JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF INFORMATION SCIENCE

AUTOMATIC PULMONARY TUBERCULOSIS BACILLI DETECTION FROM SPUTUM SMEAR MICROSCOPY IMAGE USING IMAGE PROCESSING TECHNIQUES

BY:

DIRRIBA ABDETA

NOVEMBER, 2018

JIMMA, ETHIOPIA

JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF INFORMATION SCIENCE

AUTOMATIC PULMONARY TUBERCULOSIS BACILLI DETECTION FROM SPUTUM SMEAR MICROSCOPY IMAGE USING IMAGE PROCESSING TECHNIQUES

A Thesis Submitted in Partial Fulfillment of the Requirements for Degree of Masters of Science in Information Science (Information and Knowledge Management)

BY:

DIRRIBA ABDETA

Principal Advisor: Worku Jimma (PhD)

Co-advisor: Chala Diriba (Assistant Professor)

November, 2018

Jimma, Ethiopia

JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF INFORMATION SCIENCE

AUTOMATIC PULMONARY TUBERCULOSIS BACILLI DETECTION FROM SPUTUM SMEAR MICROSCOPY IMAGE USING IMAGE PROCESSING TECHNIQUES

BY:

DIRRIBA ABDETA

As members of the Board of examining of the MSc. thesis open defense examination of the above title, we members of the board (listed below), read and evaluated the thesis and examination.

Name	Title	signature	Date
_____	Chair Person	_____	_____
_____	Principal Advisor	_____	_____
_____	Co-Advisor	_____	_____
_____	External Examiner	_____	_____
_____	Internal Examiner	_____	_____

DECLARATION

I declare that this thesis is my original work and it has not been presented for a degree in any other Universities. All the material sources used in this work are duly acknowledged.

_____

Dirriba Abdeta

November, 2018

This thesis has been submitted to the department of Information Science for examination with our approval as University Advisors:

Principal Advisor: Worku Jimma (PhD) …….

Co-Advisor: Chala Diriba (Assistant Professor)……...

November, 2018

Jimma, Ethiopia

DEDICATION

This work is dedicated to my father who died after one month when I was graduated my BSc. and my lovely wife Buzunesh Saketa

ACKNOWLEDGE

First and for most, I would like to give all glory and honor to the Gracious God who provided me everything to finish this research. I gratefully, acknowledge my advisors Worku Jimma (PhD) and Chala Diriba (Ass. Professor) for their unreserved, genuine comments and advise throughout this research work. I express my warm thanks to you, the team case office and experts of office the National Tuberculosis Reference Laboratory Ethiopian Public Health Institute for your kindly help and support. Especially, Dr. Mangistu, this research would not be successful without your help on the on sample data collection of sputum smear images. I would like to acknowledge Ambo University and Jimma University Department of Information Science for facilitating and provided me with assistants required for this thesis. I would also like to express my gratitude to all my family who helped me in different ways to accomplish this thesis. Finally, I would like to thank all my friends for their moral support and encouragements throughout my study.

Table of Contents

List of Tables

List of Figures

LIST OF ACRONYMS AND ABBREVIATIONS

AFB: Acid Fast Bacilli

APTBBD: Automatic Pulmonary Tuberculosis Bacilli Detection

ANN: Artificial Neural Network

ATS: American Thoracic Society

CDC: Centers for Disease Control and Presentation

CMYK: Cyan, Magenta, Yellow and Black

DIP: Digital Image Processing

EPHI: Ethiopian Public Health Institute

EHNRI: Ethiopian Health & Nutrition Research Institute

EPTB: Extra Pulmonary Tuberculosis

FEMoH: Federal Democratic Republic of Ethiopia Ministry of Health

GUI: Graphical User Interface

HIV: Human Immunodeficiency Virus

ISTBC: International Standard for Tuberculosis Care

MATLAB: Matrix Laboratory

MTB: Mycobacterium Tuberculosis

NTM: Non Tuberculosis Mycobacteria

PC: Personal Computer

PTB: Pulmonary Tuberculosis

PTB+: Smear Positive Pulmonary Tuberculosis

PTB-: Smear Negative Pulmonary Tuberculosis

RGB: Red Green Blue

SVM: Support Vector Machine

SWC: Step Wise Classification

WHO: World Health Organization

ZN: Ziehl-Neelsen

ABSTRACT

Tuberculosis is one of the deadly diseases worldwide developing countries including Ethiopia. It is caused by *Mycobacterium* tuberculosis that influenced fundamentally on human body lung in form of pulmonary tuberculosis disease. Sputum smear microscopy is the most widely used diagnostic tools in developing countries. The main aim of this study is to develop automatic PTB bacilli detection from microscopic sputum smear images using image processing techniques. In this study, an algorithm based on image processing technique is developed for identification of pulmonary tuberculosis bacilli in digital image of stained sputum smear. The techniques was used in this study, Gaussian filter to remove noise, contrast enhanced to enhance the quality of image and K-mean cluster used to separate image into region in image segment process. In addition, support vector machine and k-nearest neighbor classifiers were used to identify bacilli, which classified the computed based on combined both morphological and color features from sputum smear images in two classes which are bacilli detect and non-bacilli detect. Total sample size of image dataset of 180 from stained sputum images of PTB bacilli infected were obtained from EPHI. The accuracy performance measured shows that SVM algorithm found to be 94.4% and for KNN algorithm it was 92.6%. The results and observations show that SVM is a more suitable more than KNN classifier for the classify PTB bacilli detect from pathologists reading predefined image obtained from sputum smear images database. The accuracy, sensitivity, specificity and F-measures improved the performance of the prototype results of SVM are 94.4%, 95%, 94% and 96% respectively. The future work will look into the problem related to diagnosis of PTB drug resistant cases in the proposed algorithm by a clinical setting.

Keywords: Pulmonary Tuberculosis, Sputum Smear Microscopy, Image Processing Techniques, Automatic detection

CHAPTER ONE

1. INTRODUCTION

1.1. Background of the study

Tuberculosis (TB) is one of the leading deadliest diseases in the world, caused by bacilli called *Mycobacterium* tuberculosis. It is the top ten causes of death worldwide in 2015 (WHO, 2016). TB is the second leading cause of death from communicable diseases worldwide next to HIV (WHO, 2017). Therefore, it remains a significant global public health problem affecting about one-third of the world's population despite the implementation of preventive and control measures over the years (WHO, 2017). TB is an airborne bacteria disease. A pulmonary Tuberculosis (PTB) bacillus is a type of tuberculosis, which is a contagious, infectious disease that primarily attacks lungs. It is communicable disease and transmitted whenever individuals with active PTB by cough, sneeze, laugh, and release droplets in the air (Menzies and Khan, 2007). Effective treatment of PTB bacilli cases depends on close cooperation between patients and healthcare providers. Healthcare providers are such as doctors, nurses, pharmacists and laboratory technicians, health management and workers in health care centers. With the appropriate way of antibiotic treatment the disease can be cured. However, the diagnosis of PTB bacilli is very challenging tasks for the health care providers (Mohajan, 2015).

The most common worldwide method for diagnosis of patients' with active PTB bacilli disease is by using sputum smear microscopy, which can determine the presence of bacteria in sputum smear image (WHO, 2017). In microscopy examination, sputum smear images are used after sputum smear staining. It is used in the diagnosis of PTB bacilli is the manual microscopic examination of a Ziehl-Neelsen (ZN) stained sputum smear which is a time consuming and error-prone process (Panicker et al., 2016). Lab technician examines manual diagnosis of PTB bacilli

disease using the stained smear image under the microscopy crucially depends on the number of viable or latent mycobacteria in the sputum smear image, which are seen as red colored and rod-shaped objects.

In Ethiopia the most widely used techniques for the diagnosis of PTB bacilli is sputum smear image microscopy, specifically by using Ziehl-Neelsen (ZN) known as acid fast bacilli (AFB) stained (EHNRI, 2009). The limitation of AFB stained is manual diagnosis processes, that do not distinguish between viable and dead organism, limited sensitivity and high bacterial load 5,000-10,000 AFB /ml is required and misses less than 50% of TB cases and limited specificity (Lumb et al., 2013). Sputum smear microscopy is a common manual diagnosis for PTB bacilli detection. Therefore, it is time consuming, very tedious (laborious), subject to poor specificity (human error) and required highly trained personnel (Forero et al., 2006). The accuracy of the diagnosis decision making ultimately depends on the skill and experience of the technicians.

Medical imaging is the technique and process of creating visual representations of the interior of a body for clinical image analysis and medical intervention, as well as visual representation of the function of physiological (organ or tissue) characteristics (Thirumaran & Shylaja, 2015; Deserno, 2011; Gao, 2013). It has become an essential component in many fields of medical and laboratory research and clinical practice. Advanced image processing and analysis techniques of medical imaging are increasingly used in medicine (Deserno, 2011). Image process quality plays an important role in medical imaging and is expected to provide quantitative data useful for patient treatment and care (Deserno, 2011). The range of image processing and analysis of medical imaging is to improve the quality of acquired images, and extract quantitative information from an efficient medical imaging data and accurate manner.

Image processing is a set of technologies, which is an image data analysis and processing algorithms and tools to improve the interpretation of some image information more useful for patient treatment and care (Deserno, 2011). Image processing allows the extraction of useful parameters, and increases the likelihood of detection of small lesions more accurately (Gao, 2013). In medical image processing including medical image there are three main objectives: the first is reconstruction tomography imaging technology, the second is quality to improve the image contrast, uniformity and spatial resolution view and the third is, to extract useful diagnostic qualitative and quantitative image ready messages (Gao, 2013). Image processing uses various techniques such as image acquisition, preprocessing, segmentation, feature extraction and classification to improve the quality of images and assists a lot in diagnosing PTB bacilli to a great extent (Deun, 2002, Chitradevi and Srimathi, 2014, Medjahed, 2015). Many researchers used various image processing techniques to improving the diagnosis of PTB bacilli disease for better results (Ayas et al., 2014; Santiago-Mozos et al., 2014; Raza et. al., 2015; Goel et al., 2017).

The techniques used in this study, for image preprocessing were Gaussian filter, contrast enhanced, K-mean cluster algorithm used to separate image into region in image segment process. In addition, SVM was used to identify PTB bacilli, which classified the computed both morphological and color features from sputum smear images in two classes are bacilli positive and negative detect. The detection of PTB bacilli from sputum smear image manually i.e., under microscopy by eye is time consuming and prone to error as a lab technician or pathologist has to manually change the field of view of the microscopy several times until the entire stained sputum smear image on the glass slide is viewed. Thus this study was initiated with the main aim to

develop automatic PTB bacilli detection from microscopic sputum smear images using image process techniques. The developed system is considered as solution to the problems manually diagnosis such as the burden in clinician's workload, sensitivity, specificity and time consuming (the number of slide that can be screened) (Deun, 2002; Forero et al., 2006). In additional the proposed system can assist physicians/pathologists to diagnosis PTB bacilli at early stage. As a result, it can save time, increase accuracy and enhance sensitivity during diagnosis PTB bacilli disease.

1.2. Statement of the Problem

Tuberculosis is a communicable disease and remains a major public health problem throughout the world. In 2015 an estimated, 10.4 million people fell ill with TB cases worldwide and 1.8 million dead from the disease including 0.4 million among people with HIV, 1.1 million men, 0.5 million women and 0.2 million children (WHO, 2016). In Ethiopia incidence of TB was estimated to be 192 per 100, 000 in 2015 (WHO, 2016). Ethiopia is one of the 22 TB high burden countries and PTB bacilli remains one of the leading causes of mortality due to infectious diseases in the country. PTB is one of the forms of TB which is caused bacteria *mycobacterium tuberculosis*, and is an airborne infection that primarily affects lungs. The disease is contagious and it can spread through the air from an infected person to normal person. PTB bacilli can be treatable at early stage diagnosis. However, it has becoming more and more of serious problem, particularly in developing countries including Ethiopia (WHO, 2017). PTB accounts for 85% of all TB cases, and it is classified in smear positive PTB bacilli comprises 75-80% of PTB bacilli and smear negative PTB bacilli compromises 20-25% of PTB bacilli cases, worldwide (FEMOH, 2013).

The basic difficulties are diagnosis PTB bacilli manually screening. The two the main important sputum smear microscopy techniques used for the PTB bacilli diagnosis are Fluorescence microscopy and conventional microscopy. A fluorescent microscopy is used to examine Auramine-O stained sputum smear specimens, while conventional microscopy is used to examine Ziehl-Neelsen stained sputum (Swaminathan et al., 2010). Fluorescence microscopy is on average 10% more sensitive than conventional microscopy in detecting PTB in sputum smear. It is an expensive diagnostic method because of the high costs and its maintenance, and that is why they are used less. Conventional microscopy is the method used in developing countries, due to low cost and easy of equipment maintenance for screening PTB bacilli diagnostic tools, compared to fluorescence microscopy. Conventional microscopy is a widely used method for PTB bacilli detection in public health programs in developing countries although the specificity result detect rate is only about 50–60 % PTB bacilli cases (Singh et al., 2015; Panicker et al., 2016). In conventional microscopic examination, sputum smear images are used after sputum smear image staining. Three samples of coughed up or expectorated sputum is prepared and stained with Ziehl Neelsen stain and examined under microscope for the presence of Acid Fast Bacilli.

WHO (2017) recommended, AFB staining smear microscope to diagnosis PTB bacilli. But, number of AFB is tedious and labor intensive task. The low quality, inconsistent slide staining technique, variation in human perception, and fatigue lead to sensitivity as low as 40% missed PTB bacilli diagnosis, especially in the scanty specimens of the stained sputum which usually take longer time to identify bacilli. In Ethiopia, Ziehl–Neelsen stain (ZN), also known as AFB stain is presently used technique for PTB bacilli detection. A laboratory technician is expected to spend at least 15 min per slide, limiting the number of slides that can be screened. Manual

diagnosis for the PTB bacilli identification involves a labor intensive task and produce false results (Chang et al., 2007). This method has been around for many years and it has a shortage of properly trained technician, but they can often fail to detect PTB bacilli cases due to stress of a heavy workload and up to half of the patients suffering from PTB bacilli, delaying or even missing their diagnosis to get the faulty results (false positive bacilli) (Deun, 2002; Forero et al., 2006).

In additional, PTB bacilli are difficult to detect and quite complex process, especially in children, people living with HIV or those with extrapulmonary TB forms of disease. The diagnosis of PTB bacilli using rapid and accurate methods is a crucial step in the control of PTB bacilli disease. Therefore, the present study attempted to develop automatic pulmonary tuberculosis causing bacilli detection from sputum smear microscopy image using image processing techniques, which can reduce the burden on the pathologist or technician, reduce human error, and improve sensitivity of the test, and reduce the time required to diagnosis of PTB bacilli.

To solve problems identified, this study attempted to answer the following research questions:

- What are appropriate algorithms that are used to image processing techniques to identify PTB bacilli detect from sputum smear image was collected?
- What are the features extractions that distinguish the two classes of PTB bacilli detect?
- To what extent does developing an automatic PTB bacilli detection system using image processing techniques helps pathologists for decision making?
- What is the prototype system performance to classify PTB bacilli detected?

1.3. Objectives of the study

1.3.1. General Objective

The general objective of this study was to develop automatic PTB bacilli detection from sputum smear microscopy image using image processing techniques.

1.3.2. Specific Objectives

1. To collect patients stained sputum sample and acquire sputum smear images which can used for develop prototype system,
2. To select which is the best suitable algorithm used to classify PTB bacilli detection from sputum smear images,
3. To represent feature extract to identify PTB bacilli detect based on sputum smear images
4. To develop a prototype for automatic PTB bacilli detection system that can help pathologist for decision making,
5. To evaluate the performance of the system using specimen sample of image data.

1.4. Scope and Limitation of the Study

The scope of this study was to develop a prototype PTB bacilli detection system from stained sputum smear microscopy images by using image processing techniques. PTB is a type of tuberculosis diseases, which is the chronic infection of the primarily affects the lung caused by bacteria. But there is various bacteria species causes TB disease in *Mycobacterium tuberculosis complex* such as *Mycobacterium tuberculosis (MTB), M. bovis, M. africanum, M. microti and M. canettii* (Linda et al., 2002). PTB bacilli are caused by *Mycobacterium tuberculosis* bacteria which is the most common TB disease (Menzies and Khan, 2007). Therefore, this study was limited to PTB bacilli which are diagnosed by conventional microscopy from stained sputum smear microscopy image. For the accomplishment of this work, the stained sputum smear images were acquired from Ethiopian Public Health Institute (EPHI) during data collection period of the study was conducted from March 2018 to September 2018 G. C. The dataset has a total 180 (positive and negative) PTB infected from stained sputum smear image by using the microscopy diagnostic techniques and using Leica Microsystems microscopy connected to PC at EPHI. To identify PTB bacilli in the cases where there existed high degree of the quality of the camera, the image acquisition environment and other imaging factors may affect the result was considered at the limitation of this study. Lack of experience related to PTB bacilli computation included debris or undesired object of the bacilli sputum smear image and the bacilli overlapping with unusual morphology was limited.

1.5. Significance of the study

The developed prototype system is useful for diagnosis of PTB bacilli because of accurate and rapid in diagnosis, save time of laboratory technician, improved safety of patients (reduced health risk) at early stage of diagnosis, improved efficiency in health care outcome, and improved quality of care. It is helpful in providing the opinion to the pathologists for their finding causes as well as assisting them in making decision regarding diagnosis with increased accuracy. In addition, it was used in the health agencies, particularly remote or rural areas where there is lack of experts which results in improved efficiency, better quality, and reduced instances of clinical errors.

Moreover, the developed prototype system has an ability to store patient records for future reference, thereby allowing better and faster decision making. The stored patient records can also be sent across the Internet to experts located elsewhere for a second opinion (Panicker et al., 2016). The developed prototype study can improve remote patient diagnosis, screening and examination of stained sputum smear PTB bacilli problem at a reduced cost. Moreover, it can reduce dependencies on medical experts. The health hazard within the rural communities and emerging urban cities can be reduced. It is also expected that the output of the study result in a system that increase the speed of PTB bacilli disease diagnosis. The developed system helps for researchers who conducted relate studies to use as a reference and motivation to do more research on this area.

1.6. Methodology of the study

There are different approaches and tools used for developing a prototype automatic diagnosis system for PTB disease causing bacilli. In order to achieve the objectives of this study, the following methods and techniques were employed.

1.6.1. Research design

This study follows experimental research design. According to Goodwin (2010), experimental research design is a systematic research study in which the researcher manipulated and controlled testing to understand causal process. It is used for the researcher's most powerful tool for identifying cause and observes the diagnostic of the results. Dipanwita et al. (2005) used experimental research design to develop medical image system for diagnosis of communicable disease by knowledge acquiring from domain experts. It is important for an experimental research to establish cause and effect of phenomenon, which means, it should be that effects observed from an experiment are due to the cause.

Experimental research can be either basic or applied in its goals, and it can be conducting a laboratory test or other fields. Experimental researches that take place in the field research outside of the laboratory, including both experimental and non experimental methods. It is a way of gaining knowledge by means of direct and indirect experience and evidence gathered data. In this study, experimental method was used for image processing techniques to designing a prototype system and testing. As a result, in this study the researcher used experimental method for model building, analysis, and prototype development and testing, whereas non-experimental method was used for knowledge elicitation through discussion with experts and document review.

1.6.2. Study Area

The study focuses on Ethiopia Public Health Institute (EPHI). EPHI found in capital city Addis Ababa and it has contributed a lot for the improvement of public health and nutrition problems of the country in different names and organizational structures. Based on health and nutrition priority areas, the Institute had been setting different strategies at different times to address the public health problem. Currently the Institute is focusing on priority disease research and strengthening the national public health laboratory services in the country. EPHI National TB Reference Laboratory is responsible for the provision of high level diagnostic of tuberculosis laboratory testing services for patients and specimens referred from all Regional and Federal Health facilities. EPHI also provides information on services offered, quality assurance, laboratory operations, sample collection, transport and agreed turnaround times for end customers. It provides a high quality laboratory services and information on how to request laboratory tests of reference laboratory to the health providers and customers. EPHI was selected for this study which is proper specimen collection and handling is obtaining a valid and timely national PTB reference laboratory test results.

The main data source was used for this study image dataset of previously solved PTB bacilli disease cases. Dataset for digital image processing purpose from these areas used for prototype system testing was collected. The organization was selected with considering the seniority issue and also easy to get experienced domain experts, prevalence of the disease dataset (image data acquisition). In order to get the required information for the research and comments at different stage of experimentation and evaluation, discussion and unstructured interview was conducted with purposively selected domain experts who diagnoses and treat PTB patients at EPHI.

1.6.3. Population and sampling techniques and Size

The target populations of this study are the domain experts of the EPHI National TB Reference Laboratory staff. Both interview and documents analysis was done to acquire knowledge from domain experts. The domain experts are those who diagnosis and treat PTB patients at EPHI were interviewed. In this study purposive sampling technique was used to select domain experts for knowledge acquisition and to collect previous PTB patient cases sampled from study sites. The selection criterion of domain experts for the study is based on the professions or expertise, educational qualification level, years of experience on PTB bacilli diagnosis. A domain experts were purposively selected for the interview from EPHI due to ease access domain knows how about PTB bacilli diagnosis from stained sputum smear images. Accordingly, one doctors and two lab technician were selected from the study site for interviews to collect information about specimen sputum smear images and image data acquisition. For this study, a total sample size of image dataset of 180 (100 positive and 80 negative) from stained sputum images of PTB bacilli infected were obtained from EPHI.

1.6.4. Data collection

The sample of dataset consists of sputum smear slide used in this study was collected at EPHI. Sputum smears images were collected from stained sputum smear specimen of patients with PTB disease. A total sample of data (180) was collected from sputum smear microscopy through ZN stain process using Leica Microsystems microscopy connected to computer (PC) by domain expert. Dataset was collected using examine ZN-stained smears with a 10x100 objective lens under oil immersion views at EPHI. The image acquisition was captured from stained sputum smear slides at 100X magnifications. The pixels resolution was 696x514 pixels. The images were saved in Joint Photograph Experts Group (JPG) file format, with 24 bit per pixel, in RGB

(red, green and blue) color space. The images acquired can then be stored in the computer and can be processed in real-time or offline mode using image processing techniques. These image databases are the result of a real patient's sputum smear specimen that are prepared through ZN staining process using microscopic examination by domain experts. The method data source both primary and secondary data source were used for this study. The primary data was collected through interview from domain experts. Whereas, secondary data was collected from published articles and journals, TB program reports (WHO TB reports).

1.6.5. Implementation Tools

MATLAB is implementing tools, which stands for matrix laboratory, is a very powerful technical language (Chandrika et al., 2013). The presence of numbers of toolboxes has made MATLAB 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 (Chandrika et al., 2013). MATLAB is a powerful technical language for computing used to develop an application to improve digital image analysis (Houcque, 2005). It integrates computation, visualization, and programming environment. Furthermore, MATLAB is a modern programming language environment. This tool has a great capability on array based data processing and excellent tools for teaching and research. The basic data structure in MATLAB is the array, an ordered set of real or complex elements. This object is naturally suited to the representation of images, real-valued, ordered set of color or intensity data. MATLAB stores most images as two-dimensional arrays (i.e., matrices), in which each element of the matrix corresponds to a single pixel in the displayed image. Pixel is derived from picture element and

usually denotes a single dot on a computer display. Accordingly, MATLAB platform will be used to design and create the GUI of prototype of the system (www.mathworks.com, 2018).

1.7. Evaluation Methods

After the prototype system was developed conducting the performance of the prototype system is important. The analysis of system performance conducted in terms of classification accuracy rate, sensitivity, specificity and F-measure are evaluation measures used to evaluate the results of the prototype system using information from the collected sample of specimen smear image (Forero et al., 2006; Sadaphal, 2008; Dendere, 2009; Choudhary et al., 2013).

1.8. Ethical Consideration

Approval letter of ethical clearance was obtained from the Research and Ethical Review Board (RERB/IRB) of College of Natural Sciences, Jimma University for Ethiopian Public Health Institute. Confidentiality was ensured during the data sample preparation and sample collection and interview of domain expert; thus name and address of the patient was not be record at step and the data stained sputum smear image was used only for the research purpose.

1.9. Operational Definitions

Acid Fast Bacilli: Sputum, or phlegm, is often used to test for Mycobacterium tuberculosis, to find out if a patient has TB.

Bacilli: is a genus of gram-positive, rod-shaped bacteria that causing pulmonary tuberculosis

Detection: is the discovery of something which is supposed to be hidden.

Feature extraction: starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations.

Image: Is an array, or a matrix, of square pixels (picture elements) arranged in column and rows.

Microscope: is an instrument used to see objects that are too small to be seen by the naked eye.

Microscopy: is the science of investigating small objects and structures using such an instrument.

Negative pulmonary tuberculosis: sample sputum smear test site usually means you have not been infected with tuberculosis bacteria.

Pixel: is the smallest unit of a digital image or graphic that can be displayed and represented on a digital display device.

Positive pulmonary tuberculosis: A positive tuberculosis sample sputum smear test only tells that a person has been infected with tuberculosis bacteria (bacilli).

Pulmonary Tuberculosis is caused by the bacterium Mycobacterium tuberculosis, which is contagious. This means the bacteria is easily spread from an infected person to someone else, by breathing in air droplets from a cough or sneeze of an infected person.

Sensitivity: it is statistical measures of the performance of a binary classification of medical tests ability to correctly detect ill patients who do have the condition. In the example of medical test used to identify a PTB bacilli disease, the specificity of the test is the proportion of the people who test positive for the disease among those who have the disease.

Specificity: it is statistical measures of the performance of a binary classification of medical tests ability to correctly reject healthy patients without condition. In the example of a medical test for diagnosis a PTB bacilli disease, the specificity of test is the proportion of healthy patients known not to have the disease, who will test negative for it.

Sputum smears image: the mucus that comes up when the patient coughs. The samples are tested for PTB bacteria.

Ziehl-Neelsen Stain: It is a special bacteriological stain used to identify acid-fast organisms, mainly mycobacteria tuberculosis.

1.10. **Organization of the study**

This thesis is divided into five chapters. Chapter one discussed about background of the study, the statement of the problems, and research questions of the study, general and specific objectives of the study, scope and limitations of the study, significant of the study, and the methodology that the researcher used to conduct this study. Chapter two discussed about conceptual and review of related works to this study. In this chapter, the researcher addressed about tuberculosis, PTB bacilli and its characteristics, symptoms, factors and diagnosis tools, medical image processing, image processing techniques, and the steps of digital processing techniques and related works which are relevant for this study. In Chapter three, the designing and implementation of automatic PTB bacilli detection system includes designing the system architecture were discussed.

In Chapter four, the experimental result was used for the performance evaluation of the proposed prototype. It also includes development environment, binary image analysis, feature analysis, experimental results and discussion. Moreover, GUI implementation and the results of the all steps of automatic detection system were used to identify PTB bacilli were discussed. Finally, chapter five includes conclusion and recommendation; as well as the future research work.

CHAPTER TWO

2. LITERATURE REVIEW

2. 1. Overview of Tuberculosis

Tuberculosis (TB) is a chronic infectious disease caused by the *mycobacterium tuberculosis* and remains a major public health problem around the worldwide. About one third of the world's population infected with TB (WHO, 2016). It is mortality rates are high and is one of the top ten diseases causes of death next to HIV/AIDS as the greatest killer worldwide due to a single infectious agent (WHO, 2016). In 2015 an estimated, 10.4 million people fell ill with TB cases worldwide and 1.8 million dead from the disease including 0.4 million among people with HIV, 1.1 million men, 0.5 million women and 0.2 million children (WHO, 2016). Worldwide, 13% of TB patients have HIV co-infection, and as many as 37% have HIV co-infection in parts of African Region, which accounted for 75% of TB cases among people living with HIV worldwide (WHO, 2016). In 2016, over 95% TB deaths occur in low and middle countries. Ethiopia is among the high TB endemic country and most heavily affected by the HIV and TB infectious in the world (WHO, 2016). It ranks 11^{th} in the list of 22^{th} high burden countries and 7^{th} in Africa countries (WHO, 2016). In generally, a relatively small proportion of people infected with mycobacterium tuberculosis (MTB) will develop TB disease. However, the probability of developing TB is much higher among people infected with HIV. TB is also more common among men than women, and affects mostly adults in the economically productive age groups.

TB that primarily affects the lungs is known as Pulmonary Tuberculosis (PTB), but it can also affect the other parts of the body and known as extra PTB, for example lymph nodes, kidneys, bones, joints, etc (Menzies and Khan, 2007). The causative organisms of PTB diseases are five

closely Mycobacterium grouped in the *Mycobacteria tuberculosis complex: Mycobacterium tuberculosis, Mycobacterium bovis, Mycobacterium africanum, Mycobacterium microti, and Mycobacterium canettii.* PTB is caused mainly by M. tuberculosis (MTB), the natural reservoir of which is human beings. This is a fastidious, slow growing, strictly aerobic, lipid-rich, hydrophobic and acid–fast bacterial rod shaped (ATS, 2000).

2.1.1. Pulmonary Tuberculosis (PTB) Bacilli Characteristics

PTB is bar shaped straight or slightly tortuous. It is a non-motile, rod-shaped obligate aerobe bacterium. For this reason, in the classic case during active of PTB bacilli disease, MTB complexes are always found in the well aerated upper lobes (air sacs) of the lungs. AFB will be as a measurement of width rods are 0.2-0.6 μm (micrometers) and length 1-10 μm (micrometers) (Panicker et al., 2015). Bacilli may contain heavily stained areas called beads. PTB bacilli is a small, very slow-growing bacterium that can live only in people. It is an aerobic bacterium, meaning it needs oxygen to survive. A healthy individual is infected by inhaling the droplets, which settle and grow in the lungs resulting in the development of primary infection, which usually passes unobserved. Depending on the circumstances a person with primary infection progress after a latent period of months or years to post primary PTB, which results in more extensive involvement of the lung body.

2.1.2. PTB Bacilli Infection

PTB is a disease caused by bacteria that are spread through the air from person to person. If not treated properly, PTB disease can be serious. People infected PTB bacteria who are not sick may still need treatment to prevent PTB disease from developing in the future. The bacteria can live in the body without causing sickness is called Latent PTB (LPTB). In most people who breathe

in PTB bacteria and become infected, the body can fight the bacteria to stop them from growing. This means the immune system protects the individual from getting sick. People with latent PTB infection do not feel sick, do not have symptoms, and cannot spread PTB to others but they must be tested. If PTB bacteria become active in the body and multiply, the person will go having the latent PTB infection to being sick with PTB disease. The reason people with latent PTB infection are often prescribed treatment to prevent them from developing PTB disease. People with the bacteria have 10% lifetime risk of getting sick with PTB. When the person start showing symptoms, he/she may become contagious and have PTB (Mitruka et al., 2011; Oeltmann et al., 2014; CDC, 2014).

2.1.3. PTB Bacilli Symptoms

People with disease usually have cause one or more of symptoms and may spread bacteria to others. Based on the symptoms there are two forms namely, PTB and extra PTB (EPTB) disease. PTB has several manifestations. PTB bacteria most commonly grow in the lungs, and can cause symptoms such as have cough almost continuously (a bad cough that lasts two weeks or longer), cough up blood or sputum (mucus from deep inside the lungs), consistent fever including low-grade fevers, night sweats, pains in the chest, unexplained weight loss and have chills (cools) (Oeltmann et al., 2014, CDC, 2014). While EPTB disease may symptoms related to the part of the body that is affected. For example, TB of the spine may cause back pain; TB of the kidney may cause blood in urine; TB of the meningitis may cause headache or confusion; TB of the larynx can cause hoarseness, loss of appetite, unexplained weight loss, night sweats, fever, and fatigue (Oeltmann et al., 2014; CDC, 2014). EPTB disease should be considered in the differential diagnosis of ill person who have systemic symptoms and who are at high risk factor for TB. Both PTB and EPTB diseases symptoms can be caused by other diseases.

2.1.4. PTB Bacilli Risk Factors

The risk factor for getting PTB is highest for people who are in close contact with those who have active PTB (CDC, 2014). This includes being around family or friends with PTB or working in places like medical facilities or institutions that house people with PTB. These places are often correctional facilities, group homes, nursing homes, hospitals and shelters. Some people develop PTB disease soon after becoming infected (within weeks) before their system can fight the PTB bacteria (Gagneux et al, 2006). Other people may get sick years later, when their immune system becomes weak. For person whose immune systems are weak, especially those with HIV infected the risk of developing PTB disease is much higher than for person with normal immune systems (VanRie et al., 2011). People also at risk for developing PTB disease are older adults, small children, smokers, people with autoimmune disorder, such as lupus, people with lifelong conditions, such as diabetes or kidney disease, people who use drug injection, and people who are immune compromised, such as those taking chemotherapy, chronic steroids or who have HIV or AIDS (Gagneux et al, 2006; CDC, 2014). PTB is transmitted by means of invisible droplet nuclei containing the organisms that have left the reservoir during breathing, sneezing or couching and it occurs indoors where droplet nuclei can stay in the air for long time.

2.1.5. PTB Bacilli Diagnosis

PTB diagnosis is the processing of testing the existence of tuberculosis caused bacteria in patient's sputum samples. Active PTB is difficult to diagnose, especially in children and people living with HIV who have weakened in immune systems, and people who have multidrug-resistant PTB (MDR-TB) (VanRie et al., 2011). To determine if the patient has active PTB

disease, the following tests may be used. The detection of the most infectious diagnose of PTB (sputum positive pulmonary cases) is a critical step in the control of PTB in the people where in close contact. The processing of determining a PTB diagnose, which is identifying the source of infection in the people, that is, individuals who are discharging large numbers of PTB bacteria. So that they can receive prompt treatment, which in turn will cut the chain of transmission (stop the spread) and therefore lower the prevalence and mortality of PTB. The identification of people with suspected PTB is the first step is case finding. The second step involves the laboratory investigation of the PTB suspect's sputum sample to confirm those who have active PTB screening. The majority of PTB cases may be treated successfully with the appropriate course of antibiotics, but diagnosis remains a large obstacle to PTB elimination. Presently, the most common methods of diagnosis patients with active PTB is screening stained smear prepared from sputum sample microscopy images (Tapley, 2012).

2.2. Sputum Smear Microscopy

It was a first technique that was developed by German researcher Robert Koch in 1882, which he used to identify the presence of rod-shaped bacteria (Tapley, 2012). Sputum Smear Microscopy has been the primary method for diagnosis of PTB in low and middle income countries. Smear microscopy is corner stone of PTB diagnosis in resource poor setting because it is simple, inexpensive and detects most of the infectious form of PTB (FEMOH, 2013). Direct microscopy for PTB is also performed to assess the response to treatment and to establish cure or failure at the end of treatment. In addition the results are available within hours. The specificity through this technique is only about 50-60% (Lumb et al., 2013). In countries with a high prevalence of both PTB and HIV co-infection, have very low levels of the PTB bacteria in their sputum and are there recorded as sputum smear negative.

The most important methods in the diagnosis of PTB, which stained slide is examined under microscopy for signs of the PTB bacteria are called Acid Fast Bacilli (AFB) (Deun, 2002). AFB staining remains the initial step for evolution of PTB using direct microscopy examination of acid fast bacilli (AFB) in a smear. Two techniques are used for direct examination of PTB diagnostic with AFB smear microscopy are fluorescence microscopy and conventional microscopy. Fluorescence microscopy uses an Auramine based stain known as fluorochrome stain (Auramine-O or Auramine-rhodamine), while conventional microscopy uses a Carbol Fuchsin stain (Ziehl-Neelsen (ZN) or Kinyoun) using light microscopy (Swaminathan et al., 2010). Both techniques rely on the retention of stain following the application of acid fast, resulting from tight binding of the stain to mycolic lipids in the cell wall of the bacillus bacteria. The fluorescent AFB stain uses an intensive light source, such as a halogen or high- pressure mercury vapor lamp, carbol fuchsin AFB stain uses a conventional artificial light source (Swaminathan et al., 2010). Various methods of concentrating sputum based on centrifugation have been shown to increase diagnostic yield when used prior to microscopy.

In Ethiopia laboratory diagnosis of TB is usually done using acid fast bacilli (AFB) smear microscopic examination (FEMOH, 2013). Microscopic examination is used for diagnosis, monitoring and defining cure rate of treatment. Three sputum specimens must be collected and examined in two consecutive days (spot-early morning-spot) (FEMOH, 2013). Sputum smear microscopy examination is the three sputum specimens must be examined by Ziehl-Neelsen (acid fast) staining technique. PTB positive is diagnosed when at least two smear results are positive for AFB or one sputum specimen is positive with additional x-ray abnormality. Laboratory tests at national standard diagnostic algorithm indicated in figure below is used for the diagnosis of TB cases in Ethiopia.

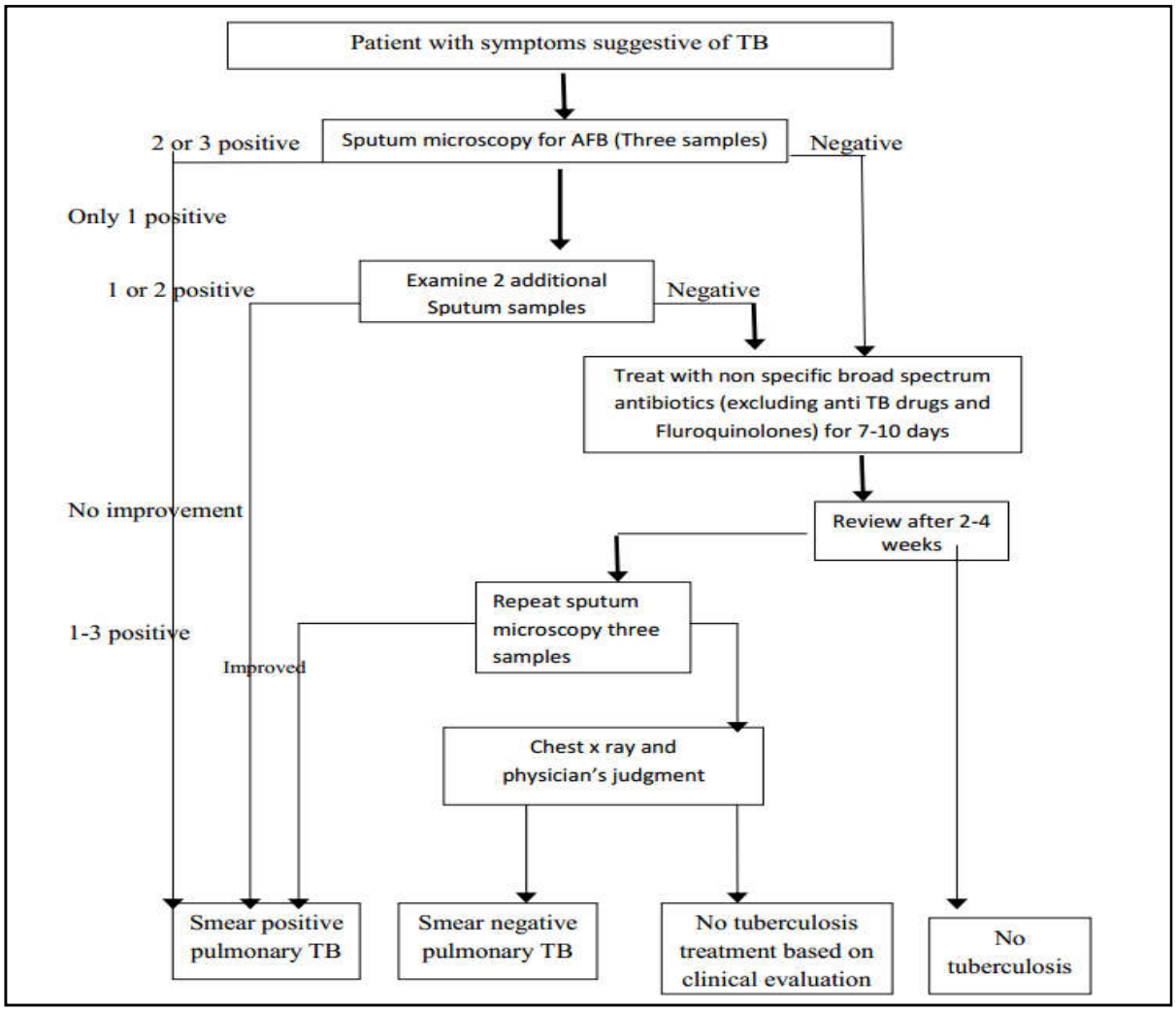


Figure 2 1: Sputum smears microscopy examination (FEMOH, 2013)

Based on the national guideline, PTB cases classifications (FEMOH, 2013) are defined in the following: smear positive and smear negative PTB

Smear Positive Pulmonary TB (PTB+):- It refers to the following:

- ✓ A patient with at least two initial sputum smear examination which were positive for acid-fast bacilli (AFB) by direct microscopy or

- ✓ A patient with only one initial sputum smear examinations positive for AFB by direct microscopy and chest radiographic abnormalities consistent with active PTB determined by a clinician.

Smear Negative Pulmonary TB (PTB-). It refers to the following:

- ✓ A patient having symptoms suggestive of PTB with at three sputum smear examinations which were negative for AFB by direct microscopy and with chest radiographic abnormalities consistent with active pulmonary TB (including interstitial abnormal images), and no response to a course of broad spectrum antibiotics
- ✓ A patient with three sputum smear examination taken at least two weeks apart and which were negative for AFB by direct microscopy and radiographic abnormalities consistent with PTB and decision by a clinician response to treat with a full course of anti PTB.

A sputum stain for mycobacteria is a laboratory test performed on a sample of sputum or phlegm. It's also known as an AFB stain or smear results. There are two procedures commonly used for acid fast staining: carbol-fuchsin (Ziehl Neelsen) and Fluorochrome staining (Swaminathan et al., 2010).

2.2.1. Ziehl-Neelsen method of acid-fast staining

ZN method of acid fast stain technique is used to stain mycobacterium species including M. tuberculosis and non-tuberculosis mycobacteria (NTM). Detection of AFB in stained and acid-washed smears examined microscopy may provide the initial bacteriologic evidence of the presence of mycobacteria in a clinical specimen. Mycobacteria stained with carbol fuchsin combined with phenol. In the 'hot' ZN technique, the phenol carbol fuchsin stain is heated to enable the dye to penetrate the waxy mycobacterial cell wall. In the 'cold' Kinyoun technique, stain is not heated but the penetration is achieved by increasing concentration of basic fuchsin

and phenol and incorporating a 'wetting agent' chemical. After staining, an acid decolorizing solution is applied. This removes the red dye (color) from the background cells tissue fibres and any organisms in the smear expect mycobacteria which retain the dye and are therefore referred to as acid fast bacilli (AFB). Following decolorization, sputum smear is counterstained with malachite green or methylene blue which stains the background material, proving contrast color which the red AFB can be seen (Swaminathan et al., 2010).

Reagent	Acid Fast	Non-Acid Fast
Carbol Fuchsin with heat	Red (Hot Pink)	Red (Hot Pink)
Acid Alcohol	Red	Colorless
Methylene Blue/Malachite Green	Red	Blue/Green

Table 2 1: Reagents and results (Swaminathan et al., 2010)

Results:

- ✓ AFB: Red, straight or slightly curved rods, occurring singly or in small groups, may appear beaded.
- ✓ Cells :green
- ✓ Background material: green

2.2.2. Fluorochrome staining methods

Fluorochrome dyes are used to stain the smear. A high pressure mercury vapour lamp is traditionally used to excite the dye, and make it fluoresce. Fluorescent microscopy is faster screening of smears than with ZN and 10% more sensitivity than conventional microscopy (Getahun et al., 2007; Rawat et al., 2012). Although fluorescent microscopy increases the sensitivity of sputum smear microscopy, additional data on specificity and on the clinical

consequences associated with false positive results are needed to guide implementation of this technology in high HIV prevalence settings (Getahun et al., 2007, Rawat et al., 2012). Cost constraints are the major issues with fluorescent microscopy. In report smears results, the laboratory usually provides the clinician with a rough estimate of the number of AFB detected as shown in Table 2. In our setting, clinicians are provided with the quantity report only for quick and easy interpretation. According to the latest recommendation by WHO and the national AFB microscopy laboratory manual, quantization scale for AFB smears according to stain used the result of the sputum smear should be indicated as follows:

Examination finding	Result as recorded	Laboratory result	No. of fields examined
No AFB in 100 oil immersion fields	Negative	NEG	100
1 to 9 AFB in 100 oil immersion fields	Positive	1-9(Scanty)	100
10-99 AFB in 100 oil immersion fields	Positive	+	100
1-10 AFB per oil immersion field	Positive	++	50
>10 AFB per oil immersion field	Positive	+++	20

Table 2 2: Manual AFB smear report results (American Thoracic Society (2000))

2.2.4. Limitation of AFB microscopy staining

Sputum smear microscopy has significant limitations in its performance. The sensitivity is grossly compromised when the high bacteria load 5000-10,000 AFB/ml sputum sample is required for detection and misses less than 50% PTB cases. It also has a poor track record in EPTB, pediatric (children) TB and in patients co-infected with HIV and TB (Luelmo, 2004, Lumb et al., 2013). Due to the requirements of serial sputum examination, some patients who do

not come back for repeated sputum examinations become 'diagnosis defaulters' (Harries, 1998). Some patients do not come back for check results, and are a lost to treatment and follow up. A limited resource, large numbers of samples, all combined together often reduce the observation time per slide to less than 60 seconds and this also contribute to reduction in the sensitivity of the test. Therefore, techniques for optimization of smear microscopy are under active investigation. There has been an attempt to reduce diagnostic defaulting by assessing the feasibility of diagnosing PTB by collecting two sputum samples on a single day (1-day protocol), and comparing this protocol with the national policy of collecting samples on consecutive days (2-day protocol). AFB microscopy stained smears to detect acid fast organisms such mycobacterium TB (MTB) and non-TB mycobacteria (NTM). But it is limited specificity of all mycobacteria are acid fast and does not provide species identification (CDC, 2014).

2.2.5. Sputum Sample Collection and preparation

For reliable laboratory diagnoses of sputum smear, a properly collected sputum specimen is mandatory. A PTB suspect should submit three sputum samples for microscopy (WHO, 2017, Swaminathan et al., 2010, FEMOH, 2013). The chances of finding PTB bacilli are greater with three sputum samples than with two samples or one sample. So an early morning sputum sample obtained deep from the lung (not saliva) is more likely to yield better recovery of AFB than a sample later in the day to contain PTB bacilli. It may be difficult for an out- patient to provide three early morning sputum samples. Therefore, in practice an outpatient usually provides sputum samples as follows:

- Day one, sample one: Patient provides an "on the spot" sample under supervision when she / he presents to the health institution. Give the patient a sputum container to take home for an early morning sample to be submitted for the following day.

- Day two, sample two: Patient brings an early morning sample.
- Day two, sample three: Patient provides another “on the spot “sample under supervision.

If a patient can’t give a sputum sample, health personnel may advice him to produce a good cough and bring up some sputum. An inpatient can provide three early morning sputum samples under supervision. They are two types of smears preparation; direct and indirect smear. Direct smear is prepared directly from a patient prior to processing, while indirect smear prepared from processed specimen after centrifugation (to concentrate the material).

2.3. Automatic PTB Bacilli Detection System

Automated PTB detection is the technique, method, or system of operating or controlling a process of PTB diagnosing and the bacilli identification process by highly automatic means, as by electronic devices, reducing human intervention to a minimum. Automated TB diagnostic systems bring about several benefits in the fight against the PTB disease. Some of the advantages of computer-based diagnosis are measurements and features make a diagnosis, and it helps radiologists on their diagnosis procedure for accuracy and efficiencies are:

- Computer-assisted diagnosis can be faster than manual screening.
- A large number of slides can be analyzed and hence the number of patients tested in any given amount of time is increased.
- Human involvement in the screening process is reduced. This will result in three desirable effects; less experienced staff can conduct the screening, labor costs are reduced and technicians suffer less fatigue, since the need to view the slides on the microscope is eliminated or reduced.
- Accuracy in the diagnosis may be improved; in areas of high PTB incidence like Ethiopia stretched laboratory resources require limited screening of sputum smears. Where a large

number of patients have to be tested, technicians may test slides less thoroughly than required and smear positive cases may go unnoticed. Usually, not more than 50 high power fields are examined, resulting in false negative reports.

- Automation of the diagnostic process could possibly reduce the risk of infection to laboratory staff

Automated detection ensures that the entire slide examined in a short space of time and will thus reduce cases of misdiagnosis. In essence, an automated TB detection process may allow a greater number of patients to be tested at a faster rate by fewer technicians and with higher accuracy (Dendere, 2009). There are three major potential benefits of automated diagnosing systems are discussed by Chang et al. (2007): "improved safety of patients, improved efficiency in health care outcome, and improved quality of care". Currently, automated diagnosis systems are increasingly being used in the healthcare sector, which results in improved efficiency, better quality, and reduced instances of clinical errors. The proposed automated PTB detection helps to screen the images of sputum smears for MTB, thereby saving time, improving the quality of diagnosis, and thus increasing the efficiency and productivity.

2.4. Medical images processing

Today, medical images are acquired by a range of techniques across all biological scale, which go far beyond the visible light photographs and microscope images of the early 20^{th} century. Medical image is the science of solving/analyzing medical problems based on different imaging modalities (x-ray, microscopic images, ultrasound etc.) and digital image techniques (Thirumaran & Shylaja, 2015, Deserno, 2011, Gao, 2013).

Advanced image processing and analysis techniques are increasingly used in medicine. In medical applications, image data is used to collect the details of the imaging of the patient's process, whether it is a disease process or physiological process (Gao, 2013). The information provided in medical imaging has become an important part of today's patient care. The medical images display information on the characteristics of the structure, organs, and physiological characteristics (Deserno, 2011). The range of image processing and analysis of medical applications is to improve the quality of acquired images, and extract quantitative information from an efficient medical imaging data and accurate manner. Image quality plays an important role for image of organs in medical imaging expected to provide quantitative data useful for patient treatment and care (Deserno, 2011). Medical imaging needs highly trained technicians and clinicians to determine the details of image acquisition as well as to analysis the results.

2.5. Image Processing Techniques

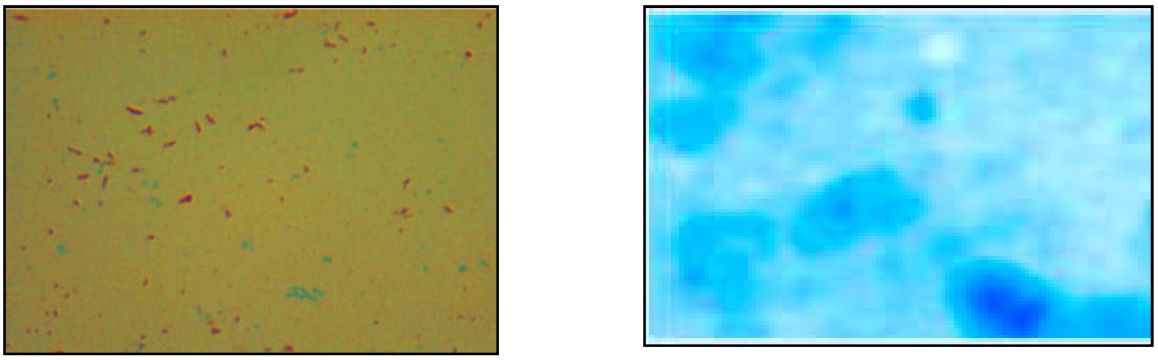
Image processing is a set of technologies, image data analysis and processing algorithms and tools to improve the interpretation of some image information more useful. Image processing allows the extraction of useful parameters, and increases the likelihood of detection of small lesions more accurately (Gao, 2013). Digital images are composed of image display an array of discrete picture elements (pixels) in formed from the picture and elements. In the digital image each pixel has an intensity value and the address of the location. In medical image pixel values to display the record counts in it. Benefit compared to the simulation of a digital image from the digital image data for further computer processing. By the increasing use of direct digital imaging systems for medical diagnostics, digital image processing becomes more and more important in health care (Deserno, 2011). Image Processing is used in biomedical image processing means the provision of digital image processing for biomedical sciences.

Image processing techniques is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is among rapidly growing technologies (Chitradevi and Srimathi, 2014). It forms core research area within engineering and computer science disciplines too. Image Processing systems are becoming popular due to easy availability of powerful personnel computers, large size memory devices, graphics software etc. (Deserno, 2011, Gao, 2013). A digital image is an array of real numbers represented by a finite number of bits. The principle advantage of Digital Image Processing (DIP) methods is its versatility, repeatability and the preservation of original data precision. Many researches in tuberculosis were used different image processing technique in the process of enhancing the sputum smear microscopic images to have better bacilli segmentation also it was used in corporation with pattern recognition to facilities the process of classifying in input images in categorical form. Generally, any automatic PTB detection systems include five major processes: image acquisition, image preprocessing, segmentation, feature extraction and classification (Chang et al., 2007; Thirumaran & Shylaja, 2015; Panicker et al., 2016; Khutlang et al., 2010; Ayas et al., 2014; Chitradevi and Srimathi, 2014).

Figure 2 2 : Steps of digital image process (Panicker et al., 2016)

2.5.1. Image Acquisition

In Ziehl-Neelsen stained sputum smear PTB bacilli detection, image acquisition is an activity of acquiring images from sputum smear microscopy image through Ziehl Neelsen (ZN) staining process using Leica Microsystems microscopy connected to PC designed for this purpose. A image or a camera can be attached to the eyepiece of the microscope to capture the digital. These images can then be stored in the computer and can be processed in real-time or offline mode using image processing techniques. Image dataset are the result of a real patient's sample specimen that is prepared by considering the experience and qualification of lab technician. Microscopic image acquisition has been done using microscope under different control environment. It is worth mentioning that adjusting resolution through different objective lens level is a key step before acquiring microscopic image. Further details and sample images regarding the different level of visibility under different objective lenses.



A B

Figure 2 3: Sputum smear image positive of PTB (A) and sputum smear image negative of PTB (B)

2.5.2. Image preprocessing

Image preprocessing is a method to convert an image into digital form and perform operations done on images at the initial level where both input and output are intensity images. The RGB

images are first converted to HSV color models and then converted to gray scale images. The images acquired may have poor illumination. Hence the contrast of the images is improved by equalizing the histogram of the image. Background elimination is done as the images may have artifacts in the background. The aim of preprocessing is an improvement of the PTB infected ZN Stained image by suppressing unnecessary distortions or enhancing major image features important for further recognition processes. Preprocessing activities in automatic PTB detection systems, a set of sequential processes enhancing image, creating binary image, removing objects, removing cells at boundary has been applied to the original image after image acquisition. This step is to make the acquired image for the PTB diagnosis system. There are three main reasons for image preprocessing are to resizing the image; to reduce the noise and last to enhance the image contrast to be analyzed as it would be an extremely slow and very computationally expensive for MATLAB to process the whole image as it is originally given by the camera. All unnecessary information such as the object of not interest are removed after implementing this task the images is also resized to the large image dimensions so that to be easily fitted and processed (Khutlang et al., 2010, Ayas et al., 2014). In image preprocessing is there various techniques to improving the perception of information in images for human viewers. It is provide enhanced input for other automated image processing techniques. In this study the researcher used different techniques for enhancing the stain sputum smear image such as Gaussian filter and contrast stretching respectively ZN stained image has more complex ground detail. Thus the object (PTB Bacteria) must be separated by the ground before identifiable. Gaussian filter was implemented to diminish the effects of camera noise and spurious pixel value (Rachna and Mallikarjuna, 2013, Santiago-Mozos et al., 2014). It also reduces the sputum smear image details and finally contrast stretching was done to improve the contrast in an image without distorting

relative level of intensities. In this study, the Gaussian and Contrast Stretching algorithms was used. Gaussian filtering technique was used to remove noise while, Contrast Stretching technique was used for enhancing the stain sputum smear images. The Gaussian algorithm is suited algorithms used for remove noise in the preprocessing stages.

2.5.3. Image Color representation

A digital image is composed of a rectangular array of pixels. Digital images can be represented as gray-level or colour images. In a gray-level image, also called an intensity image, the brightness of a pixel is represented by the intensity value of the image at the pixel location. The intensity values range from 0, which represents black, to a maximum value, which represents white. The maximum gray-level value for an *n*-bit system is 2^n -1. Intensity values between 0 and a maximum value represent shades of gray that increment towards white. Colour images are a combination of intensity images, which are interpreted differently by different colour models. A color space is a mathematical representation of a set of colors. The three most popular color spaces are RGB (used in computer graphics); YIQ, YUV, or YCbCr (used in video systems); and CMYK (used in color printing). However, none of these color spaces are directly related to the intuitive notions of hue, saturation, and brightness. This resulted in the temporary pursuit of other models, such as HSI and HSV, to simplify programming, processing, and end- user manipulation (Gonzalez and Woods, 2002, Dendere, 2009).

RGB color space: The red, green, and blue (RGB) color space is widely used throughout computer graphics. Red, green, and blue are three primary additive colors (individual components are added together to form a desired color) and are represented by a three-dimensional, Cartesian coordinate system). The indicated diagonal of the cube, with equal

amounts of each primary component, represents various gray levels (Gonzalez and Woods, 2002, Dendere, 2009).

HSV Color Space: The HSV (Hue, Saturation and Value) color spaces were developed to be more "intuitive" in manipulating color and was designed to approximate the way humans perceive and interpret color. It was developed when colors had to be specified manually, and is rarely used now that users can select colors visually or specify Pantone colors (Gonzalez and Woods, 2002, Dendere, 2009).

L*A*B color space: A Lab color space is a color-opponent space with dimension L for lightness and 'a' and 'b' for the color-opponent dimensions, based on nonlinearly compressed (e.g. CIE XYZ color space) coordinates. The L*a*b* color space includes all perceivable colors, which means that its gamut exceeds those of the RGB and CMYK color models (for example, ProPhoto RGB includes about 90% all perceivable colors). One of the most important attributes of the L*a*b*-model is device independence (Gonzalez and Woods, 2002, Dendere, 2009). Generally, when compared the three colors RGB, HSV and L*A*B and found that better segmentation performance had been achieved by using the HSV color model. In this study, HSV color model used for PTB a bacillus was segmented from image acquired was used Leica Microsystems microscopy at EPHI. Because HSV space color is much larger than the extent of computer displays, human vision, a bitmap images represented as lab requires more data per pixel to obtain the same precision as an RGB or CMYK bitmap. This color spaces is the best suited to many digital images of sputum smear image manipulates than the RGB, which is useful for sharpening images and the removing artifacts or noise in JPG images or image acquired from Leica Microsystems microscopy connected to PC.

2.5.4. Image segmentation

Image segmentation is the process of separating objects in an image into distinct, homogenous region to identify objects and borders for object description and recognition (Gonzalem and Woods, 2002, Dendere, 2009). Segmentation is often the critical step in image analysis. The approaches was used to segmentation methods are region based. Region based techniques rely on discontinuities in image values between distinct regions, and the goal of the segmentation algorithm is to accurately demarcate the boundary separating these regions. Region-based techniques rely on common patterns in intensity values within a cluster of neighboring pixels. Segmentation is an important step in image processing which aids extraction of information and attributes from images for image understanding and interpretation. The distinct characteristics of ZN-stained sputum smears, which contain red bacilli against a blue and off-white background, present a useful property that can be exploited for segmentation of bacilli (Raza et. al., 2015; Panicker et al., 2016; Goel et al., 2017). Color image segmentation is the most prominent feature used for detecting bacilli.

2.5.4.1. **K-Means Clustering segmentation methods**

The K –means cluster method is referred to as the regions and the goal of the segmentation algorithm is to group regions according to their anatomical or functional roles (Gonzalem and Woods, 2002). K-means clustering is a set of observations image into subsets (clusters). K-means clustering algorithm applied for segmenting the ZN-stained sputum smears image and it was done using Euclidean distance measure for performing image segmentation (Dendere, 2009; Khutlang, 2010). The function of k-means partitions data into k mutually exclusive cluster, and returns the index of the cluster to which it has assigned each observation. Unlike hierarchical clustering, K-means clustering operates on actual observations, and creates a single level of

clusters (Sadaphal, 2008). The dissimilarities mean that k-means clustering is often more suitable than hierarchical clustering for large amounts of data. K-means treats each observation in data as an object having a location in a space. It finds a region partition in which objects within each clusters are as close to each other as possible, and as far from objects in other clusters as possible. One can choose from three different distances measures, depending on the kind of data you are clustering. Each cluster in the region partition is defined by its member objects and by its centroid or center. Centroid for each cluster is the point to which is the sum of distances from all objects in that cluster is minimized. K-means cluster computes the cluster centroids differently for each distance measure, to minimize the sum with respect to the measure that is to be specified. For this study, color based segmentation using K-mean clustering algorithms was used. K-mean clustering is a partitioning method. It can control the details of the minimization using several optional input parameters to k-means including ones for the initial values of the cluster centroids, and for the maximum number of iterations. By default, k-means uses the k-means ++ algorithms for the cluster center initialization and the squared Euclidean metric to determine distances (Khutlang, 2010).

2.5.5. Image Feature Extraction

The extraction of image features is the fundamental step for image classification. Feature extraction starts from an initial set of measured data and builds derived values (features). It is intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations (Gonzalem and Woods, 2002; Dendere, 2009). Feature extraction is related to dimensionality reduction. The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial

data. After the segmentation step is processed not only the bacilli is segmented, different structure can also appeared which have the same shape, colour and textual properties as bacilli bacteria in microscopy. There are various types of features for image classification's aim as follow as color, shape and texture features of pixels. In these for this study, color and shape is the most important over the texture feature.

A. Shape features

Shape is an important indication human being to identify and recognize the real world objects, which is to encode simple geometrical forms of lines in different directions. Shape features extracts techniques calculate the shape features only from the extracts the region boundary shape from entire region. PTB stained sputum smear image that are to be analyzed are characterized by a large diversity of remains in terms of both shape and size. The most important features used to detect PTB are a shape level feature gained from color structure. In this study, shape level feature algorithms are used to group the nearest-neighbour connected pixels to describe the shape features. The shapes descriptors are were investigated to choose the best characteristic of PTB bacilli. Bacilli have a 1-10 μm in length and 0.2-0.6 μm in width and, presenting rod shaped. Thus, classification used consists of a single class of bacilli and a rejection class. It is evaluated for bacilli characterization are eccentricity, perimeter, area, compactness/solidity, major and minor axis lengths, and circularity was used (Ayas et al., 2014; Raza et.al, 2015; Panicker et al, 2016; Goel et al., 2017).

B. Color feature

Color is the most widely used visual features in object identification. While we perceive only a limited number of gray levels, our eyes is able to distinguish thousands of colors and a computer can represent even millions of distinguishable colors in practice. Color has been successfully

applied to recognize images, because it has very strong correlations with the underlying objects in an image. Moreover, color feature is robust to background complication, scaling, orientation, perspective, and size of an image (Ayas et al., 2014; Raza et.al, 2015; Panicker et al, 2016; Goel et al., 2017).

2.5.6. Image Classification

Image classification analyzes the numerical properties of various image features and organizes data into categories (Gonzalem and Woods, 2002, Dendere, 2009). Classification algorithms typically employ two phases of processing: training and testing. In the initial training phase, characteristic properties of typical image features are isolated and, based on these, a unique description of each classification category, i.e. training class, is created. In the subsequent testing phase, these feature-space partitions are used to classify image features. The description of training classes is an extremely important component of the classification process (Dendere, 2009). A classification method is the final stage in any diagnosis system where each unknown model is assigned to a category. In the previous research studies, there were different classification methods used in tuberculosis bacilli identification such as K-nearest neighbor (KNN) algorithm and support vector machine (SVM) algorithm (Sadaphal, 2008, Veropoulos, 2001) to classify PTB bacilli. For this study, the researcher compare and select the best suitable algorithms (KNN and SVM) results was used to identify PTB bacilli in sputum smear image based on computed morphological and color features in two class PTB bacilli and non-PTB bacilli detected.

2.5.6.1. Support Vector Machine Classification

A SVM performs classification by constructing an N-dimensional hyper plane that optimally separates that the data in two categories. SVM models are closely related to neural networks.

Support Vector Machine (SVM) models are a close cousin to classical multilayer perceptron neural networks (Raza et.al, 2015; Panicker, 2016). In SVM, an attribute, and a transformed attribute that is used to define the hyper plane is called a feature. The task of choosing the most suitable representation is known as feature selection. A set of features that describes one case (i.e., a row of predictor values) is called a vector. So the goal of SVM modeling is to find the optimal hyper plane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyper plane are the support vectors. An SVM classifies data by finding the best hyper-plane that separates all data points of one class from those of the other class. The best hyper-plane for an SVM means the one with the largest margin between the two classes. The support vectors are the data points that are closest to the separating hyper-plane; these points are on the boundary of the slab. The following figure illustrates these definitions, with + indicating data points of type 1 and – indicating data points of type –1. Support vector machine (SVM) is an algorithm and is popularly used in many pattern recognition problems, including texture classification. In SVM, the input data is non-linearly mapped to linearly separated data in some high dimensional space providing good classification performance. SVM maximizes the marginal distance between different classes. The division of classes is carried out with different kernels in below figure (www.mathworks.com, 2018).

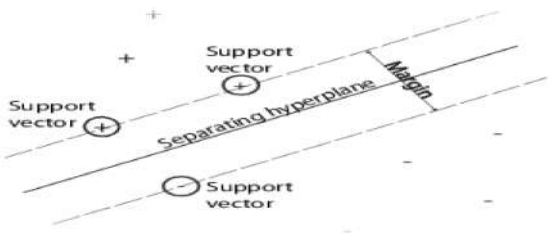


Figure 2 4: SVM process.

In classifiers algorithms, SVM is one of the advanced statistical methodologies suitable algorithm to identify PTB bacilli such as neural network. Neural network suffer from a limited ability to handle data and can only analyze the form two or three dimensions. For this study the best suitable algorithms a support vector machine (SVM) was used to identify PTB bacilli in sputum smear image, which classified the computed both morphological and color features from sputum smear images in two class PTB and non-PTB bacilli detected (Khutlang et al., 2010; Santiago-Mozos et al., 2014, Raza et.al., 2015; Panicker, 2016). A main advantage of SVM classification is that SVM performs well on datasets that have many attributes, even when there are only a few cases that are available for the training process. However, several disadvantages of SVM classification include limitations in speed and size during both training and testing phase of the algorithm and the selection of the kernel function parameters.

2.5.6.2. K-Nearest Neighbor (KNN) Classification

KNN algorithm is a method for classifying bacilli objects based on closest dataset in the feature extraction represent. KNN algorithm is among the simplest of machine learning algorithms. Training dataset process for this algorithm only consists of storing feature vectors and labels of the training images. In the classification process, the unlabelled query point is simply assigned to label of its KNN. Typically the object is classified based on the labels of its *k* nearest neighbors by majority vote. If *k*=1, the object is simply classified as the class of the object nearest to it.

When there are only two classes, *k* must be an odd integer. However, there can still be attaches when *k* is an odd integer when performing multiclass classification. After we convert each image to a vector of fixed-length with real numbers, we used the most common distance function for KNN which is Euclidean distance:

$$d(x,y) = ||x - y|| = \sqrt{(x-y).(x-y)}$$

$$= (\sum_{i=1}^{m}((xi - yi)^2))1/2$$

Where x and y are histograms in $x = R^m$ shown in figure below visualizes the process of KNN classification.

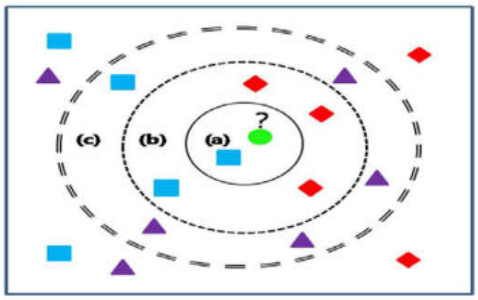


Figure: 2.5. KNN Classification

At the query point of the circle depending on the *k* value of 1, 5, or 10, the query point can be a rectangle at (a), a diamond at (b), and a triangle at (c).

A main advantage of the KNN algorithm is that it performs well with modal classes because the basis of its decision is based on a small neighborhood of similar objects. Therefore, even if the target class is multi-modal, the algorithm can still lead to good accuracy. However a major disadvantage of the KNN algorithm is that it uses all the features equally in computing for similarities. This can lead to classification errors, especially when there is only a small subset of features that are useful for classification. KNN classifies data based on the distance metric whereas SVM need a proper phase of training. Due to the optimal nature of SVM, it is guaranteed that the separated data would be optimally separated into two classes. Generally, KNN is used as multi-class classifiers whereas standard SVM separate binary data belonging to either of one class. For a multiclass SVM, One-vs-One and One-vs all approach is used. In One-vs-one approach, we have to train n*(n-1)/2 SVMs: for each pair of classes, one SVM.

2.6. Related Works

Number of researchers used using image processing techniques, to contributed detection of PTB from stained sputum smear image of microscopy that passed through ZN acid fast staining procedure. PTB is a curable disease and it can be controlled. But to control and cure the disease, correct diagnosis is needed. This disease needs continuous monitoring and proper guideline. The guidelines of WHO (2017), suggests to diagnosis PTB by screening of Ziehl-Neelsen (ZN) stained specimen under light (conventional) microscopy stained specimen. But this technique is sensitive as it needs highly trained specialists and it has a high false negative rate. In this part, previous works conducted in automatic PTB detection using image processing techniques are critically reviewed. Gaps that exist in previous works for automatic PTB detection are also identified.

Veropoulo et al. (1999) stated that a method of developing an automatic for the detection of PTB bacilli in fluorescence microscopy clinical specimen, principally sputum smears to improve the diagnostic process. In the study, the researchers were used image processing techniques and neural networks classifiers for identifying PTB bacilli. The developed system showed a sensitivity of 93.9% and specificity of 79.4% was achieved using neural network classifier to identify bacilli. As there are usually fairly numerous PTB bacilli in the sputum of patients with active pulmonary PTB. The overall diagnostic accuracy for sputum smear positive patients was expected to be very high. The potential benefits of the study discussed based on automated screening for PTB are rapid and accurate diagnosis, increased screening of the population, and reduced health risk to staff processing sputum sample.

Rechna et al. (2013) presenting an algorithm based on the image processing techniques for identification of PTB bacteria in sputum smear image from digital microscope. In the study the authors discussed the availability of expertise, time and cost are the human intervention based examination developed technique results has been a good accuracy and efficiency. The method used in the study was based on Otsu thresholding and K-means clustering approach. The performance of clustering and thresholding for segmenting PTB bacilli detection in body section was also compared.

Divekar et al. (2012) stated that, developed automated detection of tuberculosis on sputum smeared slides using a stepwise classification (SWC) algorithm to remove different types of false positives, one type at a time, and to increase the detection of PTB bacilli at different concentrations. The author was used the fluorescence microscope approach methods based on the Shannon cofactor expansion on Boolean function for classification, both bacilli and non-bacilli objects are first analyzed and classified into several different categories including scanty positive, high concentration positive, and several non-bacilli categories. The morphological and contrast features were used for extracted based on a prior clinical knowledge. The SWC is composed of several individual classifiers in the study. Individual classifier to reduce false positive is developed through minimization from a binary decision tree to classify different types of true and false positive based on feature vectors. Finally, the detection algorithm was tested on 102 independent confirmed negative and 74 positive cases. A multi-class task analysis showed high accordance rate for negative, scanty, and high concentration as 88.24%, 56.00%, and 97.96%, respectively.

Chang et al. (2007), proposed an algorithm for automated TB detection in smear images taken by digital microscopes such as Cell Scope, a novel low-cost, portable device capable of bright

field and fluorescence microscopy. Automated processing on such platforms could save lives by bringing healthcare to rural areas with limited access to laboratory-based diagnostics. Though the focus of the study was the application of automated algorithm to Cell Scope images, the method may be readily generalized for use with images from other digital fluorescence microscopes. The algorithm applies morphological operations and template matching with a Gaussian kernel to identify TB object candidates. Then moment, geometric, photometric, and oriented gradient features were used to characterize these objects and perform discriminative, support vector machine classification. Then the algorithm was tested on dataset was collected at clinics in 594 images corresponding to 290 patients. The object level classification is highly accurate, with Average Precision of 89:2% -2:1%. For slide-level classification, the algorithm performed at the level of human readers, demonstrating the potential for making a significant impact on global healthcare.

Sadaphal, et al. (2008) proposed an approach used conventional microscopy which is based on the idea of using clear *Mycobacterium tuberculosis* browse images to characterize AFB color. The study focused on Automated, multi-stage, color-based Bayesian segmentation identified possible TB objects, has been used to remove artifacts by shape comparison and color-labeled objects as 'definite', 'possible' or 'non-TB', by passing photo micrographic calibration. Superimposed AFB clusters, extreme stain variation and low depth of field were challenges. Their novel method facilitated electronic diagnosis of TB, permitting wider application in developing countries where fluorescent microscopy is currently inaccessible and unaffordable. An algorithm was used to recognize Acid-Fast Bacilli under wide latitudes of staining, magnification and resolution. No quantative results were provided in this study, but the algorithm appeared to perform acceptable based qualitative observation.

Several previous studies on diagnosis methods for PTB detection have been described in the related work (section 2.6). These methods can be divided into two approaches: using fluorescence and conventional microscope. Image viewed using a fluorescence microscope is more sensitive to PTB bacilli and the screening processing can be conducted quickly under lower magnification, compared to the conventional microscope. Fluorescent microscopy is faster screening of smears than with ZN and 10% more sensitivity than conventional microscopy (Getahun et al., 2007; Rawat et al., 2012). However, it is expensive and difficult to maintain, thus limiting the use of fluorescence microscope in low and medium income countries including Ethiopia. Therefore, recent works in detecting PTB bacilli used sputum smear images acquired from conventional microscope (Sadaphal, et al. 2008 and Khutlang et al., 2010). The specimens are stained sputum smear conventional microscope using Ziehl-Neelsen (ZN) staining procedure to visualize the bacilli. According WHO (2017) and American Thoracic Society (2000) recommended the manual AFB smear report results the quantization scale for actual number of AFB stain observed the sputum smear negative (no AFB seen per 100 field). The automatic detection techniques can save the time and cost involved, with reduced human error.

Recent work by Osman et al. (2012), proposed a light microscopy methods that combined K-means clustering and thresholding algorithms for segmenting the PTB bacilli in sputum smear image. Thresholding was used to eliminate the unwanted pixels and K-means clustering was used to classify the possible PTB bacilli pixels. In conventional microscopy images, the bacilli are not easily separated from background with a thresholding operation. In this case, for bacilli segmentation, color space techniques were used. The work had compared three color models;

RGB, HIS and C-Y for better segmentation performance had achieved. The limitation of pixel classification approaches that they cannot incorporate or region of interest information about the bacilli detected from the other objects. In low-income countries, the most common method of pulmonary tuberculosis diagnosis is visual identification of rod-shaped PTB bacilli in sputum smears by conventional microscopy.

In low-income countries including Ethiopia the most common method of PTB bacilli is visual identification of bacilli characteristics of rod-shaped PTB bacilli detect in sputum smear images by using conventional microscopy. Manually observation is done by lab technician to identify bacilli in conventional microscopy which will take 2-3 hours for a single slide. Some of the bacilli appear stained in deep-red and in some cases they appear in pale-red making challenging tasks for observer. However, the intensity distribution PTB bacilli and background are often varying from image to image, due to manually ZN stained methods. Although, several computational methods for PTB bacilli detection have been developed, accurate PTB bacilli detect and non bacilli detect classification still remains a big challenges.

In this study, developed prototype system for PTB bacilli detection image acquired from the conventional microscopy by using Leica Microsystems microscopy connected to PC at EPHI. Image acquired represent HSV color model was used to recognize image then, Gaussian filter technique was used to reduce noised and Contrast stretching was used to improve the quality of bacilli image. And also K-mean clustering was used image segmented, feature extraction represented to facilitate SVM and KNN algorithm was used to classifying PTB bacilli detected from the other objects for better evaluation system performance achieved.

CHAPTER THREE

3. DESIGN AND IMPLEMENTATION OF DETECTION SYSTEM

Automatic detection of a PTB bacilli object and classifying into its appropriate classes of diagnosis become a wide area of this study. The design and implementation involve the actual development of procedures that can be applied for automatic PTB bacilli detection system to assist pathologists in decision making. Therefore, having all the necessary steps and the knowledge from image acquisition using digital image processing and domain know how from the experts is vital. The next task is coding the image analysis and classifying process. For this study, MATLAB R2016a (9.0.0.341360) image processing tools framework was used to develop the prototype which runs on a personal computer. The detection of PTB bacilli passes through procedures that applied to different steps, in which each steps must be categorized into predefined process based on represent feature extraction.

3.1. Designing the system architecture of APTBBD

Automatic pulmonary tuberculosis bacilli detection (APTBBD) system is overall the concepts of the framework for any vision algorithm for PTB bacilli identification from the sputum smear microscopy images during the assisting pathologist's decision making. The developed system consists of five steps namely, image acquisition, preprocessing, segmentation, feature extraction and classification. First, the digital images are acquired from the environment using a Leica Microsystems microscopy connected to PC. Second, image preprocessing techniques are applied to the acquired image to remove noise in order to have an analysis towards constant color characteristics, the images are normalized. Third, an image segmentation technique is an important step in image preprocessing to extract useful features of information from the images

for human understand and interpretation. After this, several analytical discriminating techniques were used to classify the images according to identify the PTB bacilli detected or not detected.

The architecture of the APTBD system in figure 3.1 represents how the prototype works during the PTB bacilli detection. As the new query (problem) is detection of the PTB disease, it starts with acquiring images of the sputum smear microscopy images using a Leica Microsystems microscopy. In order to remove noises that occur during the image acquisition step, by applied image processing techniques like filter algorithms and contrast enhanced. Then, the features that are best suited to represent the image are extracted from the image acquired using an image analysis techniques. Based on the extracted features the training and testing data are used to identify are extracted.

Finally, support vector machine learning pattern identifier is selected to classify an image into its classes of diagnosis results. Figure 3.1 shows the architecture of the developed system based on PTB detection.

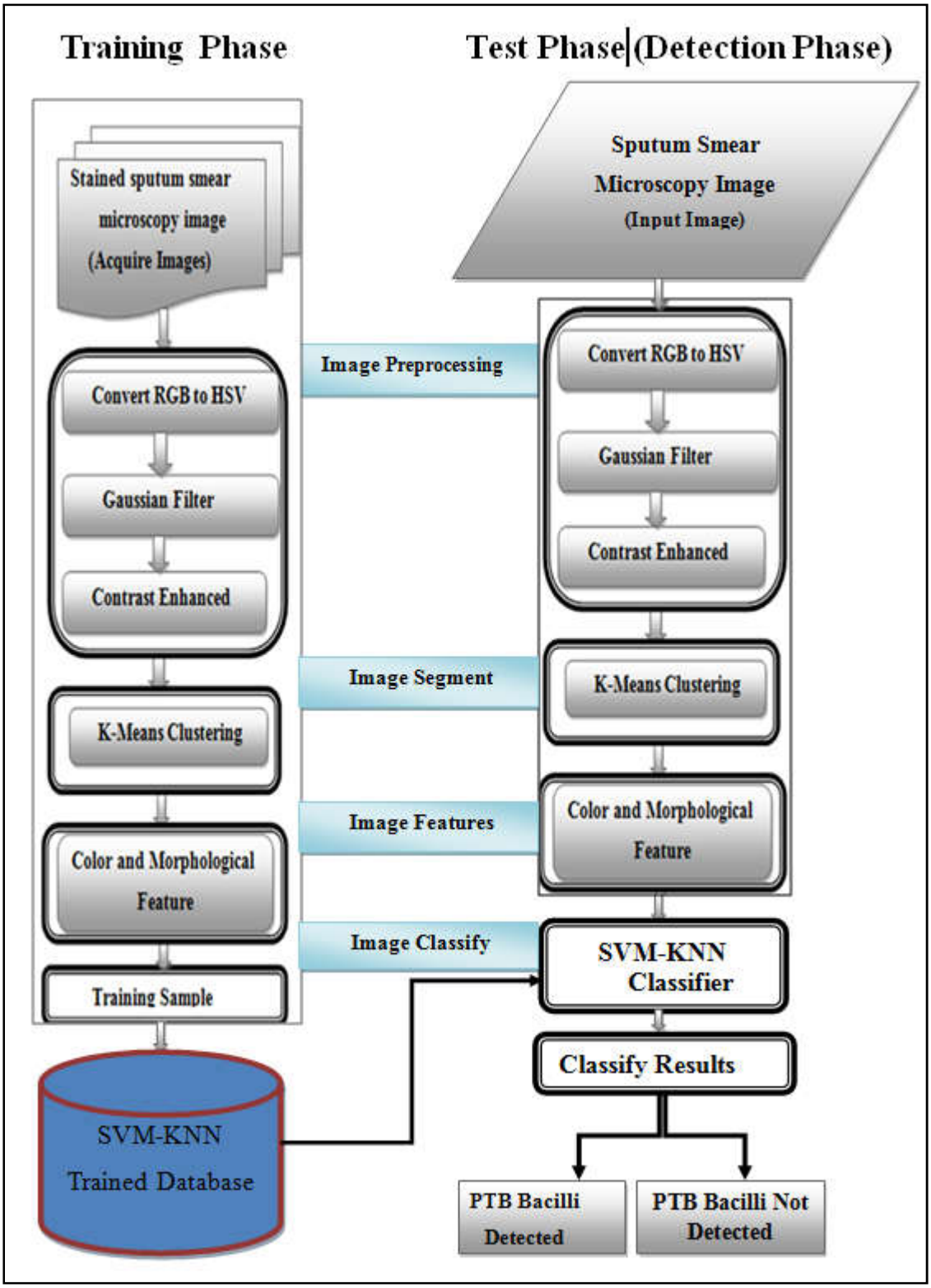


Figure 3 1: Developed Prototype System Architecture

3.1.1. PTB Bacilli Image Acquisition

A total of 180 (negative and positive) stained smear images prepared from slides specimens of patients of PTB bacilli collected from ZN stained sputum smear image were acquired by using conventional microscopic diagnostic techniques from the National Tuberculosis Reference Laboratory at Ethiopian Public Health Institute. Images were taken using Leica Microsystems microscopy which is connected to PC (figure 3.1.1). The resolution of pixel was 696*514. The images were stored in Joint Photograph Experts Group (JPG) file format, with 24 bits per pixel, in color. Some ZN stained sputum smear images were excluded due to bad images resolution that had been occur when capturing the images. Each sputum smear image is 696x514 pixels, covering a 1024x756 µm field of view at the smear.

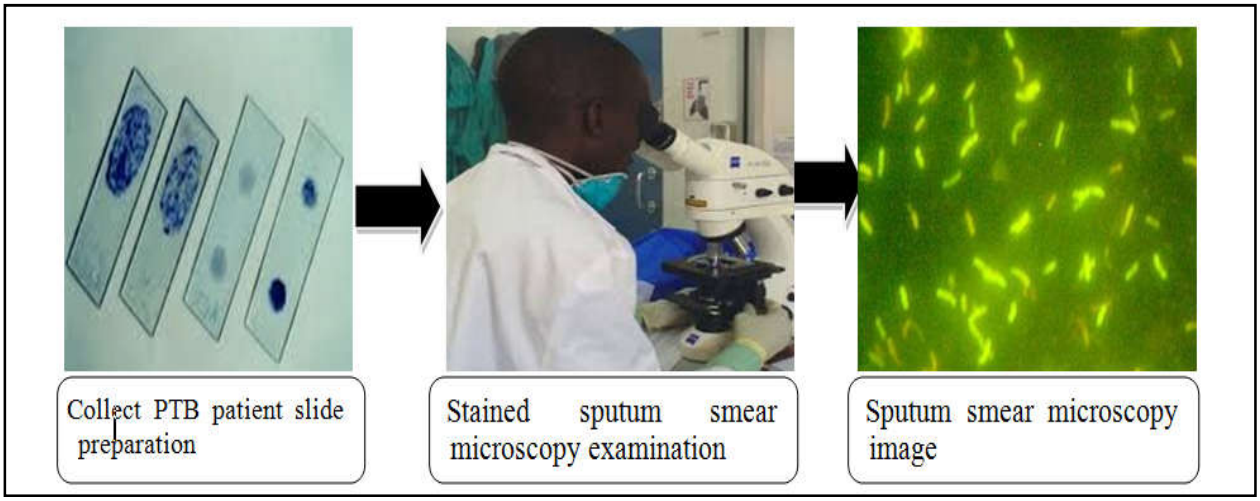


Figure 3 2: Sputum smear microscopy image from sample of specimen

3.1.2. PTB Bacilli Image preprocessing

After image acquired by using Leica Microsystems microscopy mounted to PC, it is essential to performs image preprocessing. Image preprocessing is a common name for operation with images at the lowest level of abstraction both input and outputs are intensity images usually

represented by a matrix of image function values (brightness). The color image segmentation step is starts with conversion of the RGB of original image to HSV color space model. Depending on the staining procedures the PTB bacilli detect suppose different colors may vary from looking under the microscopy light. HSV color space is converted from RGB color space and was used for color image segmentation for identifying the sputum smear image. It was used to enhance image when the background of enhanced image was off-white to identify bacilli image for easily segmented in the next stage. In addition, Gaussian filter, a nonlinear digital filtering technique was used to remove noise information from images. Gaussian filtering was widely used in digital image preprocessing because, under certain condition, it preserves edges while removing noise. Therefore, Gaussian filter in color image has been applied on separate channel and recombined to a single channel.

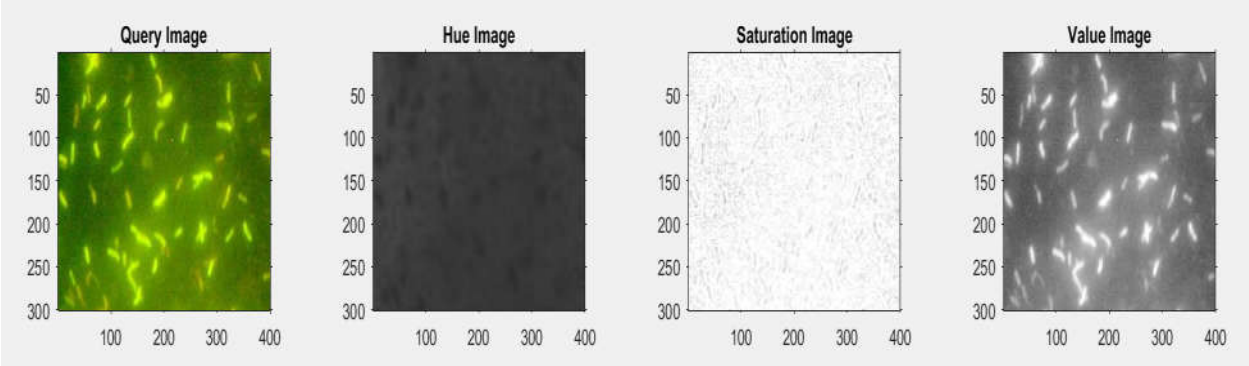


Figure 3 3: Image converted original image to HSV color model

This step generates the PTB bacilli from the sputum smear images either detected or not detected that contains a minimum noise compared with the original image acquired from sputum smear microscopic image and it passes the image through to the next step of the training phase that is image segmentation. The basic important of image preprocessing is to improve the quality of image and to reduce the undesired portion from the background of the images. Most of preprocessing techniques applied the intensity value of the region pixel for obtaining the

brightness intensity value of the input images. There is dominant reason for the requirement processes from the image preprocessing steps: Image quality improvement, noise reduction, contrast enhanced, and correction of missing or wrong pixel values, optimal preparation of data for segmentation and elimination of acquisition specific artifacts. Additionally, image preprocessing depends on mainly on the quality of image acquired from the image acquisition from Leica Microsystems microscopy connected to PC. The main aims of image preprocessing is to contain unwanted noise and to enhanced image features important from further analysis point of view, and the most of the time specific in nature depending upon the type of noise present in image.

3.1.3. PTB Bacilli Image Segmentation

After image preprocessing processed, image segmentation is very important to extract good quality features for classification image. The basic function of image segmentation are clustering pixel having same intensity value from the whole image regions, separate region or object of desired part of the original image, hiding the undesired region or objects. Image segmentations are applied in many image preprocessing areas such as recognition of object, object occlusion, estimation of the boundary, editing, or query from images dataset.

Overall, the algorithms for image segmentation are based on two basic important properties of intensity values which are discontinuity and similarity. Discontinuity property does a partition region of the image based on the sharp intensity values changes. While similarity property does a partition of the region image in to regions that are like in order to the specified criteria. In sputum smear images, segmentation process must be done accurately for the medical image analysis task which is especially for computer aided diagnosis. Different methods have been

discussed in literature, to segment sputum smear image process. Various methods for sputum smear image segmentation are presented which is a K-mean clustering method.

3.1.3.1. K-Mean Clustering segmentation

Clustering is a method divided a set of data into a specific of groups of image pixels. K-means clustering analysis is the assignment of a set of region in to subsets (clusters) which is the observation of region in the same clusters is similar in the sense. It was used for classification a given object based on a set of features into k number of classes. The classification of object is done by minimizing the squares of the distance between the object and the corresponding cluster. The algorithm for k-means clustering:

1. Pick center of K cluster, either randomly or based on some heuristic.
2. Assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center.
3. Again compute the cluster centers by averaging all of the pixels in the cluster. Repeat steps 2 and 3 until convergence is attained.

The separate characteristics of stained sputum smears, which contain red bacilli against a blue and off-white background, present a useful property that can be exploited for segmentation of bacilli. In literature review, previous researches showed that a more convincing segmentation performance has been achieved by using clustering methods mainly K-means clustering for segmentation the bacilli (Raza et. al., 2015; Panicker et al., 2016; Goel et al., 2017). That is why K-means clustering for segmentation the stained sputum smear images were used in this study. After completed the segmentation process, RGB images with black background and pink bacilli

was obtained, then it was converted to black undesired objects and white image shows the bacilli detected to facilitate for the feature extraction process.

K-means clustering is the way to separate of stained sputum smear images. K-means clustering treats each object as having a location in space. It finds partitions such as the sputum smear image within each cluster are as close to each other as possible and as far objects in other clusters as possible. K-means clustering requires specifying the number of cluster to be partitioned and a distance metric to quantify how close two objects are to each other. Using K-means to cluster objects into three clusters, using the Euclidean distance metric. Label every pixel in the image using the results from k-means cluster. For every object in our input, k-means returns an index corresponding to a cluster. Label every pixel in the images with its cluster index. Using pixels label create images that segmented by color, to separate objects in black backgrounds and bacilli images.

3.1.4. PTB Bacilli Image Feature extraction

Image analysis is the process of extracting meaningful information which is used to identify the unique features of sputum smear images that are used for classification of images into different categories. An image feature extraction is identifying the characteristics or attributes of a bacilli image. This image analysis reduces the complexity in classification problems by measuring certain properties or feature that identifies one input patterns from desired objects (another). For analysis of PTB, two classification parameters are identified. Image features have a major importance in image classification. There were morphological (shape) features and color features. In this research these two features were considered, because their structure forms like

shape and size. Hence, the classification system was based on morphology and color analysis, which considers an assessment of human visual inspection as starting point.

3.1.4.1. Morphological Feature

Morphology is the geometric of images. In this study, it is the size and shape characteristics of PTB bacilli image. It can be obtained from the analysis of binarization images. From morphology of bacilli the following features were extracted from the binarization images as described in the previous section. The researcher included area, eccentricity, compactness, perimeter, roundness, major length axis, minor length axis and EquivDimeter. These features were used to identify PTB bacilli detected based on the size and shape characteristics.

Area: area is the number of pixels inside the region covered by a bacillus detected, including the boundary region. It is measured in square pixels. Where r is radius

$$\boldsymbol{A = \pi r^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

Eccentricity (E): the eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1. An ellipse whose eccentricity is 0 is actually a circle, while an ellipse whose eccentricity is 1 is a line segment.

$$\mathbf{E} = \sqrt{\mathbf{1 - (b/a)^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

Where a is major length axis and b is minor length axis

Perimeter (P): The length of the outside boundary of the region covered by the bacilli.

Compactness: it provides a measure of how closely the shape of the bacilli approaches a circle, and it is the ration the perimeter and area of the bacilli.

$$\mathbf{Comp} = \boldsymbol{P^2/A} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3$$

Roundness(R): it measures the degree of roundness (circularity) of the shape of bacilli.

$$R = \frac{4\pi A}{P^2} \text{.....4}$$

Where *A* is area a bacilli detected region in the image and *P* is the perimeter.

Major Axis Length (Major): It is the distance between the end points of the longest line that could be drawn through the PTB bacilli region. The major axis end points are found by computing the pixel distance between every combination of border pixels in the PTB bacilli boundary and finding the pair with the maximum length.

Minor Axis Length (Minor): It is the distance between the end points of the longest line that could be drawn through the PTB bacilli while maintaining perpendicularity with the major axis.

EquivDimeter: It is the diameter of a circle having the same area as the area a PTB bacilli region and computed as:

$$D = \sqrt{4A}/\pi \text{.....5}$$

Where A is the area of a PTB bacilli region in the sputum smear image

As a summery, the feature extraction likes area, compactness and eccentricity consist basic features which characterizes the objects properties of an image as states in (Chang, 2012). This can be used for the PTB bacilli classification of positive or negative of stained sputum smear images.

3.1.4.2. Color features

Color is one of the features of the bacilli they have different color variation of each PTB bacilli type and color analysis computed by taking the mean value of RGBs (Red, Green and Blue) components and mean value of HSVs (Hue, Saturation and Value) components. RGB color spaces are widely used and are normally the default color space for storing and representing

digital images. The RGB color space is used by computers, graphics cards and monitors. Graphics file format stores RGB images as 24-bit images, where the red, green and blue components are 8-bit for each (Tink and Ajoy, 2005). The HSV color space is more intuitive to how people experience color than the RGB color space. HSV are the common perceptual descriptors of light sensation that hue is an attribute of light that distinguish one color from the other while saturation is describe the amount of whiteness of a light source in a given image. The value is measure of the brightness of a given image ((Tink and Ajoy, 2005). Normalized RGB is a representation that is easily obtained from the RGB values by a simple normalization procedure.

$$R = \frac{R}{R+G+B} \ , G = R = \frac{G}{R+G+B} \ , R = \frac{B}{R+G+B} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 6$$

$$H = \arccos\left\{\frac{[(R-G)+(R-B)]/2}{\left[(R-G)^2+(R-B)(G-B)\right]^{1/2}}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 7$$

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 8$$

$$V = \frac{1}{3}(R+G+B) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 9$$

Therefore, to compute the mean value of each component of these color model spaces, the researcher used MATLAB 2016a to split each component. Since MATLAB has built in function to convert RGB to HSV or HSV to RGB color spaces by coded. By using the built in function of MATLAB RGB, which is RGB split to red, green and blue components. Hence, the color features are extracted by computing the mean value of RGB and HSV of stained sputum smear images. That is, the mean value of red, green blue, hue, saturation and value color are computed

from each component. As summery, there were fourteen features (eight morphology features and six color features) was used for sputum smear image classification of PTB bacilli detected.

3.1.5. PTB Bacilli Image Classification

Image classification is a final stage in pattern recognition system where each unknown patterns is assigned to a category. As described in section 2.5.6, pattern recognition is the study of how machines can observe the environment, learn to distinguish patterns of interest, make sound and reasonable decision about the categories of the patterns. Patterns are any entity or object. For PTB bacilli images are patterns.

In classification, the objective was to categorize the objects in the bacilli images from a set of measurements of the object. The measured values are the features of patterns. A set of similar objects or patterns possessing more or less identical features are said to belong to a certain category called classes (PTB bacilli positive or negative detected). The images classification model has three main components. They are representation of image features, learning and testing for semantic categories using these representations and the classifiers. As describe in section 2.5.6 a classifier is a program that takes input feature vectors and assigns it to one of a set of designed classes (identify the PTB bacilli in sputum smear images).

3.1.5.1. Feature representation

Features or attributes are values measured from stained sputum smear images. As it was discussed in the previous section, they were two basic features of PTB bacilli from sputum smear images, namely morphological and color features. Fourteen (14) features (eight morphological features and six color features) were used to identify PTB bacilli in sputum smear images from

sputum smear slide in two classes PTB bacilli detected and non-PTB bacilli detected. Among the sputum smear images, two major classes are positive (assign value is 1) and negative (assign value is -1), were selected for this study. In this study, there were two classes and 180 sputum smear images which were the total number of dataset. In the training process the class values were provided by supervised learning method. In order to the classification accuracy of the system, testing features dataset which was not in the training dataset was used and were randomly selected by the crossvalind function which creates random partitions depends on the state of the default random stream.

3.1.5.2. Training and testing phase process

After the random selection of training and testing dataset by crossvalind () function, it need to train the classifier with 70% training sample content that represent members of the classes. It was very important to find good training sample because the quality of the training sample has direct impact on the quality of classification. The second parameter to train the classifier was a training label specification, which was a sequence of training and testing label elements. Each label elements represents a node in the training and testing dataset. The label elements must be in the order corresponding to the specified training nodes, and they each specified to which class the corresponding training node belongs.

After training model the performance of the classifier was measured using 30% of test dataset for accuracy, sensitivity, specificity and F-measure based on the following performance measuring parameters (Pawar & Ganorkar, 2016).

Accuracy: Accuracy is the most intuitive performance measures and it is simply a ratio of correctly classified observation to the total observation. If the accuracy is high the developed model was best.

Sensitivity: if the test is highly sensitive and the test result is negative, it is possible to nearly be sure that they don't have PTB bacilli detected. The true positive rate and finds all PTB disease.

Specificity: if the test result for a highly specific test is positive, it is nearly be sure that they actually have the PTB disease. The true negative rate and it finds only PTB bacilli detected.

F-measure: F-measure is the weighted average of sensitive and specificity. It is a measure of test's accuracy. Therefore, this measure takes both false positive and false negative into account. Intuitively it is not understand as accuracy, but F-measure is usually more useful than accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \ldots\ldots 10$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad \ldots\ldots 11$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad \ldots\ldots 12$$

$$\text{F-measure} = \frac{2TP}{2TP+FP+FN} \quad \ldots\ldots 13$$

Where, TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative

3.1.5.3. SVM and KNN Classifiers Algorithm

In this study, the PTB bacilli detected was used better classification performance measured used accuracy, sensitivity, specificity and F-measure. Support vector machine (SVM) is supervised machine learning is the best suitable algorithm which can be used for classification challenges. For this study, as described in the section 2.5.6.1 SVM which was mostly used in classification

problems is selected for better accuracy than the other KNN classifiers. It can used be to find optimal hyper plane as shown in figure SMV (section 2.5.6.1) to separate different categories of input data into higher dimension features space. And also SVM has advantage of fast training techniques, even with large number of input data. In these algorithms, the plot each data item as a point in n-dimensional space (where n is number of features used) with the value of each features (attributes) being the value of a particular coordinate. Then, the perform classification by finding the hyper-plane that differentiates the two classes (PTB bacilli or non-PTB bacilli detected).

KNN algorithm is the simplest PTB bacilli classification technique. The KNN is used to classify the PTB bacilli detect on the basis of most similar or closest training samples in the feature space. Majority vote of neighbors is used to classify the object. For an unknown sample and known training data, all the distances between all training set samples and unknown samples can be used both morphological and color feature calculated. The smallest distance corresponds to training set sample close to unknown sample.

3.1.5.4. PTB Bacilli Detection

As it was explain in the previous section, in the first phase images were used to create the knowledge base (image database) for detection purpose. In this phase, sputum smears images that were obtained from different sample acquired images that was used for training. This sputum smear images that passed through image preprocessing techniques were used in the training phase. In addition to image preprocessing sputum smear images were segmented using K-means clustering methods of image segmentation to identify the region of interest. After identify the region of bacilli images of interest, useful features were extracted in order to reduce the complex of the computational cost of the system. Since, fourteen different features included

morphological and color feature extracted of an image for creation of knowledge base. After necessary features of images were generated, SVM and KNN algorithms that is built during the training phase to identify the PTB bacilli into classes of PTB bacilli or Not PTB bacilli detect was used.

In this work, an image processing based system for automatic detection the PTB bacilli was developed. During the creation of knowledge base (database), images preprocessed and segmented to identify the region of interest. Then, fourteen different features from the images were extracted using morphological and color features. SVM and KNN classification is trained using fourteen features input from the individual image and two output vectors that represent two different classes of disease to represent the knowledge base. In the second phase, testing phase the knowledge base was used to identify of PTB bacilli detected.

In image segmentation techniques were essential stage of image analysis methods that determine the quality final results. Segmentation methods used to separate an image into a set of homogenous and meaningful region, such that the pixels in each partition region possess an identical set of features. From the segmented images appropriate features were extracted, which usually resulted as binary image. As described in section 3.5, some morphological and color features were identified. These features are used to identify PTB bacilli from a given stained sputum smear image to provide positive or negative classes. The following sections describe the detail implementation of image segmentation, feature extraction and classification processes.

3.2. Development Environment

The development of PTB bacilli detection system by integrating image analysis techniques needs a lot of money to invest. Starting from image acquisition, it need high quality digital camera, and well established and controlled environment to acquired images. In addition to this, image preprocessing techniques were resource intensive. They need powerful computers with high resolution processing speed, larger memory and disk capacity. The developed system was developed and tested on a PC of processor is Intel® core™i5-4200U CPU with 2.30GHz speed, memory (RAM) is 4.00GB of hard Disk capacity, with 64 bit Microsoft Windows 8.1 operating system.

3.3. Binary Image Analysis

A binary image is an image whose pixel values were changed to 0 and 1 or black and white. In this study, the white is inverted to black which is indicated the object of interest or the mass region on the image, and the black is inverted to white which indicated the other part of the bacilli image. The researcher was used MATLAB R2016a which is multi-paradigm numerical computing properties programming language environment developed by Mathwork (www.mathworks.com, 2018). It can display, edit, delete, process and analyze many formats and types of image tools.

MATLAB is powerful tool to analyze and process images. It read and loads a sequence of images which are stored in one folder one by one. It can also enhance the quality of each images, remove noises and changes the image to binary image for feature extraction purposes. As an input image read by MATLAB code using shown in figure 3.2.2., I = readimage(filename)

function, when image file name depends on the image format of file in the data store. The image formats supported by readimage function are those formats supported by imread.

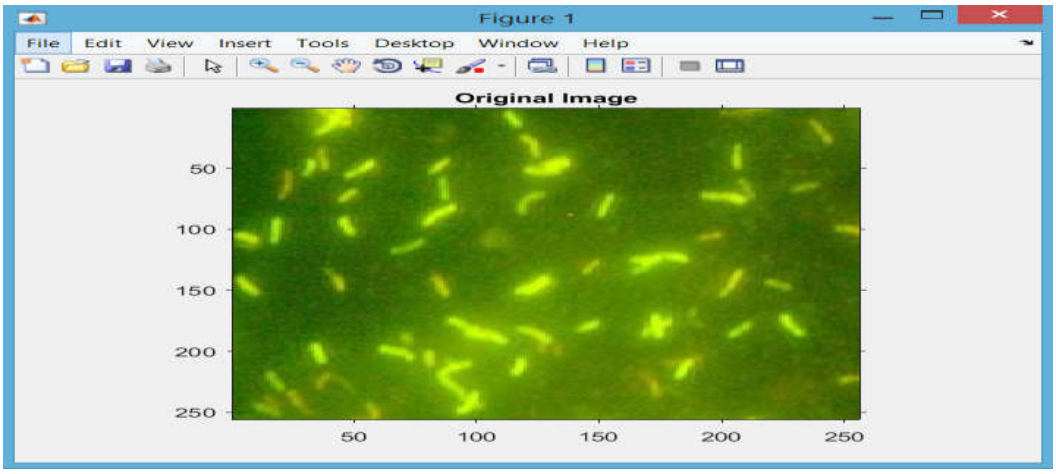


Figure 3 4: Sample of input sputum smear image

Image segmentation was done to separate each stained sputum smear images from the background by using k-means clustering techniques, which was described in section 3.1.3. First, using built in imread () function the input sputum smear image is taken then resized to [300 400]. HSV(Hue, Saturation, Value) color space is often used by people who are selecting colors from color image corresponds better to how people experience color than the RGB color space does using rgb2hsv(). The HSV color space is widely used to generate high quality of images in computer graphics. It is used to select various different colors needed for identify particularly bacilli images. Because it used to select the desired (bacilli object) color. HSV color is important for identify bacilli image objects as it gives the color according to human perception about PTB bacilli detection. HSV is very useful in image processing when we want to do histogram equalization of a color image of PTB bacilli detected based intensity components to separate bacilli objects from intensity for various reason, such as robustness to lightly changes, or remove shows undesired objects. Then compute and plot the histogram of each HSV color. Then,

calculate the average each HSV values across all of the image pixels as shown in figure 4.2. In these color components of HSV, value image color is best for identifying PTB bacilli from other undesired objects.

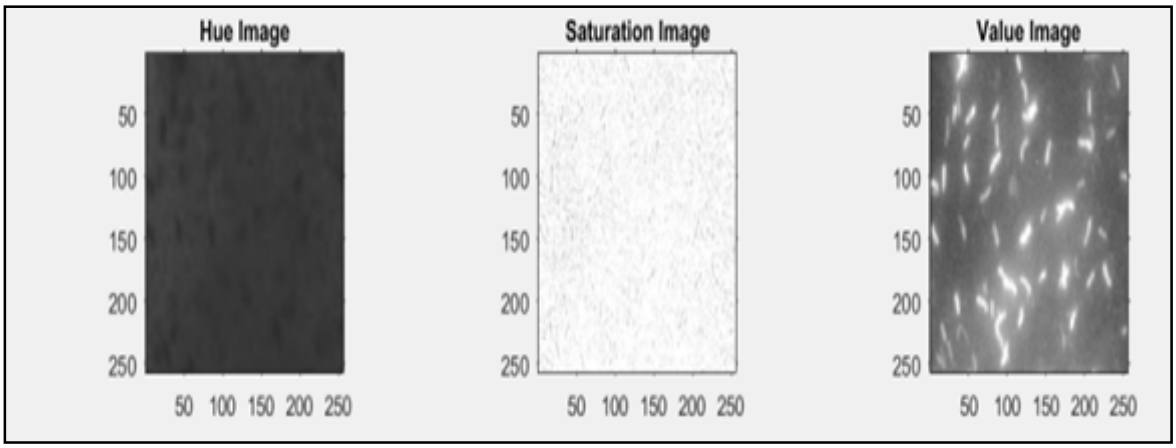


Figure 3 5: Converted RGB to HSV color space with Histogram Value

After applying HSV color, V (value) image was selected to better understand components used to represent a color for identifying PTB bacilli detected based intensity components to separate objects. Then, the next step was applying the Gaussian filter to remove noise in sputum smear image based value image color. In addition, contrast enhance was used to improve the quality of sputum smear images which is more important for user interpretation. It is differentiating in visual a property that makes bacilli objects distinguished from other objects and backgrounds. To convert Gaussian filter algorithms: G = imgaussfilt () function and to convert contrast enhanced, C = imadjust () function was used in MATLAB as shown in figure 4.3.

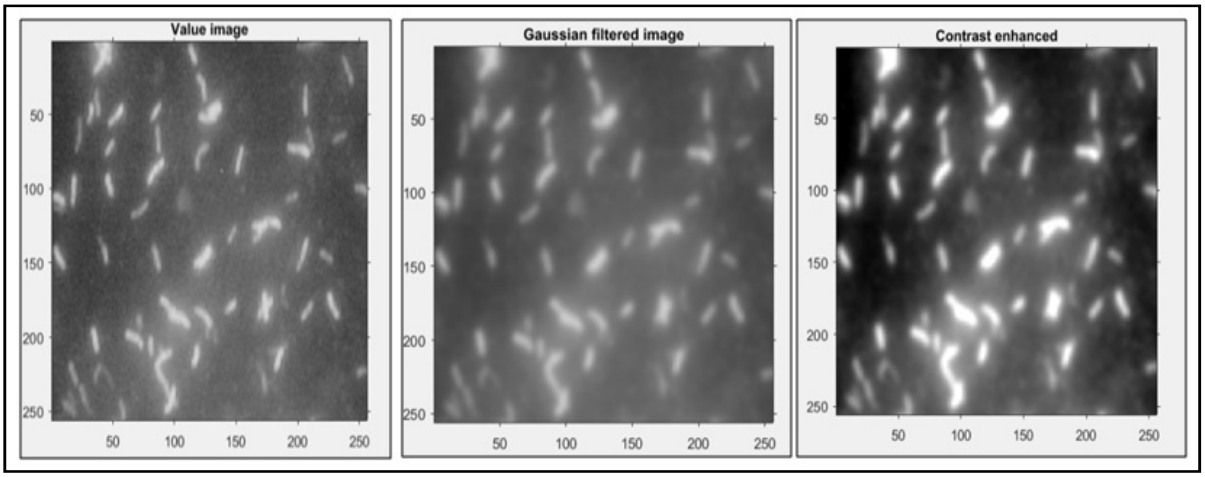
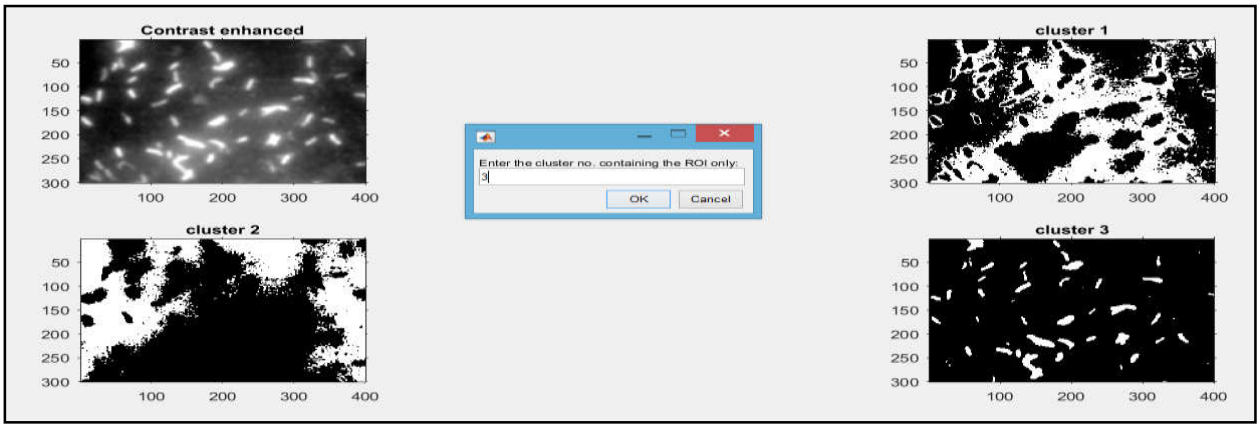
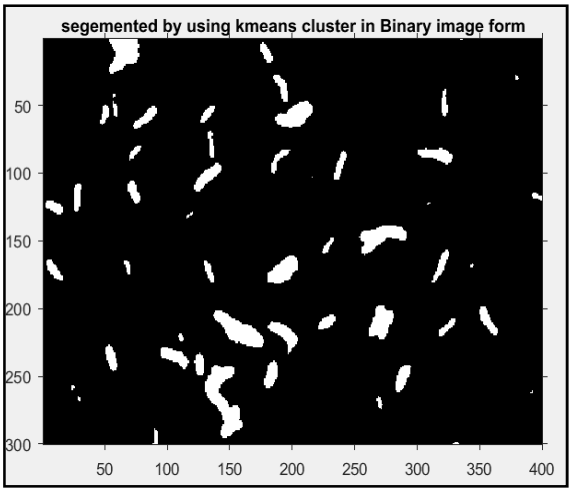


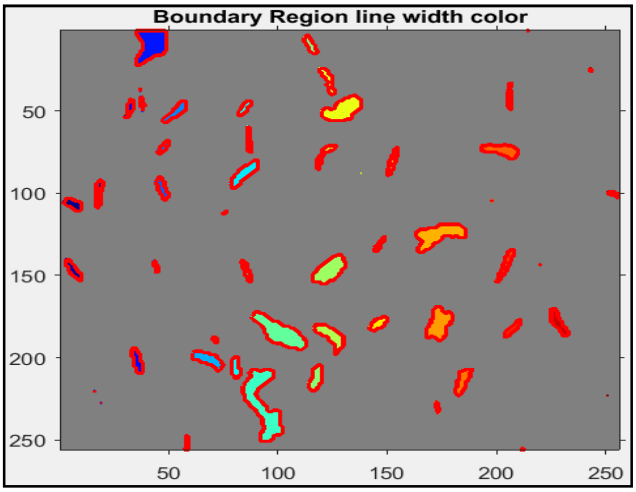
Figure 3 6: Value color image, Gaussian filter and contrast enhanced



A



B



C

Figure 3 7: A: k-means cluster used segmented image (Binary image form), B: segmented image after k-means results C: Region boundary with line color segmented

The above figure shows that the k-mean clustering algorithms is used to generate the cluster 1, 2, 3 and 4 based on the potential value of the image Figure 3.7. (A). The number of center is generated based on number of the cluster K. This centre is used as initial centre in K-means algorithm. Using the k-means algorithm, the image is segmented into K number of cluster. After segmentation of image, the image can still contain some unwanted region or noise. Although k-means has the greatest advantages of being easy to implement, it has some drawbacks. The quality of the final clustering results depends on the arbitrary selection of initial centroid. Thus, if initial centroid is randomly chosen from cluster 1-4 it will get different result for different centre. The initial centre will be carefully chosen so that we get our desire segmentation of bacilli objects as shown in figure 3.7. (B), then, was needed to consider while designed the k-means clustering. Then, the detected bacilli region of interest (ROI) for using boundary region line width color methods on the image or the region of each pixel that we are interested to compute for morphological and color feature was selected.

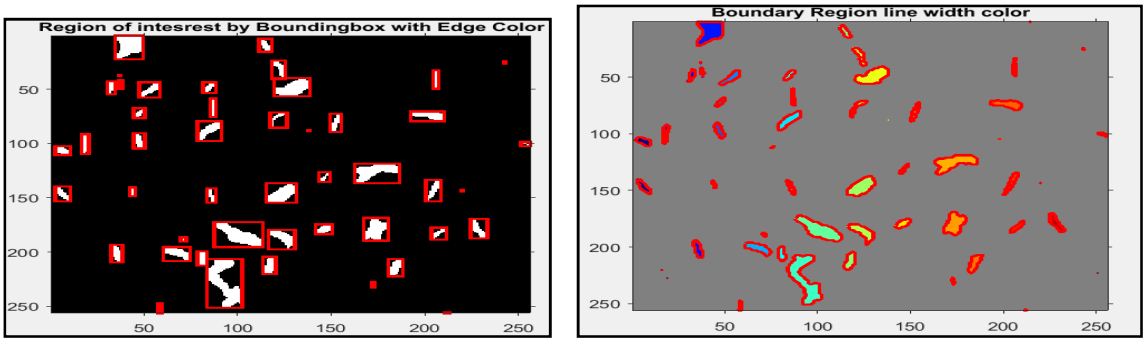
3.4. Feature extraction analysis

Image analysis involves investigation of the image data specific to identify PTB bacilli detected. Normally, the raw data of a set of images is analyzed how they can be used to extract desired information about sputum smear images. In image processing and pattern recognition, feature extraction analysis is an important step, which is a special form of dimensionality reduction. When the input data is too large to be processed and suspected to be redundant then the data is transformed into a reduced set of feature representations for identifying PTB bacilli detected.

Feature extraction was used to identify the PTB bacilli detection that contains morphological and color features described in section 3.1.5.1.

3.4.1. Morphological features analysis

A morphological feature is the size and shape characteristics of sputum smear images of PTB. As mentioned in section 3.5.1, eight morphological features were identified. They were area, eccentricity, compactness, perimeter, roundness, major length axis, minor length axis and EquivDimeter of the PTB bacilli detection. These features were computed from the image binary analysis as described in the previous section. Figure 3.8(A) shows a region of interest of PTB bacilli images that were interested to compute its morphological features based region of interest by boundary with edge color.



A B

Figure 3 8: A: morphological feature computed based on label region of interest B: morphological feature computed based on boundary region line label region of interest with color. Based on the detected region of interest as shown in figure 3.8 (B) morphological features were computed on each sputum smear image by using region properties image analysis methods in MATLAB.

The results of these eight morphological features computed the following MATLAB function: These features were computer from the binary image analysis from sputum smear image region of interest and the measured values are in pixels. As shown table 3.1, the mean value of each feature computed by used formula described in section 3.1.4.1 from each sputum smear image for the better performance using.

Variables - abdinjira

abdinjira

180x15 table

	8 Eccentricity	9 Perimeter	10 Compactness	11 Roundness	12 Equidimiter	13 MajorlenthAxis	14 MinorlengthAxis
1	'0 661691756...	'33 3349772...	'0 87798027092...	'1 182908128...	'7 061725335...	'12 59038183431...	'5 51768358436115'
2	'0 569004742...	'22 6025315...	'0 89465902222...	'1 021475344...	'4 583583996...	'8 236674775358...	'3 82484535750451'
3	'0 619066246...	'15 5229047...	'0 92823862457...	'0 899005925...	'4 320576352...	'7 034420055812...	'3 61818541864547'
4	'0 864844804...	'22 0869285...	'0 92861644873...	'1 013951287...	'6 869699547...	'10 48182709684...	'4 92660092274658'
5	'0 701617873...	'18 3675546...	'0 90098055080...	'0 966448558...	'5 059728457...	'8 309208273306...	'3 98182020957913'
6	'0 684556453...	'23 2209607...	'0 90560822235...	'1 392171678...	'4 163836231...	'7 475088046526...	'3 64499764239196'
7	'0 667254452...	'18 7609586...	'0 91899706501...	'0 923656585...	'4 488370868...	'8 186696970439...	'3 57014561599573'
8	'0 681713091...	'19 2629044...	'0 91612733876...	'0 933436077...	'4 753274592...	'7 629323065014...	'3 99029812444216'
9	'0 795015617...	'21 4638476...	'0 81819156151...	'1 599839967...	'4 870737382...	'10 84196891387...	'3 45839115666659'
10	'0 529297235...	'16 6124810...	'0 94185069405...	'0 773072371...	'3 295737504...	'5 234310411629...	'2 94512215234305'
11	'0 601859516...	'24 7933913...	'0 91347301008...	'1 055919095...	'3 841480978...	'6 854221927935...	'3 19630121014874'
12	'0 723208413...	'13 4336250...	'0 91145476602...	'0 884524419...	'3 640125990...	'6 306160390355...	'2 87997686367360'
13	'0 452072455...	'4 65532453...	'0 95076819550...	'0 429869243...	'1 921717340...	'2 934012332038...	'1 70462143786453'
14	'0 411603550...	'4 04909698...	'0 95267091967...	'0 361848321...	'1 942251929...	'2 931019596505...	'1 67636605771673'
15	'0 397973831...	'3 92844500...	'0 95783947618...	'0 357267958...	'1 862042994...	'2 766936554375...	'1 66339218649669'
16	'0 468540918...	'8 23100000...	'0 93909192117...	'0 652233225...	'2 250535946...	'3 721707983220...	'2 03308393568711'
17	'0 451996208...	'3 93521247...	'0 94938891397...	'0 363344730...	'1 898233788...	'2 843147628085...	'1 66595415338216'
18	'0 518057181...	'20 7465285...	'0 95644005114...	'0 784300368...	'3 023000371...	'4 876411693191...	'2 78222542269830'
19	'0 613869477...	'9 34547368...	'0 94816202037...	'0 659347023...	'3 099582706...	'4 981231250352...	'2 49794322703850'
20	'0 372667296...	'3 35871814...	'0 96482848207...	'0 306623746...	'1 833213779...	'2 543929406101...	'1 65526920786936'
21	'0 615437069...	'9 11722302...	'0 95533104299...	'0 550644925...	'3 038278020...	'4 480053494803...	'0'
22	'0 523127148...	'11 8456533...	'0 95454087831...	'0 519348714...	'3 081416887...	'4 822986103941...	'2 52596034685228'
23	'0 819867915...	'12 0681666...	'0 87295800264...	'0 948604794...	'3 799613053...	'6 682413397695...	'2 81117869093805'

Table 3 1: Morphological feature value of each label region of interest

3.4.2. Color feature analysis

Color feature is a visual attributes of sputum smear images that results from the light emitted or transmitted or reflected. As described in section 3.1.4.2, we have identified six color features. They were computing the mean value of RGB (Red, Green and Blue) color components and the

mean value of HSV (Hue, Saturation and Value) color components. Therefore, to compute the mean value of each component of these color spaces we used to split each component to separate image values. To do these, Matlab coded functions was used as shown in figure 3.8 (A).

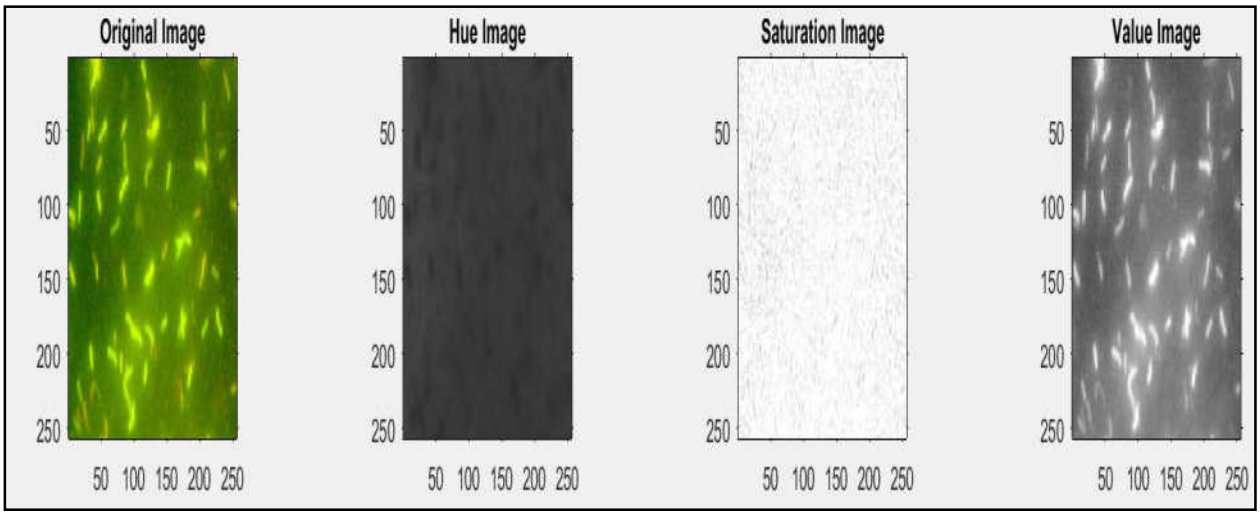


Figure 3 9: Mean value of HVS color component from original image

By using MATLAB function the RGB component was split into red, green and blue components of the results as shown in figure 3.9.

Figure 3 10: Mean value of RGB color component from original image

After each color component was split, the mean values of each component colors were computed by RGB function methods based on the specified region of interest of identified PTB bacilli as shown in table 3.2

Variables - abdinjira

abdinjira

180x15 table

	1 meanH	2 meanS	3 meanV	4 meanR	5 meanG	6 meanB
1	'0 573783...	'0 433204...	'0 890847...	'133 3901...	'185 5261...	'227 1136...
2	'0 470445...	'0 260367...	'0 655118...	'128 6047...	'151 1833...	'157 2865...
3	'0 218549...	'0 303681...	'0 863332...	'206 2403...	'200 3710...	'163 0130...
4	'0 336498...	'0 988460...	'0 257355...	'1 440750...	'65 62567...	'3 656402...
5	'0 222881...	'0 137988...	'0 915546...	'227 0696...	'210 9290...	'212 7059...
6	'0 684389...	'0 210255...	'0 751706...	'156 3016...	'153 1895...	'191 6852...
7	'0 609353...	'0 339110...	'0 833819...	'150 9731...	'158 6576...	'212 1807...
8	'0 635202...	'0 173247...	'0 879137...	'186 6370...	'193 8778...	'224 1609...
9	'0 337355...	'0 358114...	'0 886961...	'147 6035...	'226 1196...	'149 4559...
10	'0 194005...	'0 397803...	'0 405536...	'96 35054...	'100 5563...	'67 09899...
11	'0 180643...	'0 991136...	'0 447734...	'100 7273...	'107 7219...	'4 787231...
12	'0 171714...	'0 434473...	'0 546068...	'135 9071...	'137 4971...	'78 67948...
13	'0 221358...	'0 271890...	'0 511368...	'118 7728...	'130 3589...	'94 95075...
14	'0 239088...	'0 266568...	'0 513539...	'115 8334...	'130 9278...	'96 03538...
15	'0 199265...	'0 305717...	'0 477650...	'114 4381...	'121 6947...	'84 52432...
16	'0 346122...	'0 254966...	'0 526822...	'100 8759...	'134 2490...	'105 9320...
17	'0 122420...	'0 387686...	'0 555433...	'141 4528...	'126 1992...	'86 63365...
18	'0 239489...	'0 430768...	'0 537153...	'112 1490...	'124 2216...	'101 0595...
19	'0 194117...	'0 283377...	'0 515522...	'125 2688...	'131 1866...	'94 21054...
20	'0 089963...	'0 536091...	'0 613903...	'156 5161...	'117 2770...	'72 47973...
21	'0 143600...	'0 352540...	'0 531815...	'134 5976...	'126 3165...	'88 17320...
22	'0 195980...	'0 304366...	'0 511061...	'123 5043...	'129 5441...	'93 20169...
23	'0 221349...	'0 298856...	'0 486505...	'112 0792...	'124 0146...	'86 98896...

Table 3 2: Value of color feature in each label region of interest

In summary, from the above table 3.1 and table 3.2, the label of region of interest, which was computed the value of each label region of interest morphological feature and computed the mean value each component of colors (HSV and RGB) was by using MATLAB tools function. The computed mean values of combined results from the fourteen features were as shown in table 3.3.

Variables - abdinjira

abdinjira

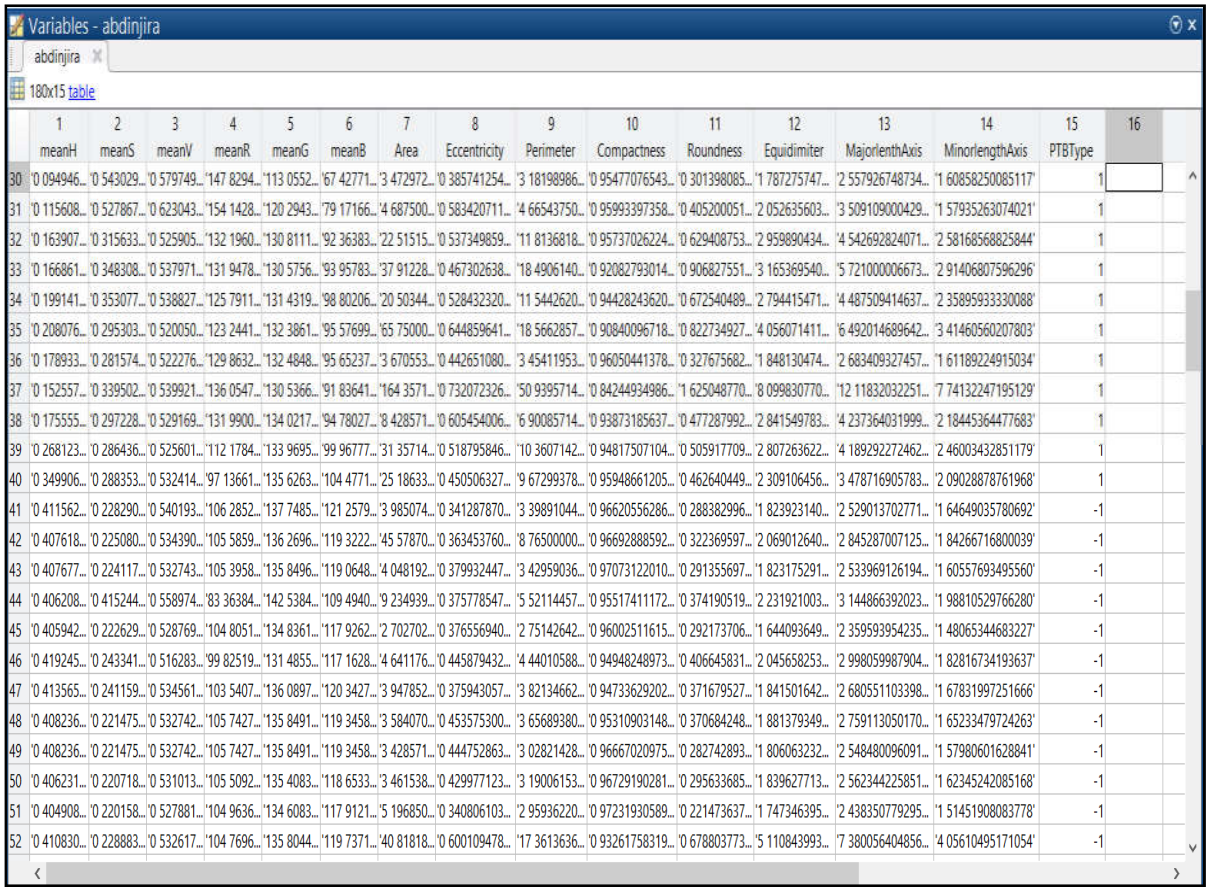
180x15 table

	1 meanH	2 meanS	3 meanV	4 meanR	5 meanG	6 meanB	7 Area	8 Eccentricity	9 Perimeter	10 Compactness	11 Roundness	12 Equidimiter	13 MajorlenthAxis	14 MinorlengthAxis
1	'0 573783...	'0 433204...	'0 890847...	'133 3901...	'185 5261...	'227 1136...	'109 2500...	'0 661691756...	'33 3349772...	'0 87798027092...	'1 182908128...	'7 061725335...	'12 59038183431...	'5 51768358436115'
2	'0 470445...	'0 260367...	'0 655118...	'128 6047...	'151 1833...	'157 2865...	'52 23423...	'0 569004742...	'22 6025315...	'0 89465902222...	'1 021475344...	'4 583583996...	'8 236674775358...	'3 82484535750451'
3	'0 218549...	'0 303681...	'0 863332...	'206 2403...	'200 3710...	'163 0130...	'26 79761...	'0 619066246...	'15 5229047...	'0 92823862457...	'0 899005925...	'4 320576352...	'7 034420055812...	'3 61818541864547'
4	'0 336498...	'0 988460...	'0 257355...	'1 440750...	'65 62567...	'3 656402...	'39 57142...	'0 864844804...	'22 0869285...	'0 92861644873...	'1 013951287...	'6 869699547...	'10 48182709684...	'4 92660092274658'
5	'0 222881...	'0 137988...	'0 915546...	'227 0696...	'210 9290...	'212 7059...	'48 08403...	'0 701617873...	'18 3675546...	'0 90098055080...	'0 966448558...	'5 059728457...	'8 309208273306...	'3 98182020957913'
6	'0 684389...	'0 210255...	'0 751706...	'156 3016...	'153 1895...	'191 6852...	'37 86274...	'0 684556453...	'23 2209607...	'0 90560822235...	'1 392171678...	'4 163836231...	'7 475088046526...	'3 64499764239196'
7	'0 609353...	'0 339110...	'0 833819...	'150 9731...	'158 6576...	'212 1807...	'35 65289...	'0 667254452...	'18 7609586...	'0 91899706501...	'0 923656585...	'4 488370868...	'8 186696970439...	'3 57014561599573'
8	'0 635202...	'0 173247...	'0 879137...	'186 6370...	'193 8778...	'224 1609...	'37 31210...	'0 681713091...	'19 2629044...	'0 91612733876...	'0 933436077...	'4 753274592...	'7 629323065014...	'3 99029812444216'
9	'0 337355...	'0 358114...	'0 886961...	'147 6035...	'226 1196...	'149 4559...	'25 58095...	'0 795015617...	'21 4638476...	'0 81819156151...	'1 599839967...	'4 870737382...	'10 84196891387...	'3 45839115666659'
10	'0 194005...	'0 397803...	'0 405536...	'96 35054...	'100 5563...	'67 09899...	'36 24050...	'0 529297235...	'16 6124810...	'0 94185069405...	'0 773072371...	'3 295737504...	'5 234310411629...	'2 94512215234305'
11	'0 180643...	'0 991136...	'0 447734...	'100 7273...	'107 7219...	'4 787231...	'59 47826...	'0 601859516...	'24 7933913...	'0 91347301008...	'1 055919095...	'3 841480978...	'6 854221927935...	'3 19630121014874'
12	'0 171714...	'0 434473...	'0 546068...	'135 9071...	'137 4971...	'78 67948...	'17 55000...	'0 723208413...	'13 4336250...	'0 91145476602...	'0 884524419...	'3 640125990...	'6 306160390355...	'2 87997686367360'
13	'0 221358...	'0 271890...	'0 511368...	'118 7728...	'130 3589...	'94 95075...	'4 866459...	'0 452072455...	'4 65532453...	'0 95076819550...	'0 429869243...	'1 921717340...	'2 934012332038...	'1 70462143786453'
14	'0 239088...	'0 266568...	'0 513539...	'115 8334...	'130 9278...	'96 03538...	'4 371237...	'0 411603550...	'4 04909698...	'0 95267091967...	'0 361848321...	'1 942251929...	'2 931019596505...	'1 67636605771673'
15	'0 199265...	'0 305717...	'0 477650...	'114 4381...	'121 6947...	'84 52432...	'4 247500...	'0 397973831...	'3 92844500...	'0 95783947618...	'0 357267958...	'1 862042994...	'2 766936554375...	'1 66339218649669'
16	'0 346122...	'0 254966...	'0 526822...	'100 8759...	'134 2490...	'105 9320...	'9 494545...	'0 468540918...	'8 23100000...	'0 93909192117...	'0 652233225...	'2 250535946...	'3 721707983220...	'2 03308393568711'
17	'0 122420...	'0 387686...	'0 555433...	'141 4528...	'126 1992...	'86 63365...	'4 415204...	'0 451996208...	'3 93521247...	'0 94938891397...	'0 363344730...	'1 898233788...	'2 843147628085...	'1 66595415338216'
18	'0 239489...	'0 430768...	'0 537153...	'112 1490...	'124 2216...	'101 0595...	'63 34285...	'0 518057181...	'20 7465285...	'0 95644005114...	'0 784300368...	'3 023000371...	'4 876411693191...	'2 78222542269830'
19	'0 194117...	'0 283377...	'0 515522...	'125 2688...	'131 1866...	'94 21054...	'12 10526...	'0 613869477...	'9 34547368...	'0 94816202037...	'0 659347023...	'3 099582706...	'4 981231250352...	'2 49794322703850'
20	'0 089963...	'0 536091...	'0 613903...	'156 5161...	'117 2770...	'72 47973...	'3 895752...	'0 372667296...	'3 35871814...	'0 96482848207...	'0 306623746...	'1 833213779...	'2 543929406101...	'1 65526920786936'
21	'0 143600...	'0 352540...	'0 531815...	'134 5976...	'126 3165...	'88 17320...	'15 69064...	'0 615437069...	'9 11722302...	'0 95533104299...	'0 550644925...	'3 038278020...	'4 480053494803...	'0'
22	'0 195980...	'0 304366...	'0 511061...	'123 5043...	'129 5441...	'93 20169...	'47 20000...	'0 523127148...	'11 8456533...	'0 95454087831...	'0 519348714...	'3 081416887...	'4 822986103941...	'2 52596034685228'
23	'0 221349...	'0 298856...	'0 486505...	'112 0792...	'124 0146...	'86 98896...	'12 58333...	'0 819867915...	'12 0681666...	'0 87295800264...	'0 948604794...	'3 799613053...	'6 682413397695...	'2 81117869093805'

Table 3 3: Computed combined both mean value of morphological features and mean value of color features

The results of mean value of the fourteen features (eight morphological features and split of color component of features) as the above table contain hue, saturation, value, red, green, blue, area, eccentricity, perimeter, compactness, roundness, equivdimeter, major length axis and minor length axis. These features were used for better classification to keep the results to identify the PTB detected or not detected. It was done column normalization on the obtained extracted features to identify the PTB bacilli. In addition, the radiologist (lab technician) reading of each sputum smear microscopic image label as PTB positive is (1) and as PTB negative is (-1) was shows in table 3.4, to each features as input to differentiate the two classes of the sputum smear images. Hence, the totals input features were fourteen and region of interest label for each

sputum smear image. These features were used to classify the PTB positive or PTB negative detected.



Variables - abdinjira

abdinjira

180x15 table

	1 meanH	2 meanS	3 meanV	4 meanR	5 meanG	6 meanB	7 Area	8 Eccentricity	9 Perimeter	10 Compactness	11 Roundness	12 Equidimiter	13 MajorlenthAxis	14 MinorlengthAxis	15 PTBType	16
30	'0 094946...	'0 543029...	'0 579749...	'147 8294...	'113 0552...	'67 42771...	'3 472972...	'0 385741254...	'3 18198986...	'0 95477076543...	'0 301398085...	'1 787275747...	'2 557926748734...	'1 60858250085117'	1	
31	'0 115608...	'0 527867...	'0 623043...	'154 1428...	'120 2943...	'79 17166...	'4 687500...	'0 583420711...	'4 66543750...	'0 95993397358...	'0 405200051...	'2 052635603...	'3 509109000429...	'1 57935263074021'	1	
32	'0 163907...	'0 315633...	'0 525905...	'132 1960...	'130 8111...	'92 36383...	'22 51515...	'0 537349859...	'11 8136818...	'0 95737026224...	'0 629408753...	'2 959890434...	'4 542692824071...	'2 58168568825844'	1	
33	'0 166861...	'0 348308...	'0 537971...	'131 9478...	'130 5756...	'93 95783...	'37 91228...	'0 467302638...	'18 4906140...	'0 92082793014...	'0 906827551...	'3 165369540...	'5 721000006673...	'2 91406807596296'	1	
34	'0 199141...	'0 353077...	'0 538827...	'125 7911...	'131 4319...	'98 80206...	'20 50344...	'0 528432320...	'11 5442620...	'0 94428243620...	'0 672540489...	'2 794415471...	'4 487509414637...	'2 35895933330088'	1	
35	'0 208076...	'0 295303...	'0 520050...	'123 2441...	'132 3861...	'95 57699...	'65 75000...	'0 644859641...	'18 5662857...	'0 90840096718...	'0 822734927...	'4 056071411...	'6 492014689642...	'3 41460560207803'	1	
36	'0 178933...	'0 281574...	'0 522276...	'129 8632...	'132 4848...	'95 65237...	'3 670553...	'0 442651080...	'3 45411953...	'0 96050441378...	'0 327675682...	'1 848130474...	'2 683409327457...	'1 61189224915034'	1	
37	'0 152557...	'0 339502...	'0 539921...	'136 0547...	'130 5366...	'91 83641...	'164 3571...	'0 732072326...	'50 9395714...	'0 84244934986...	'1 625048770...	'8 099830770...	'12 11832032251...	'7 74132247195129'	1	
38	'0 175555...	'0 297228...	'0 529169...	'131 9900...	'134 0217...	'94 78027...	'8 428571...	'0 605454006...	'6 90085714...	'0 93873185637...	'0 477287992...	'2 841549783...	'4 237364031999...	'2 18445364477683'	1	
39	'0 268123...	'0 286436...	'0 525601...	'112 1784...	'133 9695...	'99 96777...	'31 35714...	'0 518795846...	'10 3607142...	'0 94817507104...	'0 505917709...	'2 807263622...	'4 189292272462...	'2 46003432851179'	1	
40	'0 349906...	'0 288353...	'0 532414...	'97 13661...	'135 6263...	'104 4771...	'25 18633...	'0 450506327...	'9 67299378...	'0 95948661205...	'0 462640449...	'2 309106456...	'3 478716905783...	'2 09028878761968'	1	
41	'0 411562...	'0 228290...	'0 540193...	'106 2852...	'137 7485...	'121 2579...	'3 985074...	'0 341287870...	'3 39891044...	'0 96620556286...	'0 288382996...	'1 823923140...	'2 529013702771...	'1 64649035780692'	-1	
42	'0 407618...	'0 225080...	'0 534390...	'105 5859...	'136 2696...	'119 3222...	'45 57870...	'0 363453760...	'8 76500000...	'0 96692888592...	'0 322369597...	'2 069012640...	'2 845287007125...	'1 84266716800039'	-1	
43	'0 407677...	'0 224117...	'0 532743...	'105 3958...	'135 8496...	'119 0648...	'4 048192...	'0 379932447...	'3 42959036...	'0 97073122010...	'0 291355697...	'1 823175291...	'2 533969126194...	'1 60557693495560'	-1	
44	'0 406208...	'0 415244...	'0 558974...	'83 36384...	'142 5384...	'109 4940...	'9 234939...	'0 375778547...	'5 52114457...	'0 95517411172...	'0 374190519...	'2 231921003...	'3 144866392023...	'1 98810529766280'	-1	
45	'0 405942...	'0 222629...	'0 528769...	'104 8051...	'134 8361...	'117 9262...	'2 702702...	'0 376556940...	'2 75142642...	'0 96002511615...	'0 292173706...	'1 644093649...	'2 359593954235...	'1 48065344683227'	-1	
46	'0 419245...	'0 243341...	'0 516283...	'99 82519...	'131 4855...	'117 1628...	'4 641176...	'0 445879432...	'4 44010588...	'0 94948248973...	'0 406645831...	'2 045658253...	'2 998059987904...	'1 82816734193637'	-1	
47	'0 413565...	'0 241159...	'0 534561...	'103 5407...	'136 0897...	'120 3427...	'3 947852...	'0 375943057...	'3 82134662...	'0 94733629202...	'0 371679527...	'1 841501642...	'2 680551103398...	'1 67831997251666'	-1	
48	'0 408236...	'0 221475...	'0 532742...	'105 7427...	'135 8491...	'119 3458...	'3 584070...	'0 453575300...	'3 65689380...	'0 95310903148...	'0 370684248...	'1 881379349...	'2 759113050170...	'1 65233479724263'	-1	
49	'0 408236...	'0 221475...	'0 532742...	'105 7427...	'135 8491...	'119 3458...	'3 428571...	'0 444752863...	'3 02821428...	'0 96667020975...	'0 282742893...	'1 806063232...	'2 548480096091...	'1 57980601628841'	-1	
50	'0 406231...	'0 220718...	'0 531013...	'105 5092...	'135 4083...	'118 6533...	'3 461538...	'0 429977123...	'3 19006153...	'0 96729190281...	'0 295633685...	'1 839627713...	'2 562344225851...	'1 62345242085168'	-1	
51	'0 404908...	'0 220158...	'0 527881...	'104 9636...	'134 6083...	'117 9121...	'5 196850...	'0 340806103...	'2 95936220...	'0 97231930589...	'0 221473637...	'1 747346395...	'2 438350779295...	'1 51451908083778'	-1	
52	'0 410830...	'0 228883...	'0 532617...	'104 7696...	'135 8044...	'119 7371...	'40 81818...	'0 600109478...	'17 3613636...	'0 93261758319...	'0 678803773...	'5 110843993...	'7 380056404856...	'4 05610495171054'	-1	

Table 3 4: Labelling of PTB bacilli detected positive and negative assigned by pathologists

3.4.3. Graphical User Interfaces of PTB Bacilli Detection

A graphical user interface (GUI) is a set of techniques and mechanisms used for interactive communication between a program and a user. GUI has been designed for the user action to display the PTB bacilli detection results. It gives the user a better perspective of the operation that they can perform. GUI of PTB bacilli detection can make programs easier to use by proving them with a consistent appearance and with intuitive controls like button, boxes, axis and menu. In this study the researcher developed a GUI using guide user simple to browse image and

analysis the display results of the PTB bacilli detected or not. The user can browse image, by clicked button of the components at any location.

After loading of the image is done, PTB bacilli detected button is displaying the image processed performed and segmented values. Finally, the presented result gives the user a better view about each processed whether PTB bacilli positive or negative at click of the button. Generally, GUI can be used for identify PTB positive or negative after analyzed. The same GUI can be used to image processed by altering the callbacks. Using the GUI based programs allows us to change the parameters without rewriting the program and allows fast and efficient detection of PTB. Figure 4.8, shows the designed GUI where the results can be displayed clearly i.e., PTB bacilli positive or negative.

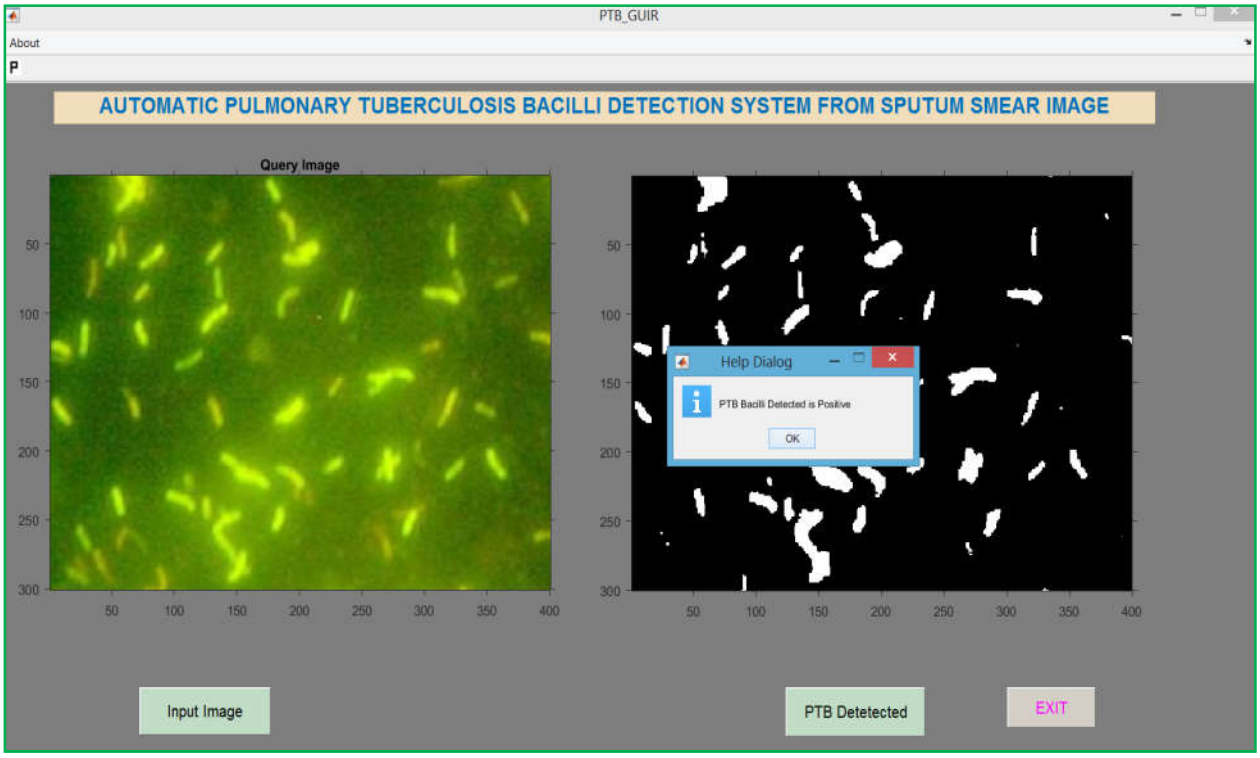


Figure 3.11: GUI interface of APTB bacilli detection

Generally, the automatic detection PTB bacilli disease variety has five steps of image processing techniques. They were image acquisition, preprocessing, segmentation, and feature extraction and classification. In the first step image acquisition is the processing of capture of acquiring of an image from stained sputum smears slides collected were taken from National Tuberculosis Reference Laboratory at EPHI. The number acquired images from each of the two categories of the PTB positive and negative detected. In the second step is an image preprocessing technique manipulating images in order to removing unwanted (undesired) noise and enhance an image quality from the image acquired. Hence, image preprocessing is used for improving the visual appearance of images to human viewer and preparing images for image segmentation the region of interest. The researcher considers different environment factors like lighting, camera resolution in order to minimizing the tasks in image preprocessing. The geometry of the viewing situation, i.e. the relative position of the sources and camera with respect to the objects of interest, usually also has a major impact on the contrast between the object and their background.

The third step is image segmentation techniques essential stage of image analysis methods that determine the quality final results. Segmentation methods used to separate an image into a set of homogenous and meaningful region, such that the pixels in each partition region possess an identical set of properties or attributes. The forth step, from the segmented images appropriate features were extracted, which usually resulted as binary image analysis. Computed both morphological and color features were identified by SVM and KNN classifiers. When compared, SVM is the best suitable algorithm was used to identify PTB bacilli from a given stained sputum smear image to provide PTB bacilli or PTB bacilli detect classes.

CHAPTER FOUR

4. EXPERIMENT RESULTS AND DISCUSSION

In image processing techniques, the first step is the acquisition of an image. During the image acquisition of the image the researcher considered different environment factors like lighting, camera resolution in order to minimizing the tasks in image preprocessing. The geometry of the viewing situation, i.e. the relative position of the sources and camera with respect to the bacilli objects of interest, usually also has a major impact on the contrast between the bacilli object and their background as described in section 2.5.1. Image processing is the critical step to reduce noise and enhance quality of image after acquired images in machine version system from stained sputum smear microscopic images through ZN staining process.

4.1. Experimental Results

Experimental result ensures the realization of the developed system architecture. It is an integral part of the development of PTB bacilli detection system. The experiment is carried out by using image processing techniques PTB bacilli by implementing algorithms in MATLAB platform.

In the previous section, computation of morphological and color features were described in details. In total, fourteen features (eight morphological and six color features) were identified. These features were used to classify different sputum smear images of PTB bacilli detected. In this, it was designed that the experimental scenarios to test the classification performance by taking the extracted features of sputum smear images. The classification was tested by using SVM and KNN classifiers algorithms in order to get a more accurate result. In order to do that, the train classifiers, a set of training sputum smear images was required, and the classes label

where it belongs to, 180 sputum smear images were taken from EPHI from the predefined two types of PTB bacilli includes positive and negative sputum smear images.

There were two basic phases of pattern classification. They were training and testing phases described in section 3.1.5.2. The researcher was used SVM and KNN classifier which is a well known algorithm to identify PTB in a given class, based on training data. The basic ideas is that the classifier takes a set of training content representing known example of class and by performing statistical analysis of the training content, using the knowledge from the training content to decide to which classes other unknown content belongs. In this study, was used the classifier to gain knowledge base (database) content based on the statistical analysis performed during training. Hence, it needs to design the classifier by partitioning the total dataset into training and testing dataset. From the total dataset of each sputum smear image type, 70% was used for training and to build classification model, and the remaining 30% of the total was used testing data. From the total of 180 dataset, 126 were used for training and 54 were used for testing. In general, a classifier has some input features based on the scenario of the designed experiment and some output features.

4.2. SVM classifier result

Support vector machine (SVMs) classifier model are fundamentally a binary (two-class) classification algorithm. It is widely used for classification tasks due to their extent generalization properties and their computational efficiency. SVMs are based on the concepts of decision planes that define decision boundaries. A decision plane is one that separates between a set of bacilli objects having different class memberships. In this study, the bacilli objects belong to either green or red. The separating line defines a boundary on the right side of which all

objects are green and to the left of which all objects are red. Any new object (white circle) falling to the right is labeled was classified, as green or classified as red should be it fall to be the separating line shown in figure 4.1. SVM classifier was used for this study on the classification of PTB bacilli detected. The experimentation was conducted under the following scenario based on the feature extracted. In all cases, it was used SVM classifier algorithm on the selected view of sputum smear images under sample of dataset. As a mentioned before, 70% of these dataset was used for training and 30% was used for testing purpose for the scenario.

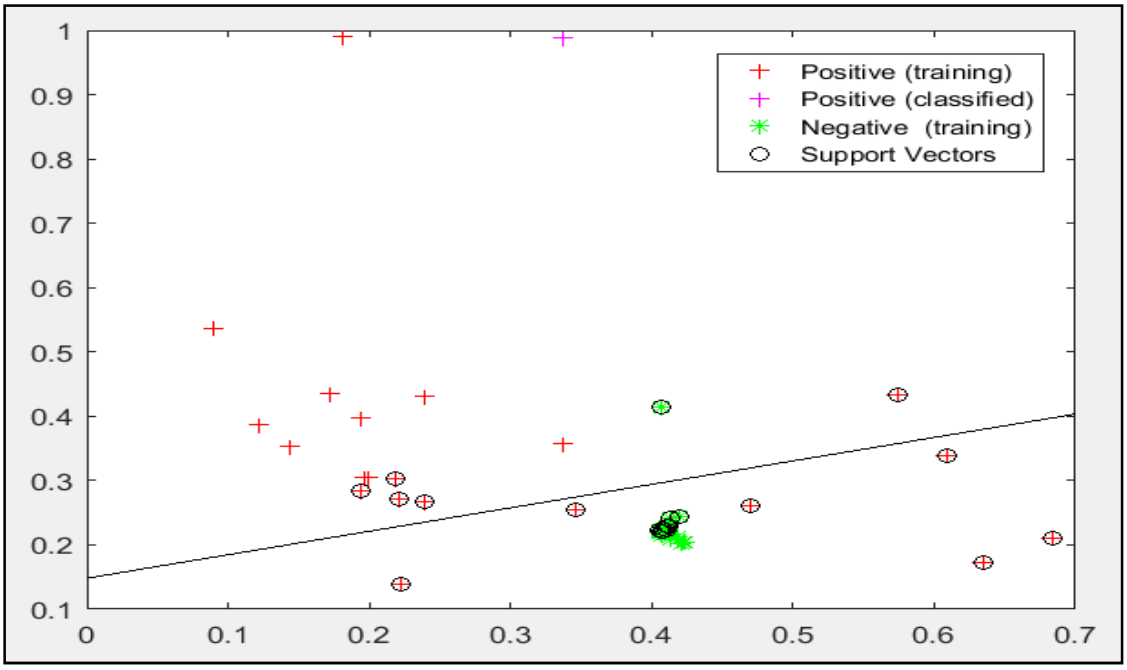


Figure 4 1: The classification result of SVM classifier model scenario of PTB bacilli positive and negative

4.3. KNN Classifier Result

KNN is a classification algorithm that classifies the given data by how close the data are related to each other. Distance calculation methods like Euclidian distance are used to find the cohesive the data are in the given dataset. KNN was used as a classifying model that has been the

capability to identify PTB bacilli and applied to the feature vector constructed to classify the PTB bacilli detected. The PTB bacilli object class is identified then the recognition/classification is completed. If the PTB bacilli object class is not known after KNN classification then KNN is applied to reduce the number of classes into two, then the distance matrix computed during the KNN is converted to a kernel matrix using the kernel trick. The experimentation was conducted by segregating the dataset into different number of training and test images based on the features extracted. For each images in training and testing, the preprocessing, feature extraction, feature reduction process and feature vector construction process. In this study, we it was used KNN classifier algorithm on the selected view of sputum smear images under sample of dataset which is mentioned before, 70% of these dataset was used for training and 30% was used for testing purpose for the scenario.

Comparing between SVM and KNN classifier algorithms Results

KNN is classified data based on the distance metrics whereas SVM classifier needs a proper phase of training. Due to the optimal nature of SVM, it is guaranteed that the separated data would be optimally separated. Generally, KNN was used as multi class classifiers whereas standard SVM separate binary data belonging to either of one class. SVMs look more computationally intensive, once training of data is done, that model can be used to predict classes even when it come across new unlabelled data. However, in KNN, the distance metric is calculated each time it come across a set of new unlabelled data. Hence, in KNN it always has to define the distance metric. SVMs have two major cases in which PTB bacilli classes might be linear separable or non-linear separable. When the PTB bacilli are non-linearly separable, it used kernel function such as Gaussian basis function or polynomial. SVM is more suitable algorithms for PTB classification other than KNN classifier.

4.4. System Performance Evaluation

As it was presented in details in the previous section, the experiments were conducted under scenario by using extracted features of the sputum smear images. The experimental results were used SVM and KNN classifier using holdout validation at 30% percentage held out were display results shows over the scenario and their performance summarized in table 4.1 and table 4.2.

The total number of dataset was 180 sputum smear images. In this study, there were two output classes, because the predefined sputum smear images of PTB bacilli were positive and negative. Classifying the test images into PTB bacilli negative or positive is required to evaluate the performance of the system assigning the image into category done by domain experts (pathologist) and the domain experts were selected from EPHI. As indicated in table 4.1, the results of both SVM and KNN classifier using both computed morphological and color features alone showed that from the tested dataset of 54 sputum smear images. There were fourteen features that is combined both (8 morphological features and 6 color features) including the dataset predefined by radiologist reading label of each sputum smear images as PTB bacilli positive (1) and negative (-1). Finally, classification performance of the prototype system computed based on table 3.4, using level set method with classifier on extracted feature of 54 (tested dataset) view of sputum smear images.

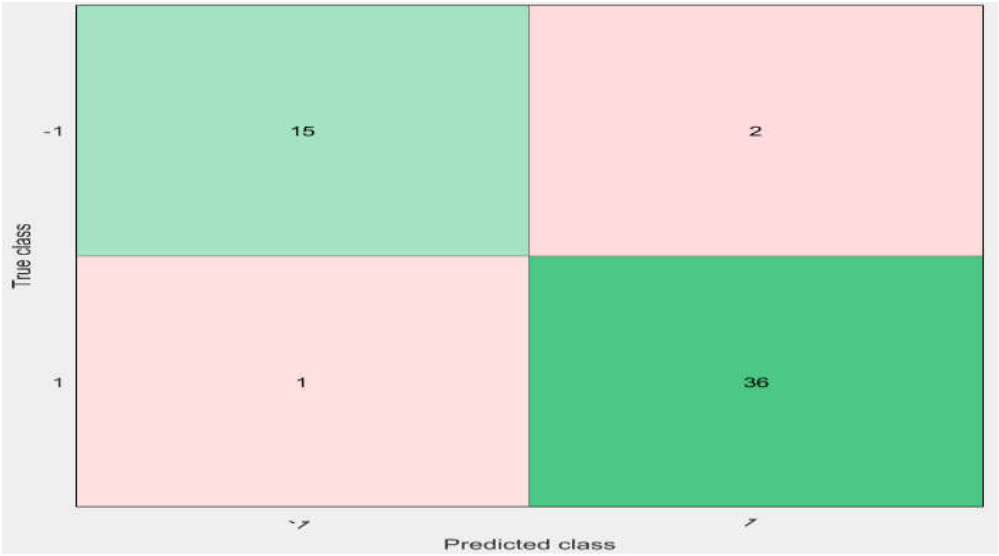
As described in section 2.2.4 most pathologists failed to identify PTB bacilli detected and misses less than 50 % due to over sight error or done manually process (Lumb et al, 2013). Therefore, in diagnosis of PTB detected, it is amenable that pathologist's skills have an important role in the results of accuracy of detecting the bacilli. In this regard, the developed system could make higher level of accuracy that depends on pathologist's skills and decision making. The researcher

develops automatic PTB bacilli detected was tested using sample of dataset selected from the ground truth and the sources mentioned previously. As described in section 3.1.5.2, the performance of the developed system can also measured using 30%, of the tested dataset for accuracy, sensitivity and specificity measures. To perform this, it was used a confusion matrix.

In this research, the system performance testing confusion matrix is a table often used to compute and describe the performance of a classification model in this study SVM classifier on a set of test data for which the truth values are known. There are two possible predicted groups' positive (1) and negative (-1). The classification made a total of 54 predictions, which means 54 patients with suspected PTB bacilli were tested for the presence of the PTB bacilli detected or not detected.

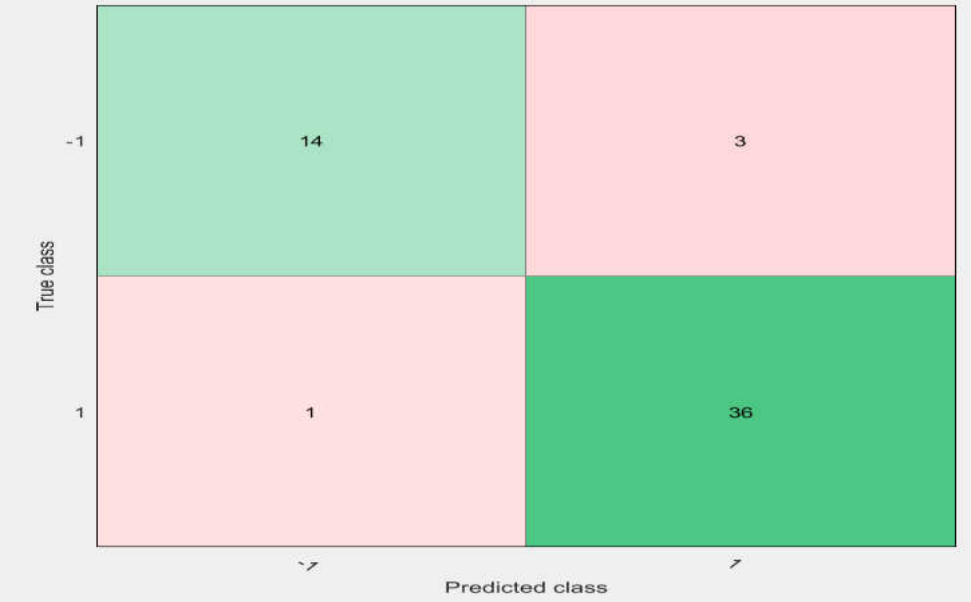
Confusion matrix has four categories: True Positive (TB), False Positive (PF), False Negative (FN) and True Negative. True positive is bacilli images that were identified by the domain expert as correctly identified and also classify results identify by the prototype system as correctly. False positive occur when incorrectly image data inputted into the system and the system is give result as classify result correctly. That means some incorrect images may be retrieved by the system as relevant. True negative is incorrectly images identify incorrectly by the prototype system and expert domain. This is the image when incorrectly detected were inputted and the proposed system classify negative decision i.e. PTB bacilli not detected. False negative is the images when incorrectly identify images were inserted in the system for testing and the prototype system classify positive result.

Table 4 1: Confusion matrix of prototype system of SVM classifier



As shown table 4.1, out of those 54 sputum smear images, the SVM classifier predicted as bacilli positive (1) was 36 and as bacilli negative (-1) was 15. But in reality, 37 bacilli images were as positive and 17 bacilli images were as negative based on the pathologists reading predefined. Based on the table 4.1, the researcher it can get the following results which help us to perform the PTB bacilli detection has been determine through accuracy, sensitivity, specificity and F-measure. True positives (TP) = 36, True negatives (TN) = 15, False positives (FP) = 2 and False negatives (FN) = 1. From the performance results was used by SVM algorithms, the overall detection accuracy was 94.4%, while sensitivity was 95%, specificity was 94% and F-measure was 96%.

Table 4 2: Confusion matrix of prototype system of KNN classifier



As shown table 4.2, out of those 54 sputum smear images, the KNN classifier predicted as bacilli positive (1) was 36 and as bacilli negative (-1) was 14. But in reality, 37 bacilli images were as positive and 17 bacilli images were as negative based on the pathologists reading predefined. Based on the table 4.2, the researcher it can get the following results which help us to perform the PTB bacilli detection has been determine through accuracy, sensitivity, specificity and F-measure. True positives (TP) = 36, True negatives (TN) = 14, False positives (FP) = 3 and False negatives (FN) = 1. From the performance results was used by KNN algorithms, the overall detection accuracy was 92.6%, while sensitivity was 93%, specificity was 92% and F-measure was 94.7%.

This results, indicates that using the integrated k-means clustering segmented methods on HSV color model and applied Gaussian filter and contrast enhanced to reduce noise and improve the quality images and feature extracted to identify PTB bacilli characteristics can be effective tools for detection system thinking that the presence of less quality images such as bacilli with debris

and unusual morphology is also under consideration. The accuracy performance measured for SVM algorithm found to be 94.4% and for KNN algorithm it was 92.6%. The results and observations show that SVM is a more suitable than KNN classifier. However, KNN is less computationally intensive than SVM classifier. Since, KNN is easy to implement, the classification of Multi-class data should be done with KNN. The algorithm that guarantees reliable detection in unpredictable situations depends upon the data. If the data points are heterogeneously distributed, both should work well. If data is homogenous to look at, one might be able to classify better by putting in a kernel into the SVM. For most practical problems, KNN is a bad choice because it scales badly it would take a long time (linear to the number of examples) to find K nearest neighbors.

4.5. Discussion

In detection of PTB bacilli, it is amenable that pathologists' skills have an important role in the results of accuracy detecting and computing bacilli. In this study, the proposed system could elevate the level of accuracy that depends on pathologists' skills. For the experiment the researcher was collected image data from EPHI by using Leica Microsystems microscopy connected to PC. However, some of these images were not sufficient in number and variety. Hence, it was necessary to collected actual data that is stained sputum smear images infected by the PTB bacilli from national tuberculosis reference laboratory research case team at Ethiopian Public Health Institute. The proposed novel algorithm to detect PTB bacilli disease and computed feature extracted was tested using sample data selected from the ground truth and sources of data mentioned previously. The researcher applied statistical approach measurement to test the performance of the proposed system. The performance of PTB bacilli detection has been determined through accuracy, sensitivity, specificity and F-measures. From the results, the

overall detection was used SVM classifier more suitable than KNN classifier, which represent accuracy was 94.4%, while sensitivity was 95%, specificity was 94% and F-measure was 96%.

As per the researcher knowledge, there is no local research attempts made to use sputum smear images by using image processing techniques for detection of PTB bacilli, but there are different researchers that used image processing techniques for identify PTB bacilli positive or negative. By considering the above performance results of automatic PTB detection system from sputum smear images using image processing techniques, it is important to compare with previous studies done by Rao Osama in 2016 the same area as presented in table 4.3.

Rao Osama (2016) used image processing approach which is applied to tuberculosis bacilli identification in sputum smear image conducted by tuberculosis reference Laboratory (TRL) at national laboratory of public health in Khartoum. The main objective is to enhance, segment and classify the sputum smear images for computerized process of tuberculosis identification. The results obtained lead to conclusion that the system can forecast with considerable sensitivity (83.07%) the decision of PTB bacilli identification based on hu moments and morphological features used.

Table 4 3: Comparison of the developed system with the previous studies

Author	Microscopy	Preprocessing techniques	Segmentation techniques	Feature extraction techniques	Classifier	Performance measurements and results in (%)			
						Accuracy	Sensitivity	specificity	F-measure
This study	Leica Microsystems microscopy (DM LS2) connected to PC	Convert RGB to HSV (used V image value), Gaussian filter and contrast enhanced	k-mean clustering + boundary region line with color	8 morphological features and six color features	SVM	94.4%	95%	94%	96%
					KNN	92.6%	93%	92%	94.7%
Rao Osama , 2016	ZEISS iLED microscope and NIKON D3100	De-correlation and Stretching Gaussian filter Contrast stretching	L*A*B color space + K-mean clustering	4 morphological features and hu moments from 1 to 7 feature	SVM	81%	83.07 %	66.66%	Not specified

As shown in table 4.6 above, Rao Osama (2016) achieved a higher interesting performance in sensitivity(83.07%) in comparing with this study. The result difference could be due to the increment in morphological features and color features that fully express the real working environment and the current work values difference. Since a more number of combined both morphological features and color features can increase the system performance. The current work conducted F-measure performance measurement registered 96% respectively which are not specified by the former work.

In Rao Osama (2016), work the system uses two morphological features like eccentricity and compactness for which makes its own contribution for a better accuracy than the current work which is 83.07%. Since their work is not only considers only two morphological features it can't be a representative of features for other features like for this study. As a result the current

researches filled this gab by analyzing the existing situation based applied both eight morphological and six color features combined together. In general, the overall result showed that were used combined both morphological feature and color features have more discriminating (tasteful) power than were used the previous study (Rao Osama, 2016). The discrimination power increases when combined both morphology and color features were used together.

In addition, through of the related works, Divekar et al. (2012) and Chang et al. (2008), claim that they could achieve better value of current work of the best algorithm was used for classification performance is support vector machine based on accuracy, sensitivity, specificity and F-measure improved the performance of the prototype compare with KNN classifier. The F-measure performance results indicated that the accuracy of prototype system has a high level value performance and acceptable. SVM algorithm is one of the advanced statistical methodologies and the best suitable algorithm than KNN algorithm to identify PTB bacilli. For this study, support vector machine (SVM) was the best algorithms was used to identify PTB bacilli in sputum smear image, which classified the computed morphological and color features from sputum smear image in two class like PTB bacilli detected and non-PTB bacilli detected Moreover, the mentioned works used limited number of sample images, whereas this dealt with various sample from EPHI. However, challenges are still reflected while computing to identify bacilli characteristics in the cases where there existed high degree of the quality of the camera, the image acquisition environment and other imaging factor may affect on the results and object it also overlapping. Another problems experience related to identify PTB bacilli characteristics calculation included debris or undesired object overlapping with unusual morphology. Automatic PTB bacilli detection is the technique, method, or system of operating or controlling a process of PTB bacilli identification process by highly automatic means, as by electronic

devices, reducing human intervention to a minimum. Automated TB bacilli diagnostic systems bring about several benefits in the fight against the PTB bacilli disease. Some of the advantages of computer-based diagnosis are measurements and features make a diagnosis, and it helps radiologists on their diagnosis procedure for accuracy and efficiencies. Because computer-assisted diagnosis can be faster than manual screening, the number of patients tested in any given amount of time is increased. Additionally, human involvement in the manual screening process is reduced. This was the result in three desirable effects; less experienced staff can conduct the screening, labor costs are reduced and technicians suffer less fatigue, since the need to view the slides on the microscope is eliminated or reduced. And, also greater accuracy in the diagnosis may be improved and could possibly reduce the risk of infection to laboratory staff.

CHAPTER FIVE

5. CONCLUSION AND RECOMMENDATION

5.1. Conclusion

This study, addresses how automatically identification of PTB bacilli disease is possible by using image processing techniques by effectively analysis various features of bacilli image characteristics by using computing both morphological and color features worked together. The developed automated PTB bacilli detection system can be used with low-cost for developing countries. The system was developed by using image processing techniques to images captured by conventional microscopy could save live in low resources communities burdened by PTB bacilli detected. The total sample of dataset 180 PTB bacilli (100 positive and 80 negative) was collected from sputum smear microscopy through ZN stain process using Leica Microsystems microscopy connected to computer from EPHI. The collected dataset was used for data training and evaluation.

The accurately detection and classification of the PTB bacilli detection is very important for the successful detection of PTB bacilli can this can be done by using processing techniques. The main aim of this study is to develop automatic PTB bacilli detection from microscopic sputum smear images using image processing techniques. In this study, an algorithm based on image processing technique is selected for identification of pulmonary tuberculosis bacilli in digital image of stained sputum smear. Applied Gaussian filter and contrast enhanced to remove noised and to improve the quality of image was used for image segmented. A K-means clustering algorithm was used for segmentation the characteristics of PTB bacilli detected which facilitate

for feature extracted. Fourteen features (eight morphological and six color features) were used to classify sputum smear images of PTB bacilli detected.

In this study, the experimental results indicate that SVM classifier is the best algorithm was used for identify and classify PTB bacilli detect compared with KNN classifier based on computed both morphological and color features. The tested data the results was displayed in GUI to indicate the PTB bacilli positive or negative classes. The classification PTB bacilli detection measuring the performance of the proposed system results are found. The accuracy, sensitivity, specificity and F-measures improved the performance of the prototype results of SVM were 94.4%, 95%, 94% and 96% respectively, whereas KNN were 92.6%, 93%, 92 and 94.7%. SVM classifier is indicated that a better results performance of prototype system for PTB bacilli detection. A graphical user interface (GUI) has been developed to read the image and display the results the sputum smear image is having PTB bacilli detected or not detected.

Generally, in this study the results show that by using SVM classifier is more reliable algorithms than was used to identify the PTB bacilli detection procedures has been high results. It is result shown and observed SVM has a higher accuracy than KNN algorithms classifier. The developed automatic system shows good accuracy and efficiency, and thus can be used to assist pathologists in making decision at early stage of diagnosis PTB bacilli. The problems overlapping to identify PTB bacilli objective in feature extraction represent compute validation is difficult.

5.2. Recommendation

The main objective of this research was to develop automated detection system for pulmonary tuberculosis bacilli identifying from sputum smear microscopy image using image processing techniques. Developing system using image processing techniques needs to consider the performance of the techniques used regarding to speed. Therefore, there are a number of problems to be investigated by future researchers in applying image processing techniques.

- Applying stained sputum smear images obtained by using digital microscopy as input image to acquired quality images for the systems
- Applying SVM classifier algorithm can be also used to design various mobile phone applications tools that helps in diagnosis on decision making
- Apply image processing technique PTB Drug Resistant, because it is time consuming, which the diagnosis of tuberculosis and resistant cases needs the development of much faster, efficient, patient friendly and cost effective techniques with more precision
- Apply to extract the features to design and compare an SVM, KNN and ANN for further investigation.
- The future work will look into the problem of overlapping objective in PTB bacilli feature extraction represent validation of the proposed algorithm

References

Ankitha, B M, Sanjeev, K. and Bharathi, S H. (2015). Computer Aided Diagnosis of Tuberculosis using First order Statistical approach. *IPASJ International Journal of Electronics & Communication (IIJEC),* Volume 3, Issue 4.

American Thoracic Society (ATS). (2000). Diagnostic Standards and Classification of TB in Adults and Children. *Am J Respir Crit Care Med.* 2000; vol. 161:1376 –1395

Ayas, S., and Ekinci, M. (2014). Random forest-based tuberculosis bacteria classification in images of ZN-stained sputum smear samples. *SIViP,* 8(1), 49–61

Centers for Disease Control and Prevention (CDC). (2014). Reported tuberculosis in the United States, 2013. Atlanta: US Department of Health and Human Services.

Chang, I.C., H wang, H.G., Hung, W. F., and Li, Y.C. (2007).Physicians' acceptance of pharmacokinetics-based clinical decision support systems. *Expert Syst. Appl*, 33(2), 296-303

Chitradevi, B. and Srimathi, P. (2014). An Overview on Image Processing Techniques. International *Journal of Innovative Research in Computer and Communication Engineering, 2 (11).*

Choudhary, D. Singh, A. K. Tiwari, S. Shukla, V. P. (2013). Performance Analysis of Texture Image Classification Using Wavelet Feature. *International Journal of Image, Graphics and Signal Processing (IJIGSP),*5 (1), 58-63.

Dendere, R. (2009). Segmentation of candidate bacillus objects in images of Ziehl-Neelsen stained sputum smears using deformable models, University of Cape Town February 2009.

Deserno, T.M. (2011): Fundamentals of Biomedical Image Processing. *Springer-Verlag Berlin Heidelberg* . DOI: 10.1007/978-3-642-15816-2 1,

Deun, V. A, Salim, A. H, Cooreman E, Hossain M. A, and Rema, A. (2002). Optimal Tuberculosis case detection by direct sputum smear microscopy: how much better is more?". Int J. Tuber Lung Dis 6, P 222-230

Dipanwita, B., Sagar B., Neelam P. & Nirmala S. (2011). Disease Diagnosis System

Divekar, A., Corina P., Gerrit C., Tarlochan, S., Fleming Y. M. L., and Sean K. (2012). Automated detection of tuberculosis on sputum smeared slides using stepwise classification, Proceeding SPIE Medical Imaging Conference, pp.1-9.

Ethiopian Health & Nutrition Research Institute (EHNRI).(2009). Laboratory Network Guidelines for Quality Assurance of Smear Microscopy for Tuberculosis Diagnosis. Addis Ababa, Ethiopia.

Federal Democratic Republic of Ethiopia Ministry of Health (FEMOH). (2013). Guidelines for clinical and programmatic management of TB, TB/HIV and Leprosy. Addis Ababa: FEMOH.; Fifth edition.

Forero, G. Cristobal, and M. Desco (2006). Automatic identification of mycobacterium tuberculosis by Gaussian mixture models, *J. Microsc.*, 223(2), 120–132.

Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M. (2006). Variable host-pathogen compatibility in *M. tuberculosis. Proc Natl Acad Sci USA,* 103, 2869–73.

Gao, Shirui. (2013). Research on Medical Image Processing Method Based on the Matlab. *Informatics and Management Science V, Lecture Notes in Electrical Engineering 208*

Getahun, H, Hurrington M, O'Brien R, Nunn P. (2007). Diagnosis of smear-negative pulmonary tuberculosis with HIV infection or AIDS in resource-constrained setting: informing urgent policy changes. *Lancet.,* 369 (9578), 2042-9

Goel, R., Kumar, V., Srivastava, S. and Sinha, A. K. (2017). A Review of Feature Extraction Techniques for Image Analysis. *International Journal of Advanced Research in Computer and Communication Engineering,* 6(2).

Gonzalez, R. and Woods, R. (1992). Digital Image Processing, Addison-Wesley Publishing Company

Goodwin, C. J. (2010). *Research in Psychology: Methods and Design.* USA, Western Carolina University: John Wiley & Sons, Inc. 6th ed.

Harries AD, Maher D, Nunn P. (1998). An approach to the problems of diagnosing and treating adult smear-negative pulmonary tuberculosis in high-HIV-prevalence settings in sub-Saharan Africa. Bull World Health Organ.,2(976), 651–62

Houcque, D. (2005). Introduction to Matlab for Engineering Students. McCormick School of Engineering and Applied Science (Northwestern University), Evanston, Illinois.

International Standards for Tuberculosis Care (ISTBC)(2013). Available from: http://www.who.int/tb/publications/2006/istc_report.pdf/

John C. Russ (1998)*: The Image Processing Handbook*, CRC Press, 3rd Edition.

Linda, M.P, Brosch, R., Sterwart, T., Cole, Somoskovi, A., Loder, A. (2002). Rapid and Simple Approach for identification of *M. tuberculosis* complex isolates by PCR-Based Genomic Deletion Analysis. *J. of Clin Microbiol, 1* (44): 2339 – 2345.

Luelmo F. (2004). What is the role of sputum microscopy in patients attending health facilities?. 2nd ed. Geneva: World Health Organization; pp. 7–13.

Lumb, R., Deun, A. V., Bastian, I. and Fitz-Gerald, M. (2013). The *handbooks: Laboratory diagnoses of Tuberculosis by using sputum smear microscopy*, SA Pathology from Road Adelaide South Australia, ISBN: 978-1-74243-602-9.

Mack, N., Kathleen, C., Macqueen, M., Guest, G. and Namey, E. (2005). Qualitative research methods: A data collector field guide. 1st ed. USA: Family Health International

Mathworks.com, 2018. Available: http//www.mathworks.com[Accessed: 1-May-2018]

Medjahed, S. A. (2015). A Comparative Study of Feature Extraction Methods in Images Classification. *I.J. Image, Graphics and Signal Processing*, 3, 16-23

Menzies D. & Khan, K. (2007). Diagnosis of tuberculosis infection and disease. Canadian Tuberculosis Standards. 6th ed. Ottowa: Public Health Agency of Canada, 53–91.

Mitruka K, Oeltmann JE, Ijaz K, & Haddad, M. B. (2011). Tuberculosis outbreak investigations in the United States, 2002–2008. Emerg Infect Dis. 1(17), 425–31.

Mohajan, H. K. (2015). Tuberculosis is a Fatal Disease among Some Developing Countries of the World. *American Journal of Infectious Diseases and Microbiology*, 3(1), 18-31.

Oeltmann, J. E., Click, E. S., & Moonan, P.K. (2014). Using tuberculosis patient characteristics to predict future cases with matching genotype results. Public Health Action. 1(4), 47–52.

Osman , M.K, Mashor, M.Y., and Jaafar, H. (2012). Combining Thresholding and Clustering Techniques for Mycobacterium tuberculosis Segmentation in Tissue Sections", Australian Journal of Basic and Applied Sciences, 5(12), 1270-1279..

Panicker, R.O., Sman, B., Saini, G. and Rajan, J. (2016). A Review of Automatic Methods Based on Image Processing Techniques for Tuberculosis Detection from Microscopic Sputum Smear Images. J Med Syst (2016), 1(40), 17.

Pawar, Chetan C. & Ganorkar, S. R. (2016). Tuberculosis screening using digital image processing techniques. *International Research Journal of Engineering and Technology (IRJET),3(*7).

Rachna H.B. and Swamy, M.S. M.. (2013). Detection of Tuberculosis Bacilli using Image Processing Techniques, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, 3(4).

Roa Osama A. A., (2016). Automatic Method for Tuberculosis Bacilli Identification in Sputum Smear Microscopic Images Using Image Processing Techniques. Sudan University of Science and Technology

Rawat J, Biswas D, Sindhwani G, Kesharwani V, Masih V, Chauhan. (2012). Diagnostic defaulters: an overlooked aspect in the Indian Revised National Tuberculosis Control Program. J Infect Dev Countries.

Raza, S. QaisarMarjan, M., Muhammad A., Farhana B. , Faisal S., Nasir M. R. (2015) Anisotropic Tubular Filtering for Automatic Detection of Acid-Fast Bacilli in Digitized Microscopic Images of Ziehl-Neelsen Stained Sputum Smear Samples

Sadaphal P, Rao J, Comstock GW, Beg MF. (2008). Image Processing Techniques for Identifying *Mycobacterium Tuberculosis* in Ziehl-Neelsen Stains, *International Journal of Tuberculosis and Lung Disease*, 579-582.

Santiago-Mozos, R., Pérez-Cruz, F., Madden, M.G., and ArtésRodríguez, A. (2014). An Automated Screening System for Tuberculosis. IEEE journal of Biomedical and Health informatics. 18(3):855–862. doi:10.1109/JBHI.2013.2282874.

Singh, N. et al. (2015). Diagnosis of pulmonary and extrapulmonary tuberculosis based on detection of mycobacterium antigen 85B by immune-PCR. *Diagnostic Microbiology and Infectious Disease, New Delhi (BT/PR4329/INF/22/144/2011),* 359–364.

Swaminathan, S., Padmapriyadarsini, C. & Narendran, G. (2010). HIV-Associated Tuberculosis: Clinical Update. *Clin Infect Dis.*50, 1377–1386.

Tapley, A. (2012). *Mobile Digital Imaging-based Microscopy for Tuberculosis Diagnosis, M.S. Thesis*. University of California, Berkeley.

Tinku Acharya and Ajoy K. Ray, *Image Processing Principles and Applications*, Jhon Wiley, 2005.

Thirumaran, J and Shylaja, S. (2015). Medical Image Processing: An Introduction. *International Journal of Science and Research (IJSR),*4(11), 2319-7064

VanRie, A., Westreich, D., & Sanne, I. (2011). Tuberculosis in patients receiving antiretroviral treatment: incidence, risk factors, and prevention strategies. *J Acquir Immune Defic Syndr*. 56(4), 349-55.

Veropoulos K, Learmonth G, Campbell C, Knight B, Simpson J. (1999). Automated Identification of Tubercle Bacilli in Sputum: A Preliminary Study. *Analytical and Quantitative Cytology and Histology*; 21(4): 277-281.

WHO. (2016). World Health Organization Global Tuberculosis Control Report 2016, Geneva. Retrieved from www.who.int/tb/publications/global_report/en/

WHO. (2017). World Health Organization Global tuberculosis report 2017, Geneva. Retrieved from http://www.who.int/tb/publications/global_report/en/

William K. Pratt: *Digital image processing, PIKS Scientific inside*, John Wiley, 4th Edition , 2007

Appendix A

Matlab code

```
lc;
close all;
clear all;

[filename, pathname] = uigetfile({'*.*';'*.bmp';'*.jpg';'*.gif'}, 'Pick a PTB Image File');
I = imread([pathname,filename]);
I = imresize(I,[300 400]);
imshow(I);
I2 = rgb2hsv(I);
figure(2);
imshow(I2);title('HSV');
I3 = im2bw(I2);
figure(3);
imshow(I3);
D = imgaussfilt(I2,2);
subplot(1,2,2)
figure(4);
imshow(D); title('Gaussian filtered image')
E = imadjust(D,stretchlim(D),[]);
figure(5);
imshow(E); title('Contrast enhanced');
%fill any holes
fill = imfill(E,'holes');
figure(6);
imshow(fill);
%binary images
A = rgb2gray(I);
figure(7);
imshow(A);
BI = im2bw(A);
figure(8);
imshow(BI);
%remove small objects
BI = bwareaopen(BI,60);
T = im2bw(I2,graythresh(I2));
figure(9);
imshow(T);

imData = reshape(A,[],1); %image convert to gray
imData = double(imData); % image convert from unit8 to double
[IDX nn] = kmeans(imData,4);
imIDX = reshape(IDX,size(A));
figure(11);
imshow(imIDX,[]), title('indexed Image');
figure(12);
subplot(3,2,1);
imshow(imIDX==1,[]); title('cluster 1');
subplot(3,2,2);
imshow(imIDX==2,[]); title('cluster 2');
subplot(3,2,3);
imshow(imIDX==3,[]);  title('cluster 3');
```

```
subplot(3,2,4);
imshow(imIDX==4,[]); title('cluster 4');
set(gcf, 'Position', get(0,'Screensize'));
x = inputdlg('Enter the cluster no. containing the ROI only:');
i= str2double(x);
figure(13);
imshow(imIDX==i); title('segemented image by using kmeans');
%show object boundaries
[B,L] = bwboundaries(imIDX==i,'noholes');
figure(14);
subplot(1,1,1);
imshow(imIDX==i,[]);
hold on
for k = 1:length(B)
  plot(B{k}(:,2), B{k}(:,1), 'g','LineWidth',1.45)
end
title('\fontsize{18}\color[rgb]{0.635 0.078 0.184} Detected PTB');
hold off
load datatrain.mat
xdata = meas(1:end,7:8);
group = class(1:end,1);
P = 0.3;
[Train, Test] = crossvalind('HoldOut',group,P);

TrainingSample = xdata(Train,:);
TrainingLabel = group(Train,1);
TestingSample = xdata(Test,:);
TestingLabel = group(Test,1);
figure;
hold on
scatter(TrainingSample(TrainingLabel==1,1),TrainingSample(TrainingLabel==1,2),'+g');
scatter(TrainingLabel(TestingLabel==-1,1),TrainingLabel(TestingLabel==-1,2),'.r');
xlabel('{x_1}')
ylabel('{x_2}')
legend('Positive Class','Negative Class')
title('Data for classification')
hold off
%disp(length(data))
%%
% * *Z-score Normalization*

svmStruct = svmtrain(TrainingSample,TrainingLabel,'showplot',true,...
    'kernel_function','rbf','rbf_sigma',0.1); %#ok<SVMTRAIN>
outLabel = svmclassify(svmStruct,TestingSample,'showplot',true);
sum(grp2idx(outLabel)==grp2idx(TestingLabel))./sum(Test);
```

Appendix B

Approval of Letter Pathologists

Date May 9, 2018
Ref. No RPG/182/2018

Research and Ethical Review Board (RERB/IRB) of College of Natural Sciences, Jimma University, Jimma

Ethical Clearance

Research Title: Automatic pulmonary tuberculosis detection from sputum smear microscopy image using image processing techniques
Study areas: Jimma University Specialized Hospital and Ethiopian Public Health Institute
Principal Investigator: Diriba Abdeta
Supervisor(s): Mr. Worku Jimma and Mr. Chala Diriba
Type of data collected:

- Images of patients' sputum from domain experts (treating patients) **to be collected by a personnel assigned by the concerned institute**
- Interview of domain experts

Ethical consideration: Confidential treatment of data

The RERB/IRB of College of Natural Sciences, Jimma University, has reviewed the **ethical considerations** of the above mentioned research proposal and made the following decision:

- ☒ Approved
- ☐ Approved with recommendation
- ☐ Approved on condition
- ☐ Disapproved

The decision is valid for 12 (twelve) months and the research should be conducted in compliance with the protocol/proposal approved by the RERB/IRB of College of Natural Sciences, Jimma University. Any subsequent revision/amendment of the protocol/proposal needs prior RERB/IRB approval before conduct of the research. Up on completion of the study, manuscripts or thesis work to the final/completed research project needs to be submitted to the RERB/IRB.

Date Approved by the RERB/IRB: **08/05/2018** in Minute No. **CNS/RERB/004/2018**

RERB/IRB Chair: Mulugeta Wakjira(PHD), Research and Graduate Studies Coordinator

Signature

Jimma University College of Natural Sciences

CC.

- College Dean Office
- College V/Dean Office
  Jimma University