

A computer algorithm for detection of tuberculosis bacilli in Ziehl Neelsen sputum smear images based on the adjustment of RGB primary component tones and geometric eccentricity

Christian del Carpio¹, Erwin Dianderas¹, Mirko Zimick², Patricia Sheen², Jorge Coronel³, Patricia Fuentes³, Guillermo Kemper¹

¹School of Electrical and Electronic Engineering - Universidad Nacional de Ingeniería, ²Laboratory of Bioinformatics and Molecular Biology - Faculty of Sciences- Universidad Peruana Cayetano Heredia, ³Tuberculosis Laboratory - Universidad Peruana Cayetano Heredia

{cdelcarpiod, erwin.dianderas, guillermo.kemper, mzimic, patricia.sheen, pfuentesbonilla} @gmail.com

Abstract — The present study proposes a method of automatic detection of tuberculosis (TB) bacilli from digital images of Ziehl Neelsen sputum smear bacilloscopy. The method is based on an algorithm that aims to automate the interpretation of optical microscopic images of sputum smears. According to the World Health Organization (WHO), a specialist can not analyze and process more than 20 samples per day (in order to not affect the analysis sensitivity and commit errors in diagnosis). Therefore, an automated tool as the proposed here, is an important contribution to the current efforts to fight tuberculosis. The algorithm is based on geometric eccentricity of ellipses and improvement of RGB component tones. Correspondence functions adjusted to sample preparation conditions were applied in order to improve the RGB primary component tones of the image. This allows to obtain an adequate segmentation of interest objects. For the recognition of each object as bacillus, the geometric descriptor of eccentricity of the ellipse was applied. The algorithm was validated with 66 independent sputum samples from TB patients. A sensitivity of 88.75% and a specificity of 95.5% was obtained for the diluted pellet method for sample preparation.

Keywords — tuberculosis, diagnostics, sputum smear, bacilloscopy, Ziehl-Neelsen, image processing, algorithm, digital image.

I. INTRODUCTION

One of the bacteria that still generates a high morbidity and mortality worldwide is *Mycobacterium tuberculosis*, which causes tuberculosis (TB) and leads to the patient's death if it is not treated timely [1].

In 2014, according to the Peruvian Ministry of Health (MINSA), more than 1.5 million people were detected as symptomatic respiratory, which first symptom of TB including cough for more than 15 days. Of this total, about 20 thousand cases of tuberculosis were diagnosed [2].

Sputum smear bacilloscopy is the most used test for diagnosis of tuberculosis. This test is performed by technical personnel who is responsible for the sample processing, reading, bacilli count present in the microscope image, and interpretation [3].

On the advice of the World Health Organization (WHO), a technician should not analyze more than 20 samples per day [3], in order to ensure and appropriate reliability of the diagnosis. This limitation and the fact that the number of respiratory symptomatic patients increases considerably annually, generates

a demand to develop computational digital image processing algorithms that allow automated visual analysis of samples and generate a reliable report of bacilli concentration.

Previous studies have proposed different algorithms of image processing oriented to detection, counting and recognition of *M. tuberculosis* bacilli.

HSV color modelling was used to improve bacilli segmentation and classification, although the sensitivity obtained is not indicated [4]. Contrast improvement of a Ziehl-Neelsen (ZN) stained tissue image using the histogram equalization method [5], improved the visualization and recognition of bacilli present in the image. This method was applied to tissue images but not for ZN smear. The level of sensitivity of the method is not specified. Bayesian segmentation based on color and discard of artifacts by shape comparison improved the identification of "possible tuberculous objects" and [6]. The level of sensitivity obtained is not specified either. A detection algorithm based on Gauss classifiers showed a sensitivity and specificity above 95% for the classification of tuberculosis samples [7].

As can be appreciated, most methods propose the improvement of segmentation or detection processes. However, there are few that estimates the sensitivity and specificity levels achieved. Likewise, studies are only performed on images obtained by the standard direct preparation method to process the sputum samples.

In the present work we propose an algorithm for bacilli detection and recognition based on adjustment of RGB primary components tonalities and the geometric eccentricity of the ellipse. The algorithm was designed and adjusted to obtain the best results from the three different microbiological methods of sample preparation: (1) Direct, (2) pellet and (3) diluted pellet. Besides the performance of bacilli detection and recognition, computational load generated by the algorithm was also taken into account, in order to avoid excessive hardware requirements. At the end, 66 sputum samples from TB patients were evaluated comparing the results of the algorithm with those obtained by visual inspection of specialists.

II. PROPOSED METHOD

The proposed method flowchart includes six steps (Figure 1). The images of the ZN sputum smears prepared by the specific methods (direct, diluted pellet or pellet) were digitized by a 5 Mpixels digital camera installed in an optical microscope.



Fig 1. Proposed method flowchart

The format of the acquired images was RGB (Red-Green-Blue).

A. Sample Preparation

Three types of ZN sputum smear procedures were used. (1) The standard direct smear microscopy, which consists of staining the sputum samples on a glass slide (Figure 2) without any further treatment (see Figure 3). [3]

(2) The concentrated sputum smear microscopy or pellet is the staining of the sputum samples treated for decontamination and concentration (Figure 4).

(3) The diluted sputum smear microscopy or diluted pellet is the staining of the samples from the concentrated sample that was further diluted in order to reduce the formation of conglomerates (Figure 5).

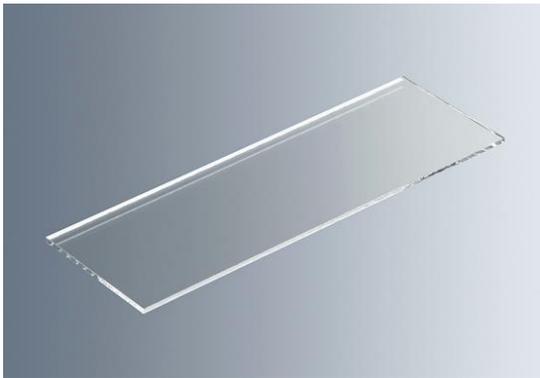


Fig 2. Glass slide foil

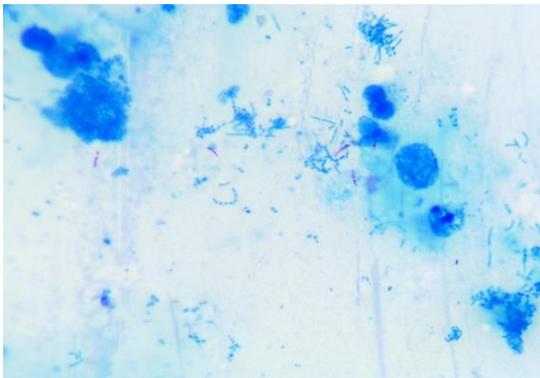


Fig 3. ZN bacilloscopy by the direct method

B. Images acquisition

Every sample for each of the three methods of preparation was digitized using a 5 Mpixels digital camera (MiniVid). The images were acquired in RGB color model (radiometric resolution of 8 bits per band) with a spatial resolution of 2592 x 1944 pixels and stored in format .bmp (Bitmap). The microscope where the camera is placed is Zeiss brand model Primo Star and a total magnification of 1000X is used.

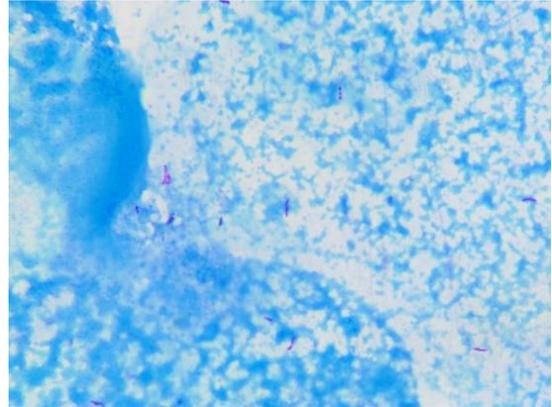


Fig 4. Image of smear microscopy ZN by pellet method

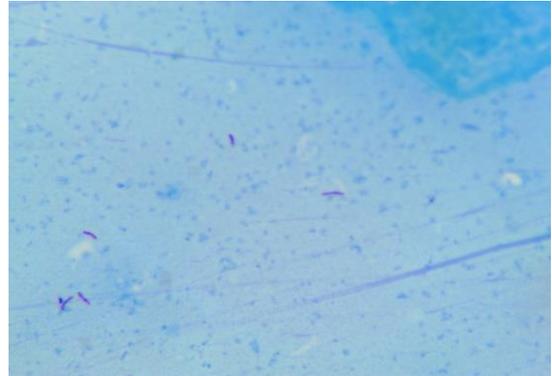


Fig 5. Image of smear microscopy ZN by diluted pellet method

C. Image enhancement

The first step is to separate the three primary color components from the image. That is, the component “R” expressed as $I_R(x, y)$, the component “G” expressed as $I_G(x, y)$ and the component “B” expressed as $I_B(x, y)$. Notably, the most relevant information referring to bacillus was found in the components “R” and “G” (Figure 6).

The transformation functions shown in equations Eq. 1, Eq. 2 and Eq. 3, were used to highlight the bacilli present in the image (Figure 7). As can be seen, component “B” is given the value of 255 to maximize the blue background by obtaining “enhanced B” $IM_B(x, y)$. In the “enhanced R” $IM_R(x, y)$ and “enhanced G” $IM_G(x, y)$ components, the bacilli can be visualized, but also various unwanted non-bacilli objects.

$$IM_R(x, y) = \begin{cases} 1.25 \times I_R(x, y), & 0 \leq I_R(x, y) \leq 204 \\ 255, & \text{otherwise} \end{cases} \quad \text{Eq.(1)}$$

$$IM_G(x, y) = \begin{cases} 0 & 0 \leq I_G(x, y) < 76 \\ 2.5 \times I_G(x, y) - 190 & 76 \leq I_G(x, y) \leq 178 \\ 255 & \text{otherwise} \end{cases} \quad \text{Eq.(2)}$$

$$IM_B(x, y) = 255 \quad \text{Eq.(3)}$$

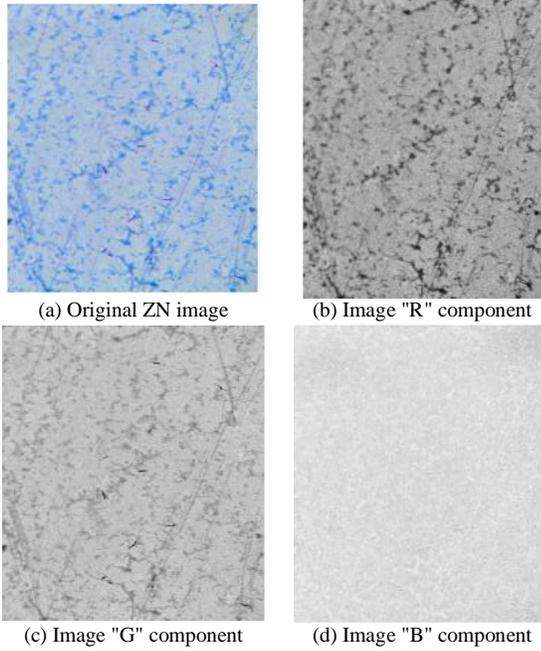


Fig 6. Components of the image with ZN staining

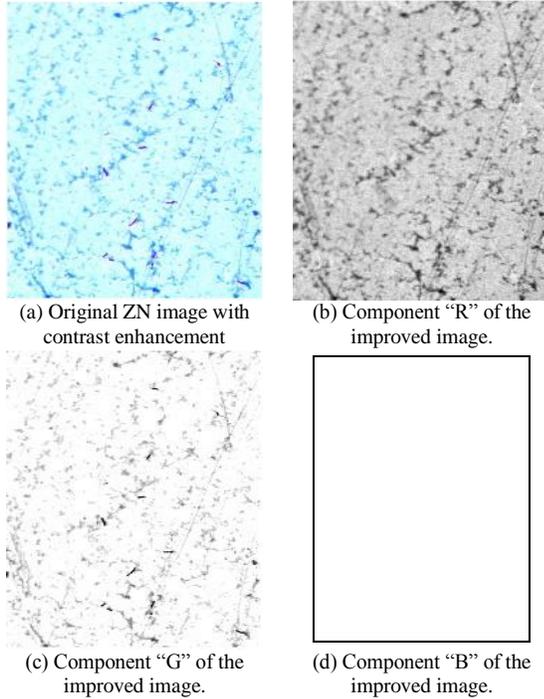


Fig 7. Components of the improved image

D. Segmentation

In order to extract the objects of interest, the improved components “R” and “G” are subtracted (Eq. 4). The result of this operation is shown in Figure 8.

$$IR(x, y) = IM_R(x, y) - IM_G(x, y) \quad \text{Eq.(4)}$$

In order to improve the visualization of all the objects present in the image, the dynamic range is reduced (Eq. 5) on the image $IR(x, y)$. The result is shown in Figure 9

$$IR_1(x, y) = \log(1 + IR(x, y)) \quad \text{Eq.(5)}$$



Fig 8. Image obtained by the difference of $IM_R(x, y)$ and $IM_B(x, y)$



Fig 9. Image $IR_1(x, y)$

E. Labeling

As can be seen, it has been possible to segment the image obtaining objects of interest and some unwanted artifacts. The obtained image is then submitted to a thresholding process (Eq. 6). The threshold used was fixed and was obtained from multiple tests performed. The result of the thresholding is the image $IB_1(x, y)$ (Figure 10).

$$IB_1(x, y) = \begin{cases} 0, & 0 \leq IR_1(x, y) < 5 \\ 255, & \text{otherwise} \end{cases} \quad \text{Eq.(6)}$$



Fig 10. Binary Image $IB_1(x, y)$

Then the labeling algorithm is applied to the binary image $IB_1(x, y)$ in order to remove labels that contain less than 50 pixels (in order to eliminate objects that by their size do not qualify to be bacilli). The image obtained from this process is $IB_2(x, y)$ (Figure 11).



Fig 11. Binary Image $IB_2(x, y)$

In order to provide solidity to the obtained objects, the morphological process of closure is applied with the structural elements shown in Fig. 12. This process is performed one after the other as indicated in equations Eq. 7, Eq. 8, Eq. 9 and Eq. 10.

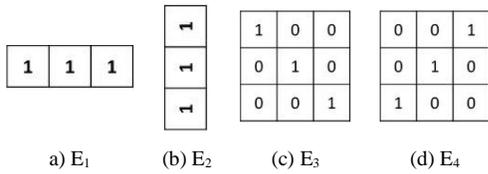


Fig 12. Structural Elements

$$IB_3 = IB_2 \bullet E_1 \quad \text{Eq.(7)}$$

$$IB_4 = IB_3 \bullet E_2 \quad \text{Eq.(8)}$$

$$IB_5 = IB_4 \bullet E_3 \quad \text{Eq.(9)}$$

$$IB_6 = IB_5 \bullet E_4 \quad \text{Eq.(10)}$$

Where IB_2 , IB_3 , IB_4 , IB_5 and IB_6 are the matrices that represent the images: $IB_2(x, y)$, $IB_3(x, y)$, $IB_4(x, y)$, $IB_5(x, y)$ and $IB_6(x, y)$ respectively. In addition, E_1 , E_2 , E_3 , and E_4 correspond to the matrix representation of the structuring elements shown in Figure 12. Finally, the symbol “ \bullet ” indicates the operation of the closing morphological process [8].

The resulting binary image $IB_6(x, y)$ is shown in Figure 13.



Fig 13. Binary Image $IB_6(x, y)$

F. Calculation of the eccentricity

The labeling algorithm was applied to the image $IB_6(x, y)$, obtaining all the candidate objects what will later be classified as if they are bacilli or not. The image of the tagged objects is shown in Figure 14. For each of these objects we

calculated the eccentricity value of an ellipse [8] using Eq. 11. The result of this process is shown in Figure 15.

$$e = \sqrt{1 - \left(\frac{b}{a}\right)^2} \quad \text{Eq.(11)}$$

Where “ e ” is the eccentricity value, “ a ” is the length of major axis, and “ b ” is the length of minor axis.

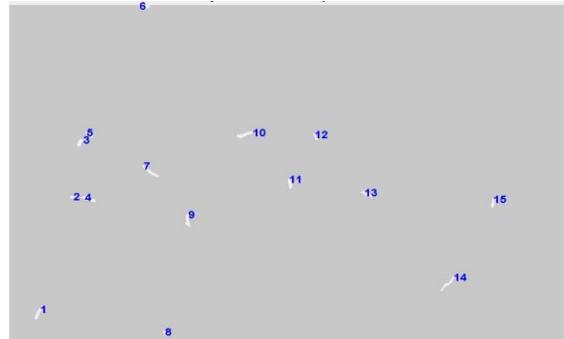


Fig 14. Image of all labeled objects

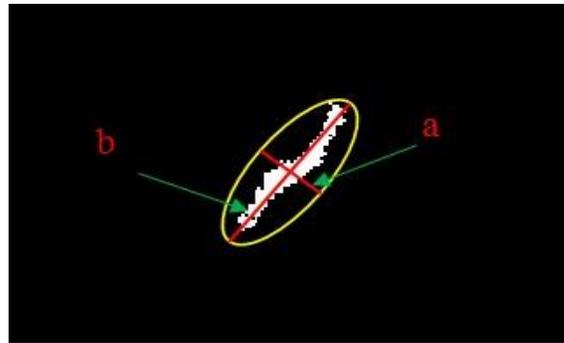


Fig 15. Calculation of the eccentricity of an object.

G. Detected bacils

The classification of a bacillus requires that the eccentricity “ e ” must be greater than 0.9. Objects with an eccentricity greater than 0.9 are shown in blue, and objects that do not satisfy this condition are shown in yellow (Figure 16).

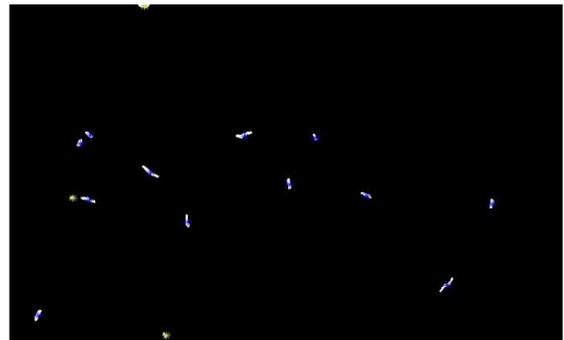


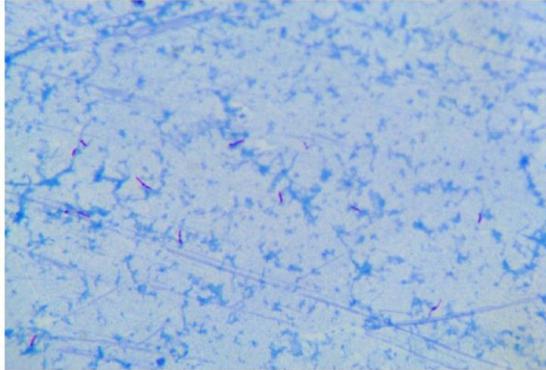
Fig 16. Image showing objects that meet eccentricity greater than 0.9

Finally, objects that do not comply with the measurement are eliminated (Figure 17).

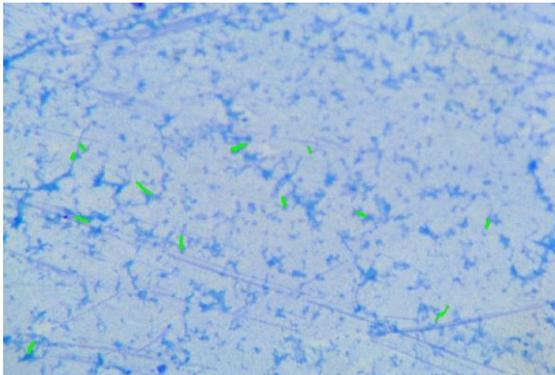
In order to enhance the visualization, the object that were detected as bacillus were overlapped with green color in the original image (Figure 18).



Fig 17. Final Image $IB(x,y)$



(a) Original Image



(b) Original image overlaid with objects detected as bacilli

Fig 18. Application of the algorithm

III. RESULTS

Sixty six images of ZN stained sputum smears were analyzed by the three methods of sample preparation, direct, pellet and diluted pellet that were obtained from 13 different patients.

Of the 66 images, 3 were negatives (without presence of bacillus) and 63 positive (with presence of at least one tuberculosis bacilli). The results of sensitivity and specificity per image are shown in Figure 19. Sensitivity values were obtained from the Eq. 12 and specificity values were obtained from the Eq. 13. The negatives results corresponds only to one patient.

$$sensitivity = \frac{VP}{VP + FN} \times 100 \quad \text{Eq.(12)}$$

$$specificity = \frac{VN}{VN + FP} \times 100 \quad \text{Eq.(13)}$$

Where “VP” means “True Positive” and are the objects that the algorithm and the specialist mark as bacilli; “VN” means “True negative” and are the objects that both (algorithm and the specialist) mark as NO bacilli; “FP” means “False Positive” and are the objects that detect the algorithm as bacilli and that are NO bacilli for the specialist; “FN” means “False Negative” and are the objects that the algorithm determined that aren't bacillus, but the specialist says that they are bacillus.

Two technicians from the Tuberculosis Laboratory of the Universidad Peruana Cayetano Heredia participated blindly in the validation of the proposed algorithm, both with extensive experience in the preparation, analysis and diagnosis of TB.

The specialists independently evaluated all the images that were submitted, without knowing the results of the software. Sensitivity and specificity were then calculated by each specialist and the average of both values was obtained (Figure 19).

The best results were obtained for the diluted pellet method. This is consistent since the treatment of the samples by this method generates cleaner images in contrast to the Pellet method and to a greater extent in contrast to the direct method.

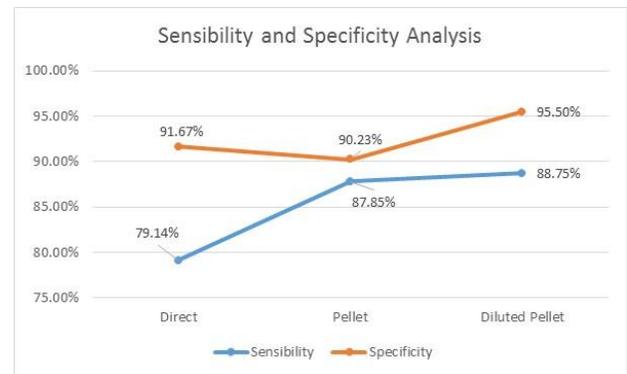


Fig 19. Sensibility and Specificity Analysis

IV. CONCLUSIONS

A sensitivity higher than 85% is achieved when the preparation methods are pellet and diluted pellet. The algorithm was able to correctly identify TB objects that are truly bacilli with 85% of acertivity (Figure 19).

Interestingly regardless what is the method of preparation of the sample, a specificity greater than 90% was achieve. That means, that the developed algorithm successfully rejects objects that are not considered bacilli by the specialist in more than 90%.

The sensitivity will indicate how well the software is detecting an object in the image as a bacillus. Greater sensitivity means that the software is more like the detection of bacilli by the specialist. Likewise, the specificity will indicate how well the software is discarding an object in the image as a possible bacillus.

Something really important in the detection of bacilli is the initial segmentation of all objects; including the total or the greater number of objects candidate to bacilli and in smaller

quantity or nullity, objects that are not bacilli. The proposed algorithm achieves this.

For the moment, the algorithm focuses on detecting single bacilli; which have an elongated shape. Because of this, the only descriptor that is being used and indicating whether or not it is bacillus, is the eccentricity of the object's ellipse. Therefore, if there were bacillus superimposed perpendicularly, or very close to them, these as a whole would have an eccentricity less than 0.9 and would not be considered as bacillus.

The improvements to be made for better validation in the detection of bacilli are the application of more parameters such as geometric descriptors, photometric measurements. Using these measures can also be used a Bayesian classifier (statistical classifier) for a better recognition.

ACKNOWLEDGMENT

The present work was carried out thanks to the research funds of the “Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica (FONDECYT)”, an initiative of the “Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica (CONCYTEC)”.

REFERENCES

- [1] CENTROS PARA EL CONTROL Y LA PREVENCIÓN DE ENFERMEDADES (2015), “**Enfermedad de tuberculosis (TB): Síntomas y factores de riesgo**” (<http://www.cdc.gov/spanish/especialescdc/sintomastuberculosis/>)
- [2] MINISTERIO DE SALUD (2016), “**Análisis de la situación epidemiológica de la tuberculosis en el Perú 2015**” Perú. (<http://www.dge.gob.pe/portal/docs/tools/tbc/asistbc.pdf>)
- [3] ORGANIZACIÓN PANAMERICANA DE LA SALUD (2009), “**Manual para el diagnóstico bacteriológico de la tuberculosis**”.
- [4] Makkapati V, Agrawal R, Acharya R (2009) **Segmentation and Classification of Tuberculosis Bacilli from ZN-stained Sputum Smear Images**. 5th Annual IEEE Conference on Automation Science and Engineering Bangalore.
- [5] Osman M.K, Mashor M.Y, Saad Z, Jaafar H (2009) **Contrast Enhancement for Ziehl-Neelsen Tissue Slide Images using Linear Stretching and Histogram Equalization Technique**. IEEE Symposium on Industrial Electronics and Applications (ISIEA 2009), October 4-6, 2009
- [6] Sadaphal P, Rao J, Comstock GW, Beg MF (2008) **Image processing techniques for identifying Mycobacterium tuberculosis in Ziehl-Neelsen stains**. Int J Tuberc Lung Dis 12: 579–582.
- [7] Khutlang R, Krishnan S, Dendere R, Whitelaw A, Veropoulos K, et al. (2010) **Classification of Mycobacterium tuberculosis in images of ZN-stained sputum smears**. IEEE Trans Inf Technol Biomed 14: 949–957.
- [8] Gonzalez, R. C, Woods, R. E. (2002), **Digital Image Processing Second Edition**, Pearson Education, ISBN 978-0201180756, New Jersey