International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

Automated Tuberculosis Classification of Chest Radiographs by Using Convolutional Neural Networks

Rahul Hooda
Punjab Engineering College, Chandigarh
Ajay Mittal
UIET, Panjab University, Chandigarh

ABSTRACT

Tuberculosis (TB) is a major health threat all over the world. Many people in the developing countries die every year due to lack of treatment and faulty diagnosis. Developing an accurate computer-aided diagnosis (CAD) system for TB detection can help in early diagnosis and containing the disease. Most of the current CAD systems use handcrafted features, however, lately there is a shift towards using deep-learning-based automatic feature extractors for TB detection. In this paper, a method for TB detection is presented which uses a deep learning architecture, ResNet. The architecture has been trained from scratch and no pre-trained weights have been used. It is a 34-layered architecture which uses skip layers. For performing the experiment, images from four different datasets namely Montgomery, Shenzhen, Belarus and JSRT datasets have been used. After manually optimizing different hyper-parameters, we have obtained validation accuracy and testing accuracy of 99.41% and 84.12% respectively.

Keywords

Deep Learning, Tuberculosis, Medical Image Processing, Chest X-rays

1. INTRODUCTION

Tuberculosis (TB) is the leading cause of death worldwide, over-leaping HIV/AIDS in the year 2016. About one-quarter of the world's population is latently infected with it, however, people with the compromised immune system such as population with HIV, diabetes or using tobacco are at a higher risk of developing active tuberculosis. Currently, there is 5-15% lifetime risk of developing active tuberculosis from latent one. According to World Health Organization's 2016 TB report [1], 10.4 million people developed TB and 1.7 million died from the disease (including 0.4 million people with HIV). Majority of the actively infected population (64 %) are from seven countries only with India leading the chart. Most of the incident cases are from south-east Asia (45%) and Africa (25%) region. While TB mostly affects men (65%) in their most productive years, disease burden among women and children is also high and increasing. In 2016, 1 million children diagnosed with TB and 250 thousand (including children with HIV) died from the disease.

Tuberculosis is a disease caused by the bacillus Mycobacterium tuberculosis. TB can be present in any part of the body but it mainly affects the lung region and is termed as *Pulmonary TB* (PTB). It is a contagious disease and can easily spread from an infected person to normal person from either coughing or sneezing. There are different radiological procedures such as chest radiography and computed radiography which can be used for primary screening of TB. However, due to cost effectiveness, low radiation dose, easier availability in health centers and wider acceptability among patients, the chest radiograph is preferred and recommended by the WHO to perform the mass screening. Mass screening is performed to the avoid majority of normal cases from undergoing costly confirmatory tests such as sputum culture and smear microscopy. Chest radiographs, colloquially known as chest x-ray (CXRs), can have different pathological patterns or manifestations inside the lung region to exhibit TB signs. These manifestations are shown in Figure 1.

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

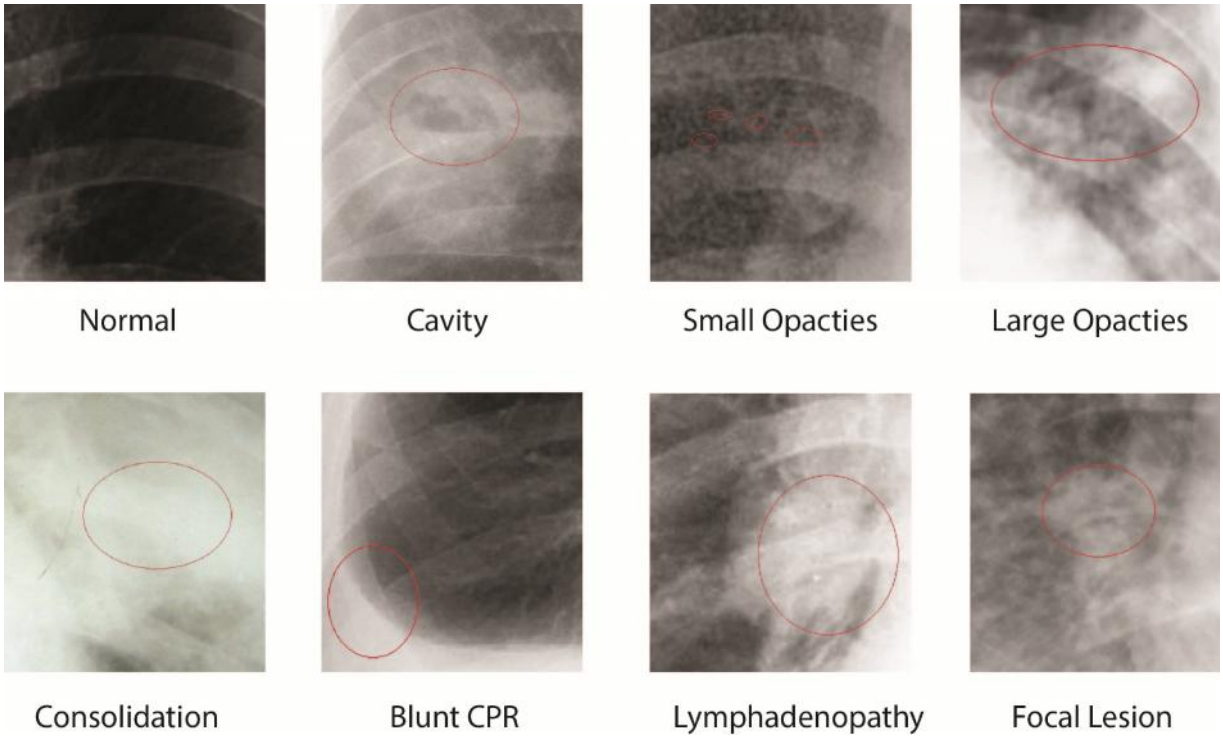

Figure 1: Different manifestations of TB in CXRs

Traditionally, CXRs were analyzed by the radiologists manually. However, due to the high burden of patients and lack of skilled radiologists at the remote location, there is a high rate of human error while analyzing the CXRs which ultimately results in many unreported cases. This leads to detection of TB at a later stage in patients and thus chances of their survival slim down. To improve TB detection during the screening process, researchers are developing computer-aided diagnosis (CAD) systems since the 1970s. These systems provide their diagnosis automatically and thus can be used in remote locations to aid the radiologists.

Initially, CAD systems consist of four phases namely pre-processing, segmentation, feature extraction, and classification. In these systems, firstly region-of-interest is segmented and thereafter important features are extracted using handcrafted methods to form a feature vector. However, after the resurrection of deep learning in 2012, researchers drifted away from this strategy. In deep-learning-based methods, segmentation is not necessarily required and best features are automatically extracted using the end-to-end architecture of these methods as shown in Figure 2. In this paper, we have applied a deep-learning-based method, which classifies CXR images into TB positive and normal images.

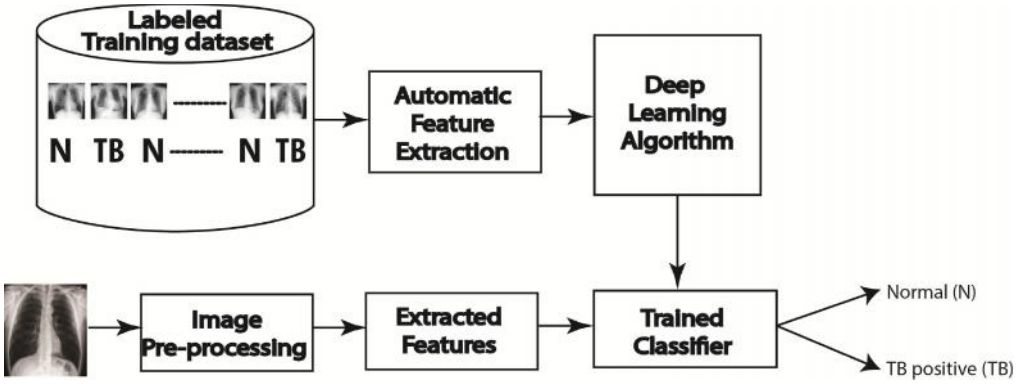

Figure 2: Deep learning-based CAD system

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

The rest of the paper is organized as follows: Section 2 presents the related work done to date. The proposed method and the datasets used are presented in Section 3. Experimental results are presented and discussed in Section 4.

2. RELATED WORK

Detection of TB is a challenging task due to the presence of different types of manifestations such as cavities, small opacities, large opacities, consolidation, focal lesions, and nodules on a CXR. In most of the research studies, hand-crafted features are extracted and a classifier is applied to classify CXR images. For detecting specific patterns on a CXR, textural and geometrical features are most prominently used.

Ginneken *et al.* [2] were the first to propose a proper method for automatic TB detection. The local texture of CXRs is extracted and analyzed to detect different abnormalities. In this method, several overlapping regions are obtained by dividing lung fields and from each region, texture features are extracted. At region level, the *k*-nearest neighbor is used to obtain classification scores and the weighted multiplier is then used to combine these values to obtain a final score. This method showed that the overall performance of the system enhances when scores from different regions are combined. Hogeweg *et al.* [3] presented another method in which scores are combined different detection systems. In this method, small circular patches are obtained from each CXR image and using these patches, features are extracted and LDA (Linear Discriminant Analysis) classifier to obtain texture score. This score is obtained at the pixel level and is later combined with clavicle detection module to remove false positive cases. Mahalanobis distance at image level is used to obtain shape abnormality score and this score is combined with previous one to obtain a final score. Hogeweg *et al.* in [4] further improved the TB detection method by including the focal abnormality in addition to textural and shape abnormalities. All the values obtained from above individual modules are combined to achieve the final score. The focal analysis is performed by using the commercially available ClearRead software.

Jaeger *et al.* [5] proposed a TB detection method in which different masks are used for lung segmentation. From the segmented portion, different shape and texture descriptors are extracted to find the pathological patterns in the CXR. For each descriptor, histogram bins are evaluated and each histogram value is considered as a feature to form the feature vector. Linear support vector machine (SVM) is used as a classifier to classify the CXR images into normal and abnormal classes. These authors in [6], also proposed a similar automated method in which two separate feature sets namely object detection based features and CBIR (Content-based Image Retrieval) based features are used, after segmenting lung boundary using graph cut segmentation method. Finally, SVM is used as the classifier for TB classification of CXRs.

In the literature, some methods are also presented which address only one specific type of manifestation such as pleural effusion. Karargyris *et al.* [7] proposed a method to detect pulmonary abnormalities by using shape and texture features. SVM classifier is used separately on shape features as well as texture features. The method showed that separate processing of feature sets leads to faster processing and also improves the final accuracy. Maduskar *et al.* [8] also proposed an algorithm to detect the pleural effusion. In this method, anatomic landmark information and refined lung segmentation are used to localize the costophrenic recess. The feature vector is formed using features which include angle, intensity and morphological information based region descriptors located around the costophrenic recess. Abnormality score is obtained by applying random forest classifier and the score indicates the severity of pleural effusion. The disadvantage of this method is that it does not detect pleural effusion when chest wall or pulmonary fissure is affected and costophrenic recess is not observed.

Recently, the research paradigm has shifted towards deep-learning-based methods for TB classification. Initial deep convolutional neural network (CNN)-based method for TB detection was proposed by Hwang *et al.* [9] using the pre-trained Alexnet network [10] and transfer learning. In transfer learning, weights of pre-trained models are used as it is for most layers and training is performed for some lower layers (known as *fine-tuning*). Another CNN-based method is presented by Cao *et al.* [11]. They have prepared a TB dataset which

includes manifestations of different types. A mobile solution has also been proposed in which a CXR can be uploaded to the server and the screening result can be obtained. They performed the experiments using the pre-trained model of GoogleNet [12] for binary as well as multi-class classification.

3. MATERIAL AND METHODS

Datasets

In this study, images from four publicly available datasets are combined to form the final dataset. These dataset includes Montgomery dataset [13], Shenzhen dataset [13], Belarus Dataset [14] and JSRT dataset [15]. The details of these datasets are mentioned below:

Montgomery Dataset: The dataset was created by U.S. National Library of Medicine (USNLM) using the services of the health department at Montgomery County (MC), USA. It consists of 138 postero-anterior (PA) CXRs collected under MC's tuberculosis screening program. The dataset contains 80 CXRs that are normal while remaining 58 CXRs have TB manifestations. The images have been acquired at two different spatial resolutions of *4020X4892* or *4892X4020* having the grayscale depth of 12 bits. In this study, all the images from the dataset have been used.

Shenzhen Dataset: This dataset was created by USNLM in association with Guangdong Medical College, Shenzhen, China. It consists of 662 CXRs, containing 326 normal CXRs and 336 TB manifested CXRs. The dataset also includes some antero-posterior (AP) CXRs in addition to PA CXRs, and thus have images of varied sizes. The size of most of the images is approximately *3000X3000* pixels. The images are available in the .jpeg format and have pixel size as 0.353 mm. All the 662 images have been used in this study to form combined dataset.

Belarus Dataset: This dataset consists of temporal CXRs and chest computed tomography (CT) scans of 420 patients. It is collected by various institutes under Ministry of Health, the Republic of Belarus to study drug-resistant TB. The spatial resolution of CXRs in the dataset is *2248X2248* pixels. In addition to the medical images, the dataset includes clinical data consisting of information about each patient. All the images are of TB infected patients. In this study, 240 CXRs images of different patients have been selected from the dataset and added to the combined dataset.

JSRT Dataset: The dataset was created by Japanese Society of Radiological Technology (JSRT) and Japanese Radiological Society. It consists of 247 PA CXRs collected from different institutions mostly from the United States of America. Of all, 154 CXR images have lung nodules, while 93 have none. The CXR images have a size of *2048X2048* pixels, with the grayscale depth of 12 bits. The CXR images are available as disk image files with .img extension. Only 93 normal images have been used in this study to balance normal and TB manifested images in the combined dataset.

Method

For performing image classification, different architectures such as AlexNet, GoogleNet, VGGNet have been proposed in the literature. All of these architectures have been presented in ImageNet Large Scale Visual Recognition Competition (ILSVRC) and have been evaluated on ImageNet dataset. Since winning ILSVRC competition in different years, these architectures have been used in many computer vision and image processing applications to solve different problems. In this study, we have used ResNet (short for *Residual Networks*) architecture proposed by Kaiming He *et al.* [16] in 2015 and won all the ImageNet challenges for classification, detection, and localization in that year. Different variants of the ResNet architecture have been presented with the different number of layers and in this study, the architecture with 34 layers have been used. The intuition behind proposing this architecture is to add extra layers and increasing the depth without decreasing the accuracy or increasing the error.

The architecture uses skip layers (also called *identity shortcut connections*) to increase the number of layers in the architecture without decreasing the performance. In shortcut connections, in addition to providing output

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

to next layer, the output of the layer is also skipped to be added to the output of another layer. In this study, the ResNet architecture has been customized to perform classification of CXR images into two classes, that is, TB negative and TB positive. The hyper-parameters of the architecture are optimized after performing various experiments.

The ResNet architecture is shown in Figure 3. It contains 34 layers which consist of 33 convolutional layers and one fully-connected layer. The first convolutional layer uses 64 7X7 filters with stride 2. Next layer is max-pool to further reduce the output size by half. After that, 32 convolutional layers are applied, which uses skip layer after the interval of two layers. All these layers use filters of 3X3 size. Two convolutional layers with skip addition constitute one block of the architecture and thus the architecture contains 16 blocks. These blocks are combined to have four sets with each having 3, 4, 6 and 3 blocks respectively. Each block uses a specific number of filters which may be different for different blocks. Therefore, every convolutional layer in the first set uses 64 filters, in the second set uses 128 filters, in the third set uses 256 filters and in the last set uses 512 filters. Batch normalization and Rectified Linear Unit (ReLU) are used before the first convolutional layer of each set. In all sets except first, convolutional with stride 2 is used as the first convolutional layer and remaining layers are with stride 1. Finally, a fully-connected layer is used to obtain the classification output for the architecture. ResNet architecture is used because it converges very fast and provides better performance than other deep architectures.

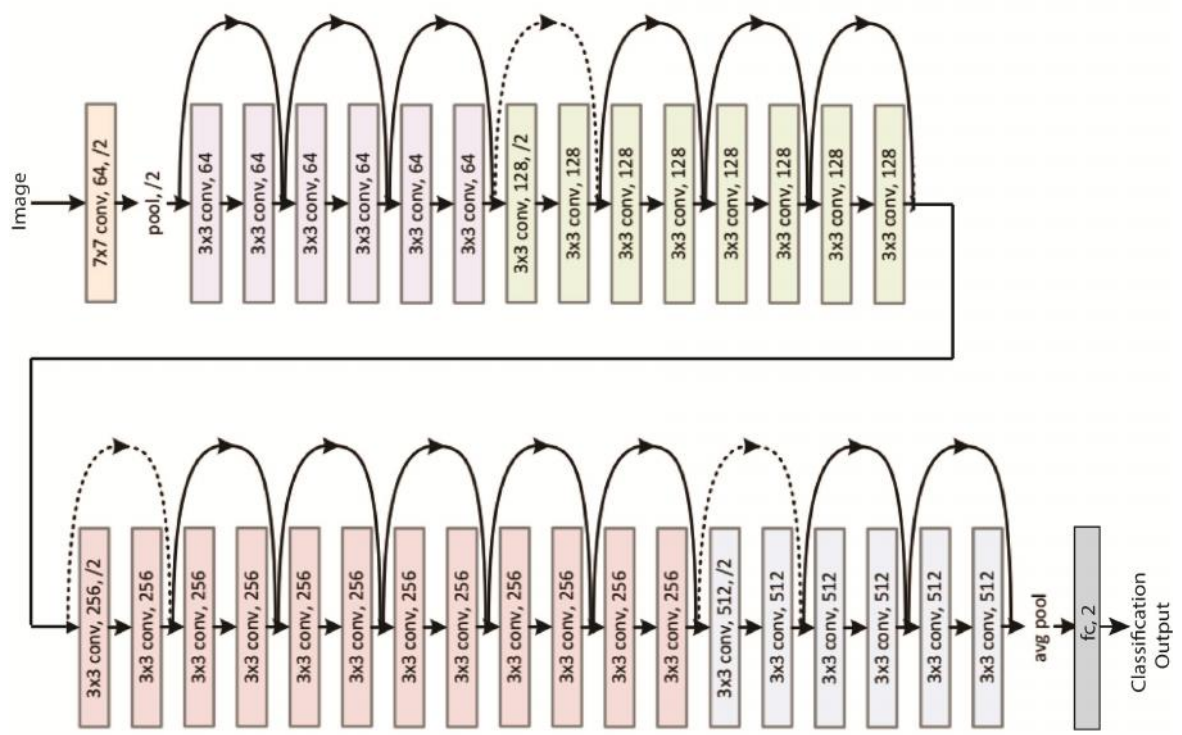

Figure 3: ResNet Architecture

4. RESULTS

Results are obtained on the combined dataset of 1133 images. Out of these, 499 are normal images and 634 are TB manifested images. The combined dataset is divided into three parts used as training, validation and test dataset in (70:15:15). This means that 793 images are used for training while 170 images each are used for validation and testing purpose. Since training images are less than the usual requirement of deep learning architectures, therefore, data augmentation techniques are applied to enhance the number of training images.

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

For each image in the training set, six different images are generated and added to the set. Out of these 6 images, three augmented images are obtained by rotating the image by 90, 180 and 270 degrees. Two augmented images are retrieved by executing horizontal and vertical flips. The last augmented image is obtained by performing histogram equalization on the input image. By obtaining the above-augmented images, the total number of training images is increased to 5551.

In this study, validation accuracy and testing accuracy are calculated. Validation accuracy is equal to the number of correctly predicted CXRs to the total number of CXRs in the validation set. Testing accuracy is equal to the number of correctly predicted CXRs to total number of CXRs in the test set. It is calculated once when the training of the architecture completes. The training images are divided into batches of size 100. Batch size is used since it is not possible to provide all the images to the architecture at once due to large memory requirements. Small batch size also helps in achieving better learning since different images are provided in each batch. The architecture is trained for 30 epochs which include 1680 steps (with the number of steps in each epoch = training images/ batch size = 56). It requires less number of epochs for training as compared to other architectures such as AlexNet and VGGNet since it has deeper architecture and uses skip layers. Each layer of the architecture has number of parameters or weights which need to be optimized. These weights are initialized to random values with zero mean and standard deviation 0.01. Loss value is calculated after each epoch using cross-entropy function. Based on the evaluated loss value, weights are updated for each layer. This process continues till the weights for each layer is optimized and loss value reduces to the minimum.

Table 1: Comparison of proposed method with other methods presented in the literature. Here, AUC stands for Area under the curve and ACC stands for Accuracy.

Authors	# of Images	Performance
Ginneken *et al.* [2]	388	AUC: 82.00%
Hogeweg *et al.* [3]	400	AUC: 86.80%
Jaeger *et al.* [5]	138	ACC: 75.00%
Jaeger *et al.* [6]	753	AUC: 90.00%
Maduskar *et al.* [8]	1267	AUC: 87.00%
Proposed Method	1133	ACC: 84.12%

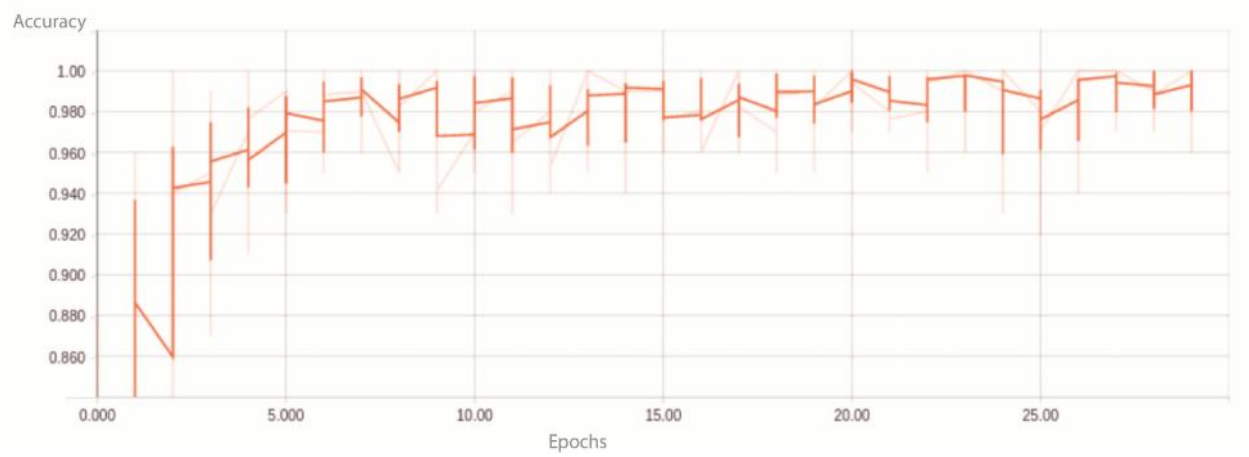

Figure 4: Training accuracy of the network after each epoch.

In addition to batch size, some other hyper-parameters such as optimization algorithm and learning rate are manually selected after performing several experiments. Adam optimizer is used as optimization algorithm

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 3
March 2018

since it performed better than other evaluated algorithms that are stochastic gradient descent (SGD) and momentum algorithm. The decaying learning rate is used in the experiments with 10^{-3} as starting learning rate and decayed to 96% after every 250 steps. The training accuracy and training loss are shown in Figure 4 and Figure 5 respectively. The validation accuracy and testing accuracy obtained by the architecture is 99.41% and 84.12% respectively. In Table I, the performance of the evaluated method is compared with other TB methods in the literature. The table shows that the performance better than most of the methods. It can be concluded that deep learning architecture shows promising results and can better the performance of handcrafted feature-based methods in different fields. In this study, Nvidia Titan X GPU processor is used to evaluate the architecture built using Tensorflow library of Python.

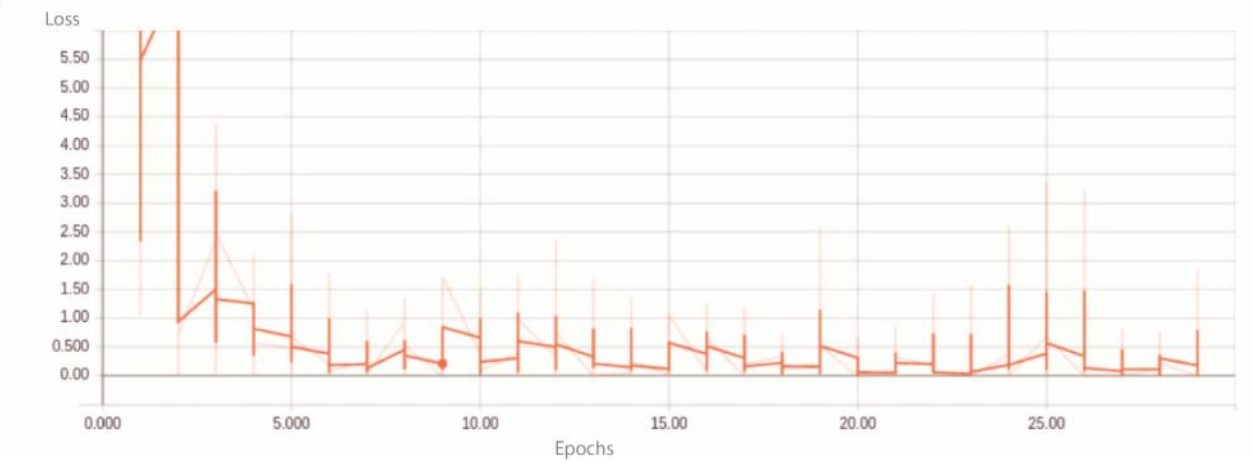

Figure 5: Training loss of the network after each epoch.

5. CONCLUSION

In this study, TB classification is performed in the CXRs using a deep architecture named ResNet. In deep architectures, best features are automatically extracted based on the training images and their outputs. The dataset for this study is prepared by combining the images from four different datasets. The number of training images is increased by using data augmentation techniques. The architecture learns features from the scratch and achieves the testing accuracy of 84.12%. The performance of the architecture is better than most of the methods presented in the literature. Future steps in the study will be to compare the performance of different architectures and use transfer learning and fine-tuning to improve the performance. It can also be tested whether providing lung segmented images to the architecture instead of original images improves the performance or not.

REFERENCES

[1] World Health Organization. "Global tuberculosis report 2016." (2016).
[2] Van Ginneken, Bram, Shigehiko Katsuragawa, Bart M. ter Haar Romeny, Kunio Doi, and Max A. Viergever. "Automatic detection of abnormalities in chest radiographs using local texture analysis." *IEEE transactions on medical imaging* 21, no. 2 (2002): 139-149.
[3] Hogeweg, Laurens, Christian Mol, Pim A. de Jong, Rodney Dawson, Helen Ayles, and Bram van Ginneken. "Fusion of local and global detection systems to detect tuberculosis in chest radiographs." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 650-657. Springer, Berlin, Heidelberg, 2010.
[4] Hogeweg, Laurens, Clara I. Sánchez, Pragnya Maduskar, Rick Philipsen, Alistair Story, Rodney Dawson, Grant Theron, Keertan Dheda, Liesbeth Peters-Bax, and Bram Van Ginneken. "Automatic detection of tuberculosis in chest radiographs using a combination of textural, focal, and shape abnormality analysis." *IEEE transactions on medical imaging*34, no. 12 (2015): 2429-2442.

[5] Jaeger, Stefan, Alexandros Karargyris, Sameer Antani, and George Thoma. "Detecting tuberculosis in radiographs using combined lung masks." In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 4978-4981. IEEE, 2012.

[6] Jaeger, Stefan, Alexandros Karargyris, Sema Candemir, Les Folio, Jenifer Siegelman, Fiona Callaghan, Zhiyun Xue et al. "Automatic tuberculosis screening using chest radiographs." *IEEE transactions on medical imaging* 33, no. 2 (2014): 233-245.

[7] Karargyris, Alexandros, Jenifer Siegelman, Dimitris Tzortzis, Stefan Jaeger, Sema Candemir, Zhiyun Xue, K. C. Santosh et al. "Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays." *International journal of computer assisted radiology and surgery* 11, no. 1 (2016): 99-106.

[8] Maduskar, Pragnya, Rick HMM Philipsen, Jaime Melendez, Ernst Scholten, Duncan Chanda, Helen Ayles, Clara I. Sánchez, and Bram van Ginneken. "Automatic detection of pleural effusion in chest radiographs." *Medical image analysis*28 (2016): 22-32.

[9] Hwang, Sangheum, Hyo-Eun Kim, Jihoon Jeong, and Hee-Jin Kim. "A novel approach for tuberculosis screening based on deep convolutional neural networks." In *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785, p. 97852W. International Society for Optics and Photonics, 2016.

[10] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." In *Advances in neural information processing systems*, pp. 1097-1105. 2012.

[11] Cao, Yu, Chang Liu, Benyuan Liu, Maria J. Brunette, Ning Zhang, Tong Sun, Peifeng Zhang et al. "Improving Tuberculosis Diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities." In *Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2016 IEEE First International Conference on*, pp. 274-281. IEEE, 2016.

[12] Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." Cvpr, 2015.

[13] Jaeger, Stefan, Sema Candemir, Sameer Antani, Yì-Xiáng J. Wáng, Pu-Xuan Lu, and George Thoma. "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases." *Quantitative imaging in medicine and surgery* 4, no. 6 (2014): 475.

[14] http://tuberculosis.by/

[15] Shiraishi, Junji, Shigehiko Katsuragawa, Junpei Ikezoe, Tsuneo Matsumoto, Takeshi Kobayashi, Ken-ichi Komatsu, Mitate Matsui, Hiroshi Fujita, Yoshie Kodera, and Kunio Doi. "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules." *American Journal of Roentgenology* 174, no. 1 (2000): 71-74.

[16] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.