

AUTOTHRESHOLDING SEGMENTATION FOR TUBERCULOSIS BACTERIA IDENTIFICATION IN THE ZIEHL-NEELSEN SPUTUM SAMPLE

Kusworo Adi *†, Rahmat Gernowo †, Aris Sugiharto* †, Adi Pamungkas †, Ari Bawono †, Nelly Mirnasri †

† Department of Physics, The Faculty of Science and Mathematics, Diponegoro University

*Master of Information Systems, Graduate Program, Diponegoro University

Jalan Prof. H. Soedarto, SH. Tembalang Semarang 50275

email : kusworoadi@undip.ac.id

ABSTRACT

Tuberculosis (TB) is an important public health issue in this world. In 1992, World Health Organization (WHO) has determined tuberculosis as Global Emergency. The caused of tuberculosis disease is *Mycobacterium tuberculosis*. This bacteria is bar shaped and has the quality of acid fast so that it's also called as Acid Fast Bacilli (AFB). In this research, an algorithm to identifying and counting the number of tuberculosis bacteria has developed by microscope imaging. Feature extraction for bacteria shape identification process using three parameter i.e. eccentricities, compactness, and metric. The training and object recognition using perceptron algorithm produce output with high accuracy. In this research obtained a correlation coefficient that is about 0.990

Keywords: neural network, image processing, tuberculosis bacteria, microscopic.

1 INTRODUCTION

TB disease commonly spread through the polluted air by mycobacterium tuberculosis which released when tubercular get a cough. This bacteria is often enter and gather in lungs and proliferate become excessive (especially in person with low body endurance), and capable to spread out through blood vessel or lymph gland so that tuberculosis (TB) became an important public health issue in the world [1,2].

Somebody is said infected TB if TB bacteria reside in his body though it's non active. Oftentimes after TB bacteria enter the body, the immunity will decreasing in controlling the bacteria. This bacteria live in the body for years in non active form. When bacteria is non active so he disease can not be spreaded to the others.

Research in the clinical laboratory field about morphology analysis of TB bacteria in Indonesia is still rated poorly. Analysis which done by the doctors

and laboratory nowadays is still conventionally. The method causing analysis result inter-doctor not always be the same. Physical condition, knowledge, carefulness, and concentration of the doctor is very determine the analysis result because observation is done directly. On the other side, the requirement of facilitation, practicability and accuracy nowadays have been something considered as the needs. The automatic screening will give some excess, such as substantial decreasing in the work load of doctor laboratory staff, testing sensitivity increasing and even better accuracy in diagnosis by increasing amount of image which can be analyzed by computer [3].

Segmentation and classification with hue color component approximation were used to TB bacteria identification. This method is developed by using color saturation thresholding in the Ziehl-Neelsen (ZN) TB bacteria image pixel [4,5]. Research about tuberculosis bacteria identification previously has been done by Siena et al. In the mentioned research, image segmentation process using color extraction manually [6]. This process will be more practical if it can be done automatically.

According to several issue mentioned above, this research is done to detecting and counting the number of tuberculosis bacteria with microscope imaging. Image of tuberculosis bacteria is obtained using a digital microscope. Color segmentation process is done automatically and identification process is done based on its morphology feature extraction. The identification process using neural network algorithm to distinguish between objects which are TB bacteria and non TB bacteria.

2 DESIGN AND IMPLEMENTATION

2.1 Identification System of TB Bacteria

Tuberculosis is a disease caused by infection of a systemic *Mycobacterium tuberculosis* germ, which

can be manifest in almost all of organ with major location in lungs which commonly is the primary infection location. Mycobacterium tuberculosis is bar shaped straight or tortuous. This bacteria has a measurement of width 0,3 – 0,6 mm and length 1-4 mm.

Pulmonary tuberculosis is a serious disease particularly in a baby and little child, malnutrition child, and child with an immunology nuisance. A considerable part of children suffering primary tuberculosis in terder years and much of them are asymptomatic and spontaneously recover without any remnant symptom. In several patient, the disease round into pasca-primary tuberculosis.

The developing process of tuberculosis identification is consist of two stage. The first stage is image processing of sputum infected TB bacteria. Then continued to the second stage that is formation neural perceptron to identify bacteria. Block diagram of the algorithm developing process of tuberculosis identification is shown in Fig. 1.

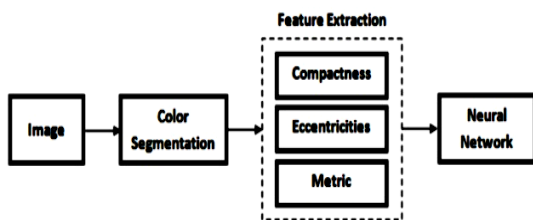


Figure 1. Block diagram of the developing process of tuberculosis identification

2.2 Color Segmentation

The object that used in this research is Ziehl-Neelsen Stain sputum sample of digital microscope image result which obtained from the Centers for Disease Control and Prevention, Public health image library. Atlanta, GA, USA: CDC, 2007 [7]. This image for about 9 samples is used as system testing. And also data set for about 929 TB bacteria shapes that is used as data of neural network training is obtained from [8,9].

Color segmentation is begun with conversion process of image color model which originally RGB (Red, Green, Blue) to HSV (Hue, Saturation, Value). Afterwards Hue component of the input image is extracted and changed to binary through auto-thresholding process using Otsu methods [10]. Then the binary image is reconstructed using closing and opening process.

2.3 Feature Extraction

Feature extraction process is begun with counting the value of eccentricities, compactness, and metric in each data training tuberculosis bacteria (929 of image). Several sample of data training which used in the training process is shown in Fig 2.

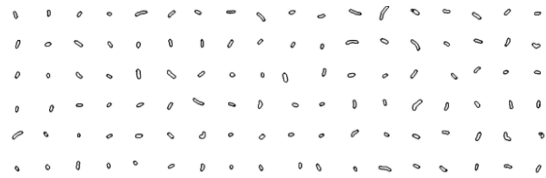


Figure 2. Several sample of data training used in the training process [8,9]

Eccentricities is a ratio between distance of foci ellipse and length of major axis [11,12].

$$e = \sqrt{1 - \left(\frac{b}{a}\right)^2} \quad (1)$$

Where e is the value of eccentricity, a is length of major axis, and b is leng of minor axis. Compactness is a ratio of perimeter square and area of an object.

$$C = \frac{P^2}{A} \quad (2)$$

Where C is the value of compactness, P is perimeter value, and A is the object area. Whereas metric is a parameter that show the roundness level of an object.

$$M = \frac{4\pi \times A}{P^2} \quad (3)$$

This three parameters value is respective saved in format .txt and will be used as input vector in the perceptron training process. Afterwards making output target vector with the same size as input vector which its value is 1. Then executed the perceptron training with input vector and output target vector made before. Result of the training perceptron is saved in format .mat. The result of this training is used to the recognition process of tuberculosis bacteria shapes in the identification process [13,14].

2.3 Neural Network Architecture

Neural network which used in this research using Levenberg-Marquardt backpropagation training function. Number of the used hidden layer is only one layer, with the number of processing elements (nodes) in each layer is as follow [13,14]:

- In the input layer, used three parameters, that is compactness, eccentricities, and metric.
- Number of element in the hidden layer is used as research variable. Test is done with range value of 5 to 50 in each range the change is 5.
- Output layer has 1 (one) element with range value of 0 to 1.
- Activation function which using sigmoid binary in the hidden layer and identities function in the output layer.

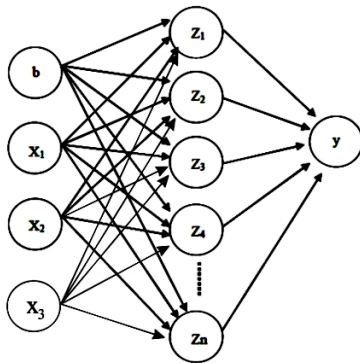


Figure 3. Model of backpropagation neural network which used in the research, with 3 input elements, n element in the hidden layer, and 1 output element.

3 RESULT

Segmentation is begun with conversion process of color model image which originally RGB to HSV. Afterwards, Hue component of the image is extracted. One of the processed image and Hue component image is shown by Figure 4 (a) and (b).

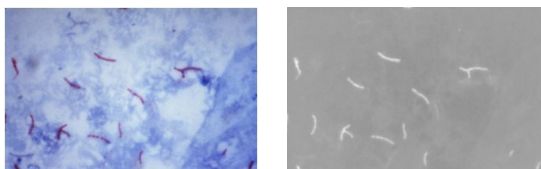


Figure 4. (a) Original image (b) Hue component image

The next process is convert Hue component image to binary through a thresholding process using Otsu method. The obtained binary image then be reconstructed using closing and opening operation. The reconstructed image is shown in Figure 5 (a).

The value of compactness, eccentricities, and metric of each object in this image is counted and these three values is became the testing data in the neural network testing. If the output value of the network is larger equal to 0.9 or smaller equal to 1.04, then the object is identified as tuberculosis bacteria. When the output value beyond the bounds of the value limit, then the object is identified as non tuberculosis bacteria. The identified image is shown in Figure 5 (b).

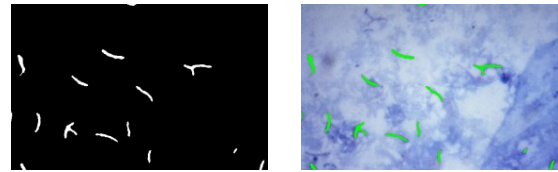


Figure 5. (a) The reconstructed image (b) The identified image

Figure 5. (a) The reconstructed image (b) The identified image

Number of the bacteria in the identified image is counted using labeling methods. The number of each piled object is counted as two, the counting result using labeling methods toward Figure 5 (b) show that the object number is 17. This number is in compliance with the result of counting manually.

Image segmentation which developed in this research is not good enough when applied to sputum image with high concentration. This is because object and background in the image has a close resemblance contrast. Sample image with high concentration is shown in Figure 6 (a) and its segmentation result is shown in Figure 6 (b).

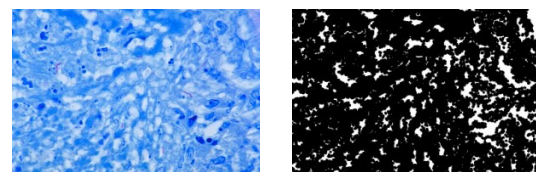


Figure 6. (a) Sputum image with high concentration (b) The segmented image

Figure 6. (a) Sputum image with high concentration (b) The segmented image

Ratio of the developed system counting result and manually counting is shown in Table 1.

Table 1. Ratio of the developed system counting result and manually counting.

No.	Counting	
	System	Manual
1	17	17
2	9	8
3	8	9
4	7	7
5	1	1
6	1	1
7	1	1

Correlation graphics between result of the developed system counting and manually counting is shown in Figure 7.

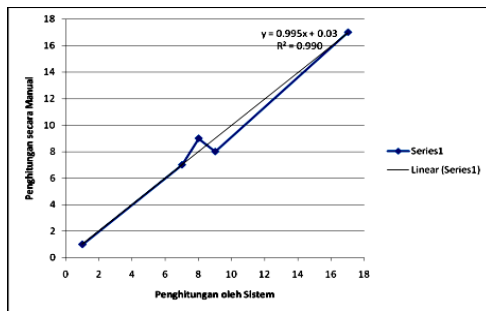


Figure 7. Correlation graphics between result of the developed system counting and manually counting

According to the graphics in Figure 7, obtained a correlation coefficient that is about 0.990.

4 CONCLUSION

According to the research, can be concluded that: The designed system is capable of detecting and counting the number of tuberculosis bacteria with microscope imaging. Feature extraction to identify bacteria shape using three parameter that is eccentricities, compactness, and metric. Training and object recognition using perceptron algorithm produce output with high accuracy. Coefficient correlation between the developed system counting and manually counting is about 0.990. This is show that Neural Network is good enough to be applied in detecting and counting the number of tuberculosis bacteria.

ACKNOWLEDGEMENT

This research was funding from the Ministry of Research and Technology Republic of Indonesia through the Insentif SINAS Program in 2012.

The authors would like to acknowledge PIMOD group for TB bacteria dataset for training neural networks.

The authors would like to acknowledge Centers for disease control and prevention (cdc) public health image library (phil) for TB bacteria image ZN .

REFERENCE

- [1]. Avicenna, 2011, " Lung tuberculosis (Lung TB)" (in English), <http://rajawana.com/artikel/kesehatan/264-tuberculosis-paru-tb-paru.html> Last access 13 April 2011 14.59.
- [2]. WHO Report 2009, 2009, *Global Tuberculosis Control 2009: Epidemiology, Strategy, Financing*, WHO Press, World Health Organization, ISBN 978 92 4 156380 2, Geneva, Switzerland.
- [3]. Forero, M. G.; Cristobal, G. and Borrego, J. A., 2003, *Automatic Identification Techniques of Tuberculosis Bacteria*, SPIE Proceedings Of The Applications Of Digital Image Processing XXVI, Vol.5203, pp. 71-81, ISBN 0-8194-5076-6, San Diego, CA, Aug. 2003, SPIE, Bellingham WA.
- [4]. Vishnu Makkapati and Ravindra Agrawal, Raviraja Acharya: *Segmentation and classification of tuberculosis bacilli from ZN-stained sputum smear images*. CASE 2009: 217-220
- [5]. Osman M.K., Mashor M.Y., Saad Z., and Jaafar H., 2010, *Colour Image Segmentation of Tuberculosis Bacilli in Ziehl-Neelsen-Stained Tissue Images Using Moving K-Mean Clustering Procedure*, 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, pp.215-220
- [6]. Siena, I., Adi, K., Gernowo, R., Mirnasari, N., 2012, *Development of Algorithm Tuberculosis. Bacteria Identification Using Color. Segmentation and Neural Networks*, IJVIPNS-IJENS, Volume 12, Issue 4, August 2012
- [7]. Centers for Disease Control and Prevention, Public health image library. Atlanta, GA, USA: CDC, 2007 <http://phil.cdc.gov/phil/home.asp>.

- [8]. Forero, M. G.; Sroubek, F. and Cristobal, G., 2004, *Identification of Tuberculosis Bacteria Based on Shape and Color*. Real-Time Imaging, Vol. 10, No. 4, (August 2004), pp. 251-262, ISSN 1077-2014.
- [9]. Forero, M. G. Cristóbal and Desco, M., *Automatic identification of Mycobacterium tuberculosis by Gaussian Mixture models*, J. Microscopy, 223, pp.
- [10]. N. Otsu. *A Threshold Selection Method from Gray-level Histogram*. IEEE Transactions on Systems and Cybernetics, 1979, Vol SMC-9 No. 1.
- [11]. R.C. Gonzalez and R.E. Woods, Digital Image Processing Second Edition. *New Jersey; Pearson Prentice Hall. 2004.*
- [12]. P. Heidenreich, L.A. Cirillo., and A.M. Zoubir, *Morphological Image Processing for FM Source Detection and Localization*. ScienceDirect, 2008, Signal Processing 89 (2009) 1070–1080
- [13]. M.H. Beale, M.T. Hagan, and H.B. Demuth, *Neural Network Toolbox™ User's Guide R2012b. USA;The MathWork Inc., 2012.*
- [14]. Bishop, C.M., 1995, *Neural Networks for Pattern Recognition*, *Oxford: Oxford University Press. ISBN 0-19-853849-9 (hardback)*

[This page is intentionally left blank]