, $L_n^{\theta}(x,\ y,\ \delta)$ where δ denotes the scale, on the order of derivative and the direction in which the derivative is computed. The rationale behind the choice for this set of filters is that they describe local image structure; they are the coefficients of a Taylor expansion of the image.

The choice of a specific filter bank is important, but not crucial. For instance, Gaussian derivatives are very similar to Gabor functions, and similar results can be obtained with an appropriate set of Gabor filters. The exact choice of orientation and scale of filters was determined empirically. Classifying Texture Feature Vectors Texture analysis yields a texture feature vector for each region in each image. The next step is to estimate the abnormality of each region based on these feature vectors. To this end, a training dataset is constructed for each region. This set contains the feature vectors of all abnormal regions, to which an equal number of normal feature vectors, from randomly selected normal images, are added. For each feature, scaling factors are computed to

normalize the mean to zero and the variance to one. The same scaling factors are applied to feature vectors before classification. Classifying Images The classification results per region may be a useful final result for a CAD scheme. Regions with a high abnormality score could be indicated while a radiologist is reading the image.

In this paper, they aggregate the regional scores into a stand-alone result: an estimate of the complete image being normal or abnormal. In principle, the results of each region could be used directly as features to classify the image as a whole using any classifier. Here we propose a classifier that multiplies the probabilities that regions are normal. The motivation for this approach is that the image is abnormal if any of the regions is abnormal.

Finally, they get the result depends on the algorithm, the method is evaluated on two databases. The first database was collected from a TB mass chest screening program, from which 147 images with textural abnormalities and 241 normal images were selected. Although this database contains many subtle abnormalities, the classification has a sensitivity of 0.86 at a specificity of 0.50 and an area under the receiver operating characteristic (ROC) curve of 0.820. The second database consists of 100 normal images and 100 abnormal images with interstitial disease. For this database, the results were a sensitivity of 0.97 at a specificity of 0.90 and an area under the ROC curve of 0.986.

[39] In their work have developed a CAD system for Computer Aided Diagnosis of Miliary Tuberculosis. They use Morphological Image Processing: In this section, granulometry is discussed as a method to analyze lung textures. And the rolling ball algorithm is discussed as a method for enhancing lung textures.

Simulating TB with a Template, Template Matching, Template Subtraction, Fourier Domain Correlation, Estimating the Threshold. They get The X-ray library at Groote Schuur Hospital in Cape Town. With the help of Prof. Steve Benin field, this set was digitized using a Lummis's scanner, which scans images at 2.8 line pairs per mm. This constitutes 65 images. Another set of images of healthy chests was also scanned in, bringing the total number of images to 120. Each image has, roughly, 2000 by 2500 pixels with a 12-bit pixel depth (saved as a 16-bit image) which means each image constitutes about 10MB of disc space. 15 of the 16 thresholds would be labeled as TB. In the TB set, 31 (94%) of the images were labeled as TB and 2 (6%) were Labeled as

healthy. In the Healthy set 19 (68%) of the images were labeled as healthy and 9 (32%) were labeled as TB. Here we put conclusion as investigation dataset has not been properly classified yet. The template is a very general shape and therefore it is possible that other textures have caused some images to be counted as positive when they have not had miliary TB present.

[1] In their work have developed a CAD system for detection and classification of lung nodules using multi-resolution MTANN in chest radiography Images. They use various steps to detect and classify the lung nodules, in the first step, a hard copy of images are converted into the soft copy of images with the help of a high-resolution scanner. These scanned images are stored in the storage drive location for additional analysis and processing during preprocessing used wiener filter Gaussian smoothing the image.

These Scanned images are imported into the computerized detection scheme system. Computerized detection scheme system performs two main tasks namely detection and classification. Detection is further classified as image acquisition, enhancement, and segmentation. They take out the features from each of the segmented nodule candidates after connected component labeling lungs based on M-ASM and watershed segmentation algorithm. To analyze the texture of the 2D image, the GLCM is broadly used. This GLCM matrix stores the co-occurrence frequencies of the pairs of gray levels which are grouped by a distance and orientation. Classifier organization depends completely on the feature selection. Feature classification includes here are contrast, energy, mean, kurtosis entropy, third moment. And they collect 30 CT (computed tomography) slices were processed. The proposed CAD system has achieved an accuracy of 73.3%, sensitivity of 88.6% and specificity of 78.33%.

Here is the summary of the related papers with our work.

Table 1: Summary of related work.

Ref.	Methods				Dataset	Result in Accuracy
	Preprocessing	Segmentation	Feature extraction	Classification		
[34]		Mask	Daubechies wavelet.	ANN	90 chest X-ray image	91.65%
[33]		Graph Cut Based	Intensity histograms (IH) Gradient Magnitude histograms (GM)	SVM	138 & 340 chest X-ray image	78.3% & 84%
[39]	Rolling ball algorithm		Granulometry Template Subtraction, Fourier Domain Correlation,	Template Matching	120 chest X-ray image	TB set 31 (94%) & 2 (6%) Healthy set 19 (68%) & 9 (32%)
[1]	Gaussian noise using a wiener filter	Watershed	GLCM	Bayesian classifier	30 CT slices	73.3%,

Comparing with the existing works discussed in the literature, and here we mentioned finding from the reviewed.

The analysis reveals several noticeable observations:

1. There is no common dataset.
2. It cannot be stated that one specific classifier performs best because each classifier has used it's in the training set and a direct comparison is not on an equal basis.
3. Most researchers on CAD system try to solve all types of tuberculosis disease with the same algorithm.

4. About half of the publication deals with TB manifestation while the other half focuses on specific manifestation.

Abnormal TB manifestation typically affects the texture and geometry of the lung in a chest x-ray. To identify these texture and geometrical feature plays a big role increase the accuracy of automatic TB screening. In most paper, the authors use common image analysis techniques like preprocessing, segmentation, feature extraction, and classification. All the reviewed publications, therefore, describe methods to extract the texture and geometric features to classify CXRs into abnormal or normal.

Subsequently, from the Table 1, those publications are reviewed in more detail which shows the best quality result. Therefore reviews serve as the starting point to our work.

From the literature, we learn about the image analysis techniques to solve the problem with the appropriate algorithm.

- We learn various preprocessing techniques appropriate for chest x-ray image. And identifying which segmentation techniques is good for chest x-ray images.
- Understand which texture feature extraction algorithm.
- Harlick texture feature extraction algorithm. Second-order statistical texture feature extraction used GLCM algorithm, it is good for diagnosis miliary tuberculosis texture feature extraction.

Finally, GLCM texture feature extraction algorithm is best fit to calculate the miliary texture features. Depend on this information we have developed the proposed solution.

CHAPTER FOUR

4. Proposed solution

The purpose of this research is to identifying the texture feature of miliary tuberculosis chest x-ray image using GLCM algorithm. This research paper focuses on second order statistical texture parameters and shows the strength of these statistical texture parameters. After the chest x-ray image download from the various website and Japanese Society of Radiological Technology (JSRT) built a database, we prepared the dataset. A set contains 20 chest x-ray images among which 10 miliary chest x-ray image and 10 normal chest x-ray, and all images are converted to JPEG format then we implement image analysis techniques. Throughout the experiment, we are concentrated on preprocessing, segmentation and texture feature extraction. Under preprocessing we change the original image to gray level 8bit image, scaling the gray level image, and apply image enhancement algorithm using histogram equalization, then normalized image intensity. While providing a good response to a clear visualization of the image and ready to image segmentation process. During segmentation, we use image thresholding techniques to binaries the image and clearly identify the lung region and the background of the image. Using this segment result apply the texture feature extraction algorithm that is GLCM texture feature extraction then diagnosis the result depend on the texture feature value either a normal chest x-ray or miliary tuberculosis chest x-ray image.

Through this research experiment, we have followed the following proposed solution as shown in figure4.1.

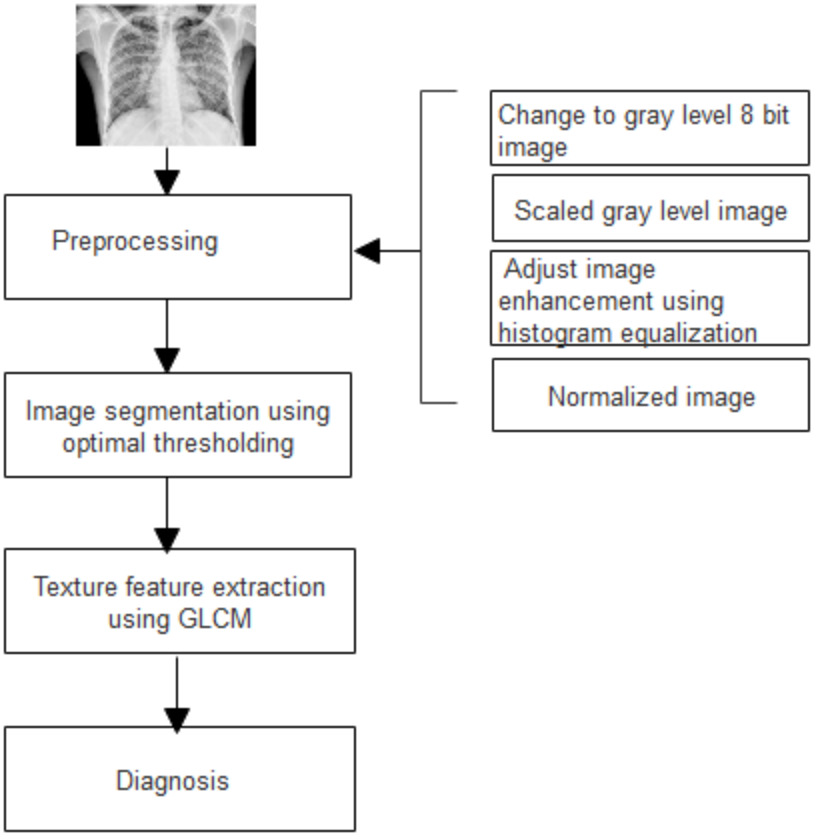


Figure 4.1: Block diagram of the proposed system

4.1. Preprocessing

In our work, the main goal of the pre-processing is to improve the image quality to make it ready to further processing by removing or reducing the unrelated parts in the background of the x-ray images [40]. It will prepare the x-ray for the next two-process segmentation and feature extraction.

In the preprocessing stage, if the image is in non-gray level then it is converted into the grayscale because the gray image reduces the processing time and produces the fast algorithms.

4.1.1. Scaled Gray Level Image

These images contain the brightness information. The number of bits that are used to represent each pixel is related to the number of different brightness levels available. The typical image contains 8 bits per pixel so there are 256 different possible gray tones (Ng) or intensity values from 0 to 255. All image changed to 2000 by 2000 pixels.

4.1.2. Contrast Enhancement

Image Contrast is the difference in the appearance of two or more parts of an image seen simultaneously. An image must have good brightness contrast for proper vision. In a low contrast image, we can't distinguish clearly between different objects. Increasing the contrast makes the light areas become lighter and dark areas become darker. Using Contrast enhancement increases the visual perception of the difference between different parts of an image [27] [30].

We use different techniques of histogram modification to improve the visual contrast of the image. The histogram of an image with the intensity levels in the range [0, L-1] is a discrete function. There are several factors which produces the noise while acquiring the image. To remove the noise content present in the image, a median filter is useful. The median filter is an order statistic filter, in which each pixel is assigned a rank [41]. This filter produces the less blurring of edges of the image. But we didn't use image filter, therefore using noise reduction may miss the miliary texture.

4.1.3. Normalization

The basis of normalization is a correlated change of brightness level in a selected region of interest to bring the model greyscale value for its 'normal' (or would-be normal) parenchyma to a standardized reference value. The selection of this reference value is determined from inspection of data. It provides a working 'norm' for a particular location within the chest x-ray images.

In this study, the decision involves either anterior or posterior regions for scans taken with the patient in the horizontal position. The anterior value is probably more reliable as a reference for patients of widely differing weights will contribute variously to the gravity and pressure effects on dependent tissue during scanning, non-dependent regions are affected much less overall. For example, a look at the average values for anterior and posterior locations reveals a small effective difference between their densities [41].

One factor informing the choice of a greyscale reference norm is the preference for adjusting brightness values upwards rather than down in order to keep all image data accessible. Any greyscale values reduced to zero lose all information. Then brightness adjustments can be found heuristically to bring corresponding regions from the data set to that value. Visual inspection of the selected region must serve to determine tissue normality. Try normalizing so that the mean is 0 and the variance is 1. This is a common technique for making intensity images invariant to

illumination changes, provided they are of the same scene [42]. If you recall from probability theory, this is performed by obtaining the Z-score:

$$z = \frac{x - \mu}{\sigma} \tag{4.1}$$

$$\mu = \text{mean}$$

$$\sigma = \text{standard deviation}$$

4.2. Image Segmentation

We use segmentation for the purpose of partitioning an image into multiple segments, so as to change the representation of an image into something that is more meaningful and easier to analyze [43]. In this work the practical application of image segmentation identifying clearly the white and black and the foreground and the background of chest x-ray image. And the choice of a segmentation technique over another and the level of segmentation are decided by the particular type of image and characteristics of the problem being considered.

4.2.1. Segmentation of the Lung Region in Chest X-ray Image

The lung and surrounding regions have a low value of intensity pixels and body pixels have high values of intensity pixels. The air between dark background and lungs is considered as having non-body pixels. The segmentation of the lung regions from the surrounding anatomical part is performed.

Thresholding is the process of converting a grayscale image to a bi-level image using an optimum threshold value T. It is a process of partitioning an image into object pixels and background pixels. An individual pixel is made an object pixel if the pixel value is greater than a certain threshold value and a background pixel.

There are two types of thresholding algorithms:

1. Global thresholding method

2. Local or adaptive thresholding method

Global thresholding methods use a single global value of threshold to partition an image into distinct regions whereas a local method uses the different local value of the threshold for different areas.

4.2.2. Iterative Thresholding

This method is based on assuming an initial value of the threshold. The best way for selecting a threshold value according to this algorithm follows these steps [5]:

1. An initial estimate T has to be taken which can be assumed to be the average of the minimum and maximum intensity value (though this is not fixed, any value can be chosen as initial T).

2. Using this value T, an image is divided into two pixels regions, T1 (with all pixel values<T) and T2 (with all pixel values>T).

3. Compute the average intensity values m1 and m2 of regions T1 and T2 respectively.

4. Compute a new value of T as

T= (m1+m2)/2

5. Repeat the above steps from 2 to 4 till difference between two successive values of T is minimal.

4.2.3. Otsu's Method

This method is considered as a variation of iterative thresholding method. It is a clustering method based upon maximizing the between-class variance. It is based upon defining well-defined threshold classes as clusters with clusters lying tightly adjacent to each other and there is a minimal overlap [44].

The within-class variance can be defined as the summation of the variance of each class as:

$$\delta^2_{within}(T) = n_B(T)\delta^2_B(T) + n_o(T)\delta^2_o(T) \qquad (4.2)$$

$$n_B(T) = sum\ of\ pixl\ values\ in\ background$$

$$n_o(T) = sum\ of\ pixl\ values\ in\ foreground$$

$$\delta^2_B(T) = variance\ of\ pixels\ in\ background$$

$$\delta^2_o(T) = variance\ of\ pixels\ in\ foreground$$

From the proved methods, the between class variance is given as:

$$\delta^2_{between}(T) = n_B(T)n_o(T)[\mu_B(T) - \mu_o(T)]^2 \qquad (4.3)$$

Where μ_B and μ_O are cluster means.

To find an optimal threshold value is to maximize the between class variance and this is relatively an easier calculation than calculation within class variance which would be minimized by maximizing between class variance. This method also works by assuming an image with a bimodal histogram [44].

The initial threshold value Ti is selected as the mean of the gray level of foreground and background pixels. To select a new threshold value, to make distinction between two different intensity regions segment the image with the current valueTi. Now Ti+1 is new threshold level which is the average of body and non-body pixels [1], [30].

$$Ti + 1 = {}^{(\mu a + \mu b)}/_{2} \qquad (4.4)$$

Where μa is the average gray levels of the body pixels and is the average of gray levels of the non-body pixels. The pixels values greater than the threshold value is set to 1 and values less than threshold value are considered as background.

4.3. Feature Extraction

After the lung region segmentation, the most difficult step is to differentiate some texture based features are calculated. The features calculated are best suited to diagnosis between the miliary/non-miliary. The features value has an information to predict the correct diagnosis. The texture based features are obtained from the gray level co-occurrence matrix (GLCM). Gray level co-occurrence matrix is the matrix that contained all the relative frequencies $p(i,j)$ with intensity value i which occurs in spatial relationship with value$j$ [27]. The total sum of the pixel values in GLCM is 1. Texture features based on GLCM such as energy, contrast, correlation and homogeneity are considered for diagnosis miliary TB and normal chest x-ray images in the proposed work.

Gray-level co-occurrence matrix (GLCM) is the statistical method of examining the textures that consider the spatial relationship of the pixels [45] [19]. The GLCM functions characterize the texture of an image by calculating how often pairs of the pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. The graycomatrix function in MATLAB creates a gray-level co-

occurrence matrix (GLCM) by calculating how often a pixel with the intensity (gray level) value i occurs in a specific spatial relationship to a pixel with the value$j$. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element (i, j) in the resultant GLCM is simply the sum of the number of times that the pixel with value$i$ occurred in the specified spatial relationship to a pixel with value$j$ in the input image [46].

A gray level co-occurrence matrix (GLCM) or co-occurrence distribution is a matrix or distribution that is defined over an image to be the distribution of co-occurring values at a given offset [45]. A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, G, in the image. The use of statistical features is, therefore, one of the early methods proposed in the image processing literature. Haralick suggested the use of co-occurrence matrix or gray level co-occurrence matrix [19], [47]. The features extracted from GLCM are Energy, Contrast, Entropy, Correlation, and Homogeneity.

After creating the GLCMs, we can derive several statistics from them using the graycomatrix function. These statistics provide information about the texture of an image. The following table lists the statistics and its description [48].

Statistic	Description
Contrast	Measures the local variations in the gray-level co-occurrence matrix.
Correlation	Measures the joint probability occurrence of the specified pixel pairs.
Energy	Provides the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment.
Homogeneity	Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

Table 2: statistics and its description source from [44]

These features are:

Contrast: - It gives the local variations in the GLCM. It determines the intensity difference between a pixel and its neighborhood. The value of contrast is 0 for a constant image [23].

$$contrast = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} (i-j)^2 \; p(i,j) \qquad (4.5)$$

Correlation: -It measures how a pixel is correlated to its neighborhood pixels. Its value lies between -1 and +1.Its value is -1 for perfectly negatively correlated image and +1 for the positively correlated image. [22], [23], [24], [49].

$$\text{Correlation} = \sum_{i=0}^{N_g}\sum_{j=0}^{N_g} p(i,j)\frac{(i-\mu x)(j-\mu y)}{\delta x \delta y} \tag{4.6}$$

Energy: - It gives the uniformity of the pixels. It is also known as second order moment. It is calculated from the sum of the square of all the elements in GLCM. The value of the energy lies between 0 and 1.If the image is constant then it has a value of 1 [24], [25].

$$Energy = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)^2 \tag{4.8}$$

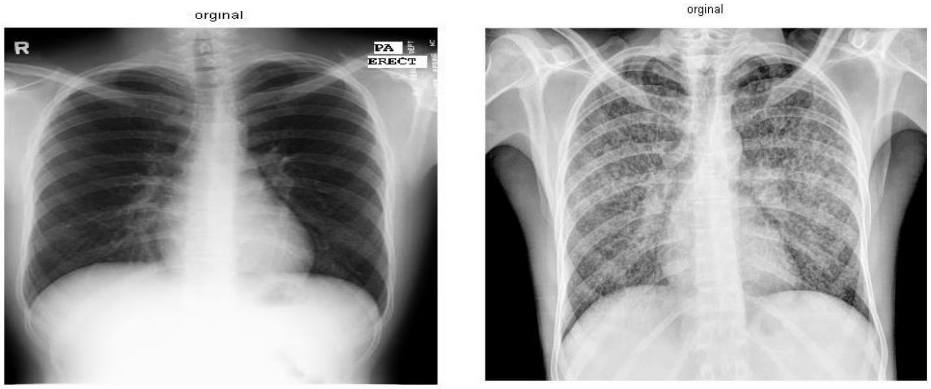
Homogeneity: It gives the distribution value of the closeness of elements of the GLCM to the diagonal of GLCM. It gives the value between the range of 0 and 1 [23], [25].

$$Homogeneity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{p(i,j)}{1+|i-j|} \tag{4.9}$$

CHAPTER FIVE

5. Result and Discussion

In this part, we discussed the result depends on of the proposed solution.

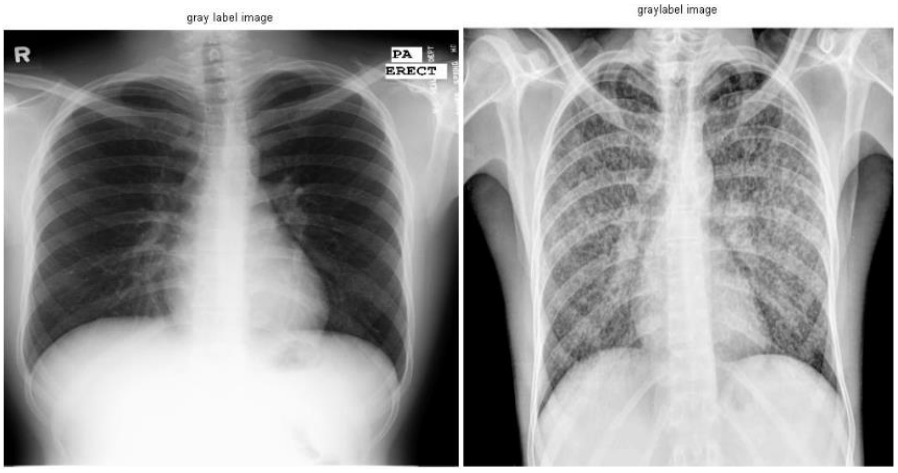


(a) (b)

Figure 5.1: Original (a) normal and (b) miliary TB image respectively

Figure 5.1 shows a normal chest x-ray image (a) and a miliary TB chest x-ray image (b) from the dataset. Images are depicted without implementing any processing step.

5.1. Gray Level Image Result



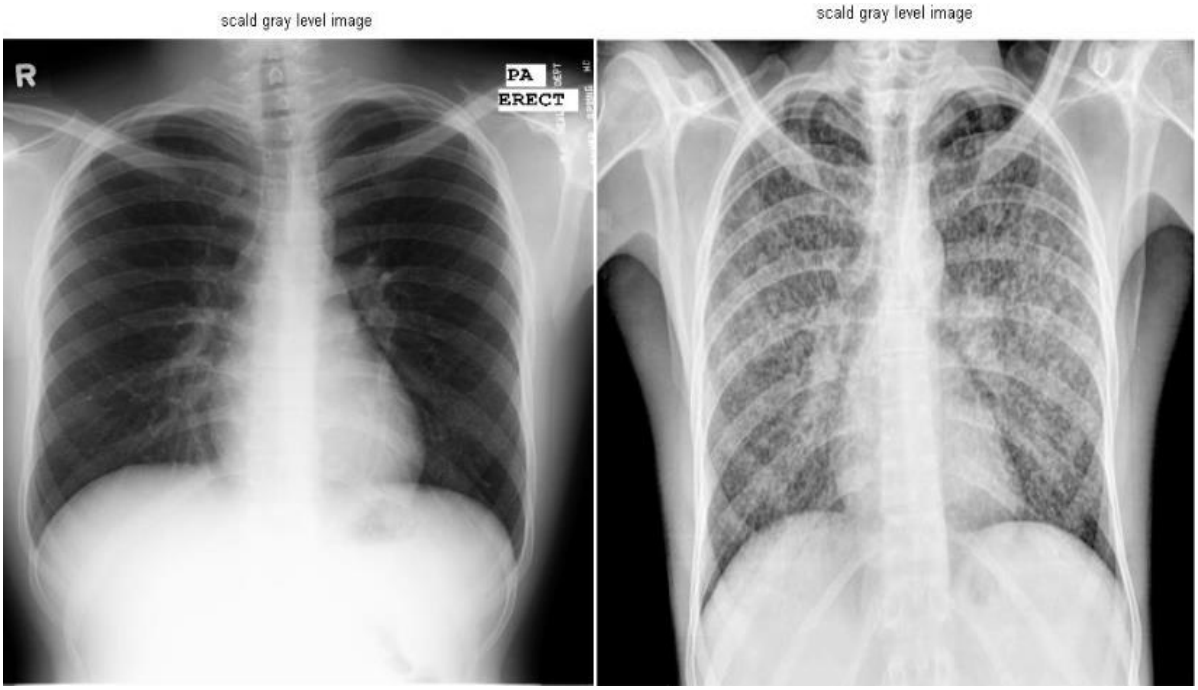
(a) (b)

Figure 5.2: Gray level image of (a) normal and (b) miliary TB image respectively

In the preprocessing stage, if the image is not grayscale then it is converted into the grayscale because the gray image reduces the processing time and produces the fast algorithms. The gray scale image is shown in figure 5.2.

These images contain the brightness information. The number of bits that are used to represent each pixel is related to the number of different brightness levels available. The typical image contains 8 bits per pixel so there are 256 different possible gray tones (Ng) or intensity values from 0 to 255.

5.2. Scaled Image Result

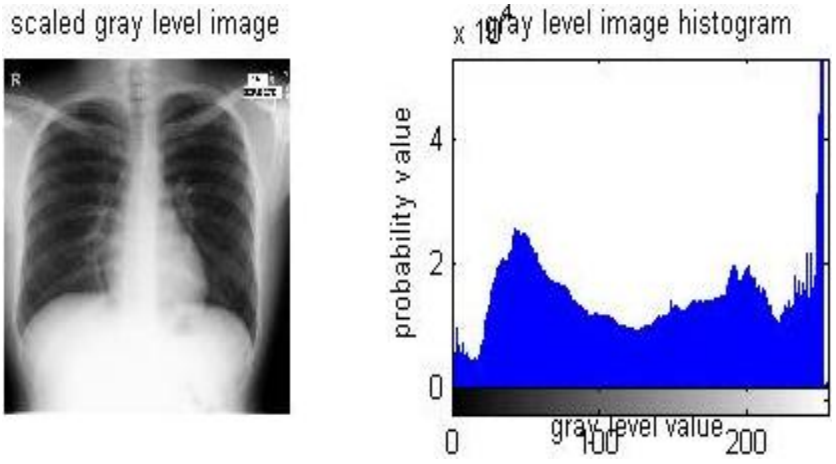


(a) (b)

Figure 5.3: Scaled image to 2000 by 2000 pixel (a) normal and (b) miliary TB image respectively

Even if the computational complexity increase due to increasing the size of the image, the chest radiographs sizes were scaled up with new dimensions 2000 by 2000 pixels and 256 numbers of gray levels (8-bits) by use of gray scale transformations were used. In order to clear and increase the size of the texture of miliary TB. As we seen in the image show some pixels are clear than the original image, this step is useful to the next preprocessing techniques.

5.3. Histogram Result of Scaled Image



(a)

Figure 5.4: Normal chest x-ray image histogram result

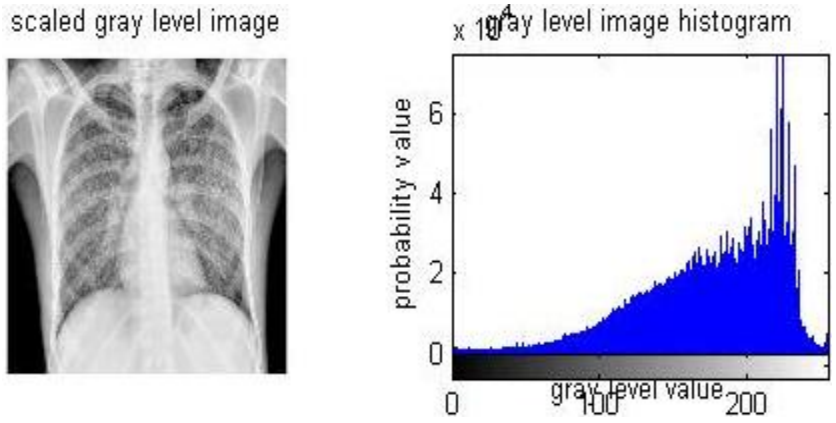


Figure 5.5: Miliary TB chest x-ray image histogram result

As showed in Figure 5.4 and Figure 5.5 the normal chest x-ray image of histogram result pixel distribution probability is closed to black from 0 to 100 and pixel distribution is close to white 200 to 255, whereas miliary tuberculosis chest x-ray image histogram result pixel distribution more close to the white.

5.4. Enhanced Chest X-ray Image Result

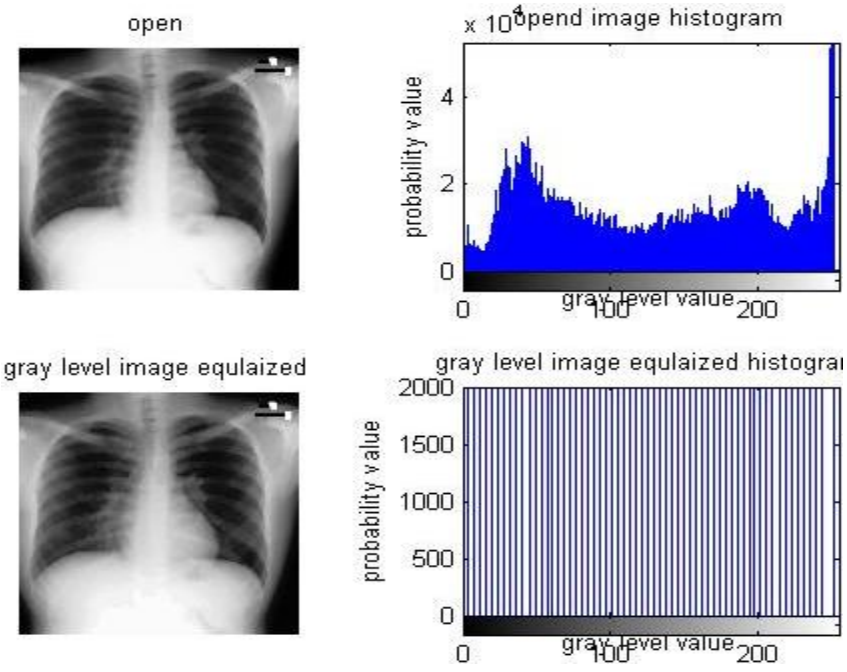


Figure 5.6: Opened and equalized normal chest x-ray image with respect to its histogram

Here implementing the rolling ball algorithm by opening the image, the image opened by 15-pixel structural element. The opened image histogram intensity versus gray level pixel distribution is range from 0 to 100 is medium and 100 to 200 is low and 200 to 255 slightly increase. When the equalized the opened image the distribution of pixel is equal and the image is enhanced.

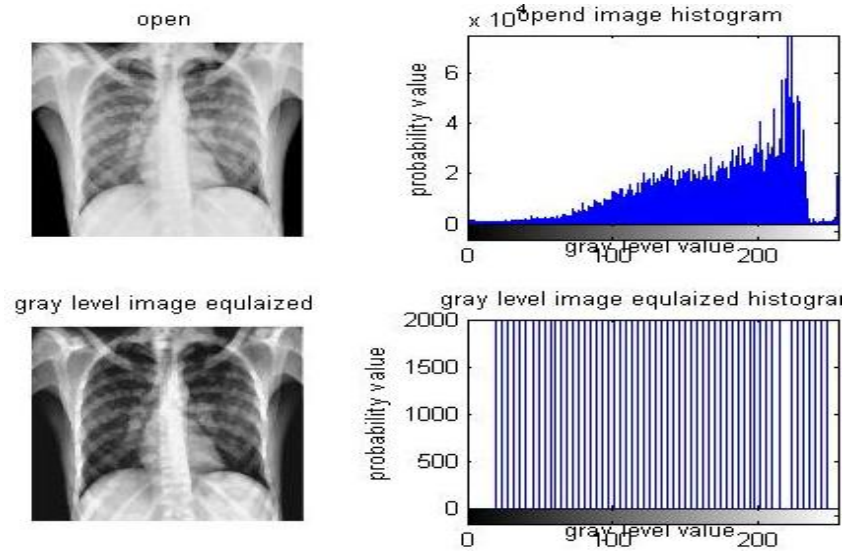


Figure 5-7: Opened and equalized miliary TB chest x-ray image with respect to its histogram

Here we implement the same algorithm like the normal image, but when we see the pixel distribution in the opened image histogram have low black value and brightness is high then equalization enhance the image and try to create an equal distribution of pixel. When equalized the image becomes dark and pixel density is high relative to the normal image.

5.5. Normalized Chest X-ray image result

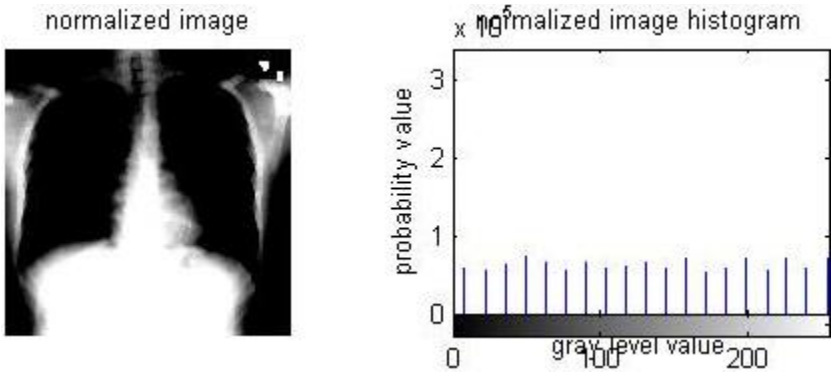


Figure 5.8: Normalized normal chest x-ray image result

Here we get a clear region of lung area, the exact lung area becomes dark and the background is white but here is some drawback that is some background area dark and this count as the lung part.

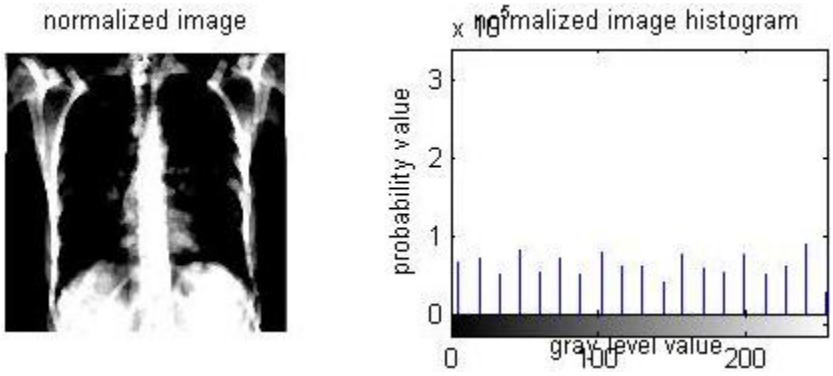


Figure 5.9: Normalized miliary TB image result

Here also implement the same algorithm like the normal image as shown in figure5.6, and when we compare the normal and miliary TB chest x-ray image the normal have no white part in the lung region whereas the miliary TB chest x-ray image has some white dot. And the histogram shows high-intensity range than miliary TB chest x-ray image comparatively.

One factor informing the choice of a greyscale reference norm is the preference for adjusting brightness values upwards rather than down in order to keep all image data accessible. Any greyscale values reduced to zero lose all information. Then brightness adjustments can be found heuristically to bring corresponding regions from each dataset to that value. Visual inspection of the selected region must serve to determine tissue normality [42].

5.6. Chest X-ray Image Segmented Result

(a) (b)

Figure 5.10: Segmented normal (a) and miliary TB (b) images

Here in segmentation part is used to identify the white and black part or (0, 1) of the lung image that used for input as the feature extraction of the image. But here in our work segmentation part is not excluding all the background of the image.

5.7. Texture Feature Result

In this work, we concentrate on second order statistical feature extraction as discussed in the proposed solution part.i.e contrast, correlation, energy, homogeneity.

Table 3: Miliary and normal chest x-ray statistical texture feature result

CXRs	contrast	correlation	energy	homogeneity
miliary1	0.003	0.9935	0.5377	0.9985
miliary2	0.0031	0.9934	0.5361	0.9985
miliary3	0.0031	0.9934	0.5331	0.9975
miliary4	0.0028	0.9938	0.5385	0.9986
miliary5	0.0028	0.9939	0.5366	0.9986
miliary6	0.0037	0.9919	0.5373	0.9981
miliary7	0.0049	0.9892	0.5363	0.9975
miliary8	0.0049	0.8892	0.5363	0.9985
miliary9	0.0022	0.9955	0.4978	0.9989
miliary10	0.0039	0.9936	0.5358	0.9985
normal8	0.0014	0.9969	0.5375	0.9993
normal2	0.0018	0.9957	0.5361	0.9957
normal3	0.0014	0.9969	0.5386	0.9993
normal4	0.0024	0.9949	0.5368	0.9988
normal1	0.0016	0.9956	0.5374	0.9956
normal5	0.0012	0.9945	0.5374	0.9987
normal6	0.0019	0.9956	0.5371	0.9991
normal7	0.0018	0.9962	0.5382	0.9991
normal11	0.002	0.9956	0.3571	0.999
normal12	0.0017	0.9962	0.5374	0.9991

Depending on the observed statistical values we can determine the follow contrast analysis.

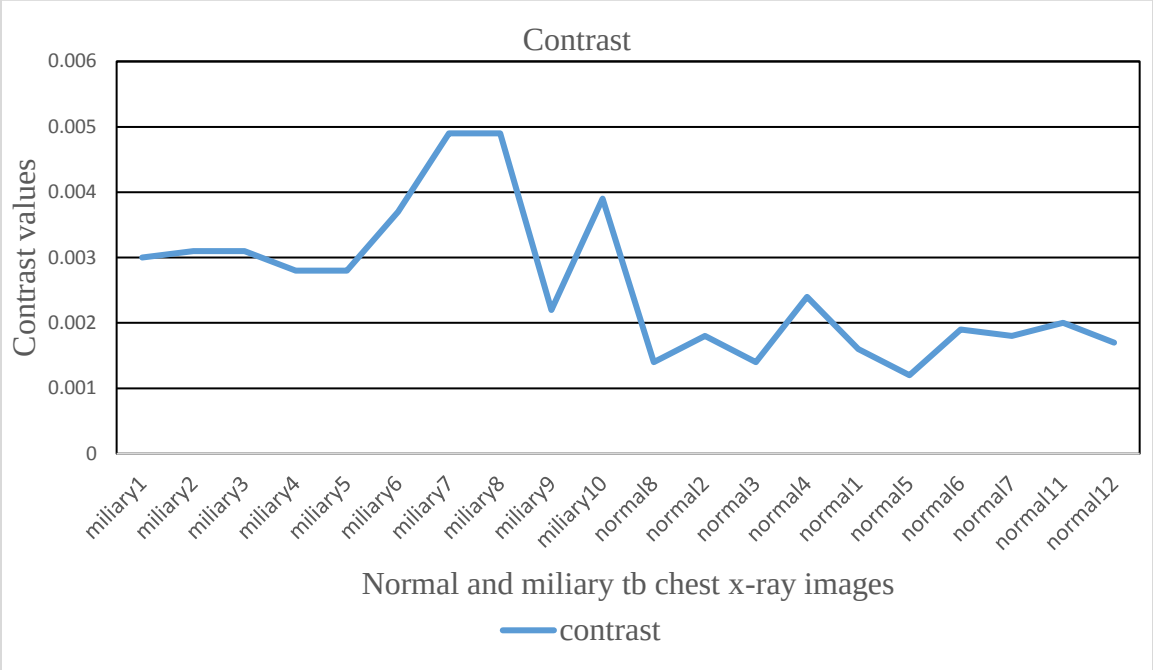


Figure 5.11 Comparison of contrast value miliary and non-miliary chest x-ray image

As shown in the above table information *contrast* gives the local variations image intensity in the GLCM. Also, it determines the intensity difference between a pixel and its neighborhood. In comparison, normal images have a lower value than miliary TB images.

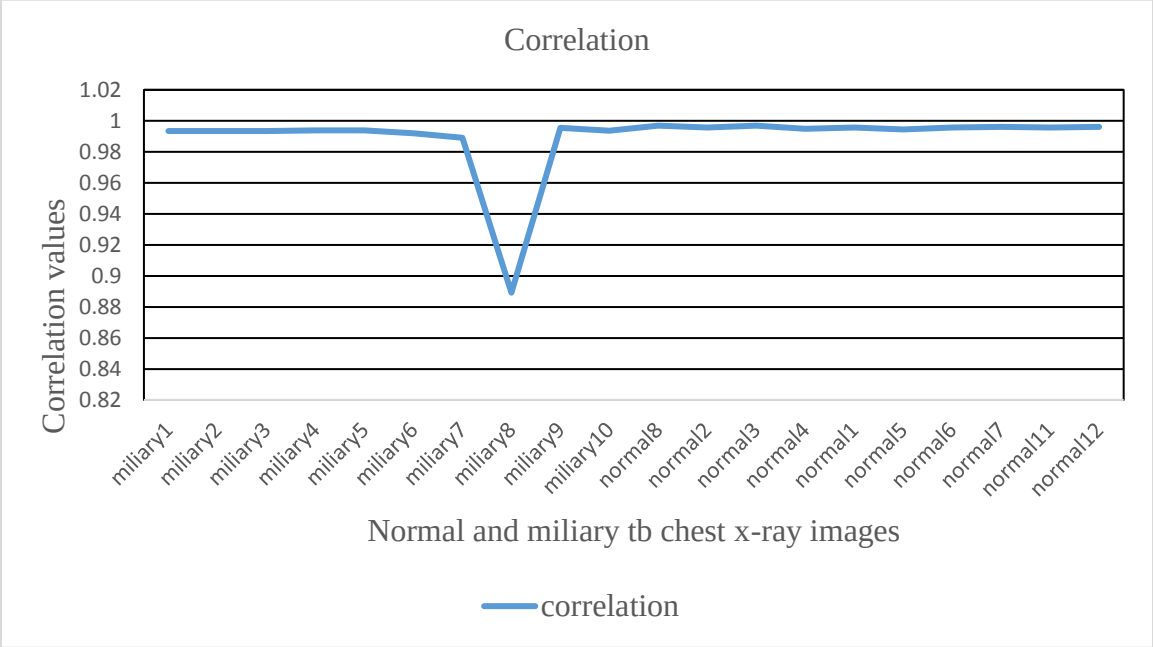


Figure 5.12: Comparison of correlations value miliary and non-miliary chest x-ray image

In the *correlation* measures, pixels are correlated to their neighborhood pixels. Its value also lies between 0 and 1, but the measured parameters lack significance to determine normal and abnormal cases.

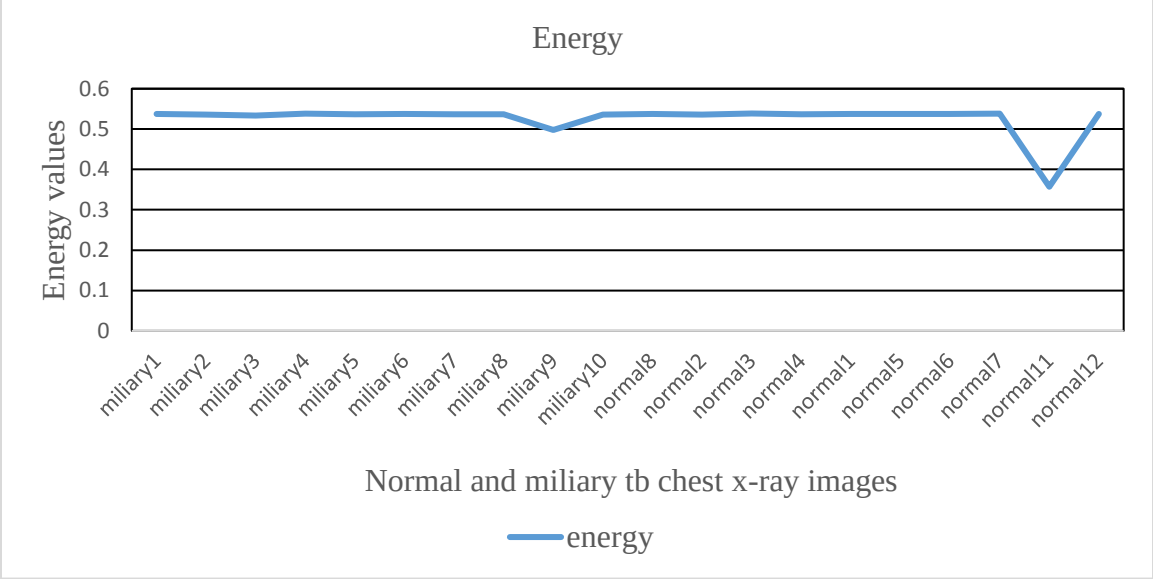


Figure 5.13: Comparison of energy value miliary and non-miliary chest x-ray image

In *energy*, we calculate the sum of the square of all the elements in GLCM. The value lies between 0 and 1 but also here discrimination between normal and abnormal cases is not possible.

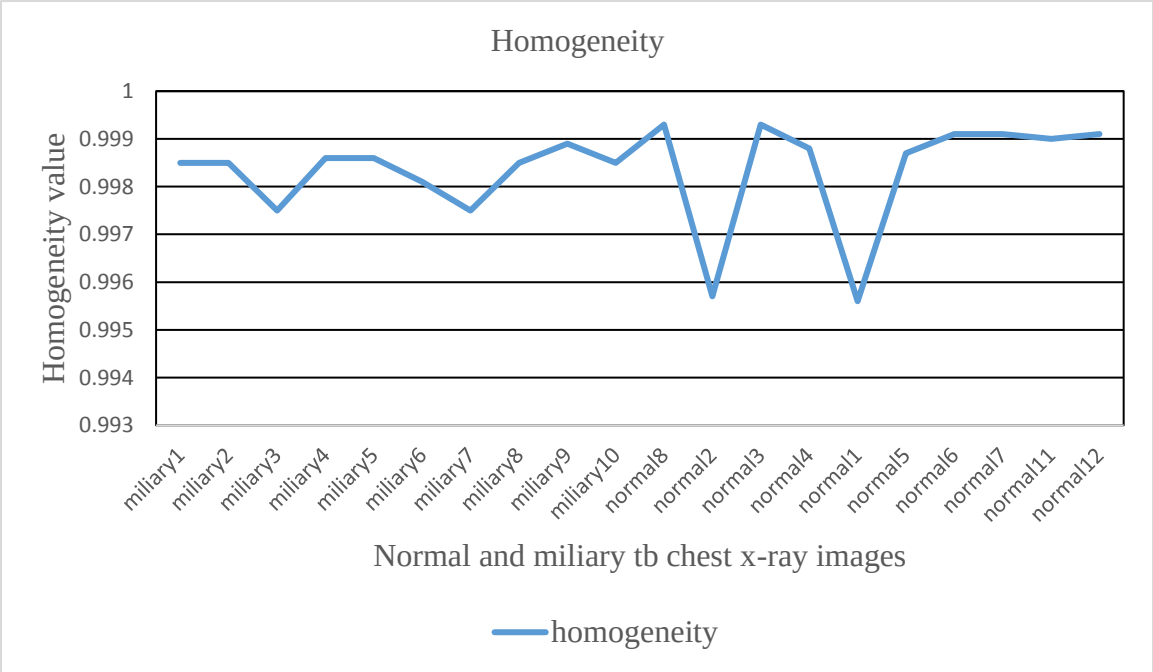


Figure 5.14: Comparison of homogeneity value miliary and non-miliary chest x-ray image

Homogeneity calculates the distribution value of the closeness of elements of the GLCM to the diagonal of GLCM. When we compare normal and miliary TB lung results, there is a slightly high value in case of normal images.
Depending on the observed statistical values we can determine the follow contrast analysis.

From this texture feature value statistical analysis, we can conclude that contrast has good information to classify the normal and miliary tuberculosis chest x-ray image. *C*ontrast value of miliary TB image higher value than normal chest x-ray image. Depend on this we can evaluate the performance of the system.

CHAPTER SIX

6. Performance Evaluation

After using GLCM functions for characterizing the texture of an image by calculating how often pairs of the pixel with specific values and in a specified spatial relationship occur in an image [26]. However, a single GLCM might not be enough to describe the textural features of the input image. For example, a single horizontal offset might not be sensitive to texture with a vertical orientation. Therefore it is essential to generate multiple GLCMs with different offset values or at different angles. MATLAB function graycomatrix is used to generate such multiple GLCMs [18], [5]. Using multiple GLCMs, second-order statistic features like, Contrast, Correlation, Energy, and Homogeneity are estimated.

These results will be presented as statistical measures such confusion matrix and correct diagnosis rate. In the above section, we define briefly these statistical measures. After that, we present the obtained results for real test of chest x-ray image texture, So Confusion Matrix is a 2-by-2 numeric array with diagnostic counts. This matrix used to measure the quality of a diagnosis system. The first column indicates the number of samples that were diagnosis as positive, with the number of true positives in the first row, and the number of false positives in the second row. The second column indicates the number of samples that were diagnosis as negative, with the number of false negatives in the first row, and the number of true negatives in the second row. Correct classifications appear in the diagonal elements, and errors appear in the off-diagonal elements. Inconclusive results are considered errors and counted in the off-diagonal elements. Table 6 shows the confusion matrix of a system that allows diagnosis two classes normal and miliary TB.

Table 4: Confusion matrix I

		Predicted	
		Negative	Positive
Actual	Negative	TN	FN
	Positive	FP	TP

TN (True Negative) – Correct Prediction as normal

FN (False Negative) – Incorrect prediction of normal.

FP (False Positive) – Incorrect prediction of abnormal.

TP (True Positive) – Correct prediction of abnormal.

From the confusion matrix shown in Table 6, Accuracy (AC) can be obtained as:

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \times 100 \qquad (5.1)$$

Sensitivity and specificity are terms used to evaluate a clinical test. The sensitivity of a clinical test refers to the ability of the test to correctly identify those patients with the disease which are calculated from equation 5.2.

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \times 100 \qquad (5.2)$$

The specificity of a clinical test refers to the ability of the test to correctly identify those patients without the disease which is calculated from equation 5.3.

$$Specificity = \frac{TN}{(TN + FP)} \times 100 \qquad (5.3)$$

As we seen in chapter five Figure 5.11 contrast is our reference to identify the normal chest x-ray image and miliary tuberculosis chest x-ray image. Therefore, according to the physician observation, we calculate the texture feature. And contrast is our reference to predict the diagnosis depend on that result we take the minimum and maximum result of the contrast, then we compare to the physicians diagnosis and our result based on that if the contrast value is greater than 0.002 we classify to miliary TB or contrast value is less than or equal to 0.002 we classify normal lung image.

Let's see the confusion matrix,

Table 5: Confusion matrix II

	Negative	Positive
Negative	9	1
Possitive	0	10

Therefore we can calculate the accuracy, specificity, and sensitivity;

Here we get accuracy 95% and specificity is 100% and sensitivity also 100%.

CHAPTER SEVEN

7. Conclusion and Future Work

7.1. Conclusion

Worldwide Tuberculosis (TB) is a serious disease, so developing Computer Aid Diagnosis (CAD) systems is important, especially for rural areas with low medical infrastructure.

In this work, we have been demonstrating the possible theoretical extensions of Gray Level Co-occurrence Matrices (GLCM). The proposed method has been discussed in detail how to detect the miliary TB texture. Preprocessing and segmentation have been implemented, and texture feature extraction and several texture features are calculated. It was found that the statistical contrast feature provides sufficient information to predict the exact diagnosis of none miliary or miliary chest x-ray images. Others, like correlation, energy, and homogeneity have no sufficient information to discriminate healthy and pathological cases. We have also discussed our result and evaluated the performance. We achieved an accuracy of 95%, specificity 100%, sensitivity 100%. Based on these data it is concluded that the gray level co-occurrence matrix provides a good basis to identify miliary TB.

A drawback of this study is that the dataset is small and hence the statistical power of the results is not large enough for generalization. For the future, it is recommended to validate the findings using a larger test image database, when available.

Second, although a number of Ethiopian hospitals have been contacted, no test images with miliary TB cases could be obtained. Most government hospitals use film chest x-ray images rather than a digital chest x-ray. Even those hospitals using digital chest x-rays like Black Lion hospital in Addis Ababa or the Adarie hospital in Hawassa have just started using digital imaging. No systematic dataset could be made available for this project. Hence the research had to be limited to a few international sources like the Japanese Society of Radiological Technology database or websites like https://ceb.nlm.nih.gov/repositories/tuberculosischest-x-ray-image-data-sets/.

7.2. Future Work

A good improvement potential can be seen in the lung segmentation algorithm by a better elimination of the background image. Furthermore, additional texture analysis algorithms, e.g. based on distinct Fourier features of Miliary TB may be evaluated with the possible use of a support vector machine (SVM). The additional information should facilitate better discrimination and hence reduce the number of false positives and false negatives. Finally, for a practical application, a graphical user interface has to be developed along with conducting a field study.

References

[1] P. K. Anita Titus, "Detection And Extraction of Tuberculosis From Chest Ct Slices Chest Ct Slices," *Ijste - International Journal Of Science Technology & Engineering,* Vol. 1, No. 11, Pp. 289-292, 2015.

[2] World Health Organization(Who), "Global Tuberculosis Report," Who, Nederland, 2015.

[3] K. G. S. A. A. N. J. Raj, "Developments In Computer Aided Diagnosis Used For Tuberculosis Detection Using Chest Radiography: Survey," *Arpn Journal Of Engineering And Applied Sciences,* Vol. Vol. 11, No. 9, Pp. 5530-5535, May 2016.

[4] A. Belew, "Prevalence Of Smear Negative Pulmonary Tuberculosis Among Ptb Suspected Patients At Adama Referral Hospital, Adama, Ethiopia," P. 16, May 2014.

[5] Textile & Engineering Institute, Ichalkaranji, India, "Texture Analysis Of Tb X-Ray Images Using Image Processing Techniques," *Journal Of Biomedical And Bioengineering,* Vol. 3, No. 1, P. 55, 2012.

[6] M. M. R. S. A. S. D. Ashish Bhalla, "Acute Lung Injury In Miliary Tuberculosis," *Indian Journal Of Tuberculosis,* P. 125, 2002.

[7] Textile & Engineering Institute, Ichalkaranji, India, "Texture Analysis Of Tb X-Ray Images Using Image Processing Techniques," *Journal Of Biomedical And Bioengineering,* Vol. Volume 3, No. Issue 1, P. 53, 16 January 2012.

[8] Who, "Global Actions And Investments Fall Far Short Of Those Needed To End The Global Tb Epidemic.," *Global Tuberculosis Report,* P. 1, 2016.

[9] A. M. &. A. S. Surendra K. Sharma, "Challenges In The Diagnosis & Treatment Of Miliary Tuberculosis," *Indian J Med Res,* Pp. 703-704, May 2012.

[10] A. M. , A. S. Surendra K. Sharma, "Miliary Tuberculosis: A New Look At An Old Foe," *Journal Of Clinical Tuberculosis And Other Mycobacterial Diseases,* P. 18, 2016.

[11] A. M. , A. S. Surendra K. Sharma, "Miliary Tuberculosis: A New Look At An Old Foe," *Elsevier,* P. 14, 2016.

[12] W. K. Pratt, Digital Image Processing, Los Altos, California: John Wiley & Sons, Inc., 2001.

[13] M. R. T. A. N. S. Lingayat, "Spatial Domain Enhancement Techniques For Detection Of Lung Tumor From Chest X-Ray Image," *International Journal Of Application Or Innovation In Engineering & Management (Ijaiem),* Vol. 2, No. 8, P. 279, 2013.

[14] P. C. A. P. B. Chandrika, "Design And Development Of Pulmonary Tuberculosis Diagnosing System Using Image Processing Techniques And Artificial Neural Network In Matlab," *International Journal Of Electronics And Communication Engineering & Technology (Ijecet),* Vol. 4, No. 2, Pp. 368-369, 2013.

[15] S. M. F. A. S. R. T. M. T. A. Redouan Korchiynel, "A Combined Method Of Fractal And Glcm Features For Mri And Ct Scan Images Classification," *An International Journal (Sipij),* Vol. 5, No. 4, P. 87, 2014.

[16] N. G. D. K. Jaspinder Kaur, "Segmentation And Feature Extraction Of Lung Region For The Early Detection Of Lung Tumor," *International Journal Of Science And Research (Ijsr),* Vol. 3, No. 6, P. 2328, 2012.

[17] A. B. M. K. C. Aditya Goyal, "A Comprehensive Review Of Image Smoothing Techniques," *International Journal Of Advanced Research In Computer Engineering & Technology,* June 2012.

[18] S. K. B. M. T. H. R. K. D. A. A. V. Bram Van Ginneken, "Automatic Detection Of Abnormalities In Chest Radiographs Using Local Texture Analysis," *Ieee Transactions On Medical Imaging,* Vol. 21, Pp. 139-148, February 2002.

[19] F. Albregtsen, "Statistical Texture Measures Computed From Gray-Level Co-occurrence Matrices," 2008, P. 1.

[20] T. S. S. Sumathi Ganesan, "Classification Of Medical X-Ray Images For Automated Annotation," *Journal Of Theoretical And Applied Information Technology,* Vol. 63, Pp. 591-592, 31 May 2012.

[21] D. K. S. K. Anuradha, "Comparison Of Feature Extraction Techniques To Classify Oral Cancers Using Image Processing," *International Journal Of Application Or Innovation In Engineering & Management (Ijaiem),* Vol. 2, No. 6, Pp. 458-459, 2013.

[22] N. Z. A. V. Pawar, "Glcm Textural Features For Brain Tumor Classification," *Ijcsi International Journal Of Computer Science,* Vol. 9, No. 3, Pp. 355-358, 2012.

[23] H. D. Alaa Eleyan, "Co-Occurrence Matrix And Its Statistical Features As A New Approach For Face Recognition," *Turk J Elec Eng & Comp Sci,* Vol. 19, No. 1, Pp. 99-102, 2011.

[24] R. U. F. Fahim Irfan Alam, "Optimized Calculations Of Haralick Texture Features," *European Journal Of Scientific Research,* Vol. 5, No. 4, Pp. 546-549, 2011.

[25] M. B. K. Dr. S.A.Patil, "Lung Cancer Classification Using Image Processing," *International Journal Of Engineering And Innovative Technology (Ijeit),* Vol. 2, No. 3, Pp. 38-40, 2012.

[26] D. B. Biswajit Pathak, "Texture Analysis Based On The Gray-Level Co-Occurrence Matrix Considering Possible Orientations," *International Journal Of Advanced Research In Electrical, Electronics And Instrumentation Engineering,* Vol. 2, No. 9, Pp. 4207-4208, 2013.

[27] H. D. Alaa Eleyan, "Co-Occurrence Matrix And Its Statistical Features As A New Approach For Face Recognition," *Turk J Elec Eng & Comp Sci,* Vol. 19, No. 1, P. 99, 2011.

[28] R. U. F. Fahim Irfan Alam, "Optimized Calculations Of Haralick Texture Features," *European Journal Of Scientific Research,* Vol. 5, No. 4, Pp. 544-551, 2011.

[29] O. O. V. V. V. G. C. S. H. D. J. O. D. A. M. D. J. N. A. Hiram Madero Orozco, "Automated System For Lung Nodules Classification Based On Wavelet Feature Descriptor And Support Vector Machine," *Madero Orozco Et Al. Biomedical Engineering Online,* Pp. 8-10, 2015.

[30] D. B. Biswajit Pathak, "Texture Analysis Based On The Gray-Level Co-Occurrence Matrix Considering Possible Orientations," *Ijareeie,* Vol. 2, No. 9, P. 4206, September 2013.

[31] B. R.Nithya, "Comparative Study On Feature Extraction Method For Breast Cancer Classification," *Journal Of Theoretical And Applied Information Technology,* Vol. 33, P. 222, 2011.

[32] D. K. S. K. Anuradha, "Comparison Of Feature Extraction Techniques To Classify Oral Cancers Using Image Processing," *International Journal Of Application Or Innovation In Engineering & Management (Ijaiem),* Vol. 2, No. 6, P. 458, Jun 2013.

[33] A. K. S. C. L. F. J. S. F. C. Stefan Jaeger, "Automatic Tuberculosis Screening Using Chest Radiographs," *Ieee Transactions On Medical Imaging,* Vol. 33, No. 2, Pp. 233-243, February 2014.

[34] I. R. H. D. M. P. Amil Ahmad Ilham, "Wavelet Analysis For Identification Of Lung Abnormalities Using Artificial Neural Network," *Makassar International Conference On Electrical Engineering And Informatics (Miceei),* Pp. 281-284, November 2014.

[35] E. A. J. Melendez, "Multiple-Instance Learning For Computer-Aided Detection Of Tuberculosis," *Medical Imaging,* Pp. 1-6, 2014.

[36] R. M. &. T. F. K.A.G. Udeshani, "Statistical Feature-Based Neural Network Approach For The Detection Of Lung Cancer In Chest X-Ray Images," *International Journal Of Image Processing (Ijip),* Vol. 5, No. 4, Pp. 425-432, 2011.

[37] T. S. S. Sumathi Ganesan, "Classification Of Medical X-Ray Images For Classification Of Medical X-Ray Images For," *Journal Of Theoretical And Applied Information Technology,* Vol. 63, No. 3, Pp. 590-595, 2014.

[38] A. N. Hrudya Das, "An Efficient Detection Of Tuberculosis From Chest X-Rays," *International Journal Of Advance Research In Computer Science And Management Studies,* Vol. 3, No. 5, Pp. 149-153, May 2015.

[39] A. K. A. G. D. Jager, "Computer Aided Diagnosis Of Miliary Tuberculosis," *Researchgat,* Pp. 1-5, Feb 2016.

[40] J. V. A. G. K. Shafeena Basheer, "Computer Assisted X-Ray Analysis System For Detection Of Onset Of Tuberculosis," *International Journal Of Scientific & Engineering Research,* Vol. 4, No. 9, P. 608, 2013.

[41] F. M. A. B. Baharudin, "The Statistical Quantized Histogram Texture Features Analysis For Image Retrieval Based On Median And Laplacian Filters In The Dct Domain," *The International Arab Journal Of Information Technology,* Vol. 10, No. 6, Pp. 2-6, 2013.

[42] "Wikipedia," 31 August This Page Was Last Modified On 31 2016, At 09:10.. [Online]. Available: Https://En.Wikipedia.Org/Wiki/Normalization_(Image_Processing). [Accessed 9 Jan 2017].

[43] A. N. Hrudya Das, "An Efficient Detection Of Tuberculosis From Chest X-Rays," *International Journal Of Advance Research In Computer Science And Management Studies,* Vol. 3, No. 5, Pp. 149-153, May 2015.

[44] R. M. Sheenam Bansal, "A Comparative Analysis Of Iterative And Ostu's Thresholding Techniques," *International Journal Of Computer Applications,* Vol. 66, Pp. 45-46, March 2013.

[45] R. K. Ada, "Feature Extraction And Principal Component Analysis For Lung Cancer Detection In Ct Scan Images," *International Journal Of Advanced Research In Computer Science And Software Engineering,* Vol. 3, No. 3, P. 188.

[46] D.-C. H. A. L. Wang, "Simplified Texture Spectrum For Texture Analysis," *Journal Of Communication And Computer,* Vol. 7, No. 8, Pp. 45-53, 2010.

[47] B. R.Nithya, "Comparative Study On Feature Extraction Method For Breast Cancer Classification," *Journal Of Theoretical And Applied Information Technology,* Vol. 33, No. 2, Pp. 220-225.

[48] Mathworks Group, "Texture Analysis Using The Gray-Level Co-Occurrence Matrix (GLC M)," Mathworks, 15 Aug 2010. [Online]. Available: Https://Www.Mathworks.Com/Help/Images/Gray-Level-Co-Occurrence-Matrix-Glcm.Html. [Accessed 10 Jan 2017].

[49] S. S. K. S. T. C.Nageswara Rao, "Cooccurrence Matrix And Its Statistical Features As An Approach For Identification Of Phase Transitions Of Mesogens," *International Journal Of Innovative Research In Science Engineering And Technology,* Vol. 2, No. 9, P. 4553, 2013.

[50] T. A. Pham, "Optimization Of Texture Feature Extraction Algorithm," *Computer Engineering,* Pp. 1-2, 2010.

[51] R. M. Sheenam Bansal, "A Comparative Analysis Of Iterative And Ostu's Thresholding Techniques," *International Journal Of Computer Applications,* Vols. 66– No.12, Pp. 45-46, March 2013.

Appendices

Appendix I

This is some matlab functions that we have been used in our experiment.

```
%read an input image%
f = imread('normal11.jpg');
imshow(f)
title('orginal');

f1=imresize(f,[2000,2000]);
f2 = rgb2gray(f1);
figure, imshow(f2)
title('gray level image');
figure;
 %f2 = rgb2gray(f1);
%f2=imresize(f,[2000,2000]);
 subplot(2,2,1);
 imshow(f2)
 title('scaled gray level image');
  %%histogram
% h = imhist(f2);
 subplot(2,2,2);
 imhist(f2);
 %plot(h);
 %imshow(h);
 xlabel('gray level value');
 ylabel('probability value');
 title('gray level image histogram');
 %%histogram equalization
 figure;
se = strel('disk', 15);
fo = imopen(f2, se);
subplot(2,2,1);
 imshow(fo)
 title('open');
 subplot(2,2,2);
 imhist(fo);
  xlabel('gray level value');
 ylabel('probability value');
 title('opened image histogram');
 im2 = histeq(fo);
 subplot(2,2,3);
  imshow(im2);
 title('gray level image equalized');
 subplot(2,2,4);
```

```
 imhist(im2);
 axis([0 255 0 2000])
 xlabel('gray level value');
 ylabel('probability value');
 title('gray level image equlaized histogram');

 %// converted to double.
im1 = im2double(im2);
%// Create normalized images
im1Norm = (im1 - mean(im1(:))) / std(im1(:));
%// Convert back to uint8
im1Norm = im2uint8(im1Norm);
%//Side by side comparison
%// Left column is the original
%// Right column is the processed image
figure;
subplot(2,2,1);
imshow(im1Norm);
title('normalized image');
subplot(2,2,2);
imhist(im1Norm);
xlabel('gray level value');
 ylabel('probability value');
title('normalized image histogram');
 level = graythresh(im1Norm);
 BW = im2bw(im1Norm, level);
 figure,imshow(BW)
 title('segmented image');
 glcm = graycomatrix(BW)
stats = graycoprops(glcm)
%plot(glcm, stats, 'Contrast', 'Correlation', 'Energy',  'Homogeneity');
```

Appendix II

This is miliary tuberculosis chest x-ray image and normal chest x-ray image that we used in the experiment.