

Pulmonary tuberculosis bacilli detection in sputum smear microscopy images using image processing techniques

Diriba Abdeta¹, Chala Diriba², Worku Jimma^{2*}

Abdeta D, Diriba C, Jimma W. Pulmonary tuberculosis bacilli detection in sputum smear microscopy images using image processing techniques. *AGBIR*.2024;40(1):887-898.

Tuberculosis remains a global health threat, particularly in developing countries like Ethiopia, where *Mycobacterium tuberculosis* causes a significant impact, primarily affecting the lungs in the form of pulmonary tuberculosis disease. Sputum smear microscopy stands as the predominant diagnostic tool in such settings. This study aims to develop a K-Nearest Neighbor classifier model for the detection of pulmonary tuberculosis bacilli in microscopic sputum smear images. The study employed image processing techniques to identify pulmonary tuberculosis bacilli in digital images of

stained sputum smears. K-Nearest Neighbor classifiers distinguish between two classes: Bacilli detection and non-bacilli detection. The image dataset, comprising 180 stained sputum images of pulmonary tuberculosis bacilli infections, was sourced from the Ethiopian Public Health Institute. The model's performance metrics, including accuracy, sensitivity, specificity and F-measure, demonstrate an impressive average accuracy of 92.6%. The developed model exhibits a sensitivity of 93%, specificity of 92% and an F-measure of 94.7%, highlighting its robust performance in pulmonary tuberculosis bacilli detection.

Key Words: Pulmonary tuberculosis, Sputum smear microscopy; Machine learning; K-nearest neighbor; Model

INTRODUCTION

Pulmonary Tuberculosis (PTB) bacillus, a highly contagious and infectious form of tuberculosis, primarily targets the lungs. Transmission occurs through the release of airborne droplets when individuals with active PTB cough, sneeze, or laugh (Heyd, 2020). Collaborative efforts between patients and healthcare professionals, including physicians, nurses, pharmacists and lab technicians, are crucial for its effective treatment [1]. While antibiotics offer a method of treatment, identifying PTB bacilli poses a significant challenge for medical professionals.

The prevalent global method for diagnosing active PTB bacilli disease is sputum smear microscopy, a technique that detects bacteria in sputum smear images [2]. Manual microscopic examination of Ziehl-Neelsen (ZN) stained

sputum smears is the common approach, but it is time-consuming and prone to errors [3]. Lab technicians rely on this method, examining stained smear images under microscopy to identify viable or latent mycobacteria as red-coloured, rod-shaped objects.

In Ethiopia, the primary technique for diagnosing PTB bacilli is sputum smear image microscopy, specifically using Ziehl-Neelsen (ZN) or Acid-Fast Bacilli (AFB) stained. The whole process of diagnosis of Tuberculosis (TB) cases in Ethiopia, a laboratory test national standard, is depicted in Figure 1, which takes long time. However, AFB-stained processes are limited by manual diagnosis, inability to distinguish between viable and dead organisms, low sensitivity, high bacterial load requirements, and limited specificity. Sputum smear microscopy, though widely used, is time-consuming, tedious, subject to poor specificity and demands highly trained personnel [4].

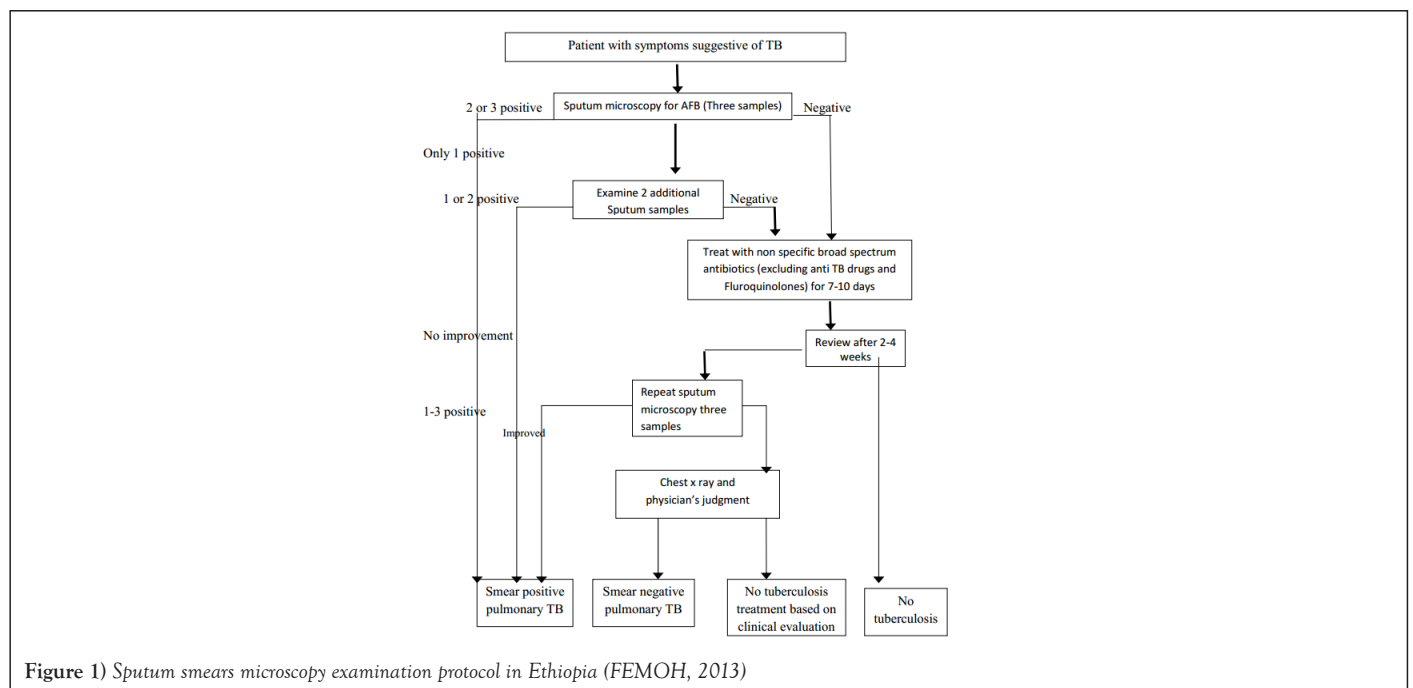


Figure 1) Sputum smears microscopy examination protocol in Ethiopia (FEMOH, 2013)

¹Department of Computer Science, Ambo University, Oromia, Ethiopia; ²Department of Information Science, Jimma Institute of Technology, Jimma University, Oromia, Ethiopia

Correspondence: Worku Jimma, Department of Information Science, Jimma Institute of Technology, Jimma University, Oromia, Ethiopia, E-mail: worku.jimma@ju.edu.et

Received: 26-Dec-2023, Manuscript No. AGBIR-23-123888; **Editor assigned:** 28-Dec-2023, Pre QC No. AGBIR-23-123888 (PQ); **Reviewed:** 11-Jan-2023, QC No. AGBIR-23-123888; **Revised:** 18-Jan-2024, Manuscript No. AGBIR-23-123888 (R); **Published:** 25-Jan-2024, DOI:10.35248/0970-1907.24.40.887-898



This open-access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits reuse, distribution and reproduction of the article, provided that the original work is properly cited and the reuse is restricted to noncommercial purposes. For commercial reuse, contact reprints@pulsus.com

Medical imaging, a crucial process for visualizing internal organs and tissues, has become indispensable in clinical analysis and intervention [5]. Advanced image processing and analysis techniques enhance the quality of medical imaging, providing quantitative data for patient care [6]. Recognizing the limitations of manual PTB bacilli detection, it is paramount to develop an automatic detection system using image processing techniques, addressing issues such as clinician workload, sensitivity, specificity and time consumption [4]. Such a system can assist physicians and pathologists, offering early PTB bacilli diagnosis, saving time, increasing accuracy and enhancing sensitivity. Hence, the study aims to develop a PTB bacilli detection model from sputum smear microscopy images using the K-Nearest Neighbor Classifier.

MATERIALS AND METHODS

Different approaches and tools were employed to develop an automatic diagnosis system for PTB disease-causing bacilli. Experimental research design was employed to develop and test a prototype model using experimental methods. The dataset, comprising images of previously diagnosed PTB bacilli cases, was obtained from the Ethiopian Public Health Institute (EPHI). This image dataset was crucial for prototype system testing. EPHI was chosen for

its expertise, experienced domain experts and the availability of the dataset.

The target populations included domain experts from the EPHI National TB Reference Laboratory staff. Purposive sampling was employed to select domain experts based on their professions, educational qualifications, and years of experience in PTB bacilli diagnosis. The total sample size for the image dataset was 180 (100 positives and 80 negatives) from stained sputum images of PTB bacilli-infected patients. Knowledge acquisition process involved interviews and document analysis. The architecture of the developed system based on PTB detection is depicted in Figure 2.

Sputum smear images were collected from stained sputum smear specimens of PTB patients at EPHI. The Leica Microsystems microscopy, connected to a computer, facilitated image acquisition at 100x magnifications. The images were saved in Joint Photograph Experts Group (JPEG) file format, with 24 bits per pixel, in Red, Green and Blue (RGB) colour space. Both primary and secondary data sources were used, with primary data collected through interviews with domain experts and secondary data obtained from published articles, journals and TB program reports. Furthermore, Matrix Laboratory (MATLAB) software was used to implement the proposed prototype and the performance evaluation employed metrics such as accuracy rate, sensitivity, specificity and F-measure.

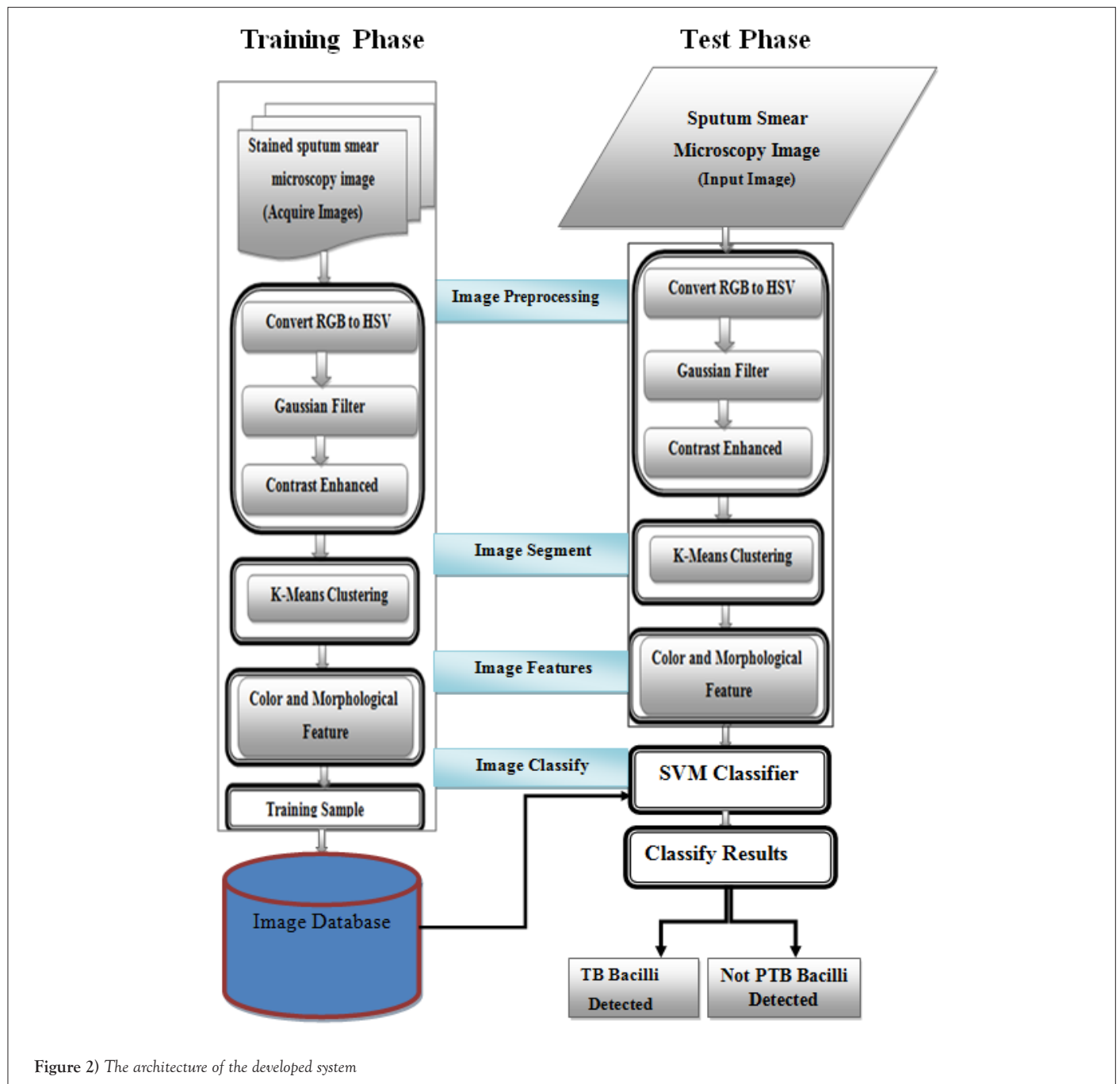


Figure 2) The architecture of the developed system

Design and implementation of automatic pulmonary tuberculosis bacilli detection system

The automatic detection and classification of objects into their respective diagnostic classes constitute a focal point in this study. The design and implementation phase involve the development of procedures tailored for an automatic PTB bacilli detection system, intending to aid pathologists in decision-making. This necessitates a comprehensive understanding of the entire process, encompassing image acquisition through digital image processing and domain expertise contributed by experts.

The subsequent task involves coding the image analysis and classification processes. For this study, MATLAB R2016a (9.0.0.341360) version of image processing tools framework was employed to construct the prototype, executable on a personal computer. The PTB bacilli detection process comprises distinct procedures categorized into predefined steps based on feature extraction.

Designing the system architecture of automatic pulmonary tuberculosis bacilli detection system

The Automatic Pulmonary Tuberculosis Bacilli Detection (APTBBDD) system encapsulates the overarching framework for any vision algorithm dedicated to PTB bacilli identification in sputum smear microscopy images, aiding pathologists in decision-making. The developed system comprises five integral steps, namely image acquisition, image preprocessing, segmentation, feature extraction and classification.

In the initial step, digital images are acquired from a slide using Leica Microsystems microscopy connected to a Personal Computer. Subsequently, image preprocessing techniques are applied to remove noise and normalize the images for consistent colour characteristics. The third step involves image segmentation, a critical preprocessing step to extract informative features for human understanding and interpretation. Following segmentation, various discriminating techniques are applied to classify images, determining the presence or absence of PTB bacilli.

The architecture of the Automated Pulmonary Tuberculosis Bacilli Detection (APTBD) system, illustrated in Figure 2, delineates how the system operates during PTB bacilli detection. The process initiates with the acquisition of sputum smear microscopy images using Leica Microsystems microscopy. To mitigate noise introduced during image acquisition, image processing techniques such as filter algorithms and contrast enhancement are applied. Subsequently, features best suited to represent the image are extracted using image analysis techniques. Based on these extracted features, training and testing data are employed for identification purposes. This systematic approach ensures the efficient operation of the APTBD system, facilitating accurate and reliable PTB bacilli detection.

Image acquisition

A total of 180 stained smear images, comprising both negative and positive cases of Pulmonary Tuberculosis (PTB), were obtained from slides prepared with specimens from PTB patients. These specimens were collected using Ziehl-Neelsen (ZN) stained sputum smear microscopy, a conventional diagnostic technique employed at the National Tuberculosis Reference Laboratory of the Ethiopian Public Health Institute. The images were captured using Leica Microsystems microscopy, connected to a PC, as depicted in Figure 3. The pixel resolution was 696×514 and the images were stored in Joint Photograph Experts Group (JPEG) format, with 24 bits per pixel, in colour. Some ZN stained sputum smear images with poor resolution were excluded from the dataset. Each sputum smear image covered a $1024 \times 756 \mu\text{m}$ field of view at the smear.

Image preprocessing

Following the image acquisition using Leica Microsystems microscopy, preprocessing became essential. Image preprocessing operates at the lowest level of abstraction, working with intensity images represented by a matrix of function values. The colour image segmentation began with the conversion of the original RGB image to the HSV colour space model. Given variations in colours resulting from different staining procedures for PTB detection, the HSV colour space was employed for colour image segmentation, facilitating the identification of sputum smear images (Figure 4). This conversion enhanced the image when the background of the enhanced image was off-

white, aiding in the easy segmentation of bacilli images in the subsequent stage. Additionally, a Gaussian filter, a nonlinear digital filtering technique, was applied to remove noise information from the images while preserving edges.

This preprocessing step generated PTB images, either detected or not detected, with minimal noise compared to the original acquired images. The processed images then proceeded to the training phase's next step: Image segmentation. The primary goals of image preprocessing included image quality improvement, noise reduction, contrast enhancement, correction of missing or wrong pixel values, optimal data preparation for segmentation and elimination of acquisition-specific artifacts. Image preprocessing's effectiveness is closely tied to the quality of images acquired from Leica Microsystems microscopy connected to the PC. The overarching objective is to eliminate unwanted noise and enhance image features crucial for further analysis with specific techniques tailored to the nature of noise present in the image.

Image segmentation

After the image preprocessing stage, image segmentation is crucial for extracting high-quality features essential for image classification. Image segmentation serves the fundamental functions of clustering pixels with the same intensity values from entire image regions, separating regions or objects of interest and eliminating undesired regions or objects. It finds applications in various image preprocessing areas, including object recognition, handling object occlusion, boundary estimation, editing and querying from image datasets.

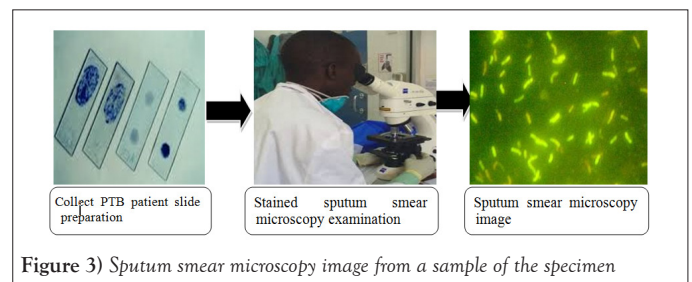


Figure 3) Sputum smear microscopy image from a sample of the specimen

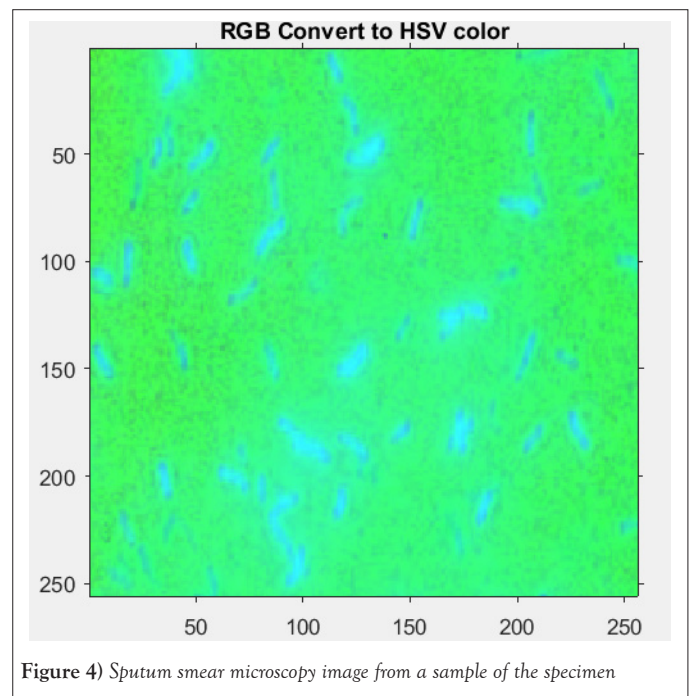


Figure 4) Sputum smear microscopy image from a sample of the specimen

Algorithms for image segmentation typically rely on two essential properties of intensity values, i.e., discontinuity and similarity. Discontinuity-based algorithms partition image regions based on sharp intensity value changes, while similarity-based algorithms partition image regions that are alike according to specified criteria. In the context of sputum smear images, accurate segmentation is crucial for medical image analysis tasks, especially in computer-aided diagnosis. Various methods out there for segmenting

sputum smear images of K-means clustering method is one.

K-mean clustering segmentation

Clustering serves as a method to group a set of data into specific clusters, and K-means clustering analysis involves assigning a set of regions into subsets (clusters) where observations in the same cluster share similarities. This method is employed for the classification of a given object into k number of classes, minimizing the squares of the distance between the object and the corresponding cluster. The K-means clustering algorithm involves the following steps:

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

Stained sputum smears, with their distinct characteristics of red bacilli against a blue and off-white background, present an exploitable property for segmentation of bacilli. Previous research indicated that K-means clustering is effective for segmentation of bacilli in stained sputum smear images. In this study, K-means clustering was employed for segmentation, resulting in RGB images with a black background and pink bacilli. The output was then converted to an image where undesired objects appeared black and the detected bacilli were displayed in white, facilitating the subsequent feature extraction process.

K-means clustering treats each object as having a location in space, aiming to find partitions where objects within each cluster are as close to each other as possible and as far from objects in other clusters as possible. The method requires specifying the number of clusters and a distance metric to quantify the proximity of objects. Using K-means to cluster objects into three clusters with the Euclidean distance metric, every pixel in the image was labelled according to the results from the K-means cluster. This pixel labelling allowed for the creation of segmented images based on colour, distinguishing objects on black backgrounds and bacilli images.

Image feature extraction

Image analysis involves extracting meaningful information to identify unique features of sputum smear images used for classification. Feature extraction in image analysis identifies the characteristics or attributes of a bacilli image, reducing the complexity in classification problems by measuring properties that distinguish one input pattern from another. For the analysis of PTB, two classification parameters were identified, namely, morphological (shape) features and colour features. Morphological features pertain to the geometric aspects of images, specifically the size and shape characteristics of PTB bacilli images. These features were extracted from binarization images, including area, eccentricity, compactness, perimeter, roundness, major length axis, minor length axis and EquivDiameter. The classification system was built upon morphology and colour analysis, considering an assessment of human visual inspection as the starting point. These features were crucial in identifying PTB bacilli based on their 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

$$A = \pi r^2 \dots\dots\dots (1)$$

Eccentricity (E): 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 a circle, while an ellipse whose eccentricity is 1 is a line segment.

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

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.

$$= \sqrt{4\pi A / P^2} \dots\dots\dots (3)$$

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

$$R = 4\pi A / P^2 \dots\dots\dots (4)$$

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.

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

$$D = \sqrt{4\pi A} \dots\dots\dots (5)$$

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. This can be used for the PTB bacilli classification of positive or negative of stained sputum smear images.

Colour features

Colour is a distinctive feature of bacilli, showcasing variations in colour among different types of PTB bacilli. Colour analysis is computed by determining the mean values of Red, Green and Blue (RGB) components and Hue, Saturation and Value (HSV) components. The RGB colour space is extensively used and is typically the default colour space for storing and representing digital images. In the RGB colour space, computers, graphics cards and monitors utilize 24-bit images, where the red, green and blue components each have 8 bits [7].

On the other hand, the HSV colour space offers a more intuitive representation of how people perceive colour compared to the RGB colour space. In HSV, hue distinguishes one colour from another, saturation describes the amount of whiteness in a light source within a given image and value measures the brightness of an image [7]. This perceptual description aligns more closely with human colour perception.

Normalized RGB is a representation easily obtained from RGB values through a straightforward normalization procedure. This normalization allows for a standardized representation of colour information, facilitating consistent analysis across images. The combination of RGB and HSV colour components provides a comprehensive perspective on the colour variations of PTB bacilli, aiding in the identification and classification of different bacillus types based on their unique colour characteristics.

$$R = \frac{R}{(R+G+B)}, G = \frac{G}{(R+G+B)}, B = \frac{B}{(R+G+B)} \dots\dots\dots (6)$$

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

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

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

Therefore, to compute the mean values of each component in the RGB and HSV colour spaces, MATLAB 2016a was employed. MATLAB offers built-in functions for converting between RGB and HSV colour spaces, simplifying the process. The RGB image was split into red, green and blue components, and mean values of RGB and HSV were calculated for stained sputum smear images. The computed mean values included red, green, blue, hue, saturation and value colours from each component. In summary, a total

of fourteen features (eight morphological and six colour features) were utilized for the classification of sputum smear images related to PTB bacilli detection.

Image classification

Image classification is the final stage in a pattern recognition system, involves assigning unknown patterns to specific categories. In the context of PTB bacilli detection, the patterns are the images themselves. The classification process entails categorizing objects in bacilli images based on measured values or features. These features are essential for identifying PTB bacilli patterns, classifying them into positive or negative categories. The image classification model comprises three main components: (i) Representation of image features, (ii) learning and (iii) testing for semantic categories and classifiers that make decisions about the identified patterns, such as PTB bacilli in sputum smear images.

Feature representation

Features, or attributes, are values measured from stained sputum smear images, including morphological and colour features. A total of fourteen features (eight morphological and six colour features) were employed for identifying PTB bacilli in sputum smear images, classifying them as either PTB bacilli detected or non-PTB bacilli detected. The two major classes selected for this study were positive (assigned a value of 1) and negative (assigned a value of -1). These classes represented PTB bacilli detected and non-detected, respectively. The dataset comprised 180 sputum smear images and a supervised learning method was used for training the classifier. The training process involved 70% of the dataset, while the remaining 30% was used for testing the classification accuracy.

Training and testing phase process

The training and testing phases involved randomly selecting datasets using the crossvalind function, which creates random partitions. The classifier was trained with 70% of the dataset, emphasizing the importance of high-quality training samples for accurate classification. Training labels were specified to correspond to the order of training nodes. After training the model, its performance was evaluated using the remaining 30% of the dataset, measuring accuracy, sensitivity, specificity and F1 score. These performance metrics provided a comprehensive assessment of the classifier's effectiveness in PTB bacilli detection. Accuracy indicated the ratio of correctly classified observations to the total observations, sensitivity measured the true positive rate, specificity gauged the false positive rate and the F1 score represented a weighted average of sensitivity and specificity. These metrics collectively determined the accuracy and reliability of the developed model for PTB bacilli detection in sputum smear images.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (10)$$

$$Specificity = \frac{TN}{TN+FP} \dots\dots\dots (11)$$

$$Sensitivity = \frac{TP}{TP+FN} \dots\dots\dots (12)$$

$$F1\ score = \frac{2TP}{2TP+FP+FN} \dots\dots\dots (13)$$

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

Support vector machine classifier

For superior classification performance in PTB bacilli detection, accuracy, sensitivity, specificity and F1 score were chosen as key metrics. The Support Vector Machine (SVM), a supervised machine learning algorithm versatile in both classification and regression tasks, was selected for its potential to deliver higher accuracy compared to other classifiers. SVM excels in finding an optimal hyperplane in higher-dimensional feature space, effectively separating distinct categories within input data. Additionally, SVM is known for its efficient training capabilities, even with a large volume of input data. The algorithm plots each data point as a point in an n-dimensional space, where n represents the number of features, with the goal of finding a hyperplane that effectively differentiates between the two classes, PTB bacilli

and non-PTB bacilli detected.

RESULTS AND DISCUSSION

Classifying results

The initial phase involved using images to establish a knowledge base (image database) for classification. Sputum smear images, subjected to image preprocessing techniques and K-means clustering for segmentation, constituted the training dataset. Fourteen features, comprising morphological and colour features, were extracted from each image to create a robust knowledge base. Subsequently, SVM classification trained on the fourteen input features and two output vectors representing positive and negative classes was applied to identify PTB bacilli in the testing phase.

Implementation of automatic pulmonary tuberculosis bacilli detection

This section details the implementation of the PTB bacilli detection process, encompassing five key image processing steps: Image acquisition, preprocessing, segmentation, feature extraction and classification. The first step, image acquisition involves capturing images from stained sputum smear slides, obtained from the National Tuberculosis Reference Laboratory at the Ethiopian Public Health Institute (EPHI). The second step employs image preprocessing techniques to enhance image quality by eliminating unwanted noise. Consideration is given to various environmental factors, such as lighting and camera resolution, to optimize the visual appearance of images. This critical preprocessing step minimizes noise and enhances image quality after acquisition. Image segmentation, the third step, is a pivotal stage in image analysis. Segmentation methods divide images into homogeneous and meaningful regions, with pixels in each region sharing identical properties. The segmented images yield binary results, from which morphological and colour features are extracted. These features play a crucial role in identifying PTB bacilli and assigning positive or negative classifications. In general, the process involves capturing images, refining them through preprocessing, segmenting them to extract relevant features and employing SVM classification to identify PTB bacilli. The developed system demonstrates a comprehensive approach to automatic detection of PTB bacilli, leveraging advanced image processing techniques and machine learning algorithms for accurate and efficient diagnosis.

Development environment

The development of the PTB bacilli detection system, integrating advanced image analysis techniques, requires a substantial investment. From image acquisition, necessitating a high-quality digital camera and a controlled environment, to resource-intensive image preprocessing techniques, the system demands powerful computing resources. The developed system was implemented and tested on a PC with an Intel® Core™ i5-4200U CPU, operating at 2.30 GHz speed, 4.00 GB of RAM, and a 64-bit Microsoft Windows 10 operating system.

Binary image analysis

Binary image analysis involves transforming an image's pixel values into 0 s and 1 s, essentially black and white. In this study, white was inverted to black, indicating the object of interest or the mass region, while black was inverted to white, representing other parts of the bacilli image. MATLAB R2016a, a multi-paradigm numerical computing programming language environment developed by MathWorks (www.mathworks.com, 2018), was used for binary image analysis. MATLAB proves to be a robust tool for image analysis and processing. It can read, load, edit, delete, process and analyse various image formats. In the presented study, MATLAB read and loaded a sequence of images stored in a folder one by one. It enhanced the quality of each image, removed noise and converted the image into a binary format for subsequent feature extraction purposes. The MATLAB code utilized the readimage (filename) function, with the image file name depending on the format of the file in the data store. The readimage function supports image formats compatible with imread. The sample image loaded into the MATLAB is presented in Figure 5.

Image segmentation

Image segmentation plays a crucial role in isolating stained sputum smear images from the background and in this study, K-means clustering techniques were employed for this purpose. The process initiated with the utilization

of the built-in imread function to capture the input sputum smear image. The choice of the HSV (Hue, Saturation, Value) colour space is based on its closer correspondence to human colour perception compared to the RGB colour space. The conversion to HSV was facilitated by the rgb2hsv function.

HSV colour space proves valuable in generating high-quality images in computer graphics, especially in selecting distinct colours relevant to identifying bacilli images. Its significance lies in its ability to represent the desired colour (bacilli object) effectively. The HSV colour space is particularly pertinent to identifying bacilli images, aligning with human perception in PTB bacilli detection. When performing histogram equalization of a colour image for PTB bacilli detection based on intensity components, HSV is advantageous due to its robustness to slight changes in lighting conditions and its capability to eliminate undesired objects. The subsequent steps involved calculating and plotting the histogram for each HSV colour component. The average HSV values across all image pixels were then computed and plotted, as depicted in Figure 6. Among the HSV colour components, the value of the image colour emerged as particularly effective in distinguishing PTB bacilli from other undesired objects.

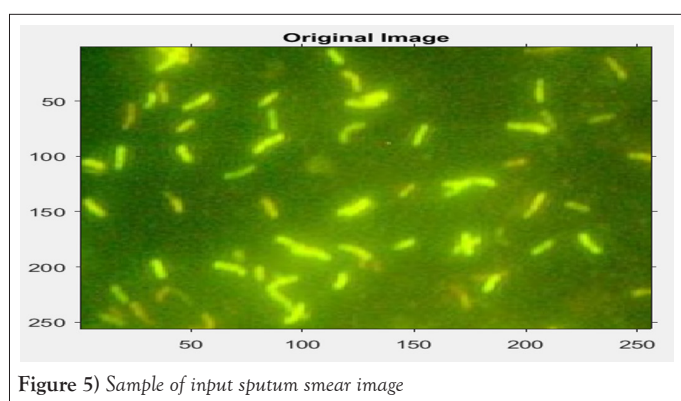


Figure 5) Sample of input sputum smear image

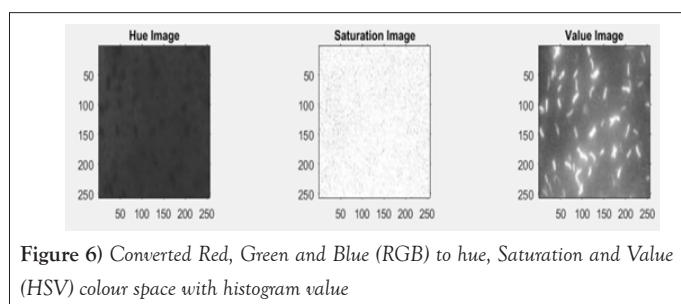


Figure 6) Converted Red, Green and Blue (RGB) to hue, Saturation and Value (HSV) colour space with histogram value

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 noises in sputum smear image based value image color. In addition, contrast enhancement 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: $G = \text{imgaussfilt}$ function and to convert contrast enhanced, $C = \text{imadjust}$ function was used in MATLAB as shown in Figure 7.

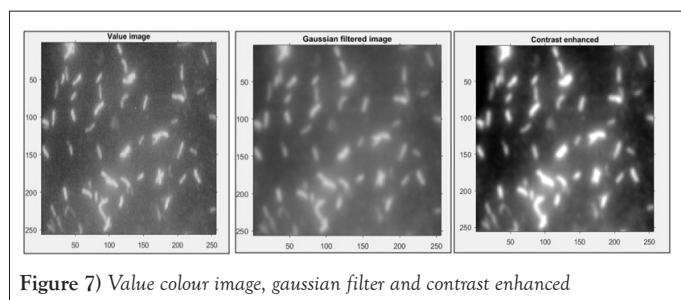


Figure 7) Value colour image, gaussian filter and contrast enhanced

Figure 8A 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. The number of centers is generated based on number of the cluster K. This center is used

as the initial center in K-means algorithm. Using the kmeans algorithm, the image is segmented into K number of cluster. After segmentation of the image, it 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 center. The initial center will be carefully chosen so that we get our desire segmentation of bacilli objects as shown in Figure 8B. 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 in computing for morphological and color feature was selected.

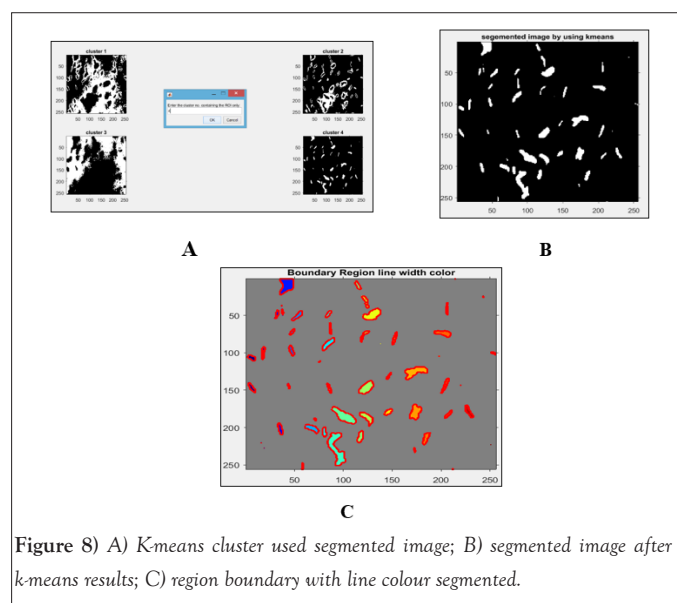


Figure 8) A) K-means cluster used segmented image; B) segmented image after k-means results; C) region boundary with line colour segmented.

Morphological features analysis

A morphological feature is the size and shape characteristics of sputum smear images of PTB. Eight (8) 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. Figure 9A shows a region of interest of PTB bacilli images that were interested in computing its morphological features based region of interest by boundary with edge colour.

Based on the detected region of interest as shown in Figure 9B, 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 MATLAB function presented in Figure 10. 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 Figure 10, the mean value of each feature computed by used formula, from each sputum smear image for a better performance.

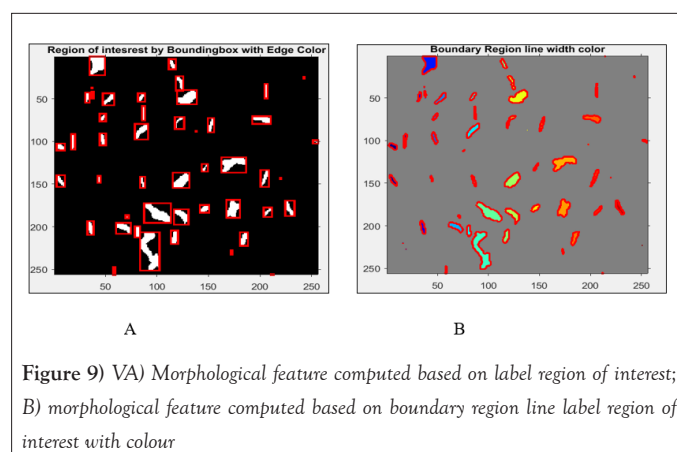


Figure 9) VA) Morphological feature computed based on label region of interest; B) morphological feature computed based on boundary region line label region of interest with colour

	8	9	10	11	12	13	14
	Eccentricity	Perimeter	Compactness	Roundness	Equidimeter	MajorlenthAxis	MinorlenthAxis
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'

Figure 10) Morphological feature value of each label region of interest

Color feature analysis

Color features are visual attributes of sputum smear images that results from the light emitted or transmitted or reflected and six color features were identified. They were computing the mean value of Red, Green and Blue (RGB) color components and the mean value of Hue, Saturation and Value (HSV) 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 were used as shown in Figure 9A. After each colour component was splitted, the mean values of each component colours were computed by RGB function methods based on the specified region of interest of identified PTB bacilli as shown in Figure 11.

In summary, from Figures 10 and 11, the label of region of interest, which computed the value of each label region of interest of the morphological feature and computed mean value of each component of (HSV and RGB) colour was again computed by using MATLAB tools function. The computed mean values of combined results from the fourteen features were as shown in Figure 12.

The results of mean value of the fourteen features (eight morphological features and six split of color component of features) as shown in the table contain hue, saturation, value, red, green, blue, area, eccentricity, perimeter, compactness, roundness, equidimeter, 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. Column normalization was done 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 Figure 13, to each feature as input to differentiate the two classes of the sputum smear images. Hence, the total 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.

Experiment results

Experimental results ensure the realization of the developed system. It is

an integral part of the development of PTB bacilli detection system. The experiment was carried out by using image processing techniques PTB bacilli by implementing algorithms in MATLAB platform.

First the input image was resized by using MATLAB function imresize with 300 × 400 pixel rates. The Figure 14A shows the resized the original images query. The HSV colour space converted was applied for essential to segment sputum smear images, the process of choosing the suitable colour space for better automatic detection of PTB bacilli affect with the staining of smear image and the procedure that was used for acquisition the bacilli image as shown Figure 14B. In this preprocessing phase the Gaussian filter was used for reduced noised from bacilli image as shown in Figure 14C and contrast enhanced the bacilli image was enhanced by with adjustment as shown in Figure 14D. In segmentation phase was done by implementing K-means clustering algorithm with the K value of 3. The Figure 14E demonstrated the segmented region of interest. After the segmentation process, the segmented image was converted to black and white image to facilitate the extraction of features. The Figure 14F demonstrated the conversion of RGB bacilli image into binary bacilli image and then fourteen features were extracted from the resultant image. The dataset is grouped into two classes PTB bacilli positive or PTB negative.

The computation of morphological and color features was 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 study, it was designed that the experimental scenarios test the classification performance by taking the extracted features of sputum smear images. The classification was tested by using SVM classifiers algorithms to get a more accurate result. To do so, the train classifiers, i.e., a set of training sputum smear images was required, and the classes label where it belongs to. For this purpose, 180 sputum smear images were taken from EPHI from the predefined two types of PTB bacilli includes positive and negative sputum smear images Figure 15.

	1	2	3	4	5	6
	meanH	meanS	meanV	meanR	meanG	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...

Figure 11) Value of colour feature in each label region of interest

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	meanH	meanS	meanV	meanR	meanG	meanB	Area	Eccentricity	Perimeter	Compactness	Roundness	Equidimeter	MajorlenthAxis	MinoriLengthAxis
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 87997686376360'
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...	'117 2770...	'72 47973...	'3 895752...	'0 372667296...	'0 615437069...	'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'

Figure 12) Computed combined both mean value of morphological features and mean value of colour features

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	meanH	meanS	meanV	meanR	meanG	meanB	Area	Eccentricity	Perimeter	Compactness	Roundness	Equidimeter	MajorlenthAxis	MinorlenthAxis	PTBType	
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 35895933300088'	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 7690...	'135 8044...	'119 7371...	'40 81818...	'0 600109478...	'17 3613636...	'0 93261758319...	'0 678803773...	'5 110843993...	'7 380056404856...	'4 05610495171054'	-1	

Figure 13) Labelling of Pulmonary Tuberculosis (PTB) bacilli detected positive and negative assigned by pathologists

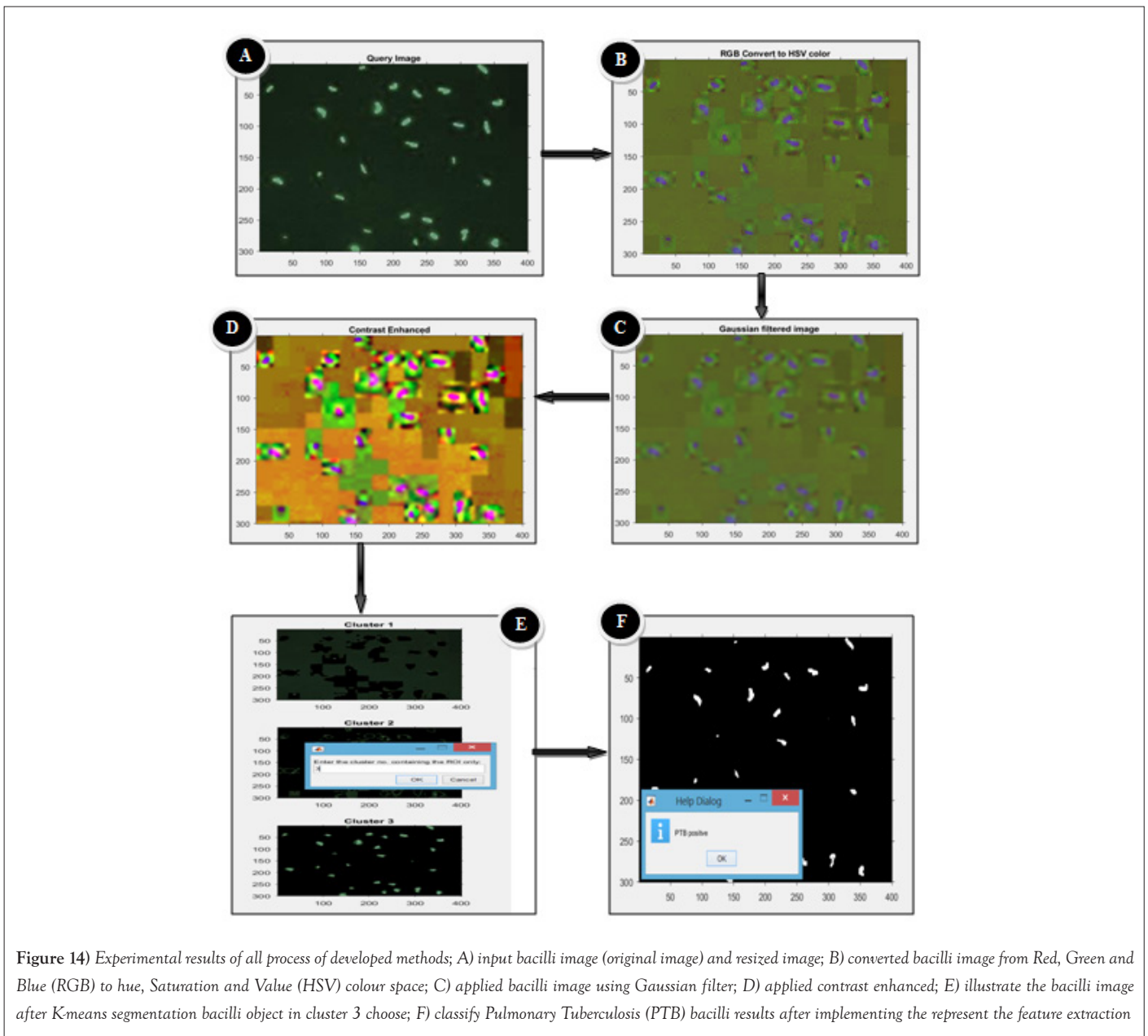


Figure 14) Experimental results of all process of developed methods; A) input bacilli image (original image) and resized image; B) converted bacilli image from Red, Green and Blue (RGB) to hue, Saturation and Value (HSV) colour space; C) applied bacilli image using Gaussian filter; D) applied contrast enhanced; E) illustrate the bacilli image after K-means segmentation bacilli object in cluster 3 choose; F) classify Pulmonary Tuberculosis (PTB) bacilli results after implementing the represent the feature extraction

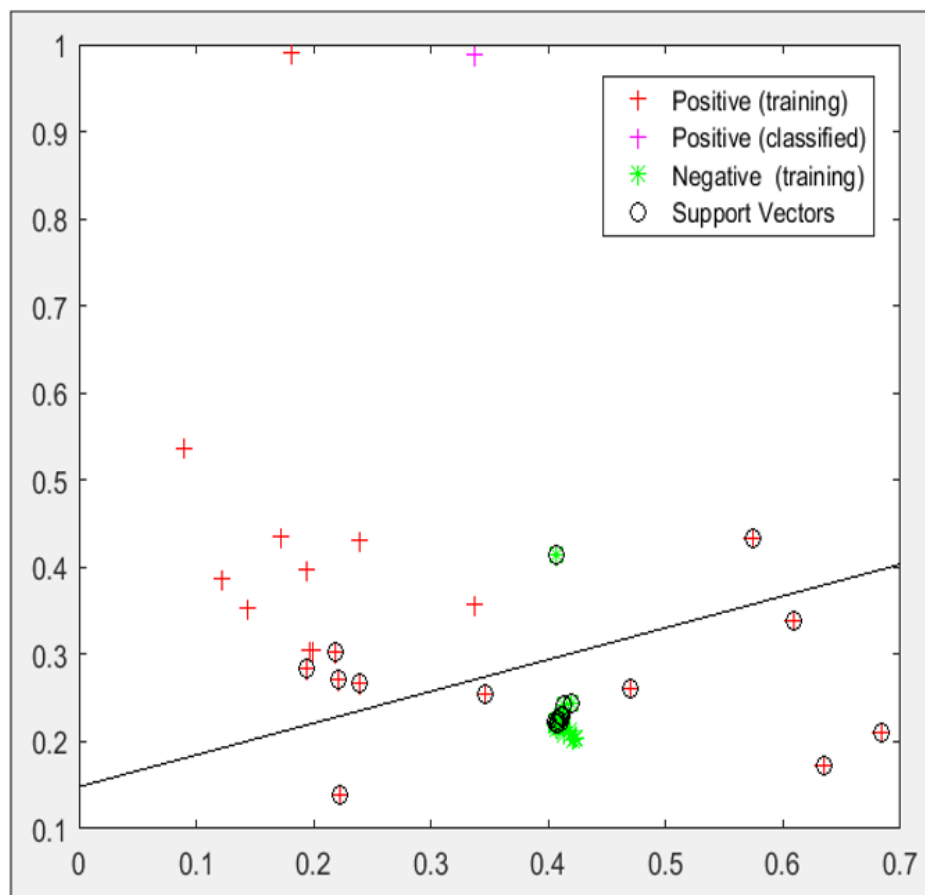


Figure 15) The classification result of SVM classifier scenario of PTB bacilli positive and negative; Note: (+): Positive (training); (+): Positive (classified); (*): Negative (training); (o): Support vectors

There were two basic phases of pattern classification, namely training and testing phases. The researchers used SVM classifier which is used a well-known algorithm to identify PTB in a given class, based on training data. The basic ideas are 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, a classifier was used to gain knowledge base (database) content based on the statistical analysis performed during training. Hence, to design the classifier partitioning the total dataset into training and testing dataset was necessary. 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 for testing purpose. From the total of 180 datasets, 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 [8].

Support vector machine classifier result

Support Vector Machines (SVMs) are foundational binary classification algorithms renowned for their extensive generalization properties and computational efficiency. The algorithm is particularly favored for classification tasks owing to its capability to establish decision boundaries through the concept of decision planes. Decision planes serve as boundaries that distinguish between objects belonging to different class memberships. In the context of this study, the objects were categorized as either green or red.

The decision plane functions as a separator, delineating a boundary where objects to the right side are classified as green, and those to the left side are classified as red. Any new object, represented as a white circle, falling to the right of the separating line would be labelled and classified as green, while falling to the left would result in a classification as red. In this study, the SVM classifier played a pivotal role in classifying PTB bacilli detection.

Experimental scenarios were designed based on the features extracted. For each scenario, the SVM classifier algorithm was applied to the selected views of sputum smear images within the dataset. As previously mentioned, 70% of this dataset was allocated for training, and the remaining 30% was reserved for testing purposes in each scenario. This approach ensured the robust training and evaluation of the SVM classifier in the context of PTB bacilli detection [9].

Analysis of the findings

The experiments were conducted under different scenarios by using extracted features of the sputum smear images. The experimental results of SVM classifier using 30% percentage splitter were shown over the scenario and their performance summarized in Table 1.

The dataset for this study comprised a total of 180 sputum smear images. The dataset was divided into two classes, distinguishing between PTB bacilli-positive and PTB bacilli-negative sputum smear images. The classification of test images into PTB bacilli-negative or -positive was conducted by domain experts (pathologists) from EPHI, who assessed and labelled each image.

The SVM classifier was employed with both morphological and colour features, encompassing a total of fourteen features (eight morphological and six colour features). The testing dataset consisted of 54 sputum smear images, and the classification performance of the proposed system was evaluated based on the results obtained from the SVM classifier and the ground truth.

Pathologists often face challenges in identifying PTB bacilli, with an error rate exceeding 50% due to oversights or manual processing errors. The developed system, however, demonstrated a higher level of accuracy, influenced by the skills and decision-making abilities of pathologists.

The system's performance was tested using a sample dataset derived from the ground truth, and the results were measured using accuracy, sensitivity, and

specificity. The confusion matrix, a standard tool for evaluating classification models, was utilized for this purpose. The SVM classifier made a total of 54 predictions, with 36 images predicted as positive and 15 as negative. However, the actual readings by pathologists indicated 51 positive and 3 negative images.

The evaluation metrics, including accuracy, sensitivity, specificity and F1 score, were calculated based on the confusion matrix. The results, presented below, provide a comprehensive assessment of the PTB bacilli detection system's performance. These metrics serve as crucial indicators to gauge the effectiveness of the system in accurately identifying PTB bacilli in sputum smear images.

From the performance results, the overall detection accuracy was 94.4%, while sensitivity was 95%, specificity was 94% and F1 score was 96%. As per the researchers' knowledge, there is no local research attempts made to use sputum smear image by using image processing techniques for detection of PTB bacilli, but there are different researches that used image processing techniques for identify PTB bacilli positive or negative elsewhere. 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 Osama (2016) the same area [10]. 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, Sudan. The main objective of the study was 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, a finding which is in line with our studies outcome. The detail comparison of our findings with that of Osama is presented in Table 2.

Upon comparing the findings with Rao Osama's study conducted in 2016, it is evident that Osama achieved a lower sensitivity of 83.07%, compared to the current study, which is 95%. This discrepancy may be attributed to the increased incorporation of morphological features and colour features in the present work, providing a more comprehensive representation of the real working environment.

The current research introduces a combined set of eight morphological features and six colour features, resulting in an impressive F1 score of 96%. Notably, the former study did not specify F1 score performance measurements. Additionally, Osama's work utilized two morphological features, namely eccentricity and compactness, contributing to its accuracy of 83.07%. However, it is important to note that focusing on only two morphological features may not fully represent the diversity of features considered in the current study [11].

The current research addresses this gap by comprehensively analyzing the existing situation, incorporating a broader spectrum of features. The combination of eight morphological and six colour features exhibits greater discriminating power than the previous study [10]. The synergistic effect of combining both morphology and colour features enhances the system's ability to discriminate between different classes. Moreover, the performance metrics of the support vector machine classifier in the current work, including accuracy (94.4%), sensitivity (95%), specificity (94%) and F1 score (96%), showcase significant improvements compared to the values reported in previous study. This highlights the advancement and enhanced performance of the prototype developed in the current research.

TABLE 1
Scenario for diagnostic result of testing data using Support Vector Machine (SVM) classifier

Actual/predicted class	P	N	Total
T	36	15	51
F	1	2	3
Total	37	17	54

TABLE 2
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	F1 score
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	Support Vector Machine (SVM)	94.40%	95%	94%	96%
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	Support Vector Machine (SVM)	81%	83.07%	66.66%	Not specified

CONCLUSION

The development of an automated Pulmonary Tuberculosis (PTB) detection system, utilizing low-cost image processing techniques, holds immense potential for aiding in the diagnosis of PTB in resource-constrained regions. This study presents a novel algorithm for the automatic identification of PTB bacilli in stained sputum smear images, providing a valuable contribution to PTB diagnosis. The system leverages conventional microscopy, allowing for the utilization of low-cost equipment prevalent in developing countries. The dataset, sourced from the Ethiopian Public Health Institute (EPHI), forms the basis for algorithm training and evaluation. By accurately detecting and classifying PTB bacilli, the system aims to enhance the efficiency of PTB diagnosis, particularly in areas with limited access to high-quality diagnostic resources.

The algorithm employs image processing techniques, including Gaussian filtering and contrast enhancement, to preprocess and improve the quality of acquired images. K-means clustering facilitates image segmentation, extracting essential features for PTB bacilli identification. Fourteen features, comprising eight morphological and six colour features, are employed for the classification of PTB bacilli in sputum smear images. The study employs a Support Vector Machine (SVM) classifier, leveraging both morphological and colour features for PTB detection. The developed Graphical User Interface (GUI) serves as a user-friendly platform for image input and results display, indicating whether PTB bacilli are detected or not.

In terms of performance evaluation, standard image processing metrics, including accuracy, sensitivity, specificity and F1 score, are utilized. The proposed system demonstrates commendable results, with accuracy, sensitivity, specificity and F1 score measuring at 94.4%, 95%, 94% and 96%, respectively. The promising outcomes indicate the system's potential as an effective tool to aid pathologists in the early-stage diagnosis of PTB bacilli.

ACKNOWLEDGEMENTS

The authors acknowledge Jimma University for providing a small research fund and EPHI and the domain experts for taking part in this study.

REFERENCES

1. Kim CJ, Kim Y, Bae JY, et al. Risk factors of delayed isolation of patients with pulmonary tuberculosis. *Clin Microbiol Infect.* 2020;26(8):1058-1062.
2. World Health Organization. Global tuberculosis report 2013. World Health Organization; 2013.
3. Panicker RO, Soman B, Saini G, et al. A review of automatic methods based on image processing techniques for tuberculosis detection from microscopic sputum smear images. *J Med Syst.* 2016;40:1-3.
4. Forero MG, Cristobal G, Desco M. Automatic identification of *Mycobacterium tuberculosis* by gaussian mixture models. *J Microsc.* 2006;223(2):120-132.
5. Thirumaran J, Shylaja S. Medical image processing-an introduction. *Int J Sci Res.* 2015;4(11):1197-1199.
6. Deserno TM. Fundamentals of biomedical image processing. *Biomed Imag Proc.* 2010;2010:1-51.
7. Acharya T, Ray AK. Image processing: principles and applications. Wiley. 2005;2005:48.
8. Deserno TM. Fundamentals of biomedical image processing. *Biomed Imag Proc.* 2010;2010:1-51.
9. Heyd AT. Management of latent tuberculosis infection among an inner-city population with psychosocial barriers to treatment adherence. 2020.
10. Altayeb RO. Automatic method for tuberculosis bacilli identification in sputum smear microscopic images using image processing techniques (Doctoral dissertation, Sudan University of Science and Technology).
11. Mohajan H. Tuberculosis is a fatal disease among some developing countries of the world. *Am J Infect Dis.* 2015;3(1):18-31.