Automatic identification of *Mycobacterium tuberculosis* in Acid-Fast stain sputum smears with image processing and Neural Networks

S. P. Fang[1,]*, Hui-Jine Hsu[2], Li-Lin Hung[2], and Yue-Sheng Wu[1]

[1]*Department of Electronic Engineering, Southern Taiwan University, Tainan 710, TAIWAN*
[2]*Department of Pathology, Chi Mei Medical Center, Tainan 710, TAIWAN*
[*]*spfang@mail.stut.edu.tw*

Abstract: Tuberculosis is a serious global infectious disease and caused 1.3 million deaths in 2008. This paper employs digital image processing technology and Neural Networks to automatically identify *Mycobacterium tuberculosis* in Acid-Fast stain sputum smears. We propose six parameters according to the color and shapes of the bacilli, and obtain 87.2% sensitivity and 80.0% specificity without Neural Networks and 97.8% sensitivity and 72.4% specificity with Neural Networks.

1. Introduction

Tuberculosis (TB) is a highly infectious disease that caused 1.3 million deaths worldwide and 762 deaths in Taiwan in 2008 [1, 2]. Its threat is getting severe due to the symptom complexity, combination with HIV, and drug resistance of *Mycobacterium tuberculosis*. Accurate identification of the TB bacillus is the first step and key point to stop and control the disease. Medical laboratories identify the TB bacillus with two routines, culturing and sputum smear microscopic examining. The TB bacillus growth is pretty slow and culturing the bacillus normally needs about three months. Microscopic Examining sputum smear can give results in a short time and has two methods: Fluorescence and Acid-Fast (Ziehl-Neelsen) stain.

Forero et al. [3-5] and Veropoulos et al. [6,7] presented automatic identification techniques of the TB bacillus in fluorescence microscopic images. Preparing fluorescence microscopic slides is more expensive than preparing Acid-Fast stain slides. It might be an obstracle in developing countries, which need immediate measures to control tuberculosis. Furthermore, the fluorescence method is constrained by its high false positive because some other bacteria appear to be fluorescent.

Examining Acid-Fast stain slides is a time consuming and exhausting task because medical technologists should check 300 different microscopic views in a single slide, resulting in high false negative due to quick review. However, little work has been done on automated identification for Acid-Fast stain. Sadaphal et al. presented the automated identification by examining axis-length ratio, eccentricity, and area [8]. This paper proposes more parameters and employs Neural Networks to automatically identify the TB bacillus in Acid-Fast stain sputum smears.

2. Methods

Mycobacterium tuberculosis is red and rod-like in Acid-Fast stains. We extract the red figures from microscopic images taken by an Olympus DP70 digital camera with 4080×3072 pixels and $1,000 \times$ magnification, as shown in Fig. 1. The white figures of Fig. 2, a binary format image, are the original red ones of Fig.1. They pass through the color examining and are defined as candidate figures of TB bacilli for further examining.

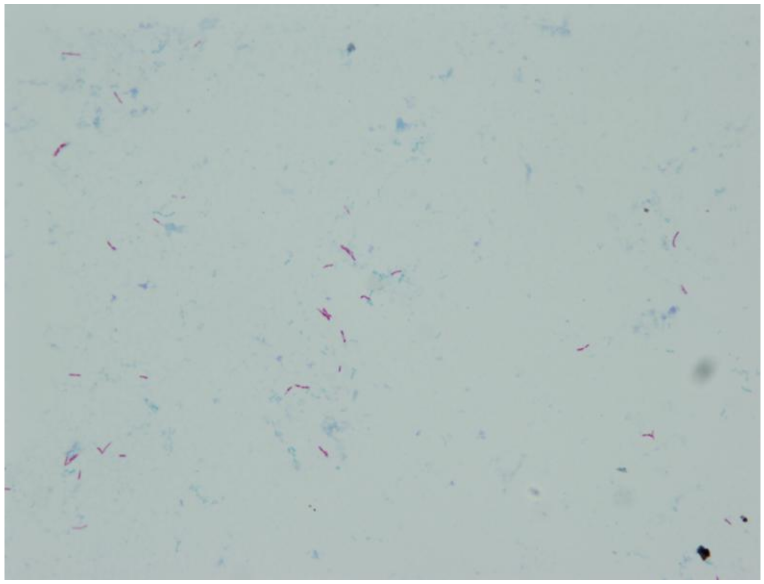

Fig. 1. An Acid-Fast stain sputum smear microscopic image. The red rod-like figures are the TB bacilli.

Fig. 2. The binary image of Fig. 1 processed by red figures extracted from background. The white figures are candidate figures of the TB bacilli.

Fig. 3 shows the identification flow chart of the candidate figures. The TB bacilli are three dimensional non-rigid objects and their shpaes are not unique after being processed on two dimensional slides. Some red spots reside in slides that are not prepared well, as shown in Fig. 4. Moreover, the bacteria in Fig. 4 are not the TB bacilli, although they have similar shapes. All of the above tell us that the candidate figures should be checked carefully.

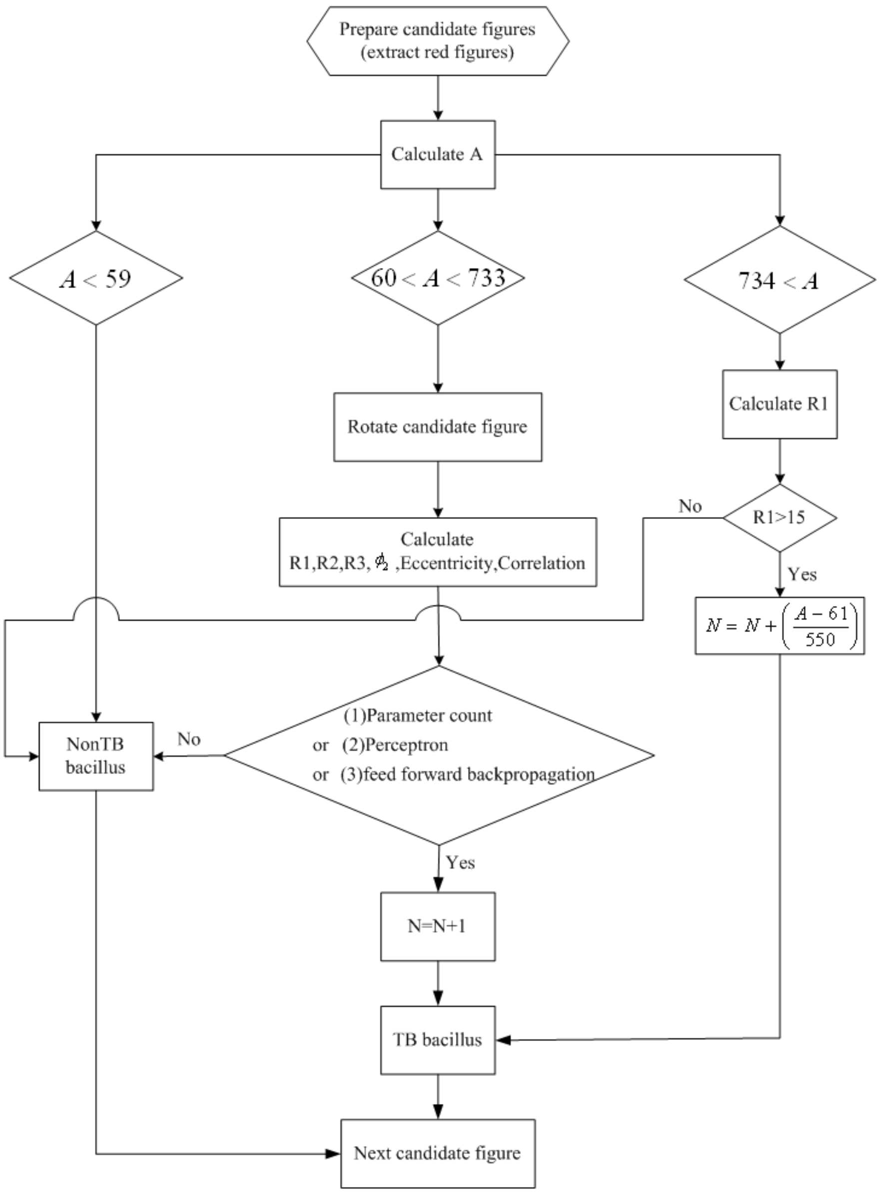

Fig. 3. The identification flow chart of candidate figures.

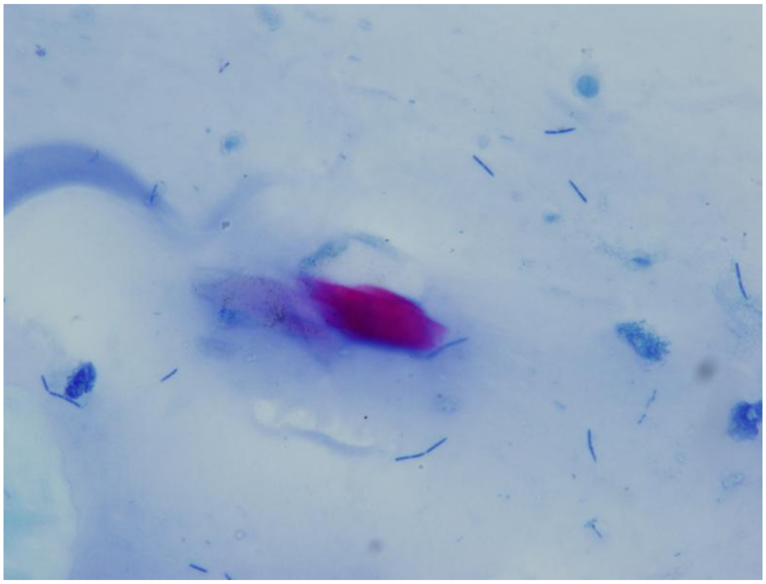

Fig. 4. The red spot is a residue of Acid-Fast stain. The bacteria here are not the TB bacilli, although they are rod-like.

We took 44 Acid-Fast stain microscopic images. Experienced medical technologist identified the images and found 553 TB bacilli. The mean area for an isolated bacillus is 683 pixels and its standard deviation is 652 pixels. The candidate figures are examined by their areas first. The criterion of the area for an isolated TB bacillus is between 60 and 733 pixels. Figures with pixels smaller than 60 will not be considered as TB bacilli, and figures with pixels larger than 734 pixels will be checked by (perimeter)^2/area to be classified into Acid-Fast stain residues or TB bacilli. Large area TB bacillus figures are estimated their numbers of TB bacilli by (A-61)/550, the 61 pixels is the overlaping area of two TB bacilli.

Most candidate figures of sputum smears are isolated. We propose six parameters to identify TB bacilli. They are R1= (perimeter)^2/area, R2 = long axis/short axis, R3 = μ_{02}/μ_{20}, rotational moment invariant ϕ_2 [9], eccentricity, and correlation with a 14 $\times$ 49 pixels rectangle, where

$$\phi_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \quad (1)$$

$$\mu_{pq} = \iint (x - x_c)^p (y - y_c)^q f(x, y) dxdy, \quad (2)$$

(x_c, y_c) is the center coordinate of the candidate figure, and $f(x,y)=1$ if (x,y) is the figure and $f(x,y)=0$ if (x,y) is background. The 14 $\times$ 49 pixels rectangle might not be a standard TB bacillus microscopic figure, but it is a good similarity. The candidate figures should be rotated to the same direction as the rectangle before processing correlation. Table 1 shows the mean values and standard deviations of the six parameters.

Table 1. The mean values and standard deviations of the six parameters of 553 TB bacilli.

	R1	R2	R3	ϕ_2	eccentricity	correlation
mean	28.9	4.84	27.6	1.83×10^{10}	0.954	0.599
standard deviation	11.5	2.02	21.7	3.91×10^{10}	0.0670	0.170

We studied the identification with these parameters in three different processes: 1. parameter count, 2. Perceptron, and 3. feed-forward backpropogation scheme of Neural Network.

The criterions of parameter count are selected according to the rod-like shape and distinction with circle-like Acid-Fast stain resides. Table 2 shows the criterions of the six parameters of the TB bacilli.

Table 2. The criterions of the six parameters of TB bacilli.

	R1	R2	R3	ϕ_2	eccentricity	correlation
lower limit	5.12	0.585	0	1.00×10^{5}	0.839	0.243
upper limit	52.1	8.71	70.2	1.75×10^{11}	1.00	1.00

With the parameter count method, we classify the candidate figures into TB bacilli as they satisfy the parameter count number. For example, in the count≧2 colummn of Table 3, the figures are identified as TB bacilli if two or more parameters of the six parameters are in the criterions. Identification accuracy is defined by sensitivity=TP/(TP+FN) and specificity=TN/(TN+FP), where TP is true positive, FN is false negative, TN is true negative, and FP is false positive.

Table 3. The results of the parameter counts.

	count=6	count≧5	count≧4	count≧3	count≧2
TB bacilli/Non TB bacilli	553/254	553/254	553/254	553/254	553/254
TP	482	538	547	550	553
FN	71	15	6	3	0
TN	203	132	102	101	101
FP	51	122	152	153	153
sensitivity	87.2%	97.3%	98.9%	99.5%	100 %
specificity	80.0%	52.0%	40.2%	39.8%	39.8%

We employ Matlab Neural Network Toolbox to implement the Perceptron and the feed-forward backpropagation scheme of Neural Network. The Perceptron has six neurons and its accuracy is 100% sensitivity and 39.8% specificity. The feed-forward backpropagation has one hidden layer of eight neurons and one output layer of four neurons. The result is 97.8% sensitivity and 72.4% specificity. Table 4 is the results of the Perceptron and the feed-forward backpropagation Neural Network.

Table 4. The results of the Perceptron and the feed-forward backpropagation Neural Network.

	Perceptron	feed-forward backpropagation
TB bacilli /Non TB bacilli	553/254	553/254
TP	553	541
FN	0	12
TN	101	184
FP	153	70
sensitivity	100%	97.8 %
specificity	39.8%	72.4 %

Center for Disease Control of the U.S. suggests that one Acid-Fast slide needs to be checked 300 views with 1,000× magnification, each observing time is 3 sec [10]. The mean elapse times for parameter count, Perceptron, and feed-forward backprogation are 0.705 sec, 0.687 sec, and 0.691 sec respectively (CPU is Intel Core i5 750, and Operation System is Windows XP).

3. Discussion and future work

We have developed automated identification techniques for *Mycobacterium tuberculosis* in Acid-Fast stain sputum smears. With parameter count method, we find large count number decisions, strict decision conditions, result in low sensitivity and high specificity. Medical examination reports give “negative” and “positive” with valence, instead of the precise number TB bacilli. This situation might accept high sensitivity at the expense of low specificity. Our Neural Network schemes did not seem to help the accuracy too much. Future work could focus development on robust identification techniques for various preparing skills from different medical technologists.

4. Acknowledgement

The project was sponsored by Chi Mei Medical Center under the Contract No. 130980190.

References and links

1. http://www.who.int/mediacenter/factsheets/fs104/en/index.htm.
2. http://www.cdc.gov.tw/ct.asp?xItem=27032&ctNode=1947&mp=1.
3. M. G. Forero, F. Sroubek, and G. Cristobal, “Identification of tuberculosis bacteria based on shape and color,” in *Real-Time Imaging*, **10**, Issue 4, 251-262 (2004).

4. M. G. Forero, G. Cristobal, and J. Alvarez-Borrego, “Automatic identification techniques of tuberculosis bacteria,” SPIE **5203**, 71-81 (2003).
5. M. G. Forero, G. Cristobal, and M. Desco, “Automatic identification of *Mycobacterium tuberculosis* by Gaussian mixture models,” in *Journal of Microscopy*, **233**, Issue 2, 120-132 (2006).
6. K. Veropoulos, C. Campbell, and G. Learmonth, “Image processing and neural computing used in the diagnosis of tuberculosis,” in *Intelligent Methods in Healthcare and Medical Applications* (Digest No. 1998/514), IEE Colloquium on, 8/1-8/4 (1998).
7. K. Veropoulos, C. Campbell, G. Learmonth, B. Knight, and J. Simpson, “The Automated identification of Tubercle Bacilli using image processing and Neural computing techniques,” in *Proceedings of the 8th International Conference on Artificial Neural Networks*, vol. 2, (Skovde, Sweden, 1998), pp. 797-802.
8. P. Sadaphal, J. Rao, G. W. Comstock, and M. F. Beg, “Image processing techniques for identifying Mycobacterium tuberculosis in Ziehl-Neelsen stains,” in *The International Journal of Tuberculosis and Lung Disease*, **12**, No. 5, 579-582 (2008).
9. Y. S. Abu-Mostafa and D. Psaltis, "Recognitive Aspects of Moment Invariants," in *IEEE Transactions of Pattern Analysis and Machine Intelligence*, **PAMI-6**, No. 6, 698-706 (1984).
10. R. W. Smithwick, *Laboratory Manual for Acid-Fast Microscopy*, 2nd, Ed., (U. S. Department of Health, Education, and Welfare, Atlanta, GA, 1976).