

**FINAL REPORT FORM**

Principal Investigator: Benjamin Castaneda, Ph.D.

Title: Associate Professor & Director of the Medical Image Research Lab

Institution of Affiliation: Pontificia Universidad Catolica del Peru

Address: Av. Universitaria 1801, San Miguel, Lima, Peru

Telephone: Office (+51)16262000 ext. 4720 Cell (+51)1999096966  
Fax (+51) 16262854 / (+51) 16262087

E-Mail Address: castaneda.b@pucp.edu.pe

Title of Project:

Developing Computerized Screening for Early Detection of Tuberculosis in Peru

Project Dates: July 1<sup>st</sup> 2009 – September 30<sup>th</sup> 2010

Approved Funding: \$ 10,540.00

Monies Spent: \$ 10,568.15

Monies Due the Fdtn: \$ 0.00

Report Due Date: November 15<sup>th</sup>, 2010

A PDF copy of this report was sent that has my approval to be posted to the Lindbergh Web site as appropriate  yes  no



Signature of P.I.: \_\_\_\_\_

Date Submitted: November 15<sup>th</sup>, 2010

08/09

## 2. Summary

The spreading of Tuberculosis (TB) worldwide is worrisome. According to the World Health Organization, it killed an estimated 1.1 - 1.7 million people in 2008. During the same year, Peruvian authorities reported as many new cases per 100,000 inhabitants as were reported in India and various African countries. To control this epidemic, it is necessary to improve its diagnosis. Currently, the most useful diagnostic tool is the direct examination of smears of sputum which is a time consuming and observer dependent procedure. Thus, we developed a software algorithm to analyze microscopic images of smears of sputum for detection of Mycobacteria responsible for the disease. We defined a new staining protocol, and acquired and annotated 24,900 images following our acquisition protocol. Initial results over a small subset of the image database reported a performance of 94% in sensitivity, 76% in specificity, and 84% in accuracy. A more refined version of this algorithm will offer a faster and more accurate diagnosis by screening more fields in less time. This will improve the productivity of health workers who may spend more time attending patients. Also, clinicians may offer earlier diagnosis even in rural areas, leveling-off limitations based on expertise, time and equipment. Patients will obtain faster diagnosis and better monitoring of treatment. This work benefits the general population by improving the quality of their living environment while reducing the chances of transmission.

### 3. Discussion

#### A. Problem

Despite efforts to stop the spreading of TB worldwide, it killed an estimated 1.1 to 1.7 million people and there were approximately 8.9 to 9.9 million incident cases in 2008. In Peru, the same year, there were approximately 100 to 299 new cases per 100,000 inhabitants. The best strategy to stop the epidemic of TB is to isolate and treat infectious patients before they spread the disease. The most common and transmissible form of TB is pulmonary TB. Its diagnosis in Peru and most of the world is based upon the Directly Observed Therapy Short-course strategy. Thus, patients with clinical signs or those who have been exposed to the disease are asked to provide two samples of sputum. Smears of sputum are stained and microscopically screened for AFB. The reading quality depends upon a technician who in a 6-hour working day has to read up to twenty-four smears. In areas of highest incidence like the laboratories of large hospitals or referral centers, this tedious process slows down the efficiency of the strategy and it is dependent on the experience and capability of the technician.

Based on this assay, patients with the infection will be classified as smear-positive. They will receive subsidized treatment and enter integral programs that will include the monitoring of their families. When a sample is identified as negative but the patient is infectious (smear-negative, culture-positive), he will not receive proper treatment and will keep spreading the disease. In smear-positive patients, other samples of sputum must be requested after some weeks to follow up treatment responses. In smear-negative patients, extra-samples are examined to confirm initial negative readings. Even if the number of patients is low, the number of samples to analyze can be elevated. Therefore, laboratories of referral hospitals cannot meet the demands for TB screening. Furthermore, the quality and quantity of the tests could be compromised since technicians cannot dedicate enough time to properly examine each sample.

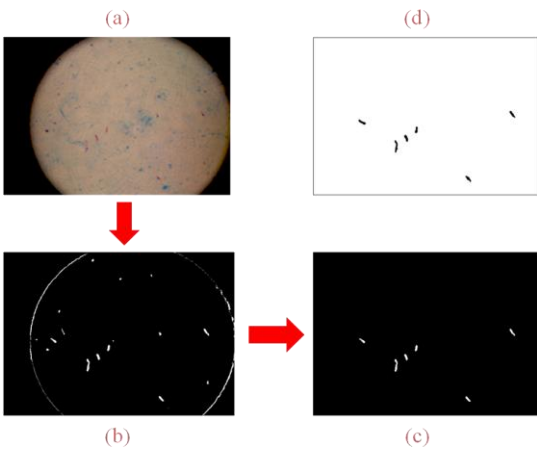
#### B. Solution

Limited resources for the training of laboratory staff, supervision, and quality control of reagents and procedures are challenges currently faced in Peru and other countries with low Human Development Index. We proposed to automate this task by creating an annotated library of images and using machine learning techniques to analyze digital images of the microscopic fields of the samples. This project aimed to increase the number of smears observed per day while providing an accurate diagnosis. By diagnosing more patients and better monitoring the effectiveness of treatment, the quality of the human environment would improve, especially in the poorest places of the world. Also, by having prompt and adequate diagnosis, better programs for public health management could be executed.

In collaboration with the National Hospital “Dos de Mayo” in Lima, Peru, we conducted the following tasks:

##### 1) Initial proof of concept of the whole procedure

In order to have preliminary data, we performed an initial test in which 300 images were collected by manually adjusting a 7-megapixel digital camera (FE-270, Olympus Inc.) into one of the eye pieces of a binocular microscope (CX41, Olympus Inc.). Using a PC, each image was processed based on color information, edge detection and size to establish objects of interest. Some of these objects were filtered further based in heuristics including size, eccentricity and color. The remaining objects were classified as “bacillus” or “not a bacillus” using a machine learning technique (Support Vector Machines) based on the mathematical representation of their shape. These results were compared to manually segmented images which were considered as ground truth. These images were annotated by a technician with 15 years of experience. Figure 1 shows a summary of the image processing techniques applied to a microscopic image.



**Figure 1.** Summary of the image processing applied to a single microscopic image. The original image (a) is pre-processed following a combination of edge detection, morphological color thresholding and area/size based filtering. The resulting image (b) contains several objects which are candidates for bacilli. Each one of these candidates is analyzed further by a machine learning algorithm which compares shape characteristics of the candidates against a database. The output is an image which only contains the objects classified as bacilli (c). The resulting image is compared to the manual segmentation of an expert technician (d) to assess the performance of the algorithm.

Results were encouraging (see next Section) and suggested that a new protocol for staining the smears may reduce the number of objects to analyze (and therefore, false positives), and improve the performance of image processing algorithms while keeping the diagnostic value.

## 2) Analysis of the diagnostic value of the addition of methylene blue as the last step in Ziehl–Neelsen (ZN) staining

In this task, we compared the outcome of TB diagnosis with and without the addition of methylene blue. Images from the slides of 8 patients with TB were acquired at the Hospital Nacional Dos de Mayo. Each slide was processed following the ZN staining but without adding methylene blue. 100 images per patient were taken at this point. Subsequently, the slide was rinsed, methylene blue was added and 100 images more were acquired. An expert technician, with 15 years of experience, annotated both sets of images. TB diagnosis (-,+,++,+++) and number of bacilli were compared for both cases. Results suggested that a correct diagnosis may be obtained without the use of methylene blue in the staining process.

## 3) Definition a new acquisition protocol

A new protocol for acquisition of images was designed. Besides the important difference in the staining, the protocol established precisely the settings of the microscope Olympus CX-41 (amount of illumination, magnification, lens) and the digital camera Olympus E-510 10 MP (exposure time, picture format, ISO number). Three images with different focus depth were taken per field of view in order to compensate for any out of focus distortion. 200 fields were acquired per patient.

## 4) Acquisition and annotation of microscopic images database

Following the new protocol, a database of approximately 25000 digital images was created. Each image was annotated by an expert technician.

## 5) Development of Image Processing Algorithms

The processing of the digital images still followed the scheme presented in (1) (see Figure 1), however, learning from the initial experience, the algorithms for detection and counting of bacilli were improved. Specifically, new mathematical representations for color (YIQ color space) and shape (Hu, Fourier, Zernike and own-derived descriptors) were introduced.

## 6) Assessment of image quality.

It is important to assure the quality of the images to be processed in order to avoid analyzing stained smears which have been poorly prepared. We proposed a color-based quality index. This metric compares the average color inside a bacillus with its surrounding area. When the colors are very similar, it implies that the staining process was not performed properly.

## 7) Evaluation of Performance

By comparing the results of the algorithm and the annotation performed by the expert technician, performance was evaluated in terms of accuracy, sensitivity and specificity.

## C. Results

The following results were achieved:

- 1) An **annotated database of images** compose of:
  - a) 300 microscopic digital images following ZN staining
  - b) 6000 microscopic digital images for 10 patients: Half of the images followed ZN staining and the other half followed ZN staining without methylene blue.
  - c) 18600 microscopic digital images with the new protocol derived for this project. It contains 600 images for 31 patients. According to the Ministry of Health of Peru, a patient can be diagnosed as: (-), no bacillus detected in 100 fields; (P), 1 – 9 bacilli detected in 100 fields; (+), 10 – 99 bacilli detected in 100 fields; (++) , 1 – 10 bacilli detected for one field in 50 fields; and (+++), more than 10 bacilli detected for one field in 20 fields. The 31 patients we chose were diagnosed as: 6 (-), 4 (P), 7 (+), 7 (++) and 7 (+++). We will continue to increase the number of patients for this database until images for 35 patients, 7 per diagnosis, are acquired.  
To the best of our knowledge, this is the largest database of this type.
- 2) A **stand-alone software** has been developed in Matlab. Initial results show 93.6% sensitivity, 75.7% specificity, and 83.7% accuracy. Although these results are better than others reported in the literature (75.6% in accuracy, 76.0% in sensitivity and 73.8% in specificity), we have not achieved our goal of obtaining over 98% in specificity. However, there are still more techniques to evaluate in order to improve the algorithm's performance. It is important to mention, that these results were obtained using only a fraction of the database and that a complete test has yet to be done. In terms of reducing the time of analysis, the software can test one sample in 20 minutes. Although, this time seemed considerably long, we expect that it would be reduced 10 times (~2 min) when the software will be migrated from Matlab to C++.
- 3) A **user-guide** for the stand-alone software in Spanish. Currently the software is very simple and it is thought to be used by an engineer. Therefore, it is not ready to be used in the field. Thus, when the software is re-written to be used in a clinical setting, a new software guide should be re-written too.
- 4) **Reports.** Documentation indicating the performance of the algorithms has been included as parts of several Theses to obtain the degree of Electrical Engineer by students at PUCP.
- 5) **Theses.** The following theses are under preparation:
  - a) “Detección de bacilos de tuberculosis en muestras de esputo utilizando procesamiento de imágenes,” David Kanashiro (expected next March).

- b) “Análisis de características de forma del bacilo de Koch para detección automática de TB en imágenes digitales,” Javier Ticona (expected next February).
- c) “Detección de bacilos de tuberculosis presentes en muestras de esputo por medio descriptores de forma,” Gustavo Aguilar (expected next December).
- d) “Detección de Tuberculosis en prueba de esputo utilizando procesamiento de imágenes,” Gerardo Valladares (expected next February).

## 6) Publications.

Two abstracts were presented at the International Conference of the Union against Tuberculosis and Lung Disease, in Berlin, November 2010. One abstract was presented at the IEEE Pan American Health Care Exchanges International Conference in Lima, March 2010. Please see the section of Publications in this report for more information.

Although overall results are encouraging, further improvements are necessary in order to propose a new clinical protocol based on the automatic counting of bacilli.

## D. Application

A faster and more accurate TB diagnosis will be possible since more fields will be screened in less time. This will improve the chances of successful treatment. Besides, analyzing more sputum samples will allow clinicians to do better monitoring. Also, they will be able to spend more time in other activities like culturing samples for further validation. Finally, it can reduce the risk of transmission in health centers. Laboratory staff of rural medical centers will be benefit too. Since they are responsible for laboratory examinations of all types of samples, the time saved by automated diagnosis will improve their overall performance. Besides, if the staff does not have the expertise, the time or the equipment to analyze sputum samples, a digital camera and a computer with access to internet can assist the staff to do early diagnosis of TB. Finally, faster diagnosis will improve the quality of the whole environment, especially in the poorest places of the world, by reducing the chances of TB transmission.

Specifically, this project has made the following important contributions:

- 1) A large annotated database has been implemented which will be used to improve the state of the art on image processing algorithms to count bacilli.
- 2) An algorithm has been implemented into a stand-alone software with encouraging results. However, its clinical implementation requires a better performance in terms of specificity and time.
- 3) Awareness has been raised in the Peruvian (biomedical) engineering community about the important problem of TB. In addition, the engineering community has made contact with organizations dedicated to fight TB. This is an important step since a multidisciplinary approach is needed to solve this problem.

## E. Future

This project has opened a series of important questions and challenges that need to be address in future work:

- 1) Validation of the new staining procedure. The database will be tested to analyze the clinical diagnosis value of the new staining protocol in a larger scale.
- 2) Segmentation of the database. The database has been annotated, however, to use it to improve the image processing algorithms; each bacilli should be segmented manually.
- 3) Test of new algorithms to improve accuracy, sensibility, specificity and speed to count bacilli.
- 4) Comparison of the inter- and intra- observer variability between results from image processing algorithm results and human observers.

- 5) Implementation of low cost hardware for tuberculosis. Hardware should be built to automatically prepare the smears and to control the microscope.
- 6) Educational implications. Once a final software is developed and evaluated in a clinical setting, it would be also important to evaluate its educational potential.

The results of the present or related-work are being used as preliminary data for a grant proposal to be presented to the Peruvian National Council of Science and Technology. The new proposal will address a new staining method, completion of the database and stand-alone software, and the creation of a device for automatic sample elaboration.

#### F. Other

It is important to mention that the research is currently on its way to build a low cost hardware to control the microscope funded by the 2010 Annual Research Award from PUCP. This is a required complement of the Lindbergh Foundation Grant, since it aims to build the hardware which will interact with the software algorithm to count bacilli automatically. Without one or the other, the project would be incomplete. Figure 2 shows two hardware prototypes that have been implemented.

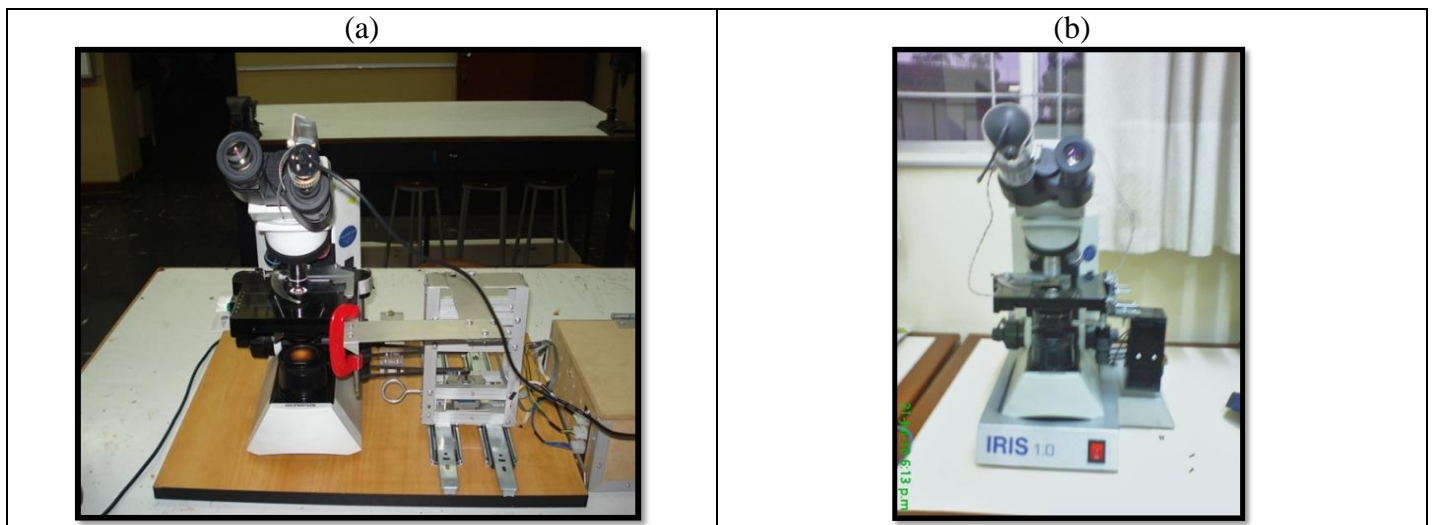


Figure 2. First (a) and second (b) prototypes built at PUCP to control the microscope through a computer.

4. **Personnel (one page total):**

A. In a few sentences, briefly describe the work done on the project by all personnel listed in the original application. Calculate or estimate the number of hours worked on the project for each.

- Benjamin Castaneda (500h). Directed the Project and coordinated the image processing group.
- Luz Huaroto (500h). Led the acquisition of images and oversaw the whole project.
- Fanny Casado (200h). Coordinated the Biomedical Group.
- Agustín Julio Tanta (200h). Expert technician in bacilloscopy. Responsible for annotating the images.
- Kristians Diaz (1800h). Worked on the image processing algorithms
- Andrea Palma (240h). Worked on the acquisition and annotation of microscopic images (Intern)
- Gerardo Valladares (240h). Worked on the machine learning algorithms. (Intern)
- Víctor David Gavilán (1500h.). Worked on the acquisition and annotation of microscopic images (Intern)
- David Kanashiro (200h). Worked on the acquisition and annotation of microscopic images (Intern)
- Javier Ticona (200h). Worked on the acquisition and annotation of microscopic images and machine learning algorithms.
- Gustavo Aguilar (600h). Worked on the machine learning algorithms.

B. Indicate the current work and/or future plans of all personnel now that the work funded by the Lindbergh grant has been completed.

- *Benjamin Castaneda*. Director of the Medical Imaging Research Laboratory at PUCP. He will continue to research in methods for diagnosis of TB and other diseases (Cancer, Leishmaniasis).
- *Luz Huaroto*. Director of the Laboratory for Mycobacteria at Hospital Nacional Dos de Mayo. She will continue to work and improve the diagnosis methods for TB in the hospital.
- *Fanny Casado*. About to finish her Ph.D. in Environmental Medicine. She is currently looking for a job and research opportunities. There is a good possibility that she will join PUCP.
- *Agustín Julio Tanta*. Continue to work at the Laboratory for Mycobacteria at Dos de Mayo hospital.
- *Kristians Diaz*. He entered the Master program on (Medical) Image Processing at PUCP. Afterwards, he will pursue a Ph.D. degree overseas.
- *Andrea Palma*. She joined an airline company as a service engineer.
- *Gerardo Valladares*. He will apply to the Master Program on Image Processing at PUCP.
- *Víctor David Gavilán*. He entered the Master program on (Medical) Image Processing at PUCP. Afterwards, he will pursue a Ph.D. degree overseas.
- *David Kanashiro*. He joined a company in tele-communications
- *Javier Ticona*. Worked on the acquisition and annotation of microscopic images
- *Gustavo Aguilar*. Currently looking for a job and research opportunities.

5. **Equipment:** Include a description of all equipment purchased (\$250+), purchase price, useful life, and suggested disposition.
- Laptop (E105-S1602, Satellite, Toshiba). Price: 864 USD, 4 years (2009-2012), Continue to be used for research to improve diagnosis of Tuberculosis at the Medical Image Research Lab at PUCP.
6. **Funding:** State new sources, time frame, and amount of funding received to continue this project or related work. List any funds awarded following notification of your selection as a Lindbergh Grant Recipient until the present time.
- Annual Research Award at PUCP (2010-2011) – 10K USD
7. **Accounting:**
- A. Complete the Financial Report Form attached. See file attached
  - B. Copies of all approved requests for changes in the budget. See file attached
  - C. Submit copies of all receipts/vouchers for any single items whose cost is over \$250. See file attached
  - D. Please feel free to include a narrative to accompany the Financial Report Form if an explanation is necessary.
  - E. The final \$1,000 of the grant amount awarded, less any unspent funds due the Foundation, will be remitted to the organization or person stated in the original Letter of Agreement. Any unspent funds in excess of the \$1,000 must be returned to the Foundation.

8. **Publications/Publicity/Photos, Etc.:**

A. Enclose copy(ies) of all publications/publicity generated regarding your grant/project. Please include the date and publication where it appeared, and the circulation of the publication, if available. Publications submitted subsequent to filing the Final Report Form that pertain to the grant project should also acknowledge support of the Foundation. One copy should be mailed to the Foundation.

Find attached the abstracts and posters for:

- Benjamin Castañeda, N. Gustavo Aguilar, David Kanashiro, Javier Ticona, Roberto Lavarello, Fanny L. Casado, Luz M. Huaroto, “Automated Tuberculosis Screening Using Image Processing Tools,” *The International Journal of Tuberculosis and Lung Disease*, vol. 14, n. 11, pp. s345.
- Benjamin Castañeda, N. Gustavo Aguilar, Gerardo E. Valladares, A. Julio Tanta, Fanny L. Casado, Luz M. Huaroto, “Analysis of the diagnostic value of the addition of methylene blue as the last step in Ziehl–Neelsen staining,” *The International Journal of Tuberculosis and Lung Disease*, vol. 14, n. 11, pp. s347.
- Benjamin Castañeda, N. Gustavo Aguilar, David Kanashiro, Javier Ticona, Roberto Lavarello, Luz M. Huaroto, “Automated Tuberculosis Screening Using Image Processing Tools,” *Proceedings of the IEEE Pan American Health Care Exchanges International Conference*, Lima, Peru, March, 2010, pp. 111.

Future scientific articles based on this project will acknowledge the support of the Foundation.

- B. Include one copy of a select group of photos (5-10) that tell your story and DVD clippings (if available) for permanent Foundation use in publicizing your work. See attached.

### C. Credits

This project and its support from the Lindbergh Foundation were mentioned over several presentations and news reports:

- Rochester Review, November-December 2009, Vol. 72, No. 2  
<http://www.rochester.edu/pr/Review/V72N2/cn-grad.html>
- Rochester Center for Biomedical Ultrasound Annual Report 2009
- PC World Perú, Perú, April 2009
- Presentations at Universities:
  - o Universidad Nacional de Trujillo, Perú, May 2009
  - o Unviersidad Católica Santa María, Perú, June 2009
  - o Universidad Nacional San Antonio Abad del Cuzco, Perú, July 2009
  - o Universidad Nacional de la Amazonía Peruana, Perú, July 2009
  - o Universidad Nacional de Piura, Perú, November 2009
  - o Rochester Institute of Technology, Rochester, USA, January 2010
  - o Pontificia Universidad Católica del Perú, Perú, May 2010
  - o Colegio de Ingenieros del Perú, Perú, May 2010
  - o Oncosalud, Perú, November 2010
  - o Peruvian Society of Radiology, Perú, November 2010

9. **Copies:** Please submit a complete original (items 1-9) of this report and a second copy of items 1-6.