

PAPER • OPEN ACCESS

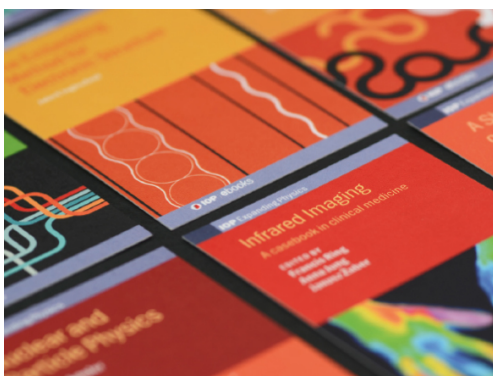
An Automated Intelligent Identification and Counting System Procedure for Tuberculosis

To cite this article: Fatin Atiqah Rosli *et al* 2019 *J. Phys.: Conf. Ser.* **1372** 012079

View the [article online](#) for updates and enhancements.

Recent citations

- [Automatic Bacillus anthracis bacteria detection and segmentation in microscopic images using UNet++](#)
Fateme Hoorali *et al*



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

An Automated Intelligent Identification and Counting System Procedure for Tuberculosis

Fatin Atiqah Rosli¹, Mohd Yusoff Mashor^{*1}, Siti Suraya Md Noor²

¹School of Mechatronic Engineering, Universiti Malaysia Perlis, 02600, Perlis, Malaysia.

²Department of Microbiology, School of Medical Science, Universiti Sains Malaysia Kubang Kerian, Kelantan, Malaysia

*yusoff@unimap.edu.my

Abstract. Tuberculosis (TB) is an infectious disease caused by Mycobacterium Tuberculosis or TB Bacilli. Currently, the classification of TB bacilli is carried out by microbiologist by using Ziehl-Nielsen (ZN) stained smear sputum slide under a light microscopy. However, the manual evaluation is time-consuming and lead to slow decision. Furthermore, the sensitivity is less due to incline of human error which lead to inaccurate conclusion. Therefore, this study proposes an intelligence identification and counting system to detect the presence of TB bacilli in the ZN-stained smear sputum image. This system is designed to identify the presence of TB bacilli and count the number of TB bacilli by applying digital image processing and artificial intelligence techniques. In image acquisition, there are 70 samples images of ZN-stained smear sputum image were collected from Hospital Universiti Sains Malaysia (HUSM) Kubang Kerian, Kota Bharu, Kelantan, Malaysia. The image processing technique consists of contrast enhancement, segmentation, and feature extraction. The contrast of original image was enhanced by the combination of global enhancement, local enhancement and Contrast Limited Adaptive Histogram Equalization (CLAHE). Then, the enhanced image was segmented using color thresholding and the features were extracted consists of on 18 colour features, 15 shape features and 5 texture features. Afterward, the features underwent feature selection to select the relevant features by using Neighborhood Component Analysis (NCA) and ReliefF Analysis. The study showed that there are relevant features were chosen by ReliefF at feature weight more than 0.004 including (8 colour features, 11 shape feature and 3 texture features) for improving the performance and accuracy of Multilayer Perceptron (MLP) trained by Scaled Conjugate Gradient (SCG). For classification process, MLP, k-Nearest Neighborhood (k-NN) and Support Vector Machine (SVM) are used with 6 folds cross-validation. It was found that MLP has the highest of accuracy, sensitivity and specificity with 93.8%, 93.4% and 94.1% respectively.

1. Introduction

Tuberculosis (TB) is an infectious disease transmitted from a person to another person by coughing or breathing in the air containing bacteria called Mycobacterium Tuberculosis or TB Bacilli. It can be divided into Pulmonary TB (PTB) and Extra-Pulmonary TB (EPTB). TB mainly target the lungs, but also can affect any part of the body such as bone and joints, kidneys, brain, and lymph node.



According to The World Health Organization (WHO) [1], TB is the top 10 leading causes of death worldwide in 2017. Estimated 10 million people were infected by TB and 1.6 million of people die from the disease. There were 5.8 million of men, 3.2 million of women and 1 million of children are suspected in the overall 10 million that were infected with TB. It shows that more than half of the overall cases are represented by men.

Nowadays, manual diagnostic procedure is commonly used to detect the presence of TB by microbiologist. In many countries, sputum smear microscopy is the main tool for laboratory diagnosis of TB [2]. Although it requires simple laboratory, this procedure reveal low detection rate and less sensitivity for determining a positive TB [3]. Moreover, the diagnosis under the microscope can lead to eye pain and poor posture during screening because it takes 15 minutes per slides and the microbiologist require to confirm hundreds of slides may per person. The smear microscopy method is categorised as manual diagnostic procedure. Currently, the manual diagnostic procedure is quite outdated compared to other current technology worldwide, include in every field. Thus, artificial intelligence (AI) is explored to develop an autonomous system in detection of TB bacilli

Samuel *et al.* [4] proposed a system using Inception V3-SVM model in transfer learning and fine-tuning for classification of infected and non-infected by TB Bacilli. The system had the limitation in low availability of the dataset and decreased the sensitivity because unable to support adaptive learning. The average accuracy of the classifier is 95.05%.

Panicker *et al.* [5] has proposed a system to detect TB Bacilli by using pixel classification for detection TB Bacilli. Pixel classification specifically used the convolution neural network (CNN) in the architecture of the network which extracted a patch in images and probability value was generated to determine the presence of bacilli. The architecture of this classifier consisted of 2 convolution layers contained 32 filter maps, 3 convolution layer contained 64 filters and 1 convolution layer with 128 filters.

Mithra *et al.* [2] claimed that Gaussian-Fuzzy-Neural network (GFNN) classifier had achieved better performance than another model for diagnosing of TB Bacilli specifically in determining images into the bacilli and few bacilli. Besides that, the counting bacilli in the overlapping bacilli were computed by GFNN. The GFNN model was used as a classifier to classify bacilli into three categories which were few bacilli, overlapping bacilli and non-bacilli. Moreover, the model was also used for counting the bacilli by computing few bacilli and overlapping bacilli.

The current research consists of two parts which is image processing and classification. Image processing consists of contrast enhancement, segmentation and filtering. Then, the segmented images undergo feature extraction and feature selection. After that, the selected features become inputs for classifiers to determine the best classifier by evaluating accuracy, sensitivity and specificity.

2. Methodology

Figure 1 shows the flowchart for the methodology. The process involved image pre-processing, contrast enhancement, segmentation, filtering, feature extraction and classification. The ZN-stained smear sputum slides were acquired from Hospital Universiti Sains Malaysia (HUSM) Kubang Kerian, Kota Bharu, Kelantan, Malaysia.

2.1. Image Pre-Processing

In pre-processing method, the background and foreground were adjusted in order to get a clear view of TB Bacilli and partially eliminate the undesired object in the images such as a bluish colour of sputum and other artefacts. The foreground is enhanced so that the purple reddish TB Bacilli become clearer. This method has three parts which are global contrast enhancement, local contrast enhancement and contrast limited adaptive histogram equalization (CLAHE).

2.2. Segmentation

The color thresholding is implemented in order to segment the non-reddish colour and reddish colour of TB bacilli. The colour model implements the *LAB* colour space to adjust the pixels so that the

reddish colour was segmented. This method uses Color Thresholding Apps provided by MATLAB R2017a. The outputs are the RGB image and binary image. This is due to `regionprops()` requires binary image to proceed. The RGB image of colour thresholding is used only for display, however the binary image is used to further filtering the unwanted objects within the image. The unwanted objects areas as known as artefacts. Then, the original colour of the image was retrieved based on the resulted segmented binary image.

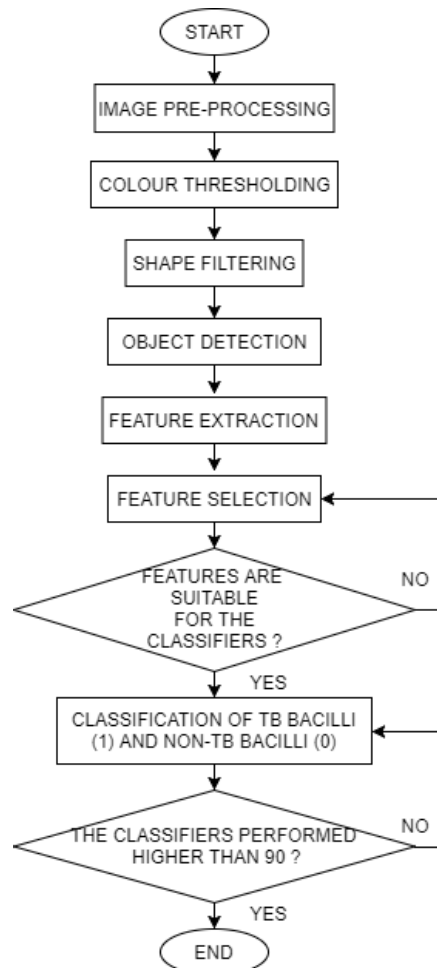


Figure 1 The flowchart of methodology

2.3. Feature Extraction and Feature Selection

2.3.1. Feature Extraction

A total of the 38 features have been extracted, namely 18 colour, 15 shape and 5 texture. The colour features were extracted from the mean, variance and standard deviation of RGB colour and HIS colour space. The shape features were extracted from `regionprop()` function which are perimeter, area, solidity, diameter and eccentricity. Other shape features are compactness, circularity, the ratio of major over minor axis and 7 Hu's Moment. Lastly, the texture features were extracted by using `grayprops()` where the RGB segmented image converted into grayscale image. These features are correlation, energy, contrast, homogeneity, and entropy.

2.3.2. Feature Selection

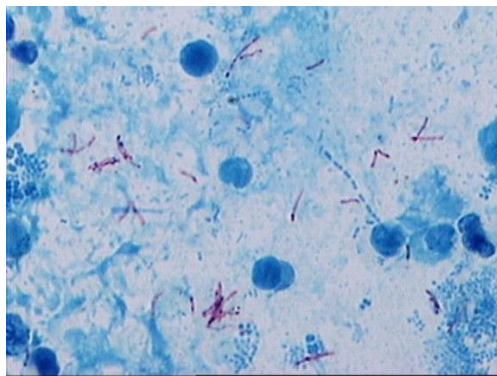
There are 38 extracted features from the previous process have undertaken feature selection to avoid overfitting. There are two types of feature selection methods that were used in this study, namely Neighborhood Component Analysis (NCA) and ReliefF Analysis. In MATLAB, ReliefF and NCA are the built-in function. For ReliefF, the function is *relief()* that give feature weight as the output. Usually, the feature weight which assigns as zero, that indicates the irrelevant feature and whereas NCA uses function *fscnca()* that will give rank and feature weight as outputs. In NCA, the feature weight assigns either zero or negative value is the irrelevant feature.

2.4. Classification

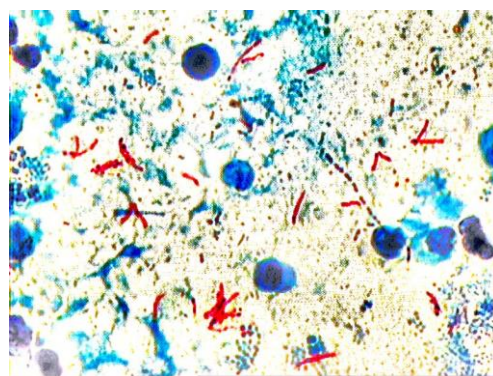
After extracting and selecting the features, the subsets of data were used for classification. This classification used 3 models of classifier for detecting TB bacilli in the segmented image. These 3 models are Multilayer Perceptron (MLP), K-Nearest Neighbour (K-NN) and Support Vector Machine (SVM). Each classifier has different approach in order to classify different objects existent within the image. All classifiers will implement k-fold cross validation. The analysis is further when the data split into training and testing by using k-fold cross validation approach. The number of k will be varied in order to train the data and evaluated the accuracy, sensitivity and specificity of the classifier. The best classifier has the highest of these three evaluation parameters.

3. Experimental Result and Discussion

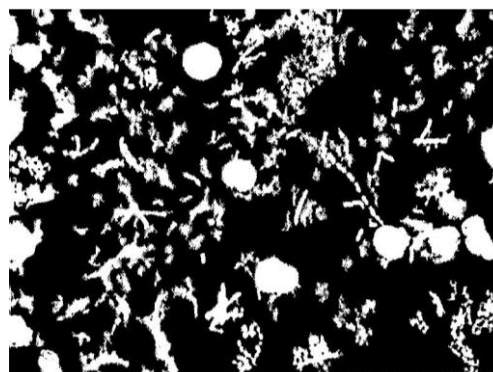
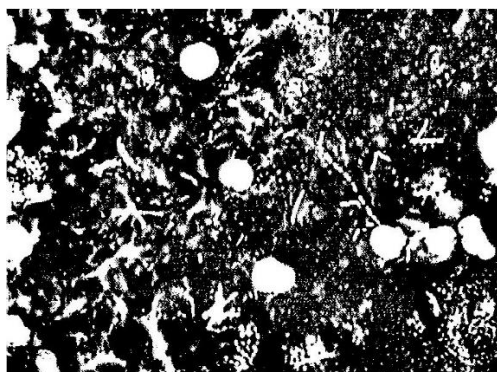
The procedure was evaluated using 70 images of ZN-stained smear sputum that were collected from Hospital Universiti Sains Malaysia (HUSM) Kubang Kerian, Kota Bharu, Kelantan, Malaysia. Figure 2 shows images to illustrate the process of image enhancement, filtering and segmentation where the image was segmented until only TB bacilli and other reddish colour of objects were remained in the image. The images of the process are shown in Figure 2(a) until Figure 2(g)



(a) Original Image



(b) Enhanced Image



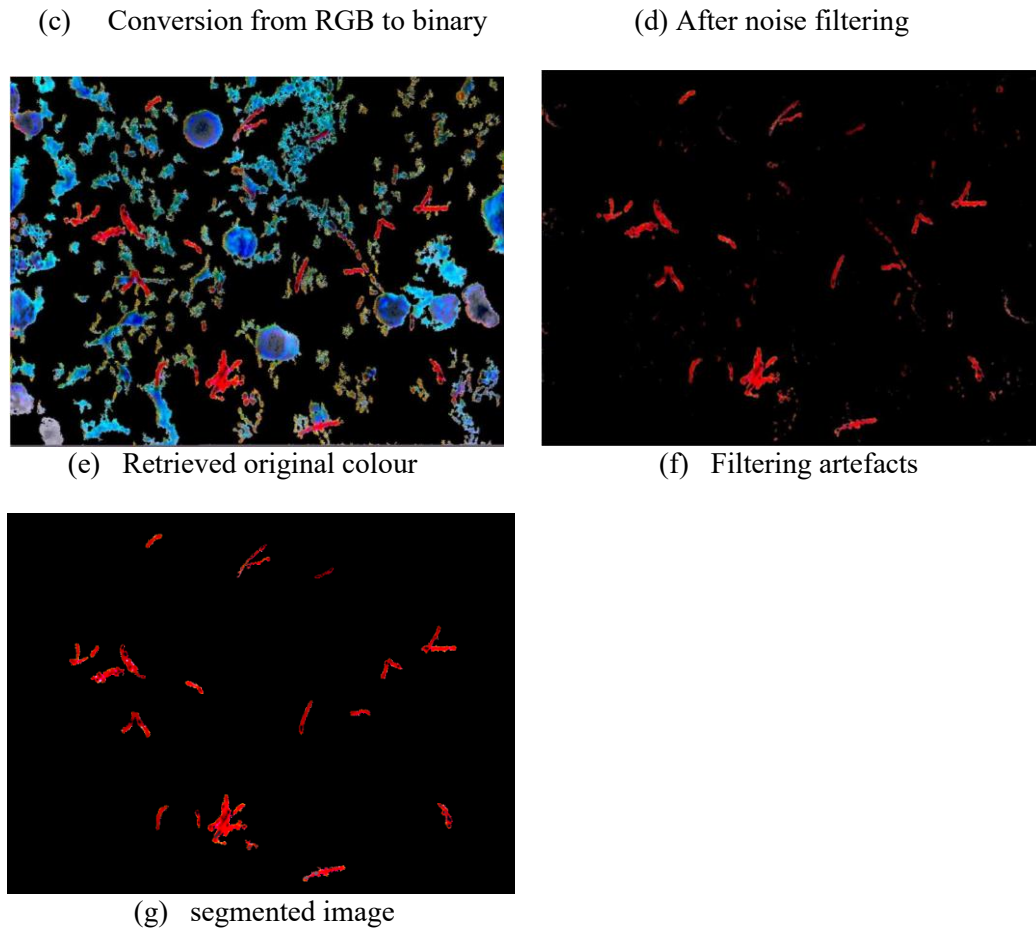


Figure 2 The TB bacilli image segmented process

There are 38 features that have been extracted in this study. However, the features have to be selected in order to improve the performance of classifier. Neighbour Component Analysis (NCA) and ReliefF were used for this task. As a result, 22 features have been selected with the lowest MSE by ReliefF as shown in Figure 3. Table 1 shows the performance of the three classifiers, evaluated using 6 folds cross-validation. The best performance is MLP 15 hidden layer that yielded the accuracy, sensitivity and specificity of 93.8%, 93.4% and 94.1% respectively.

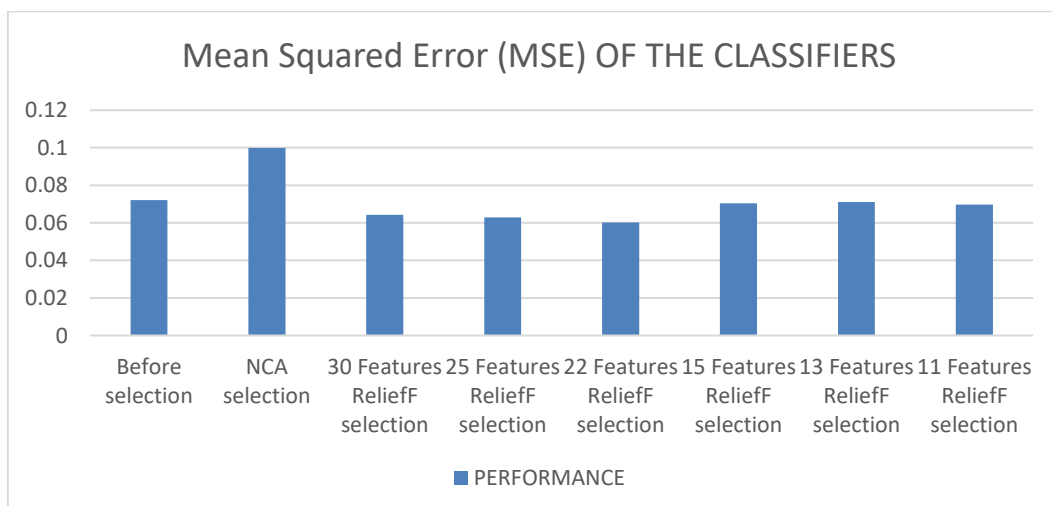


Figure 3. The performance of the classifiers
Table 1 Performance of classifiers

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
MLP trained by SCG	93.8	93.4	94
k-NN (k =5)	85.0	82.1	87.9
SVM (polynomial kernel function)	91.1	91.2	90.9

4. Conclusion

The proposed image segmentation procedure has successfully segmented TB and non-TB bacilli from the background sputum. After the image processing was performed, 528 objects have been detected in the images that containing TB bacilli and non-TB bacilli. There were 38 features that have been extracted in this study, NCA and ReliefF algorithms were then used to select the most suitable features. The best accuracy was produced by ReliefF with 22 selected features. The selected features are 8 colour features. MLP, k-NN and SVM were used for classification. All the classifier yield good classification performance with the accuracy of higher than 80%. MLP trained by yielded the highest accuracy, sensitivity and specificity with 93.8%, 93.4% and 94.1% respectively.

Acknowledgements

The authors would like to express the highest gratitude to Universiti Malaysia Perlis and Prof. Dr. Mohd. Yusoff Mashor as my final year project's supervisor which he had giving me completely support, guidance and motivation through all the procedure of this study. I am genuinely thankful to have support and consolation from my family and friends who teach and share their knowledge with me. Lastly, I might want to express my earnest thanks toward other people who had helped me in the advancement of the project.

References

- [1] Global tuberculosis report 2018. Geneva: World Health Organization; 2018. Licence: CC BY-NC-SA 3.0 IGO.
- [2] Organización Mundial de la Salud, *Laboratory diagnosis of tuberculosis by sputum microscopy*. 2013.
- [3] J. C. S. Ngabonziza *et al.*, "Diagnostic performance of smear microscopy and incremental yield of Xpert in detection of pulmonary tuberculosis in Rwanda," *BMC Infect. Dis.*, vol. 16, no. 1, pp. 1–7, 2016.
- [4] R. Dinesh Jackson Samuel and B. Rajesh Kanna, "Tuberculosis (TB) detection system using deep neural networks," *Neural Comput. Appl.*, vol. 4, pp. 1–13, 2018.
- [5] R. O. Panicker, K. S. Kalmady, J. Rajan, and M. K. Sabu, "Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods," *Biocybern. Biomed. Eng.*, vol. 38, no. 3, pp. 691–699, 2018.