# Automatic Bacillus Detection in Light Field Microscopy Images Using Convolutional Neural Networks and Mosaic Imaging Approach*

M. K. M. Serrão, M. G. F. Costa, *Member*, *IEEE*, L. B. Fujimoto, M. M. Ogusku and C. F. F. Costa Filho*, Member, IEEE*

*Abstract*— **Tuberculosis (TB) is one of the top 10 causes of death worldwide. The diagnosis and treatment of TB in its early stages is fundamental to reducing the rate of people affected by this disease. In order to assist specialists in the diagnosis in bright field smear images, many studies have been developed for the automatic Mycobacterium tuberculosis detection, the causative agent of Tb. To contribute to this theme, a method to bacilli detection associating convolutional neural network (CNN) and a mosaic-image approach was implemented. The propose was evaluated using a robust image dataset validated by three specialists. Three CNN architectures and 3 optimization methods in each architecture were evaluated. The deeper architecture presented better results, reaching accuracies values above 99%. Other metrics like precision, sensitivity, specificity and F1-score were also used to assess the CNN models performance.**

*Clinical Relevance*— **The presented works provides an automatic method to aid the diagnosis of tuberculosis in bright-field microscopy.**

## I. INTRODUCTION

Tuberculosis (TB) is an infectious disease that can affect a variety of organs or systems. The most common causative agent is M. tuberculosis, also known as Koch's bacillus. Tb is one of the top 10 causes of death worldwide. According to the World Health Organization, in 2017, approximately 10 million people were affected by TB disease. Of those, approximately 1.3 million with HIV negative and an additional of 300,000 people with HIV positive died. About 85% of Tb deaths in 2017 occurred in Africa and Southeast Asia region. India was responsible for 27% of global deaths from TB. [1].

Properly diagnosing and treating TB in its early stages are key actions to control and stop the spread of the disease. In recent years, in the literature, methods have been published to automate the diagnosis of TB. These process a sputum smear microscopy image using a variety of computational techniques: digital image processing [2] and [3]), traditional machine learning algorithms [4] and [5], and shallow neural networks and deep neural networks (such as the convolutional neural networks) [6], [7] and [8].

Quinn et al. [6] developed a method using Convolutional Neural Networks (CNN) that provides experimental results for three diagnostic tests: malaria (in blood smear samples); tuberculosis (in sputum smear samples) and intestinal parasites

(in a stool sample). From the sample images, the authors produced patches, which are smaller images corresponding to the size of the pathogens approached in the work, and used them for training the network. To generate the Tb bacilli-containing patches, experts have delimited a region of the image with bounding boxes. After training, the model was able to classify a small region of an image, containing or not an object of interest (bacillus). To identify pathogens in complete images, aiming to rule out multiple detections, the authors used the non-maximum suppression (NMS) algorithm.

López et al. [7] developed a method also based on CNN for bacilli´s classification. The database inputs used for CNN training and testing is comprised patches of 40x40 pixel size, extracted from smear images. 50% of these patches were positive ones (containing bacilli) and 50% were negatives (without bacilli). In the original database images, the experts marked the bacilli with a small red dot on it. Furthermore, a yellow square was drawn around a bacilli cluster and a black triangle around the doubtful bacilli. For the detection of the bacilli in complete smear images, the authors also used the NMS algorithm to rule out multiple detections of the same bacillus.

Xiong et al. [8] developed an CNN based on the Google's CIFAR-10 model to classify Tb bacilli. Similar to the previously mentioned works, the developed model was trained with 32x32 pixel patches. After applying transformations in the patches, the database used by the authors was comprised of more than 3 million samples. The database was formed from expertly labeled smear images using ASAP software.

Regarding the works previously presented, two observations can be made. The first one is the fact that the use of patches for training and testing the models. The second one is that the bacilli were not segmented manually, but were marked with bounding boxes or marked with a single point.

The method proposed in this paper also uses a CNN model to perform segmentation and identification of bacilli in brightfield smear images. Similar to previous works, the proposed method also works with patches of smear images. However, for each patch containing a bacillus, a binary image with the segmented bacillus structure was obtained. Thus, the CNN input is fed with information about the shape of the bacillus and not only about its existence or not in a patch. Another feature of this work is that the database used for training and testing consisted of mosaic images formed by a

M. K. M. Serrão, M. G. F. Costa, L. B. Fujimoto and C. F. F. Costa Filho are with Federal University of Amazonas, Amazonas, Brazil (e-mails: mkalline@gmail.com, mcosta@ufam.edu.br, olfujimoto@gmail.com, ccosta@ufam.edu.br) and M. M. Ogusku is with INPA, Amazonas, Brazil (mmogusku@inpa.gov.br)

rectangular mosaic composed of positive and negative patches. This approach was developed in order to reduce the number of duplicate bacilli in the complete image analysis, when using single patches in CNN input, and thus minimize the use of the NMS algorithm. For optimal performance, it was evaluated three CNN architectures combined with three optimization methods.

## II. METHODOLOGY

### A. Materials

This research was approved by the Human Research Ethics Committee of Instituto Nacional de Pesquisa da Amazônia (INPA), protocol 186/08. The database used in this paper was made available by the UFAM Pattern Recognition and Optimization Research Group [9]. It consists of sputum smear microscopy images stained with the Kinyoun technique, and patches of 40x40 pixels: 13,977. Patch is positive if contain a bacillus, a cluster of bacilli, fragmented bacilli or a bacillus in the final multiplication step. Patch is negative if containing no bacillus, or fragment of bacillus at the patch edges. To this work, positive patches with bacillus clusters and fragmented bacilli were removed, resulting in 9,700 positive patches. Also, all negative patches with bacilli fragments at its extremity were also removed, resulting in 25,000 samples. For each positive patch, a complementary binary image, where the bacillus was manually segmented, was obtained. In these images, the bacilli are black (gray level=0), while the background is white (gray level=255). These binary patches are the gold standard for positive patches. The gold standard of each negative patch is an entirely white patch. For CNN training, validating and testing, the positive patches is split into 50%, 25% and 25%, respectively. The same division was used for negative patches.

To train, validate and test the proposed method, mosaic images were generated from patches database. The mosaic image is composed of 100 patches, resulting in an image of 400x400 pixels, where, approximately, 50% are negative and 50% are positive ones. Figure 1 shows an example of generated mosaic image. A corresponding gold standard image is also generated. In this example, it can be observed that, of the 100 patches that make up the mosaic image, 56 are positive and 44 are negative.

A total of 5,000 mosaic images were generated, 60% for the training set, 20% for the validation set and 20% for the test set.
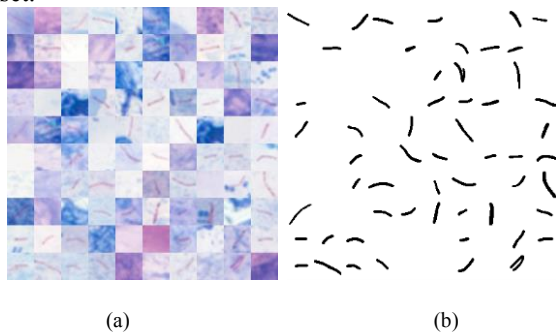


|         (a)          |          (b)          |

Figure 1. (a) Example of mosaic image and (b) its corresponding gold standard image.

### B. CNN Architecture and Training Parameters

The three CNN architectures proposed in this paper are semantic segmentation networks and were based on an architecture proposed by Miyagawa et al. [10]. It is important to emphasize that, although the task of each CNN architectures is bacillus segmentation, the main objective is not a precise bacillus segmentation, but only its detection, aiming to perform bacilli count present in the sputum smear microscopy image.

In general, the three architectures feature convolutive layer blocks consists of a sequence of 3x3 filter, batch normalization layer and ReLU layer (Conv + Batch + Relu), followed by a max pooling subsampling layer with a 2x2 filter. In oversampling convolution layers, 4x4 filter are used. In the final step, a 1x1 size convolutive layer resizes the output depth to 2, the number of classes used in segmentation. After that, the softmax and pixel classification layers are combined to classify each pixel as belong to bacillus or background. Figure 2 shows the three architectures.

The first architecture (CNN1), the largest one, has four maxpooling steps (subsampling steps). with three Conv + Batch + Relu sequences before the first two subsampling steps, four sequences before the last two, and four more sequences after the last subsampling step. After the subsampling steps there are four oversampling steps.

The second architecture (CNN2) has two maxpooling steps, with three Conv + Batch + Relu sequences, before each subsampling step. After the subsampling steps there are two oversampling steps.

The third architecture (CNN3), the smallest one, has two maxpooling steps and two oversampling steps like CNN2. However, it has only one Conv + Batch + Relu sequence before each subsampling step.

In semantic segmentation training, the output of each pixel was optimized using a logistic regression cost function.

For CNN training, three optimization methods were employed: Stochastic Gradient Descent with Momentum (SGDM), Root Mean Square Propogation (RMSProp) and Adaptive Moment Estimation (ADAM). Table 1 shows the training parameters used for CNN training. The experiments were performed using MATLAB version 2018b, a 3.2 GHz Intel i7-8700 processor computer with 16 GB RAM and 8 GB GeForce GTX 1070 GPU.

TABLE I.        TRAINING PARAMETERS

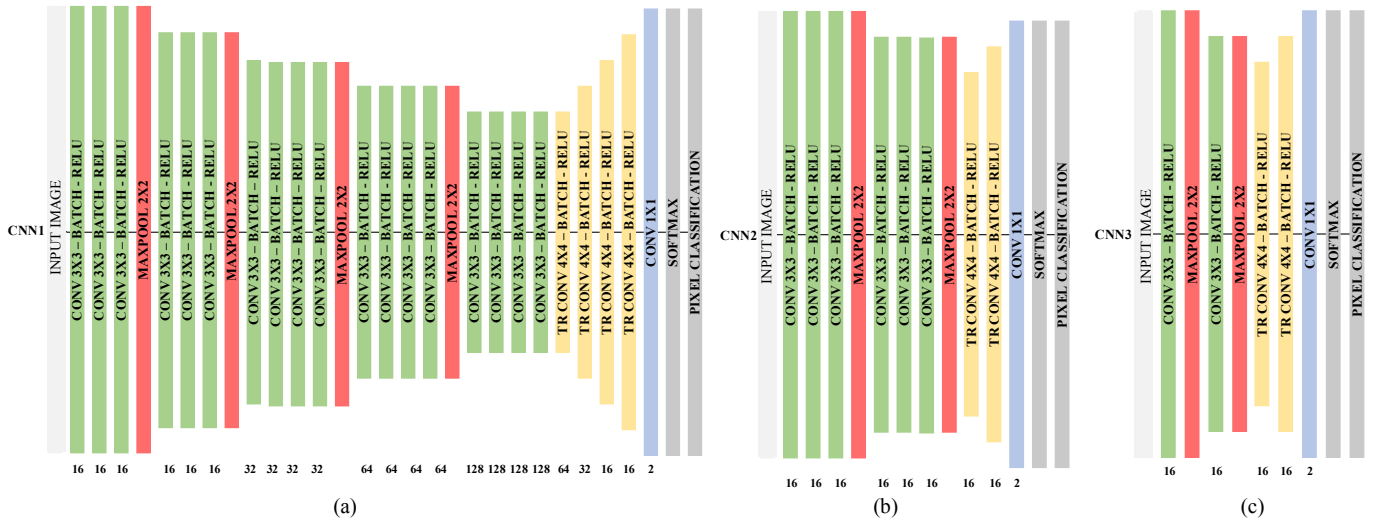| Parameters | Values |
|---|---|
| Initial learning rate | 0.001 |
| Learning rate drop factor | 0,5 |
| Maximum number of epochs | 20 |
| Lot size | 10 |
| Stopping criteria | 6000 iterations |

Figure 2. Proposed architectures: (a) CNN1 architecture; (b) CNN2 architecture and (c) CNN3 architecture

## C. Evaluation metrics

CNN performance is obtained by counting the number of segmented bacilli. In CNN output, the segmented objects are classified as bacilli after applying an area filter of 25 pixels. Very small objects are disregarded.

To evaluate the quality of each designed architecture, the following performance metrics were used: accuracy, precision, sensitivity, specificity, and F1-score. These metrics are shown in equations (1)-(5):

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precision: \frac{TP}{TP+FP} \quad (2)$$

$$Sensitivity: \frac{TP}{TP+FN} \quad (3)$$

$$Specificity = \frac{TN}{TN+FP} \quad (4)$$

$$F1-score = 2 \times \frac{Precision \times Sensitivity}{Precision+Sensitivity} \quad (5)$$

where: TP - true positives, TN - true negatives, FP - false positives and FN - false negatives.

## III. RESULTS AND DISCUSSIONS

Tables II, III and IV show the performance of the CNN1, CNN2 and CNN3 architectures, respectively, in the validation dataset. Table V shows the best performance in the validation dataset for each CNN architecture. The best performance for CNN1 was obtained with SGDM method, for CNN2, with ADAM method, and, for CNN3, with RMSProp method.

Overall, the best performance was obtained with CNN1, while the worst one was obtained with CNN3. However, it is noteworthy that CNN3 achieved the best performance with the RMSProp method for most of the metrics. The best CNN2 performance was obtained with the ADAM method. CNN1 network reached a performance over 99% for all evaluated metrics. CNN2 underperformed 1% to 2% compared to CNN1, while CNN3 underperformed 2% to 4% relative to CNN1. However, the sensitivity was above 99% in all simulations.

Figures 3 shows examples of bacilli segmentations obtained with CNN1, CNN2 and CNN3 architectures, respectively, using the SGDM optimization method in CNN1, ADAM in CNN2 and RMSProp in CNN3. As noted, for each architecture, these methods achieved the best performance for most of the metrics.

In segmented images, noise presence, corresponding to small areas, is observed, being more frequently when using CNN3. The lowest noise density is obtained with the CNN3 network. Much of the noise is removed with the area filter. Therefore, we conclude that using deeper CNNs is important to noise elimination in the bacilli segmentation.

TABLE II. CNN1 PERFORMANCE IN THE VALIDATION SET

| Optimization Method | Accuracy (%) | Precision (%) | Sensitivity (%) | Specificity (%) | F1-score (%) |
|---|---|---|---|---|---|
| SGDM | **99.649** | 99.394 | **99.908** | 99.388 | **99.650** |
| RMSProp | 99.587 | **99.530** | 99.647 | **99.527** | 99.589 |
| ADAM | 99.629 | 99.356 | 99.908 | 99.349 | 99.631 |

TABLE III. CNN2 PERFORMANCE IN VALIDATION SET.

| Optimization Method | Accuracy (%) | Precision (%) | Sensitivity (%) | Specificity (%) | F1-score (%) |
|---|---|---|---|---|---|
| SGDM | 98.054 | 96.374 | 99.811 | 96.342 | 98.062 |
| RMSProp | 98.297 | 96.838 | **99.819** | 96.807 | 98.306 |
| ADAM | **98.534** | **97.308** | 99.805 | **97.283** | **98.541** |

TABLE IV. CNN3 Performance in Validation Set.

| Optimization Method | Accuracy (%) | Precision (%) | Sensitivity (%) | Specificity (%) | F1-score (%) |
|---|---|---|---|---|---|
| SGDM | 95.681 | 91.932 | 99.886 | 91.700 | 95.744 |
| RMSProp | **97.012** | **94.388** | 99.847 | **94.282** | **97.041** |
| ADAM | 95.603 | 91.675 | **99.998** | 91.477 | 95.656 |

TABLE V. Performance in test set.

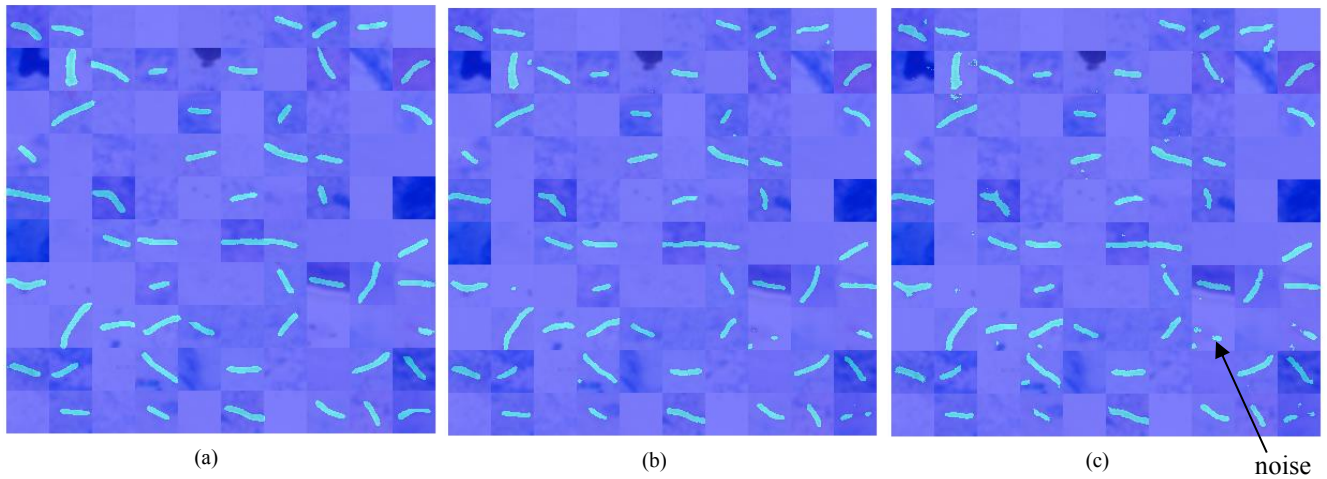| Architecture & Optimization Method | Accuracy (%) | Precision (%) | Sensitivity (%) | Specificity (%) | F1-score (%) |
|---|---|---|---|---|---|
| CNN1 - SGDM | 99.665 | 99.349 | 99.988 | 99.341 | 99.667 |
| CNN2 – ADAM | 98.284 | 96.729 | 99.908 | 96.693 | 98.293 |
| CNN3 - RMSPROP | 96.801 | 93.973 | 99.877 | 93.847 | 96.835 |



(a)  (b)  (c)  noise

Figure 3. Examples of segmented mosaic-image performed by: (a) CNN1 with SGDM method; (b) CNN2 with ADAM method and (c) CNN3 with RMSProp method.

## IV. CONCLUSION

In this paper we evaluated CNN performance in bacilli segmentation using light fields smear images. For CNN training and testing, it was proposed the use of 400x440 mosaic images, consisting of 100 positive and negative bacilli patches with dimensions of 40x40 pixels. Three architectures, with different depths, were trained with three optimization methods. The CNN1 architecture, the largest one, achieved the best performance, providing a significant noise reduction. There was no significant difference between the performance of the optimization methods.

## REFERENCES

[1] W. H. O. WHO, "Global Tuberculosis Report 2018." Geneva, p. 277, 2018.

[2] M. G. F. Costa, C. F. F. Costa Filho, J. F. Sena, J. Salem, and M. O. de Lima, "Automatic identification of mycobacterium tuberculosis with conventional light microscopy," *30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 382–385, 2008.

[3] M. I. Shah, S. Mishra, M. Sarkar, and C. Rout, "Automatic detection and classification of tuberculosis bacilli from ZN-stained sputum smear images using watershed segmentation," in *International Conference on Signal Processing*, 2016, pp. 1–4.

[4] R. Khutlang *et al.*, "Images of ZN-Stained Sputum Smears," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 4, pp. 949–957, 2010.

[5] C. F. F. Costa Filho, P. C. Levy, C. de Matos Xavier, L. B. Mendonça Fujimoto, and M. G. Fernandes Costa, "Automatic identification of tuberculosis mycobacterium," *Rev. Bras. Eng. Biomed.*, vol. 31, no. 1, pp. 33–43, 2015.

[6] J. A. Quinn, R. Nakasi, P. K. B. Mugagga, P. Byanyima, W. Lubega, and A. Andama, "Deep Convolutional Neural Networks for Microscopy-Based Point of Care Diagnostics," *Proc. Int. Conf. Mach. Learn. Heal. Care*, vol. 56, pp. 1–12, 2016.

[7] Y. P. López, C. F. F. Costa Filho, L. M. R. Aguilera, and M. G. F. Costa, "Automatic classification of light field smear microscopy patches using Convolutional Neural Networks for identifying Mycobacterium Tuberculosis," *2017 Chil. Conf. Electr. Electron. Eng. Inf. Commun. Technol. CHILECON 2017 - Proc.*, vol. 2017-Janua, pp. 1–5, 2017.

[8] Y. Xiong, X. Ba, A. Hou, K. Zhang, L. Chen, and T. Li, "Automatic detection of mycobacterium tuberculosis using artificial intelligence," *J. Thorac. Dis.*, vol. 10, no. 3, pp. 1936–1940, 2018.

[9] Grupo de Pesquisa em Reconhecimento de Padrões e Otimização, "http://dgp.cnpq.br/dgp/espelhogrupo/4186," *[Online]*, 2008. [Online]. Available: http://dgp.cnpq.br/dgp/espelhogrupo/4186. [Accessed: 21-Oct-2019].

[10] M. Miyagawa, M. G. F. Costa, M. A. Gutierrez, J. P. G. F. Costa, and C. F. F. C. Filho, "Lumen Segmentation in Optical Coherence Tomography Images using Convolutional Neural Network," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2018-July, no. c, pp. 600–603, 2018.