ICIC Express Letters
Volume **17**, Number **3**, March **2023**
 ISSN 1881-803X
pp. **367–374**

ARTEFACT SEGMENTATION IN MICROSCOPY IMAGES FOR AUTOMATIC ACID-FAST BACILLI QUANTIFICATION

Suci Aulia[1], Tati Mengko[1], Andriyan Bayu Suksmono[1]
and Bachti Alisjahbana[2]

[1]School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, West Java 40257, Indonesia
suciaulia@telkomuniversity.ac.id; tmengko@itb.ac.id; suksmono@stei.itb.ac.id

[2]Medical Faculty
Universitas Padjadjaran
Bandung, West Java 40257, Indonesia
b.alisjahbana@unpad.ac.id

Received March 2022; accepted June 2022

Abstract. *As revealed from the Global Tuberculosis (TB) Control Report of 2020, from 2013 to 2019, there was an increase of 69% in TB cases from 331,703 to an expected total of 845,000 cases in Indonesia. Conventional Microscopy (CM) and Fluorescence Microscopy (FM) are the two major sputum smears microscopy techniques used to diagnose tuberculosis. FM is a more costly diagnostic approach in view of the high price of the microscopy machine and its maintenance. For this, CM in developing countries might become a more suitable application. This study presents a new method for Acid-Fast Bacilli (AFB) quantification based on color space RGB channels in conventional microscopy images. Artefact subtraction into host image was proposed to identify the AFB using some thresholding methods. The experimental results have been confirmed by pathologists and specialists from the Medical Faculty of Padjadjaran University. The system performance has succeeded in increasing the average accuracy rate of the reference by 10.56% for a total of 228 test images on the dataset of the Public Health Laboratory in Morelia (PHLM).*
Keywords: Mycobacterium tuberculosis, Acid-fast bacilli, Color space, Microscopy artefact, Sputum smear

1. **Introduction.** The bacteriological test of Ziehl-Neelsen (ZN) stained microscopy is important for Tuberculosis (TB) identification and World Health Organization (WHO) recommends that it should examine 300 view fields to improve its sensitivity [1,2]. Unfortunately, as it is still completely manual, WHO procedure takes a long time to diagnose TB based on the International Union Against Tuberculosis and Lung Disease (IUATLD) grading scale [3-5]. Thus, sputum smear analysis remains the most widely used procedure in developing countries due to its lower cost. Acid-Fast Bacilli (AFB) detection of mycobacterium tuberculosis on sputum samples based on image processing has been studied for over a decade. It progressed from geometric shape-based to colour-based AFB detection [6-8]. Apart from the data gathered, the concern in prior research was the heterogeneity of the data collected. The thickness of the ZN staining, number of artefacts, and sputum smear size all influenced the sputum sample data.

The researchers have widely used ZN Sputum smear Microscopy image Database (ZNSM-iDB) [2] to evaluate multiple approaches for identifying AFB to date. The ZNSM-iDB scanned images from 30 FoVs where 15 FoVs were actually from positive smear samples and others were from the negative ones. The results of AFB detection utilizing

DOI: 10.24507/icicel.17.03.367

watershed included the sensitivity of 93.3% and specificity of 87%. Another research [9] used a dataset from the Laboratory of the Instituto Nacional de Pesquisas da Amazonia (INPA) to detect AFB in two steps. The first stage used Fast Non-Local Means (FNLM) to increase image quality, while the second one used a Convolutional Neural Network (CNN) for AFB classification. The results for all 120 datasets achieved 78.4% precision. The previous research by Díaz-Huerta et al. in 2019 used a different dataset from [2,9,10]. This dataset created 200 images of sputum samples in five image size variations from the Public Health Laboratory in Morelia (PHLM), Mexico. This research proposed K-means and Bayes methods to detect the AFB, resulting in the highest accuracy of 98.66% on the size of 2513×2073 pixels and achieved the average accuracy of 47.22% for all image size variations. According to the result from research before, the PHLM dataset [10] provides the best image quality compared to ZNSM-iDB and INPA-lab datasets. Therefore, in this study, we proposed a method to optimize the average accuracy of AFB detection for all image size variations using the PHLM dataset. This paper proposes a novel approach for AFB detection and quantification based on colour thresholding and object contouring. The MTB had a distinctive colour after given ZN staining on the sputum sample, namely reddish degradation. We use this condition to filter objects based on reddish degradation by filtering the host image on the red channel. From the results of host image filtering on the red channel, many non-AFB objects are detected. Then we detect the contours of each object and reduce each object outside the AFB size, or called artefacts. We call this method artefact segmentation, which has never been done in previous studies and has succeeded in calculating the number of AFB on a large scale from previous studies.

The organization of the paper is as follows. Section 2 describes the material and methods used in this study, including the database resource, the proposed system, and the AFB quantification algorithm. In Section 3, the result and discussion are briefly for each step details including results of color space-channels filtering, color histogram thresholding, object contouring, artefact segmentation and subtraction, and then validating. Section 4 concludes the paper.

2. Material and Methods.

2.1. Data collection and proposed system. The input image in this work uses the PHLM dataset [10], which provided over 200 images from 30 bacilloscopies called the host image. Because MTB had a distinctive colour after being given ZN staining on the sputum sample, namely reddish degradation, the host image was filtered based on the Red, Green, and Blue channels. Then, each layer was segmented based on the histogram colour range for each RGB channel. Several thresholding algorithms, such as Otsu, Mean, and li, were tested to detect the MTB object of each RGB layer and count the number of objects, as shown in Figure 1.

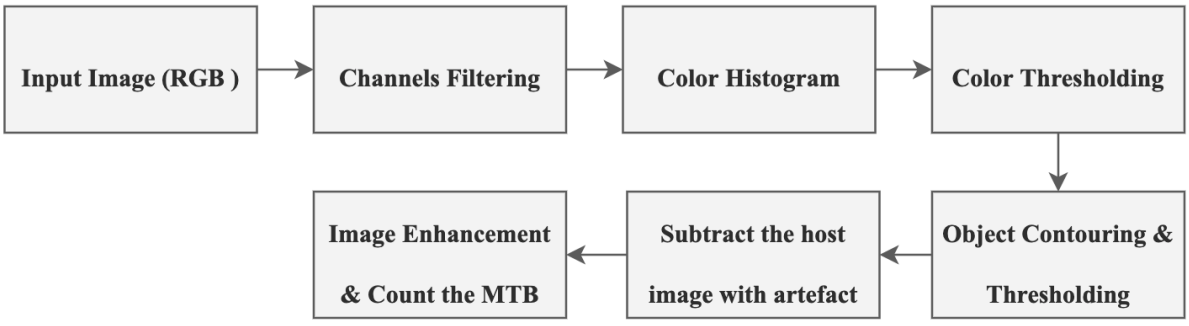


FIGURE 1. Block diagram of AFB detection and quantification system based on artefact segmentation

2.2. **AFB quantification algorithm.** Borderline on every object is a limitation or segmentation between objects, so the number of detected objects and AFBs can be calculated using the object labelling method. The steps starting from AFB objects detection to AFB objects labelling in this study are presented as follows:

Input: Sputum sample images, $\mathbf{I}(\boldsymbol{x}, \boldsymbol{y})$	
Output: Segmented AFB objects, $\mathbf{S}(\boldsymbol{x}, \boldsymbol{y})$	
Step 1	Filter $\mathbf{I}(\boldsymbol{x}, \boldsymbol{y})$ from RGB channels to each channel, then select the best channel that shows the most obvious artefact based on their color
Step 2	Implement the color histogram to set the threshold value ($\boldsymbol{th}$) using Equation (1)
Step 3	$\mathbf{O}(x, y)$: contour of each detected object
Step 4	Set the th of contour region (th_{CR}), then apply to segmenting the artefacts $\mathbf{A}(x, y)$
Step 5	Post-processing on $\mathbf{A}(x, y)$ using Otsu-th, Adaptive-th, Mean-th, and Li-th based on Equation (2) to Equation (5). Then select the best th method
Step 6	$\mathbf{S}(x, y) = \mathbf{O}(x, y) - \mathbf{A}(x, y)$
Step 7	Object contouring on $\mathbf{S}(x, y)$, and give the label for each contoured-object
Step 8	Sum all objects labelled as the number of detected AFB objects

a) *Color histogram method.* $b(x, y)$ is the thresholded image, $I(x, y)$ is the host image, and th is the thresholding value [7,11].

$$b(x, y) = \begin{cases} 1, & \text{if } I(x, y) > th \\ 0, & \text{if } I(x, y) < th \end{cases} \tag{1}$$

b) *Otsu thresholding method.* σ_w^2 is Otsu threshold value, w_0 and w_1 are the weights between two classes segmented by a threshold t. σ_0 and σ_1 are variances of the two classes [12-15].

$$\sigma_w^2(t) = w_0(t)\sigma_0^2(t) + w_1(t)\sigma_1^2(t) \tag{2}$$

c) *Adaptive thresholding method.* I_{xy} is adaptive thresholded image, (x, y) is a pixel point, f_{xy} is sum of all pixel value of an image [16].

$$I_{xy} = f_{xy} + I_{x-1,y} + I_{x,y-1} - I_{x-1,y-1} \tag{3}$$

d) *Mean thresholding method.* Mean thresholding (M) is a local threshold by calculating the mean of the intensity distribution [16] as evaluated using Equation (4):

$$M = \frac{\max(\textit{pixel value}) + \min(\textit{pixel value})}{2} \tag{4}$$

e) *Li thresholding method* ($\theta(t)$) is given in Equation (5), where h_j is the histogram of the region, μ_1 is the intensity constraint of image, N_1 and N_2 are the smallest pixel's value in the two regions, and f_i is the observed image [17].

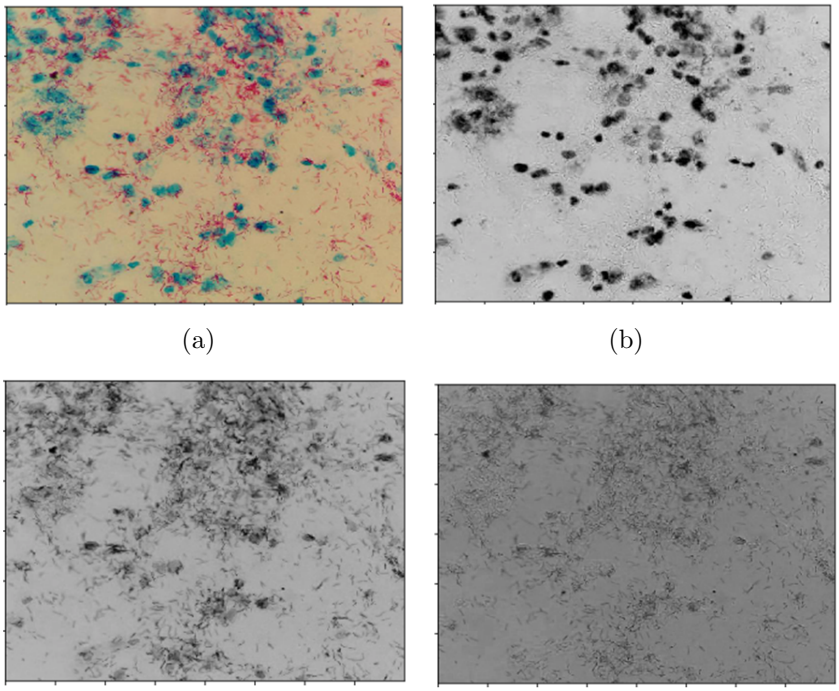
$$\theta(t) = \sum_{j<t} h_j(j - \mu_1(t))^2 + \sum_{j\geq t} h_j(j - \mu_2(t))^2, \quad \mu_1(t) = \frac{\sum_{f_i<t} f_i}{N_1}, \ \ \mu_2(t) = \frac{\sum_{f_i\geq t} f_i}{N_2} \tag{5}$$

f) The accuracy (Acc) of proposed method is evaluated using Equation (6), where C_s is the number of AFB objects counted by proposed algorithm and C_v is the validation number of AFB objects counted by the pathologists [18,19].

$$\%Acc = \left[1 - \frac{|C_s - C_v|}{C_v}\right] \tag{6}$$

3. Results and Discussion.

3.1. **Color space-channels filtering.** In this study, AFB detection and quantifying used the PHLM dataset. As shown in Figure 2, the first step was to filter the host image from the RGB channels to each channel.

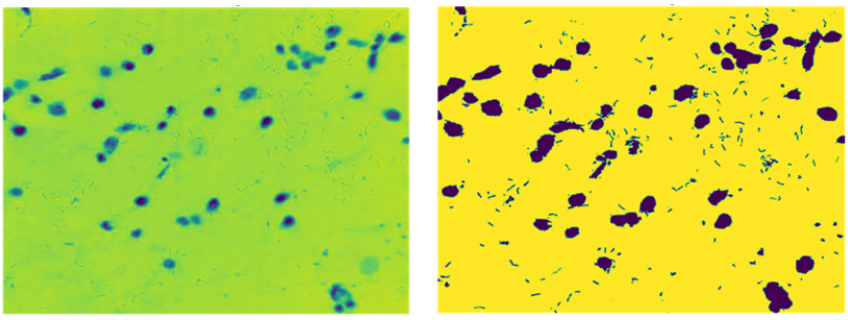


(a) (b)

(c) (d)

FIGURE 2. (a) Original image or host; (b) a-image in red channel; (c) a-image in green channel; (d) a-image in blue channel

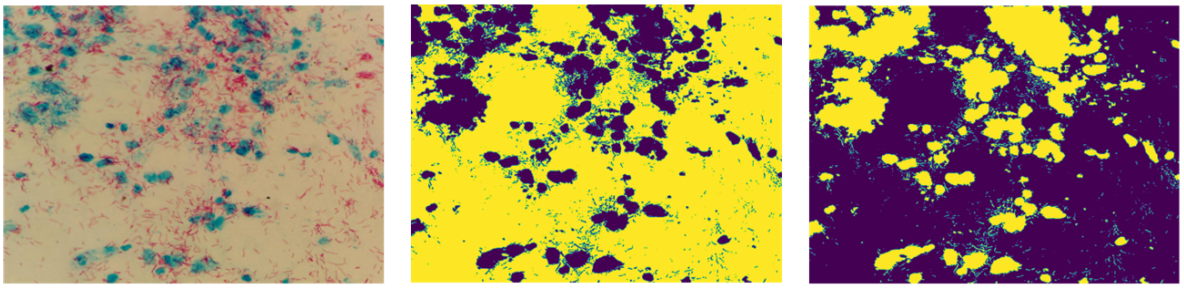
3.2. **Color histogram thresholding.** From Figure 2, the most visible objects in Human System Integration (HSI) are in Figure 2(b) (Red-ch). Thus, the next step to be used to detect AFB objects was Red-ch. This work used a color-threshold method using Equation (1) to separate the sputum sample image objects from the background. The color of the AFB object in the sputum sample image that has been given ZN staining was reddish degradation, while the other objects were relatively blue. The host image on Red-ch was filtered with $th > 150$ as seen in Figure 3.



(a) (b)

FIGURE 3. Red-ch filtering result: (a) Without colour thresholding; (b) with $th > 150$

3.3. **Object contouring.** After obtaining Red-ch from the host image "R_800 600R_imag en3.jpg" on the PHLM dataset, as shown in Figure 3, yellow background and two other objects were in the color purple and green. Then, what was to be done was the process of contouring the object on Red-ch. The goal was to define the features and shape of the detected object. The contoured object results in each object are shown in Figure 4. In Figure 4(c), the background and object colors have been reversed to make it easier visually. Thus, the large objects, namely artefacts, were on yellow color and small objects identified as AFB are on green one.

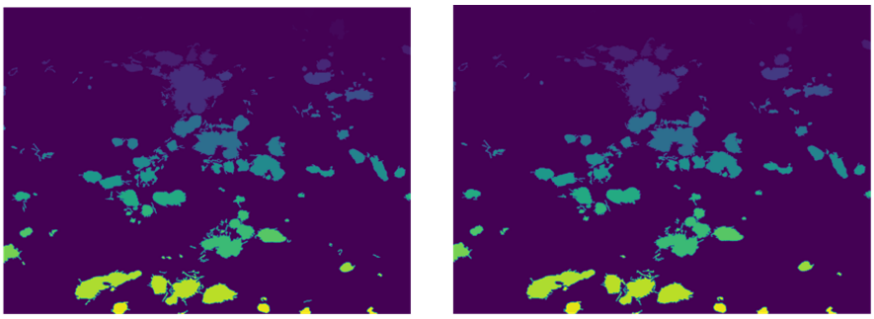


(a) (b) (c)

FIGURE 4. (a) Original image; (b) Red-ch with $th > 150$; (c) the results of the contoured object in the image-b

Borderline on each object is a limitation or segmentation between objects; then the number of detected objects and the number of AFBs can be quantified using the object labelling method. The contour image as shown in Figure 4(c) shows that the artefact object is larger than the AFB. Then, the next step is the segmentation of the artefact object.

3.4. **Artefact segmentation.** The test results of the artefact segmentation process are shown in Figure 5 by taking the threshold of contour region $th_{CR} > 50$ and $th_{CR} > 80$. Based on Figure 5, th_{CR} value was the main parameter that affected the artefact segmentation results. However, the differences in the artefact contour region made the th_{CR} value in this experiment unstable for each image. This limitation is one of the points that must be developed in further research.



$th_{CR} > 50$ $th_{CR} > 80$

FIGURE 5. Effect of th_{CR} value on artefact segmentation

3.5. **Artefact substraction.** If looking at the artefact segmentation results based on the th_{CR} value in Figure 5, the segmentation artefact images had a different colour degradation. Thus, the artefact segmentation image was corrected or smoothed before entering the artefact subtraction process using the thresholding (th) method, i.e., Otsu-th, Adaptive-th, Mean-th, and Li-th as shown in Figure 6. As shown in Figure 6, the most generated

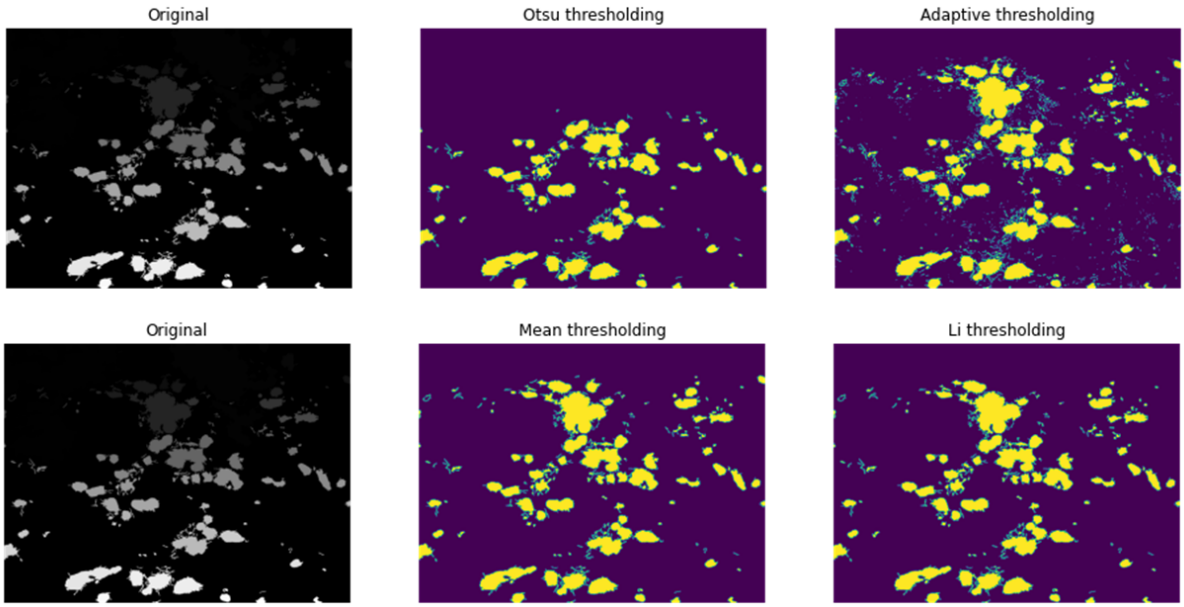
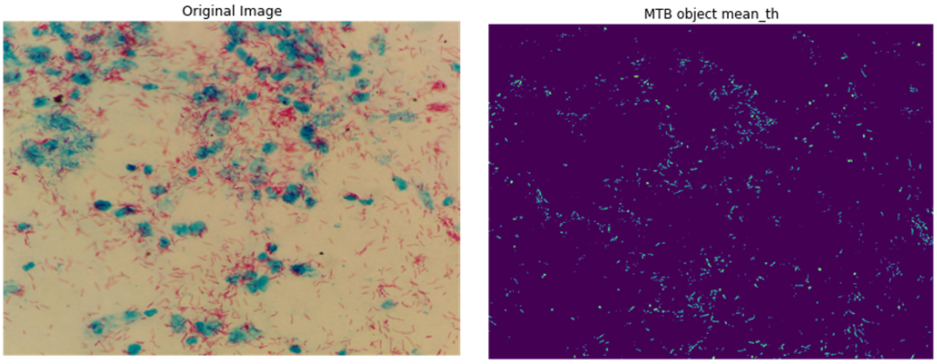


FIGURE 6. Comparison of *th* methods for artefacts segmentation in sputum sample images



(a) (b)

FIGURE 7. (a) Original image; (b) AFB detection results from the artefact segmentation method

artefact segmentation was the Adaptive-*th* and Mean-*th* methods. However, there were still many segmented AFB objects in the Adaptive-*th* method. Thus, the next Mean-*th* method was used for artefact segmentation.

3.6. **AFB detection validation testing using the color space and thresholding algorithm.** The AFB results of MTB objects detection in the sputum sample image are shown in Figure 7(b). Furthermore, the testing results in Table 1 were carried out on 228 test images from the PHLM dataset consisting of image size variations and different background colours. The cited experiment [10] provided relatively low accuracy ($< 75\%$) except for image size 2513×2073 pixels. We present the average accuracy for all image size variations in this experiment in Table 1, and the experimental results have been validated by the pathologist from the Medical Faculty, Padjadjaran University.

From Table 1, it can be seen that [10] and our proposed method achieve an average accuracy of 57.50% and 68.06%, respectively. The proposed method could provide better detection results on image data with low resolution. The use of low resolution could reduce the computational load because the processed file had a smaller size. However, it still needs further research to improve the level of accuracy at low resolutions.

TABLE 1. Testing result of AFB object detection based on image size variations on the PHLM dataset

Image size (pixels)	Resolution (Av.)	Accuracy (%)	
		[10]	Proposed method
2513 × 2073	238 kB	98.66	90.09
1920 × 1080	141 kB	70.56	77.96
1280 × 960	95 kB	59.34	70.21
800 × 600	47 kB	36.49	65.76
640 × 480	33 kB	22.47	36.32
The average		57.50	68.06

4. **Conclusions.** We have developed an automated system for identifying and quantifying the number of Acid-Fast Bacilli (AFB) contained in a ZN-stained sputum sample image. The proposed system consists of two stages; the first stage using a colour segmentation method based on the red channel colour space and the second one using the artefact segmentation and subtraction in the host image to get only AFB objects and this stage was successfully implemented using mean thresholding.

The experimental results have been confirmed by pathologists and specialists from the Medical Faculty of Padjadjaran University. The system performance has succeeded in increasing the average accuracy rate of the reference approach 10.56% for a total of 228 test images on the Public Health Laboratory in Morelia (PHLM) dataset. Each image was segmented within approximately 7 seconds using the method on [10]. Meanwhile, this experiment successfully counted AFB approximately within 4 seconds for each image. However, the artefact contour region differences made the th_{CR} value in this experiment unstable for each image. Hence, this limitation is one of the points that must be developed in further research.

Acknowledgment. The researcher would like to thank ITB, Telkom University, Universitas Padjadjaran and LPDP for their research facilities and funds.

REFERENCES

[1] N. A. Ismail et al., Performance of a novel algorithm using automated digital microscopy for diagnosing tuberculosis, *Am. J. Respir. Crit. Care Med.*, vol.191, no.12, pp.1443-1449, doi: 10.1164/rccm.201502-0390OC, 2015.

[2] M. I. Shah et al., Ziehl-Neelsen sputum smear microscopy image database: A resource to facilitate automated bacilli detection for tuberculosis diagnosis, *J. Med. Imaging*, vol.4, no.2, 027503, doi: 10.1117/1.JMI.4.2.027503, 2017.

[3] D. Zingue, P. Weber, F. Soltani, D. Raoult and M. Drancourt, Automatic microscopic detection of mycobacteria in sputum: A proof-of-concept, *Sci. Rep.*, vol.8, no.1, 11308, doi: 10.1038/s41598-018-29660-8, 2018.

[4] D. Dawson et al., Technical guide: Sputum examination for tuberculosis by direct microscopy in low income countries, *International Union Against Tuberculosis and Lung Disease*, Paris, France, 2000.

[5] L. W. Reza et al., LED-fluorescence microscopy for diagnosis of pulmonary tuberculosis under programmatic conditions in India, *PLoS One*, vol.8, no.10, doi: 10.1371/journal.pone.0075566, 2013.

[6] M. I. Shah, S. Mishra, M. Sarkar and S. K. Sudarshan, Automatic detection and classification of tuberculosis bacilli from camera-enabled smartphone microscopic images, *2016 4th International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp.287-290, doi: 10.1109/PDGC.2016.7913161, 2016.

[7] K. S. Mithra and W. R. Sam Emmanuel, Automated identification of mycobacterium bacillus from sputum images for tuberculosis diagnosis, *Signal, Image Video Process.*, vol.13, no.8, pp.1585-1592, doi: 10.1007/s11760-019-01509-1, 2019.

[8] M. El-Melegy, D. Mohamed, T. ElMelegy and M. Abdelrahman, Identification of tuberculosis bacilli in ZN-stained sputum smear images: A deep learning approach, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp.1131-1137, doi: 10.1109/CVPRW.2019.00147, 2019.

[9] C. F. F. CostaFilho, P. C. Levy, C. M. Xavier, M. G. F. Costa, L. B. M. Fujimoto and J. Salem, Mycobacterium tuberculosis recognition with conventional microscopy, *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp.6263-6268, doi: 10.1109/EMBC.2012.6347426, 2012.

[10] J. L. Díaz-Huerta, A. del C. Téllez-Anguiano, M. Fraga-Aguilar, J. A. Gutiérrez-Gnecchi and S. Arellano-Calderón, Image processing for AFB segmentation in bacilloscopies of pulmonary tuberculosis diagnosis, *PLoS One*, vol.14, no.7, e0218861, doi: 10.1371/journal.pone.0218861, 2019.

[11] C. del Carpio et al., An algorithm for detection of tuberculosis bacilli in Ziehl-Neelsen sputum smear images, *Int. J. Electr. Comput. Eng.*, vol.9, no.4, pp.2968-2981, doi: 10.11591/ijece.v9i4.pp2968-2981, 2019.

[12] A. K. Bhandari and I. V. Kumar, A context sensitive energy thresholding based 3D Otsu function for image segmentation using human learning optimization, *Appl. Soft Comput.*, vol.82, 105570, doi: 10.1016/j.asoc.2019.105570, 2019.

[13] Y. C. Liang and Y. C. Yin, Optimal multilevel thresholding using a hybrid ant colony system, *J. Chinese Inst. Ind. Eng.*, vol.28, no.1, pp.20-33, doi: 10.1080/10170669.2010.531771, 2011.

[14] A. K. Bhandari, N. Singh and I. V. Kumar, Lightning search algorithm-based contextually fused multilevel image segmentation, *Appl. Soft Comput.*, vol.91, 106243, doi: 10.1016/j.asoc.2020.106243, 2020.

[15] X. Bao, H. Jia and C. Lang, Dragonfly algorithm with opposition-based learning for multilevel thresholding color image segmentation, *Symmetry*, vol.11, no.5, doi: 10.3390/sym11050716, 2019.

[16] D. Bradley and G. Roth, Adaptive thresholding using the integral image: Read but to check with sir, *J. Graph. Tools*, vol.12, pp.13-21, 2007.

[17] C. H. Li and C. K. Lee, Minimum cross entropy thresholding, *Pattern Recognit.*, vol.26, no.4, pp.617-625, doi: 10.1016/0031-3203(93)90115-D, 1993.

[18] Y. Zhai, Y. Liu, D. Zhou and S. Liu, Automatic identification of mycobacterium tuberculosis from ZN-stained sputum smear: Algorithm and system design, *2010 IEEE International Conference on Robotics and Biomimetics*, pp.41-46, doi: 10.1109/ROBIO.2010.5723300, 2010.

[19] A. Y. Yousif, S. M. Younis, S. A. Hussein and N. M. G. Al-Saidi, An intelligent computing for diagnosing COVID-19 using available blood tests, *International Journal of Innovative Computing, Information and Control*, vol.18, no.1, pp.57-72, doi: 10.24507/ijicic.18.01.57, 2022.