Segmentation of Tuberculosis Bacilli in Conventional Microscopy Images Through Accelerated CNN Using Dilated Convolutions

Philipe Rangel Demuth, Patrick Marques Ciarelli, Jorge Leonid Aching Samatelo

[1]Programa de Pós Graduação em Engenharia Elétrica
Universidade Federal do Espírito Santo
Av. Fernando Ferrari, 514 - Goiabeiras, Vitória - ES, Brasil

philipe.demuth@ufes.br, patrick.ciarelli@ufes.br, jorge.samatelo@ufes.br

Abstract. *In this work, we propose a method to achieve microscopy image segmentation, in which a convolutional neural network (CNN) is used. The method is divided in two parts: (i) the CNN is trained for pixelwise classification of image; (ii) the training CNN is accelerated, removing the redundant operations, allowing the classification of the pixels from an entire image patch at the same time. The method was evaluated over a dataset with* 120 *images obtained using conventional microscopy in sputum smear sheets prepared according to the Ziehl-Neelsen technique. In the experimental evaluations carried out on this dataset, we obtained an accuracy of* 97.33% *and recall of* 96.30%*. The accelerated CNN is* 44 *times faster, maintaining identical prediction results. These results show that the proposed method has the potential to handle the given problem.*

1. Introduction

According to the United States Centers for Disease Control and Prevention (CDC), tuberculosis is an infectious disease caused by the Mycobacterium tuberculosis bacillus. It usually affects the lung, but it can also attack any part of the body, such as liver, spine and brain [CDC-Tuberculosis 2018].

The health problems caused by tuberculosis affect approximately 10 million people each year and is one of the ten leading causes of death in the world. In the last 5 years, it has been the world's leading cause of death by a single infectious agent, beating HIV/AIDS [WHO et al. 2017]. Despite this fact, if the diagnosis is made at the right time, most people who develop tuberculosis can be cured [WHO et al. 2017].

Diagnosis of tuberculosis disease is usually made by coloring a microscope slide using the Ziehl-Neelsen staining method, which is then analyzed by a specialist, using an optical microscope in the search for bacilli [Smart 2017]. This is a simple, quick and inexpensive technique that is extremely specific in areas with high tuberculosis prevalence [ISTC 2006]. Two methods of microscopy can be used for the detection of tuberculosis bacilli: conventional microscopy and fluorescence. Although fluorescence microscopy has a sensitivity of about 10% greater when compared to light microscopy, it is less common in developing countries because it has a higher cost [ISTC 2006].

The inspection process which includes, in addition to checking, the counting of bacilli, is usually time consuming and tiring. Therefore, a system of automatic recognition

of tuberculosis bacilli would be interesting, allowing the diagnostic process to become faster. To perform such task, a fundamental step is the segmentation of the microscopy image. Segmentation of an image is understood as the task of subdividing the image into its constituent regions. The way the subdivisions are made depends on the problem to be solved [Gonzalez and Woods 2012]. For the case of two-class (binary) segmentation, a binary image (which classifies each pixel of the image into one of two classes) is generated from the original image. In this case, it is understood as binary object any set of pixels in the binary image that correspond to a bacillus or bacilli grouping.

In this article we propose a technique of segmentation based in a convolutional neural network (CNN). Specifically, a CNN is used to operate as a predictor of classes generating a probability image, called as *heatmap*. The values of the pixels of a *heatmap* are in the range $[0, 1]$ and represent the probability that a pixel belongs to a bacillus. The most probably class of each pixel is determined using an $argmax$ function. Here, two principal differences in the training and test steps of the CNN are proposed, in short: (*i*) in the training step, the CNN is trained using sparse samples from the training set, allowing the creation of a classifier capable of give the probability of a pixel to belong to a bacillus; (*ii*) in the test step, the CNN is accelerated, removing redundant operations that otherwise would occur if we were to classify every single pixel of the image. This modification makes the image segmentation process many times faster. Experimental results on a dataset indicated that the proposed technique obtained results with an accuracy of segmentation of 97.33% and recall of 96.30%.

2. Related Work

Although the earliest microscopy image segmentation works have used the fluorescence technique, they will not be mentioned here because they involve a higher cost imaging method and therefore less used in developing countries.

The authors of [CostaFilho et al. 2012] propose to use the difference between the R (red) and G (green) color channels to segment the image using a global adaptive threshold segmentation technique. The artifacts present in the segmented image are eliminated using morphological filters, color filters, and size filters. Binary objects with less than 200 pixels are removed. The results achieved were 76.65% and 12% for sensitivity and accuracy, respectively. However, no details are given about how these values were achieved.

In [Siena et al. 2012] is used *decorrelation stretch* [Chitade and Katiyar 2010] as a pre-processing step, followed by a segmentation technique based on the clustering algorithm k*-means*. The classification of the segmented structures is done using a multilayer artificial neural network with 15 neurons in the hidden layer. The result of this work is an accuracy of 88%. Nine CDC images were analyzed (tested), while the training images were those used in [Forero et al. 2003]. In this study, quantitative results are not presented in relation to segmentation.

In [Chayadevi and Raju 2014] is presented a color-segmentation technique using the *Watershed* algorithm on $YCbCr$, HSI, and Lab color spaces. Then, binary objects are analyzed by shape characteristics, such as: area, perimeter, compactness, eccentricity, major and minor axis, a ratio between the axes and some of their invariant moments. However, no quantitative evaluation is performed.

In general, the cited works have the following restrictions: they use threshold-dependent global segmentation techniques [CostaFilho et al. 2012], and they need highly parameterized post-processing steps [Sadaphal et al. 2008] [CostaFilho et al. 2012] [Chayadevi and Raju 2014] and, in some cases, quantitative analyzes of the results are not presented [Chayadevi and Raju 2014].

Differently, the proposal of this work is based on the use of a CNN acting as a predictor of binary classes. By extracting information from the data themselves, there is no need to manually adjust the segmentation parameters: it "learns" in its training stage, generating a high accuracy in the segmentation.

In relation to the use of CNNs in the task of segmentation of medical images, in the literature one can find different applications. For example: [Chen and Chefd'Hotel 2014] make use of a CNN to target cells of the immune system present in microscope slides with tumor samples. Such cells are highlighted using immunohistochemical (IHC) staining. In this work, a correlation coefficient between the number of cells counted by the algorithm and the number of manually counted cells was of 99.49%; in [Tschopp et al. 2016] the segmentation is done using a CNN and a sliding window approach to classify every pixel and predict membranes from the nervous system of organisms like Drosophila melanogaster. The nervous system structure image is captured with a high-throughput Electron Microscopy (EM). Since the use of CNNs to classify single pixels takes a lot of redundant computations when using sliding window approach, a method to classify the pixels simultaneously of entire patches is also proposed.

3. Proposed method

The proposed method for the microscopy segmentation problem is described in this section. This method is divided in two steps, specifically:

- **Pixelwise class prediction**: Using the methodology based on patches, a CNN is trained to predict the probability of every input pixel to belong to a given class. The pixels can be either the object of interest (bacillus or group of bacilli) or the background. The already trained network is applied to every pixel of the microscopy image, generating as output a probability image.
- **CNN acceleration**: Using the same weights, the architecture of the original CNN is modified to remove redundant operations that comes when classifying every single pixel of the test image, allowing a speedup of the overall process.

Both steps will be explained on the following subsections.

The proposed methodology based on *patches* has six main steps summarized in Figure 1 and explained below:

① For each training image, patches in which the central pixel belongs to a bacillus (positive samples) or their central pixel belonging to the background (negative samples) are randomly extracted. The features of these patches will be explained in detail in Subsection 4.2.
② As Subsection 3.2 describes, the CNN is trained using the patches from step 1.
③ An entire image patch is fed to the Accelerated CNN.
④ With the weights obtained in the training step, the Accelerated CNN is an improved version of the trained CNN. The process of converting between both CNNs is described in Subsection 3.3.

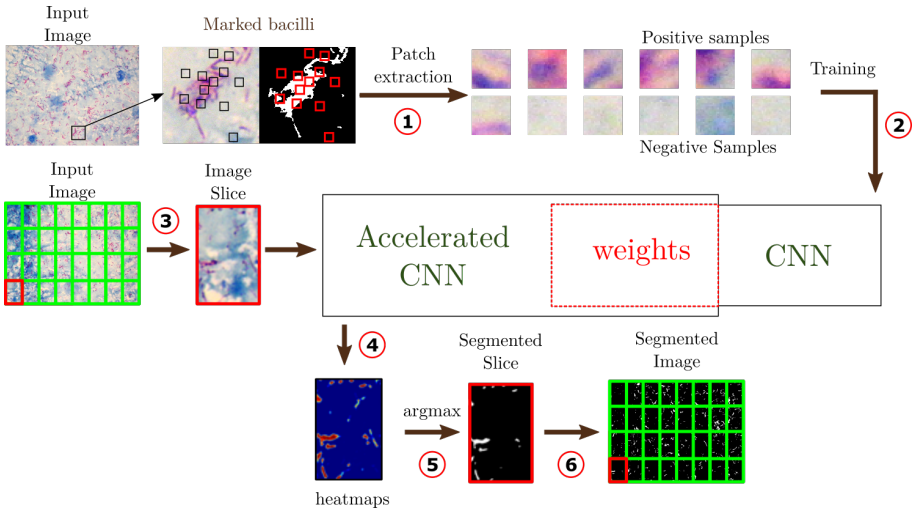


Figure 1. *Framework* used in this work.

⑤ From the probability image patch, a binary image is generated using the maximum a posteriori (MAP), that assigns for each pixel from the patch to the class that has the largest posterior probability.
⑥ The entire segmented image is a mosaic of smaller segmented patches of the input image.

3.1. CNN architectures

A CNN is similar to a classical neural network, containing layers of neurons that perform operations on the results of the previous layers, and the parameters of these operations are adjusted in a training step. CNNs have been used in problems involving image processing and computer vision, since they have the capacity to consider spatial characteristics of the image, because they are based on the convolution operation. A current summary of the subject can be found at [Gu et al. 2015] and [Rawat and Wang 2017].

The CNN architecture is composed by many layers, and two types of relevant layers are the convolutional layers and *max pooling* layers. The convolutional layers are formed by masks that slide over the pixels of the image by performing linear operations, and their outputs pass through a nonlinear function (in this work the ReLU function is used) [Nair and Hinton 2010]), which adds a greater ability for the network to "learn" a task. The amount by which each mask shifts on the image is the stride (s). Stride equals to 1 means the mask shifts one unit at a time. In turn, the layers of *max pooling* aim to reduce the size of the array by selecting the maximum value of each set of values defined by the size of the mask to be applied. Additionally, max pooling layers help to build a network that is more robust to translations in the input image.

3.2. Pixelwise class prediction

The methods based on patches have been increasingly used on the field of medical imaging, with several applications, such as image segmentation, image denoising, image registering, anomalies detection and image synthesis [Wu et al. 2015]. Neural Networks patch based solutions, as CNNs, demand a huge amount of samples, in this case images, for an adequate training. This premise is difficult to achieve in the medical field, because it is

necessary to have an expert to manually annotate the images, and this result in datasets that usually contain few images. Such limitation is circumvented using patches, since from each image several smaller regions can be extracted (patches), thus increasing the amount of training samples for the CNN. In this type of approach the image segmentation is done on each pixel separately, instead of an image at a time[Litjens et al. 2017].

In this work, the CNN used was based on the LeNet-5 [LeCun et al. 1998] network, being modified for the case where the input is a patch of dimensions $45 \times 45 \times 3$ pixels, drawn around a pixel at be classified as background or object. Such pixel belongs to a color image in the RGB color space. This architecture is easily trained and has shown that produces good classification results in datasets of images with similar dimensions [LeCun and Cortes 2010]. This CNN has 2 interpolated layers plus 2 layers of *max pooling*. The final 2 convolutional layers are used to generated the fully connected layers. The output layer, consists of 2 neurons with the *softmax* activation function, where each one of the neurons indicates the probability value of the central pixel of the *patch* be an object ($P(F)$) or background ($P(B)$). The training network structure is illustrated in Figure 2.

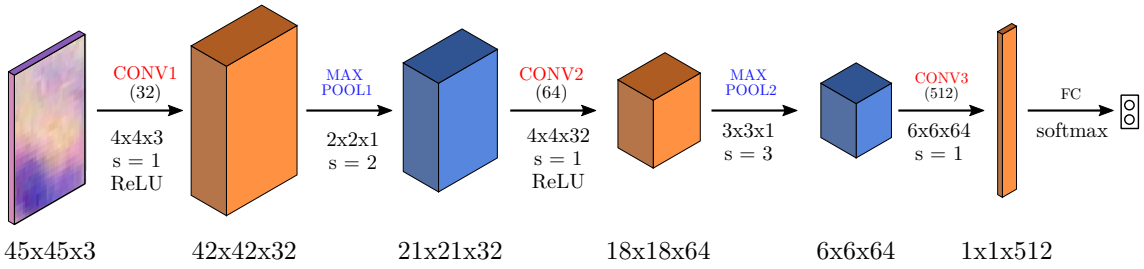


Figure 2. CNN used for the training step. In this network, s is the stride of the convolutional/pooling kernel.

3.3. CNN acceleration

Sliding window techniques for pixelwise classification have great overlap between adjacent windows, which implies in redundant computation of convolutions and poolings. To overcome this, [Li et al. 2014] and [Tschopp et al. 2016] show a method to convert the pixelwise classification CNN into an optimized Accelerated CNN, eliminating the redundant computations present in sliding window segmentation. They propose a technique called *strided kernel*, which overview can be seen in Figure 3.

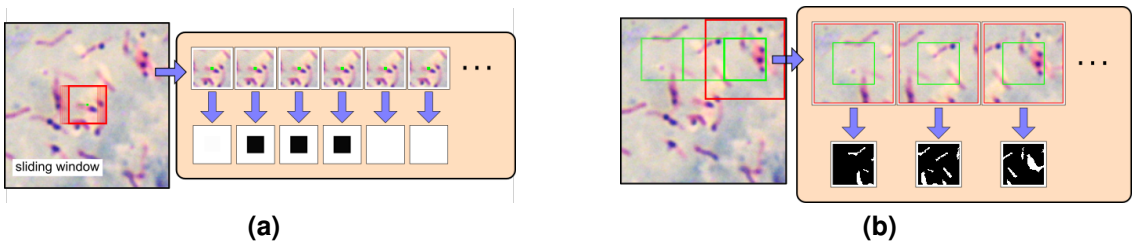


Figure 3. (a) In sliding window segmentation, each pixel of the image is classified individually; (b) in strided window segmentation a whole patch is segmented at a time. Image adapted from [Tschopp et al. 2016].

Since the CNN's layers with stride greater than 1, specially the poling layer, reduce the resolution of the input image, the main idea of the *strided kernel* technique is to modify the original convolution and pooling kernels by inserting a specific number of zeros to compensate the down-sampling. [Li et al. 2014] call these kernels d-regularly sparse kernels. In the paper presented by [Yu and Koltun 2015] these kernels are called dilated convolutions. An example of dilated convolution kernel can be seen in Figure 4.

(a) $k \in R^{3\times3}$. **(b)** k conversion result.

Figure 4. (a) 3×3 convolution kernel k and $d = 1$, whose entries are represented by colored squares. (b) Conversion of convolution kernel k in (a) to a 2-regularly sparse convolution kernel k with $d = 2$. Image adapted from [Li et al. 2014].

Following the conversion procedure presented in [Tschopp et al. 2016], the Accelerated CNN structure is presented in Figure 5.

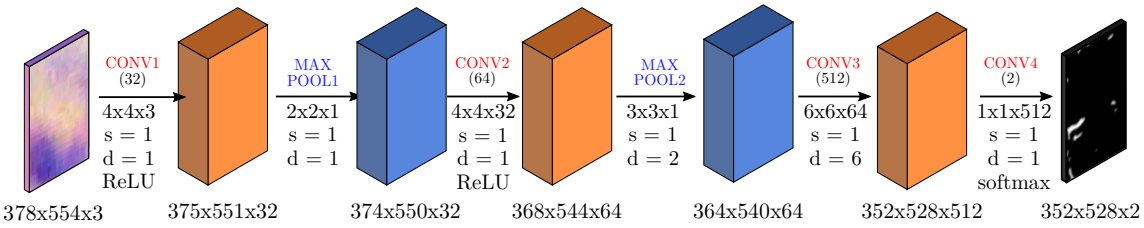


Figure 5. Accelerated CNN used for whole image patch segmentation. In this network, s is the kernel stride and d is the dilation rate.

4. Experiments and Results

4.1. Dataset

The dataset used in this work consists of 120 images of 12 patients, being 10 images of a single slide for each patient. The images were obtained using a 10 MegaPixel camera attached to a conventional microscope. The images of the slides with the sputum smear were prepared using Ziehl-Neelsen staining. The training and test set images have a fixed resolution of 2816×2112 pixels. In all images, the bacilli were identified and marked pixel by pixel by a specialist. This dataset was obtained from [Soares et al. 2015].

4.2. Implementation Features

The CNN was implemented using the *TensorFlow* library, version 1.5.0 on *Python*. In addition, the configuration of the machine used in the experiments was: (*i*) Linux operational system, Ubuntu 18.04 LTS distribution; (*ii*) Intel core i7-8700k with 6 physical cores; (*iii*) 32 GB DDR4 RAM memory; (*iv*) 2 TB data store unit (Hard Disk) + 240 GB SSD; (*v*) NVIDIA Geforce GT 1080Ti video card, with 11 GB dedicated memory.

The training and test sets were created from the dataset, each with 60 images. A validation set was separated from training patches to validate CNN's learning. This set corresponds to 10% of the total number of training patches. The set of training patches was created according to the following methodology: for each one of the 60 images of the training set, 300 patches were randomly extracted as **positive samples** and 300 patches as **negative samples**. In cases where the patch had pixels outside the image, it was filled using mirrored padding. All patches have dimensions of $45 \times 45 \times 3$ pixels.

The hyperparameters used for the CNN and the Accelerated CNN and segmentation are: (*i*) Regarding the CNN: the initial learning rate used is 0.01, with an exponential decay of 0.95. The CONV3 and CONV4 layers have its weights normalized using L2 norm. The search for these parameters was empirical. (*ii*) In relation to segmentation, if $P(F)$ is greater than $P(B)$ the corresponding pixel is considered to be foreground (F), else background (B). This is done using an *argmax* function.

4.3. Segmentation time

In Figure 6 is presented the segmentation time for the original and the accelerated CNN. From such figure, it can be seen that the segmentation mean time for the original CNN and Accelerated CNN are $105.68s$ and $2.42s$, respectively. This means that the accelerated CNN is about 44 times faster.

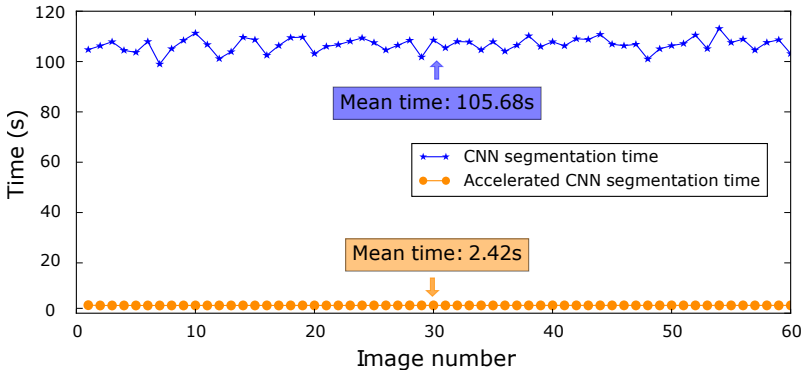


Figure 6. Segmentation time on the test set.

4.4. Segmentation results

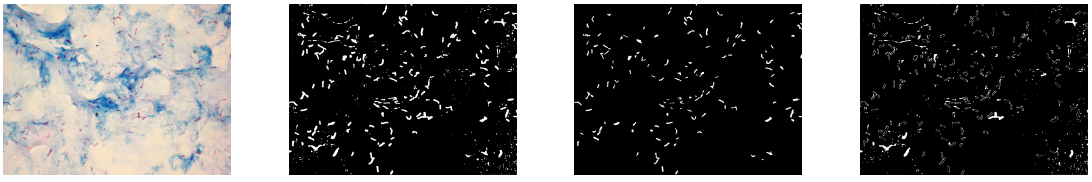
Considering that pixelwise segmentation was modeled as a binary classification problem, the evaluation used in this work is done using the following metrics: (*i*) accuracy (Ac), which evaluates the technique's ability to correctly classify pixels as object or background; (*ii*) recall (Re), which evaluates the technique's ability to correctly classify all pixels labeled as objects (bacilli); (*iii*) precision (Pr), which evaluates the technique's ability to wrongly classify few pixels from the background as object. and, (*iv*) F_1 measure which is the harmonic mean between recall and precision. These metrics are defined by the equations: $Ac = \frac{TP+TN}{TP+TN+FP+FN}$, $Re = \frac{TP}{TP+FN}$, $Pr = \frac{TP}{TP+FP}$ and $F_1 = 2\frac{Re.Pr}{Re+Pr}$, where TP are the true positives (numbers of pixels correctly classified by the CNN as object), TN are the true negatives (number of pixels correctly classified by the CNN as background), FP are the false positives (number of pixels wrongly classified by the CNN as object), and FN are the false negatives (number of pixels wrongly classified by the CNN as background).The evaluation of the metrics Ac, Re, Pr and F_1 on the test set is shown in Table 1.

Table 1. Metrics on the test set when using the original and the accelerated CNN.

	$\overline{Accuracy}$	$\overline{Recall}$	$\overline{Precision}$	$\overline{F_1\ measure}$
Original CNN	97.33	96.30	30.95	42.36
Accelerated CNN	97.33	96.30	30.95	42.36

About the values obtained, the following comments are made: (***i***) since the two networks share that same weights and the Accelerated CNN is constructed in way that mimics the original CNN output, the value of the metrics evaluated are the same in both CNNs; (***ii***) the high value of accuracy is caused by the large amount of true negatives. As the number of pixels belonging to the background is much larger than the number of pixels belonging to the bacilli, the amount of true negatives has a much greater weight in the calculation of accuracy than the true positives; (***iii***) the high recall value indicates that the technique is able to segment almost all the pixels belonging to the bacilli. In medical imaging this is preferable rather than high accuracy. The more structures are identified as bacillus, the greater the chance of detecting the disease. Although this is done with a large number of false positives (implying a low accuracy), it is less likely that a tuberculosis patient will be diagnosed as being healthy (and not taking the treatment) than the reverse; (***iv***) the low precision value indicates that several false positives are detected in the segmentation, i.e., some pixels belonging to the background are classified as bacilli; (***v***) it should be noted that the problem itself is strongly unbalanced, i.e., the number of pixels corresponding to the background is greater than the number of pixels corresponding to the bacilli (each image has a mean of 99.14% of the pixels belonging to the background and 0.86% belonging to the bacilli). Such imbalance of the classes produces that the false positives have a strong impact in the calculation of the precision; (***vi***) the low value of the measure F_1 is caused by the low precision of the technique, since the value of the recall is high.

An example of a segmented image can be seen in Figure 7. It may be noted that: (***i***) because of the high recall of the technique, it is able to segment all of the specialist-labeled bacilli as well as some unlabeled structures (see Figures 7b and 7c); (**ii**) the segmentation does not work well on the edges of the bacilli, together with the extra objects that are not in the marking, decreases the precision of the technique (see Figure 7d).



(a) **(b)** **(c)** **(d)**

Figure 7. Example of a segmented image. (a) Input image; (b) Segmented Image; (c) Ground truth; (d) Difference between (b) and (c).

5. Conclusion and future works

The main objective of this work was to propose a technique for the segmentation of microscopic images of colored sputum blades according to the Ziehl-Neelsen method. The proposed technique is based primarily on a Covolutional Neural Network using the patches

methodology and its accelerated version using dilated convolutions. Comparing the two versions, the accelerated CNN is 44 times faster than the original CNN and both achieve exactly the same results. In relation to the segmentation result, using a microscopy image dataset, an accuracy of 97.33% and a recall of 96.30% was obtained. From the values of precision and recall it was observed that the proposed technique presents: high accuracy due to the large number of background pixels, and a high recall, due to the identification of a good part of the pixels belonging to the bacilli. On the other hand, the precision was low (30.95%), due to the imbalance of the classes, since in average each image contains 99.14% of the pixels belonging to the fund and 0.86% belonging to the bacilli. Finally, in order to reduce the number of false positives, a classification step of segmented binary objects based on the morphological characteristics of these objects will be implemented in the future.

Acknowledgments

Patrick Marques Ciarelli and Philipe Rangel Demuth thank the partial funding of their research works provided by CNPq (grants 312032/2015-3 and 141434/2017-1, respectively).

References

CDC-Tuberculosis (2018). `https://www.cdc.gov/tb/`. accessed in 01/03/2018.

Chayadevi, M. and Raju, G. (2014). Automated colour segmentation of tuberculosis bacteria thru region growing: A novel approach. In *Applications of Digital Information and Web Technologies (ICADIWT), 2014 Fifth International Conference on the*, pages 154–159. IEEE.

Chen, T. and Chefd'Hotel, C. (2014). Deep learning based automatic immune cell detection for immunohistochemistry images. In *International Workshop on Machine Learning in Medical Imaging*, pages 17–24. Springer.

Chitade, A. Z. and Katiyar, S. (2010). Colour based image segmentation using k-means clustering. *International Journal of Engineering Science and Technology*, 2(10):5319–5325.

CostaFilho, C. F., Levy, P. C., Xavier, C. M., Costa, M. G., Fujimoto, L. B., and Salem, J. (2012). Mycobacterium tuberculosis recognition with conventional microscopy. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 6263–6268. IEEE.

Forero, M., Cristobal, G., and Alvarez-Borrego, J. (2003). Automatic identification techniques of tuberculosis bacteria. In *Applications of digital image processing XXVI*, volume 5203, pages 71–82. International Society for Optics and Photonics.

Gonzalez, R. C. and Woods, R. E. (2012). Digital image processing.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., and Wang, G. (2015). Recent advances in convolutional neural networks. *CoRR*, abs/1512.07108.

ISTC (2006). "international "standards for "tuberculosis "care. `http://www.who.int/tb/publications/2006/istc\_report.pdf`. accessed in 01/03/2018.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

LeCun, Y. and Cortes, C. (2010). MNIST handwritten digit database.

Li, H., Zhao, R., and Wang, X. (2014). Highly efficient forward and backward propagation of convolutional neural networks for pixelwise classification. *arXiv preprint arXiv:1412.4526*.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

Rawat, W. and Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29(9):2352–2449.

Sadaphal, P., Rao, J., Comstock, G., and Beg, M. (2008). Image processing techniques for identifying mycobacterium tuberculosis in ziehl-neelsen stains. *The International Journal of Tuberculosis and Lung Disease*, 12(5):579–582.

Siena, I., Adi, K., Gernowo, R., and Miransari, N. (2012). Development of algorithm tuberculosis bacteria identification using color segmentation and neural networks. *International Journal of Video and Image Processing and Network Security*, 12(4):9–13.

Smart, T. (2017). Background on smear microscopy in tb diagnosis. `http://www.aidsmap.com/Background-on-smear-microscopy-in-TB-diagnosis/page/1426650/`. accessed in 01/03/2018.

Soares, L. A., Côco, K. F., Salles, E. O. T., and Bortolon, S. (2015). Automatic identification of mycobacterium tuberculosis in ziehl-neelsen stained sputum smear microscopy images using a two-stage classifier. In *VISAPP (3)*, pages 186–191.

Tschopp, F., Martel, J. N., Turaga, S. C., Cook, M., and Funke, J. (2016). Efficient convolutional neural networks for pixelwise classification on heterogeneous hardware systems. In *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pages 1225–1228. IEEE.

WHO et al. (2017). Global tuberculosis report 2017. In *Global tuberculosis report 2017*.

Wu, G., Coupé, P., Zhan, Y., Munsell, B., and Rueckert, D. (2015). *Patch-Based Techniques in Medical Imaging*. Springer.

Yu, F. and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.