Automated Detection of Pulmonary Tuberculosis in Chest X-Ray Images Using Convolutional Neural Network

Ryan Paul D.J. Gozum
Department of Computer Science
College of Engineering
University of the Philippines Diliman

Jan Samuel C. Matuba
Energy Engineering Program
College of Engineering
University of the Philippines Diliman

ABSTRACT

Tuberculosis is one of the leading causes of death globally according to World Health Organization [1]. This is unacceptable considering that TB is highly curable and preventable. Majority of infected populations was from resource-poor and marginalized communities with weak health care infrastructure. Early diagnosis of TB from chest X-ray images is critical in mitigating transmission and reducing the growth rate of TB epidemic [2]. In this paper, we want to develop a simple classifier targeted to address problems with delayed diagnosis of TB. Using supervised learning approach, we will be using a pre-trained Convolutional Neural Network from ImageNet to classify Chest X-Ray images as having manifestations of tuberculosis or as normal. By fine-tuning network hyperparameters on the new dataset, we have significantly reduced the cost of computation, while still obtaining an accuracy above 80% which is at par with state-of-the-art methods.

1. INTRODUCTION

Tuberculosis (TB) is an infectious disease caused by the bacillus Mycobacterium tuberculosis. TB is a leading cause of death by infectious disease worldwide, alongside human immunodeficiency virus-acquired immune deficiency syndrome (known as HIV-AIDS) [3]. Since 1990, there have been substantial improvements with detection and treatment of TB: mortality from TB reduced by 47%. Despite the achievements, TB still ranks alongside HIV as a leading cause of death worldwide. In 2014 alone, 1.5 million people died due to TB [3].

The relative lack of radiology interpretation expertise plays a critical role in early detection of TB. As the volume of data increases, it will become increasingly difficult for them to go through each X-Ray image while maintaining the same level of accuracy and efficiency [5]. Thus, developing a reliable and accurate computer-aided diagnosis (CAD) system can support radiologists in making quantitative and well-informed decisions. It will also contribute largely to reduced mortality caused by TB [6].

Rigorous studies on postero-anterior chest radiographs from various medical databases have been performed to extract informative and discriminative knowledge for classification. Previous approaches follow a serial process such as image preprocessing, boundary segmentation, feature extraction, and classification. For example, histogram analysis of the image (e.g. histogram equalization) is usually applied to enhance the contrast at the boundaries as preprocessing, and then region-of-interests (ROIs) that are informative regions for diagnosis (i.e. lung in case of TB detection) are segmented. To extract features from them, carefully designed shape and texture descriptors are considered, and finally popular classification algorithms are trained to classify normal and abnormal images. These features have been designed based on domain-specific knowledge. However, such manually designed features are limited to describe a number of variations existing within abnormal images [6].

Recent developments in Deep Learning, particularly the use of Convolutional Neural Networks, for various computer vision tasks such as object classification and detection spurred a renewed interest in applying techniques to medical images [5]. Rather than extracting specific knowledge, CNN has its way of learning the features from the dataset through a series of convolution and pooling steps.

In this work, we show the effectiveness of using a pre-trained CNN architecture called AlexNet for object classes that are not present in the ImageNet database. The datasets used are publicly available postero-anterior chest radiographs taken from the Department of Health and Human Services, Maryland, USA, Montgomery County and Shenzhen No. 3 Peoples Hospital in China. It is verified that transferring low-level filters from pre-trained models based on large-scale general images is very effective for training. Also, dropout regularization technique with keep probability of 0.5 helps reduce overfitting caused by lack of data. CNN gives a promising result accuracy during validation of above 80%, which is at par with state-of-the-art methods.

2. BACKGROUND AND RELATED WORK

Several machine learning techniques have been proposed over the years to improve accuracy and turnaround time for object classification tasks. In previous algorithms, they follow a serial process that requires carefully designing domain-specific models to extract discriminative features from medical images. Unfortunately, this manual technique can be difficult for images that have local translations. Deep Learning, specifically CNN, has gained popularity in computer vision tasks to overcome these challenges with regards to feature extraction. Some works by groups that recently applied CNN classifier for the Shenzhen dataset include Hwang (2016) which used the AlexNet architecture. Rohilla (2017) used smaller variations of AlexNet and VGGNet and achieved 70% and 80%, respectively. Lakhani (2017) used ensembles of AlexNet and GoogLeNet classifiers to achieve classification accuracy close to 0.99. The most comprehensive work so far was done by Islam (2017) who evaluated for AlexNet, VGGNet, and ResNet for binary classification and also for multi-class classification of different tuberculosis manifestations.

3. METHODOLOGY

The model was created using Tensorflow framework for Python and visualized in Tensorboard. Training was performed in a NVIDIA Telsa K80 GPU accelerator with 12 GB Memory, 61 GB RAM, 100 GB SSD.

3.1 Transfer Learning

As described earlier, transfer learning is a technique where a network is pre-trained in a much broader dataset (commonly ImageNet) and use the weights obtained as our starting point for a new classification task. This significantly speeds up the computation as transferring of low level filters helps the model converge faster.

3.2 Description of Datasets

In our paper, we combined two publicly available datasets as described below [9].

Montgomery County X-Ray Set: X-ray images in this data set have been acquired from the tuberculosis control program of the Department of Health and Human Services of Montgomery County, MD, USA. This set contains 138 posterior-anterior x-rays, of which 80 x-rays are normal and 58 x-rays are abnormal with manifestations of tuberculosis. All images are de-identified and available in DICOM format. The set covers a wide range of abnormalities, including effusions and miliary patterns. The data set includes radiology readings available as a text file.

Shenzhen Hospital X-ray Set: X-ray images in this data set have been collected by Shenzhen No.3 Hospital in Shenzhen, Guangdong providence, China. The x-rays were acquired as part of the routine care at Shenzhen Hospital. The set contains images in JPEG format. There are 326 normal x-rays and 336 abnormal x-rays showing various manifestations of tuberculosis.

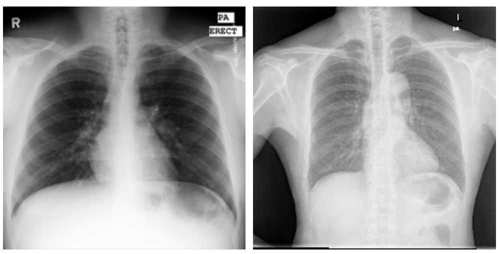

Fig. 1. MC X-Ray Set & Shenzhen X-Ray Set

3.3 Pre-Processing Step

Images from our dataset are converted to RGB format and rescaled to 227x227 as required input for AlexNet. We also performed image augmentation by applying Gaussian Blur filter with a radius of 0.7, and image rotations by 5°, -5°, 8°, and -8°. Train and Test data are partitioned into 7:1 (Dataset 1 - with augmentation) and 6:2 (Dataset 2 - no augmentation).

Table 1: Dataset partitions

	Training		Testing	
	Normal	Positive	Normal	Positive
Dataset 1	3440	3560	500	500
Dataset 2	294	306	100	100

3.4 Convolutional Neural Network

The results from related literature have shown that deep convolutional networks displayed significantly better performance compared to other models. In this project, the architecture AlexNet was explored. For simplicity, we obtained parameters from [10] that converted them from the format of Caffe to Tensorflow, both are popular Deep Learning frameworks. The constituting layers of the AlexNet architecture are shown in Fig. 2 and are also listed in Table 2.

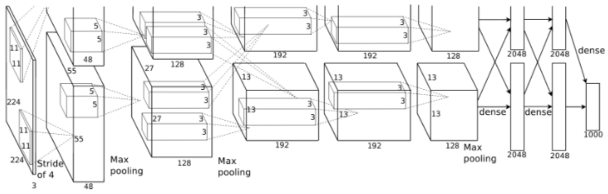

Fig. 2. Architecture of AlexNet

In this paper, our goal is to load the pre-trained weights from [10] to initial layers, while keeping the fully-connected layers fixed. We want to understand the effect of increasing the number of trainable layers in the overall accuracy of classification model.

Table 2: AlexNet Architecture

1	Convolutional 1
2	Pooling 1
3	Normalization 1
4	Convolutional 2
5	Pooling 2
6	Normalization 2
7	Convolutional 3
8	Convolutional 4
9	Convolutional 5
10	Pooling 3
11	Fully Connected 1
12	Fully Connected 2
13	Fully Connected 3
14	Softmax Classification

3.5 Performance Metrics

The main performance metric to be used is accuracy which is the ratio of number of correctly classified samples to total samples. Accuracy measure will enable direct comparison of performance of the developed model among prior studies which results have accuracy values around 90%. We also used gradient descent optimizer and cross-entropy function to calculate loss between predicted and true label in the softmax layer.

4. RESULTS AND DISCUSSION

The training is done for augmented or non-augmented datasets, and for two- or three- trained fully-connected layers, for a total of four different configurations. The accuracy and cross-entropy plots for the different sets are shown in the figures below while the respective runtimes are listed in Table 4. Then the summary of the results and the respective improvements as caused by increasing the number of trained fully-connected layers or by augmenting the dataset is shown in Table 3.

Fig. 3. Accuracy and Cross Entropy Plot for Dataset 1 (with Augmentation) when ran on batch size 100, epoch 50, with a learning rate of 0.001 and trainable layers are set to F7 and F8. Final validation accuracy: 89.00

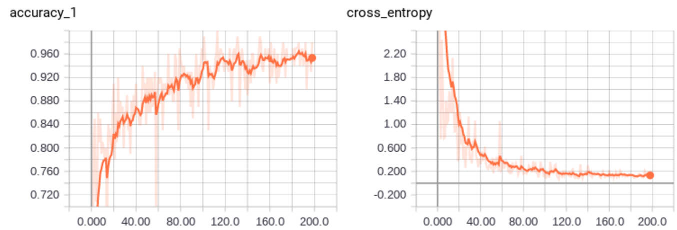

Fig. 4. Accuracy and Cross Entropy Plot for Dataset 2 (without Augmentation) when ran on batch size 100, epoch 200, with a learning rate of 0.001 and trainable layers are set to F7 and F8. Final validation accuracy: 84.50%

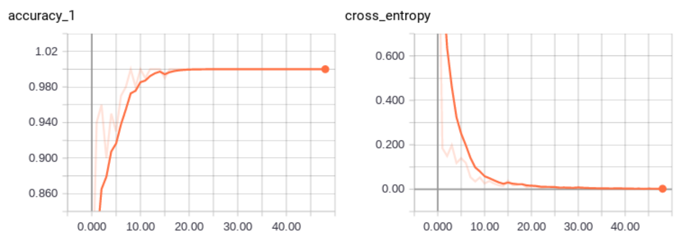

Fig. 5. Accuracy and Cross Entropy Plot for Dataset 1 (with Augmentation) when ran on batch size 100, epoch 50, with a learning rate of 0.001 and trainable layers are set to F6, F7, and F8. Final validation accuracy: 99.30%

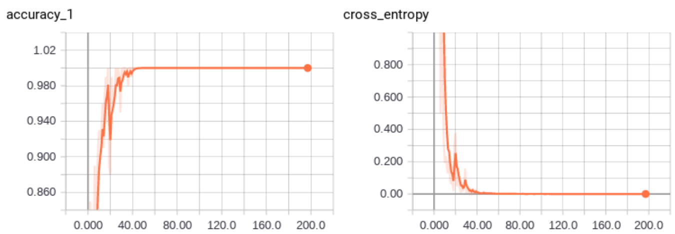

Fig. 6. Accuracy and Cross Entropy Plot for Dataset 2 (without Augmentation) when ran on batch size 100, epoch 50, with a learning rate of 0.001 and trainable layers are set to F6, F7, and F8. Final validation accuracy: 86.00%

We can observe from the figures above that using augmentation technique (as in Dataset 1) gives a higher accuracy during validation compared to the original dataset that was trained on higher epoch due to lack of training data. Increasing the number of trainable layers for the model also gives results that are above 90%, which is at par with recent studies developed on x-ray image analysis that trains data from scratch.

Table 3: Net improvement in terms of accuracy with increasing number of trainable layers

Accuracy	2-Layer	3-Layer	Net Improvement
Dataset 1	89.0%	**99.3%**	10.30%
Dataset 2	84.5%	86.0%	1.50%
Improvement	4.50%	13.30%	-

We can also conclude that employing the transfer learning technique helped speed up the computation without compromising the accuracy of classification. Training a complex network from scratch entails the usage of expensive computational resource, thus training usually take hours or days to finish. The table below shows the running time using our approach:

Table 4: Runtime

Accuracy	2-Layer	3-Layer
Dataset 1	36m 35s	35m 37s
Dataset 2	22m 44s	21m 46s

5. CONCLUSION

In this work, we successfully developed a binary classifier of tuberculosis from the Shenzhen and Montgomery County chest X-Ray image datasets, based from a AlexNet architecture. Accuracy levels of 84.5% and 86.0% were achieved for non-augmented datasets with two and three trained layers, respectively. Using data augmentation techniques including Gaussian blur and rotation, the augmented dataset achieved significantly better accuracy levels at 89.0% and 99.3% for the two and three trained layers, respectively. Therefore, increasing the number of trainable layers and augmenting the dataset and were found to improve the performance of a pre-trained AlexNet model for classifying tuberculosis.

6. REFERENCES

[1] Alcantara, M. F., Cao, Y., Liu, C., Liu, B., Brunette, M., Zhang, N., Curioso, W. (2017). Improving Tuberculosis Diagnostics using Deep Learning and Mobile Health Technologies among Resource-poor Communities in Peru. *Smart Health.*

[2] Cao, Y., Liu, C., Brunette, M., Zhang, N., Sun, T., Zhang, P. (2016). Improving Tuberculosis Diagnostics using Deep Learning and Mobile Health Technologies among Resource-poor and Marginalized Communities. *IEEE First Conference on Connected Health: Applications, Systems and Engineering Technologies.*

[3] World Health Organization (WHO), Global tuberculosis report, 2015, *http://www.who.int/tb/publications/global$_r$eport/en/.*

[4] Chang, L., Cao, Y., Alcantara, M., Liu, B., Brunette, M., Peinado, J., Curioso, W. (2017). TXCNN:

Detecting Tuberculosis in Chest X-Ray Images Using Convolutional Neural Network. *The International Conference on Image Processing*. Beijing, China.

[5] Islam, M., Aowal, M., Minhaz, A., Ashraf, K. (2017). Abnormality Detection and Localization in Chest X-Rays using Deep Convolutional Neural Networks. *Computer Vision and Pattern Recognition*.

[6] Hwang, S., Kim, H.-E., Jeong, J., Kim, H.-J. (2016). A Novel Approach for Tuberculosis Screening Based on Deep Convolutional Neural Networks. *Journal of Medical Imaging*.

[7] Lakhani, P., Sundaram, B. (2017). Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*.

[8] Rohilla, A., Hood, R., Mittal, A. (2017). TB Detection in Chest Radiograph Using Deep Learning Architectures. *International Journal of Advance Research in Science and Engineering*.

[9] Jaeger S, Candemir S, Antani S, Wng YX, Lu PX, Thoma G. Two public chest X-ray datasets for computeraided screening of pulmonary diseases. Quant Imaging Med Surg 2014;4(6):475-477.

[10] http://www.cs.toronto.edu/guerzhoy/tf_alexnet/