BME 590L Final Project: Classification of Microscopic Sputum Images With Tuberculosis Bacilli Using Neural Networks

Huisi Cai, Zhen Lin

Prof. Roarke W. Horstmeyer

04/29/2019

1 Introduction

About 1,700 million people, or one third of the world's population are , or have been, infected by Mycobacterium tuberculosis (TB) .[1] The risk of TB infection has greatly diminished for developed country due to effect control.[1] However, tuberculosis remains as one of the most common infectious diseases in developing countries due to lack of monitoring and treatment.

Tuberculosis care, including effective detection and treatment of patients, is the core element of TB control.[2] The current international TB guidelines recommend the microscopic examination of three sputum specimens for acid-fast bacilli in the evaluation of persons suspected of having pulmonary TB.[2] The process of examining the microscopic images could be time consuming and labor-intensive. Therefore, we propose to classify whether the tuberculosis is presented in the sample automatically using deep neural network training.

The goal of this project is to train the network to correctly classify the microscopic specimen as "with TB cell" or "without TB cell". There are three main components in this project: 1) pre-process the image database by cropping them into suitable size for training and labeling them correctly, 2) perform image classification on the self-established dataset with different networks and identify a network with the best performance, and 3) Add a physical layer of weights on the three color channels to improve classification accuracy. Our result shows that VGG with "tanh" activation yields the highest accuracy of 95% for validation data. We concluded that neural network is a promising way to automatically classify microscopic image with and without TB bacilli.

2 Related work

Rulaningtyas et al. proposed a method to classify TB cells by first extracting the geometry of cell and then feed these mycobacterium shape features into a backprojection neural network for classification.[3] This method focus on the elongated shape of the bacilli while we proposed to focus more on the color difference of the cell.

For this project, we used VGG, Densely-connected Neural Network (DNN), and ResNet as our training networks. VGG is a simple but powerful CNN proposed by Simonyan et al. in 2015.[4] It significantly improved prior-art configurations by increasing the depth of convolutional networks. Also in 2015, He et al. proposed a deep neural network with residual learning called ResNet,[5] which are easier to optimize and achieved better accuracies than traditional deep neural networks.

3 Methods

Our project was carried out based on the **Annotated tuberculosis image dataset**, which was obtained from the Makerere Automated Lab Diagnostics Database.[6] We first pre-processed the images from the image dataset to establish our own dataset with 2 classes, labeled as "TB" and "non-TB". Then, we performed classification on the established dataset with different networks and compared their performances. Last, we added a physical layer to

check if an improvement could be achieved by optimizing the weights on red, green, and blue color channels.

3.1 Image pre-processing

The Annotated tuberculosis image dataset consists of 928 sputum color images with bounding boxes of 3734 bacilli. The TB bacilli locations are recorded in a '.xml' file for each image. To get images with tuberculosis, we extracted the coordinates of each labeled box to find the center pixel of each TB cell. Then a 100 x 100 image is cropped around that center pixel to ensure the TB area has been fully preserved in the TB image.

To get the "non-TB" images, the bounding box areas identified as TB cells in original images are replaced with pixel of value of zeros. 10 random crops were performed for each image to get 100 x 100 non-TB images. If the image contains zero value pixel, that is, contains TB cell, or it exceeds the image boundary limits, image will not be saved.

We performed such operations on 500 original images and obtained 4584 "non-TB" images and 3492 "TB" images to feed into the networks.

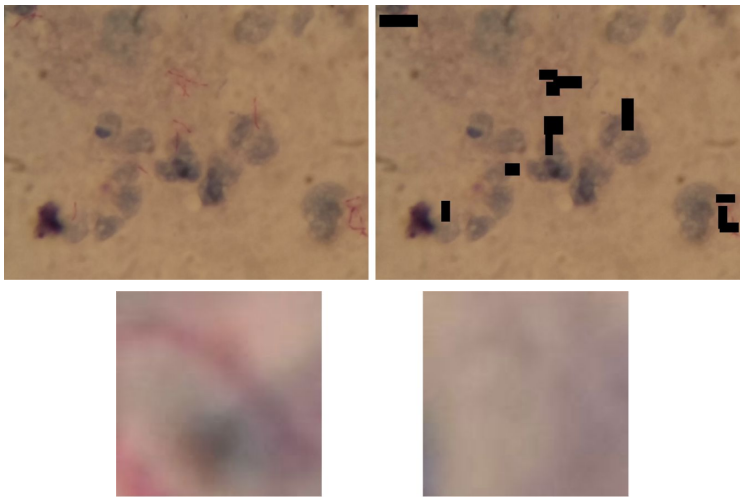

Figure 3.1 Original image from Makerere Lab (top left), Image without bacilli (top right), 100 x 100 "TB" Image (bottom left), and 100 x 100 "non-TB" Image(bottom right)

3.2 Classification on self-established dataset with different networks

After importing our "TB" and "non-TB" image dataset into tensorflow, we performed classifications with VGG, DNN and ResNet. VGG and DNN were built through 'keras.sequence()'. ResNet was built based on the official model of 'ResNet50' released in keras.applications. We applied 15 layers of densely-connected layers with output space of (*, 100) .

3.3 Optimization of weights of color channels

For the optimization of weights on color channels, we added a physical layer to the training network. Three trainable variables are created for each of the three color channels. Each weight is applied to a mask that is put on each color channel of the image. Images from training and validation dataset were then fed into a CNN to achieve the optimization of the three weights.

4 Results

4.1 Classification results of original images with different networks

In the classification of original images from the self-established dataset, we've used 3 types of neural networks, which are VGG, DNN and ResNet. The loss and accuracy during training and validation processes are plotted below. The x-axis is the number of epochs. The final loss and accuracy in both process as well as the sensitivity and specificity are shown in Table 5.1.

From Table 5.1 displayed below, all 4 networks performed well in the training data classification with accuracy >90%. VGG with 'tanh' activation has done the best job with an accuracy of 96.2%, while DNN has done the worst with an accuracy of 91.47%. However, when it comes to the validation dataset, the gap between the accuracies seemed to have grown larger. Both VGG networks have maintained their great accuracies in the validation dataset, while DNN and Resnet experienced obvious drops of accuracies.

In terms of the sensitivity and specificity, VGG networks gave great accuracy in both; DNN give high sensitivity and low specificity; and ResNet gave similar results for both.

4.1.1 VGG

In training of VGG networks, both 'relu' and 'tanh' activation method have been tested. Please see Appendix A for 'relu' method.

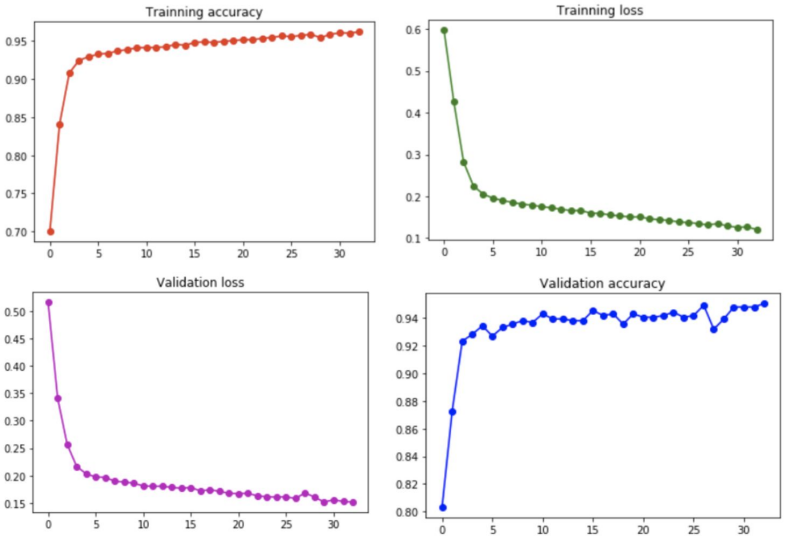

Figure 4.1 Classification result with VGG with 'tanh' activation of original images

4.1.2 DNN

In training of DNN networks, 'relu' activation have been tested.

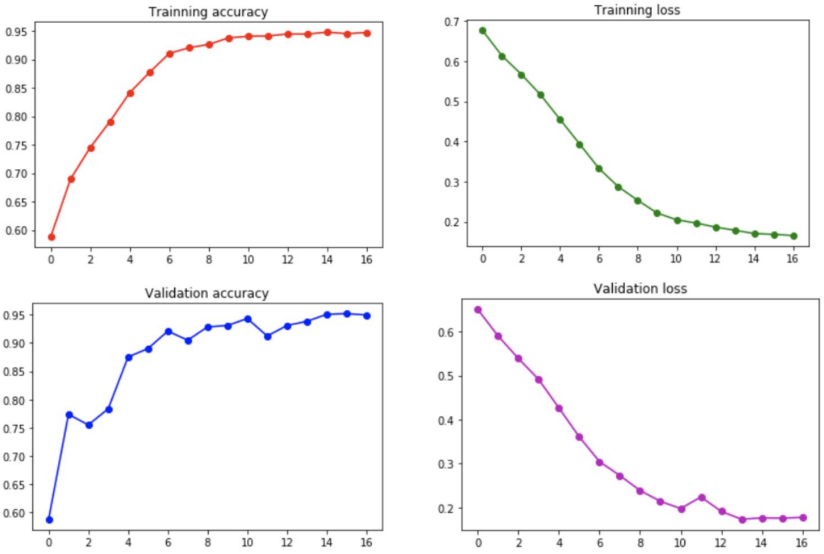

Figure 4.2 Classification result with DNN with 'relu' activation of original images

4.1.3 ResNet

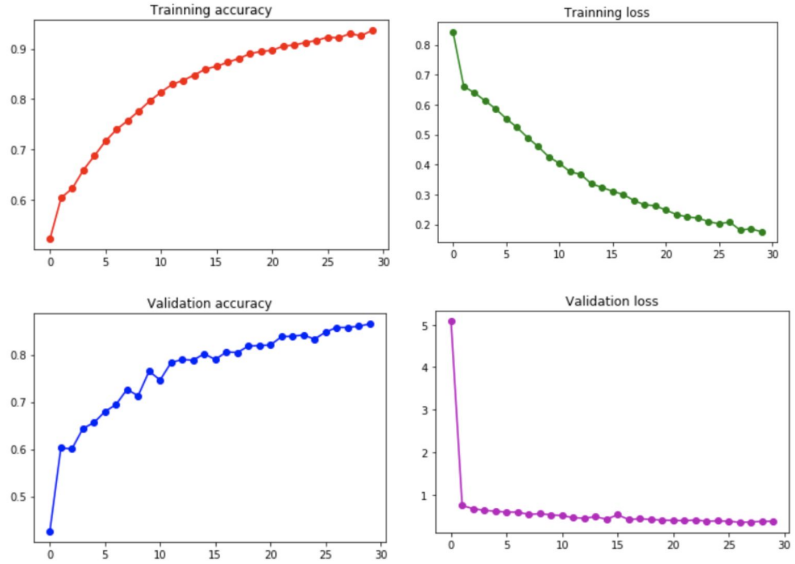

Figure 4.3 Classification results with ResNet

4.2 Classification results after trianning weights of 3 color channels

We compared the training results of the weights of all 3 color channels with and without adding the physical layer. For the untrained color channels case, the weight for each channel has been set as 1. Please see in Appendix A for the loss and accuracy plot.

For the trained color channels case, the weight for each channel before training is 1. After training, the weight of red channel becomes 0.99327266, green channel 1.0004733 and blue channel 1. Figure 4.4 shows the training loss and accuracy for every 1000 iterations.

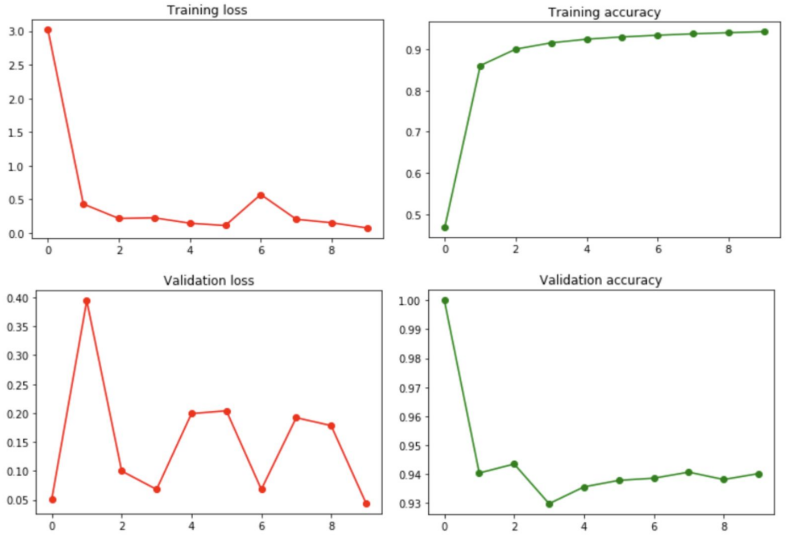

Figure 4.4 Classification result with trained weights of colors

5 Discussion

5.1 Classification results of original images with different networks

VGG performs the best compared to DNN and ResNet, while the most sophisticated ResNet performs poorly on the validation dataset. One possible reason accounting for this is that ResNet owns so many trained parameters that are over-defining the model. In the two activation methods we chose for VGG, 'tanh' activation works slightly better than 'relu' activation. In addition, VGG also yields the highest sensitivity and specificity. We are more interested in the sensitivity since it reflect the number of false negative in the results. It is crucial not to give false negative results when detecting TB cells since it acts as a screening method and should not miss any potential positives. As a result, we conclude that VGG with 'tanh' activation is the best networks for the classification purpose.

Table 5.1 Comparison of classification results of different networks

		VGG('relu')	VGG('tanh')	DNN	ResNet
Training	accuracy	0.9473	**0.9620**	0.9147	0.9362
	loss	0.1653	**0.1204**	0.2521	0.1755
Validation	accuracy	0.9493	**0.9505**	0.8948	0.8651
	loss	0.1777	**0.1518**	0.2661	0.3746
	sensitivity	0.9365	**0.9615**	0.9572	0.8632
	specificity	**0.9658**	0.9352	0.8088	0.8676

5.2 Results of classification with trained weights of color channels

It can be noticed in Table 5.2 that, the classification performances are almost the same when training with or without weighted color channels. In fact, the optimized weights for each

color channel is very close to 1, which means that physical layer have little effect in the training process. Therefore, the training results with or without the layer are similar as expected.

One reason might be that the original images are already in great quality regarding the setting of color channels, so there's no need to optimize it. And according to what we see from the original dataset, we agree that the original data is already clean and easy to distinguish from "TB" and "non-TB".

Table 5.2 Comparison between classification with and without weighted color channels

		without weights	with weights
Training	accuracy	0.9390	**0.9447**
	loss	0.0716	0.0751
Validation	accuracy	**0.9479**	0.9421
	loss	0.0771	0.0439
	sensitivity	**0.9407**	0.9364
	specificity	**0.9538**	0.9468

6 Conclusion and Future Work

Regarding the results displayed above, VGG seems to be the best choice for this classification purpose. It yields an accuracy of 95% and sensitivity of 96%. Much future work could be done to this project. We may add some noise to see if changing the weights of color channel can improve the classification performance. In addition, we could explore other physical layers such as feature extraction or illumination phase. Since we have a great dataset with bounded box for bacilli locations, we can also train neural networks to perform cell identification. In conclusion, deep neural network is a promising way to classify microscopic sputum with TB bacilli, but much work is needed in the future to turn it into an effective and accurate diagnostic tools for tuberculosis.

Reference

[1] Kochi, Arata. "The global tuberculosis situation and the new control strategy of the World Health Organization." (2001): 71-75.

[2] Mase, S. R., et al. "Yield of serial sputum specimen examinations in the diagnosis of pulmonary tuberculosis: a systematic review." *The International Journal of Tuberculosis and Lung Disease* 11.5 (2007): 485-495.

[3] Rulaningtyas, Riries, Andriyan B. Suksmono, and Tati LR Mengko. "Automatic classification of tuberculosis bacteria using neural network." *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*. IEEE, 2011.

[4] Simonyan, Karen , and A. Zisserman . "Very Deep Convolutional Networks for Large-Scale Image Recognition." *Computer Science* (2014).

[5] He, Kaiming , et al. "Deep Residual Learning for Image Recognition." (2015).

[6] "Automated Laboratory Diagnostics." *Automated Laboratory Diagnostics - AI Research Group, Makerere University*, air.ug/microscopy/.

Acknowledgement

APPENDIX A. Addition Results from Variations on Networks

Figure A.1 Classification result with VGG with 'relu' activation of original images

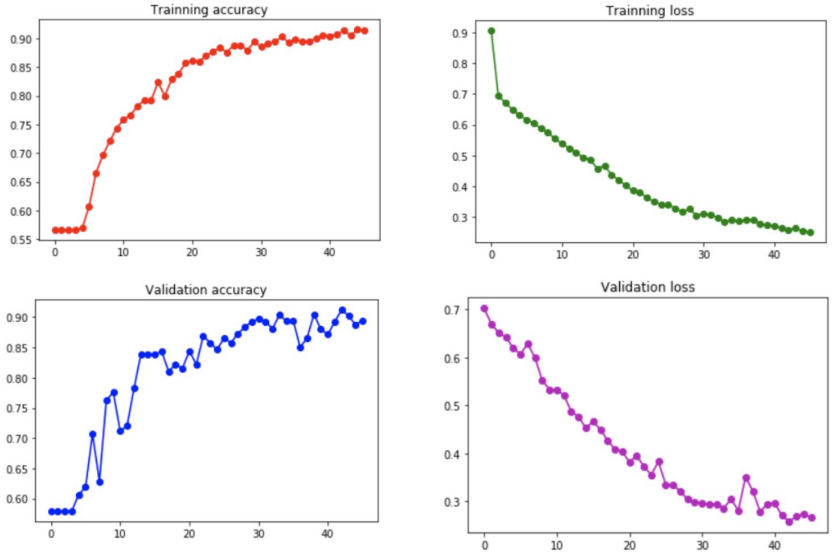

Figure A.2 Classification result with untrained weights of colors