



Diagnosing Tuberculosis Using Deep Convolutional Neural Network

Mustapha Oloko-Oba and Serestina Viriri^(✉)

School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal, Durban, South Africa
mustaphasmo@yahoo.com, viriris@ukzn.ac.za

Abstract. One of the global topmost causes of death is Tuberculosis (TB) which is caused by mycobacterium bacillus. The increase rate of infected people and the recorded deaths from TB disease is as a result of its transmissibility, lack of early diagnosis, and inadequate professional radiologist in developing regions where TB is more prevalent. Tuberculosis is unquestionably curable but needs to be detected early for necessary treatment to be effective. Many screening techniques are available, but chest radiograph has proven to be valuable for screening pulmonary diseases but hugely dependent on the interpretational skill of an expert radiologist. We propose a Computer-Aided Detection model using Deep Convolutional Neural Networks to automatically detect TB from Montgomery County (MC) Tuberculosis radiographs. Our proposed model performed at 87.1% validation accuracy and evaluated using confusion matrix and accuracy as metrics.

Keywords: Chest radiograph · Convolutional neural network · Preprocessing · Feature extraction · Tuberculosis · Classification

1 Introduction

Tuberculosis (TB) is a contagious disease that is considered worldwide as a significant source of death from a single transmittable agent as well as among the top 10 sources of death [1, 2]. The TB disease is caused by the mycobacterium tuberculosis bacillus which is easily contractible by having close contact with infected individuals. This disease mostly affects the lungs but can as well affects other parts of the body [3, 4].

World Health Organization (WHO) estimated about 10 million individuals fell sick as a result of Tuberculosis disease in 2018 which resulted in about 1.2 million deaths from the previous 1.3 million deaths recorded in 2017 [2].

Tuberculosis disease is more prevalent in developing regions and can affect both males and females but more prominent in males. Among the total number of individuals infected with tuberculosis in 2017, 1 million cases were reported in children aged less than 14, 5.8 million cases in males, and 3.2 million cases in females [1].

Tuberculosis disease is certainly curable but needs to be detected early for appropriate treatment. Several lung examination techniques are available but the Chest radiographs conversationally known as chest X-ray or CXR for short is a prominent screening tool for detecting abnormalities in the lungs [5, 6, 14]. Basically TB manifestation can be detected on CXR, however, quality CXR imaging equipment along with skilled radiologists to accurately interpret the CXR is either limited or not available in TB prevailing regions [7, 8].

A geographical report by the World Health Organization of most Tuberculosis cases for 2018 is shown in Table 1.

Table 1. Geographical TB cases (WHO, 2018)

| Region | Cases (%) |
|-----------------------|-----------|
| South-East Asia | 44 |
| Africa | 24 |
| Western Pacific | 18 |
| Eastern Mediterranean | 8 |
| The Americas | 3 |
| Europe | 3 |

Due to the deadly nature of TB disease and the rate at which it can easily be spread, WHO has laid emphasis on more proactive measures for continuous reduction of TB cases and deaths [9]. Also, the decision to embark on a mission to put an end the universal TB epidemic by the year 2030 is underway as contained in the 2019 Global Tuberculosis report [2].

The lack of skilled radiologist, high number of patients waiting in line to be screen and mostly outdated equipment which results in a high rate of errors in properly screening the CXR remain a major problem that requires prompt attention.

As a result, to profer solution to the issue of limited or lack of expert radiologist and misdiagnosis of CXR, we propose a Deep Convolutional Neural Networks (CNN) model that will automatically diagnose large numbers of CXR at a time for TB manifestation in developing regions where TB is most prevalent. The proposed model will eliminate the hassle of patients waiting in line for days to get screened, guarantee better diagnosis, performance accuracy and ultimately minimize cost of screening as opposed to the process of manual examination of the CXR which is costly, time-consuming, and prone to errors due to lack of professional radiologist and huge number of the CXR piled up to be diagnosed.

2 Related Work

The evolution of computer-aided detection and investigative systems has offered a new boost attributed to the emerging digital CXR and the ability of computer

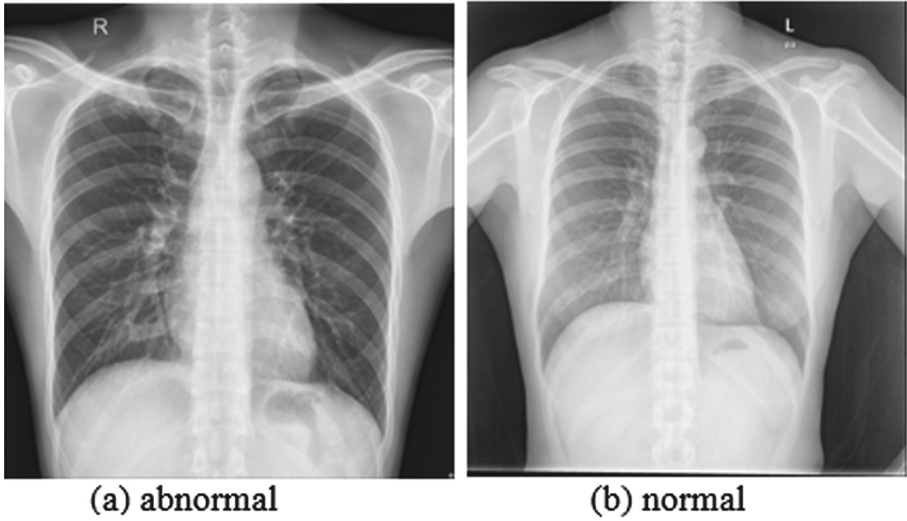


Fig. 1. Sample of the normal and abnormal chest X-ray.

vision for screening varieties of health diseases and conditions. Although much impressive research has been carried out in the last few years regarding computer-aided detection, nevertheless lots more finding is required in the field medical imaging to improve the existing methods and find convenient lasting solutions to deadly medical conditions as the case of Tuberculosis and many more.

A processing method that combines the Local Binary Pattern with laplacian of gaussian was employed in [10] for the manual detection of Tuberculosis nodules in CXR. This research centers on accentuating nodules by the Laplacian of Gaussian filter, lung segmentation, ribs suppression and the use of local binary pattern operators for texture classification. Computer-aided diagnosis system presented by [11] for screening Tuberculosis using two different Convolutional Neural Networks architectures (Alexnet and VGGNet) to classify CXR into positive and negative classes. Their experiment which is based on Montgomery and Shenzhen CXR datasets found VGGNet outperformed Alexnet as a result of a deeper network of VGGNet. The performance accuracy of 80.4% was obtained for Alexnet while VGGNet reached 81.6% accuracy. The authors conclude that improved performance accuracy can be achieved by increasing the dataset size used for the experiment.

One of the first research papers that utilized deep learning techniques on medical images is shown in [12]. The work was based on popular Alexnet architecture and transfer learning for screening the system performance on different datasets. The cross dataset performance analysis carried out shows the system accuracy of 67.4% on the Montgomery dataset and 83.7% accuracy on the Shenzhen dataset. A ConVnet model involving classifications of different manifestations of tuberculosis was presented in [13]. This work looked at unbalanced, less

categorized X-ray scans and incorporate cross-validation with sample shuffling in training their model. The Peruvian Tuberculosis datasets comprising of a total of 4701 image samples with about 4248 samples marked as abnormal containing six manifestation of tuberculosis and 453 samples marked as normal were used for the experiment to obtain 85.6% performance accuracy.

CNN has also been applied by the authors of [15] for extracting discriminative and representative features from X-ray radiographs for the purpose of classifying different parts of the body. This research has exhibited the capabilities of CNN models surpassing traditional hand-crafted method of feature extraction.

An approach based on deep learning for the classification of chest radiographs into positive and negative classes is depicted in [16]. The CNN structure employed in this work consists of seven convolutional layers and three fully connected layers to perform classification experiments. The authors compared three variety optimizers in their experiments and found the Adam optimizer to perform better with validation accuracy of 0.82% and loss of 0.4013.

Other methods that have been utilized for TB detection and classification includes: Support Vector Machine [17, 18], K-Nearest Neighbor [19], Adaptive Thresholding, Active Contour Model, and Bayesian Classifier [20], Linear Discriminant Analysis [21].

It is evident from the related work that more effort is required in dealing with the Tuberculosis epidemic that has continued as one of the topmost causes of death. In view of this, we have presented an improved performance validation accuracy concerning to detecting and classifying CXR for TB manifestation.

3 Materials

3.1 Datasets

The Montgomery County (MC) CXR dataset was employed in this research. The MC dataset is a TB specific dataset made available by the National Library of Medicine in conjunction with the Department of Health Services Maryland, U.S.A for research intent. This dataset composed of 58 abnormal samples labeled as “1” and 80 normal samples labeled as “0”. All samples are of the size 4020 by 4892 pixels saved as portable network graphic (png) file format as shown in Fig. 1. This dataset is accompanied by clinical readings that give details about each of the samples with respect to sex, age, and manifestations. The dataset can be accessed at <https://lhncbc.nlm.nih.gov/publication/pub9931> [23].

3.2 Preprocessing

Since deep neural networks are hugely dependent on large data size to avoid overfitting and achieve high accuracy [22], we performed data augmentation on the MC dataset as a way of increasing the size from 138 samples to 5000 samples. The following types of augmentation were applied: horizontal left and right flip with a probability = 0.3, random zoom = 0.3 with an area = 0.8, top and bottom

flip = 0.3, left and right rotation = 0.5. Other preprocessing task employed here includes image resizing, noise removal and histogram equalization. The data augmentation procedure used in this work is not such that gives room for data redundancy.

4 Proposed Model

A model based on Deep Convolutional Neural Network (CNN) structure has been proposed in this work for the detection and classification of Tuberculosis. CNN models are based upon feed-forward neural network structures for automatic features selection and extraction as a result of taking advantage of the inherent properties of images. The depth of a CNN model has an impact on the performance of the features extracted from an image. CNN models have many layers but the Convolutional layer, MaxPooling, and the Fully connected layer are regarded as the main layers [15]. At the time of model training, diverse parameters are optimized in the convolution layers for extracting meaning features before is been pass on to the fully connected layer where the extracted features are then classified into the target classes which in this case is “normal and abnormal” classes.

Our proposed CNN structure is composed of feature extraction and features classification stages. The feature extraction stage consists of convolution layers, batch normalization, relu activation function, dropout, and max pooling while the classification stage contains the fully connected layer, flatten, dense and a softmax activation function. There are 4 convolution layers in the network for extracting distinct features from the input image with shape $224 \times 224 \times 3$ that is passed to the first convolutional layer learning 64, 3×3 filters, the same as the second convolutional layer. Both the third and fourth convolutional layers learn 128, 3×3 filters. Relu activation function and batch normalization were employed in all the convolutional layers but only the second and fourth layer uses max pooling with a 2×2 pooling size and 25% dropout. The fifth layer which is the fully connected layers output 512 feature that is mapped densely to 2 neurons required by the softmax classifier for classifying our images into normal and abnormal classes. The detail representation of our proposed TB detection model is presented in Table 2.

4.1 Convolution Layer

At each convolution layer, the feature maps from the preceding layer convolute with kernels which are then fed through the ReLU activation function to configure the feature output maps. Also, each output map can be formulated with respect to several input maps. This can be mathematically written as:

$$y_j^i = f \left(\sum_{i \in N_j} y_i^{l-1} * M_{ij}^l + a_j^l \right)$$

where y_j^i depicts the j^{th} output feature of the l^{th} layer, $f(\cdot)$ is a nonlinear function, N_j is the input map selection, y_i^{l-1} refer to the i^{th} input map of $l-1^{th}$ layer, M_{ij}^l is the kernel for the input i and output map M in the j^{th} layer, and a_j^l is the additive bias associated with the j^{th} map output.

Table 2. Details representation of our proposed TB detection model

| Layers | Output size | Filter |
|-------------|----------------------------|--------------|
| INPUT IMAGE | $224 \times 224 \times 3$ | |
| CONVO1 | $112 \times 112 \times 64$ | 3×3 |
| ACTN | $112 \times 112 \times 64$ | |
| BATCHNORM | $112 \times 112 \times 64$ | |
| CONVO2 | $112 \times 112 \times 64$ | 3×3 |
| ACTN | $112 \times 112 \times 64$ | |
| BATCHNORM | $112 \times 112 \times 64$ | |
| MAXPOOL | $56 \times 56 \times 64$ | 2×2 |
| DROPOUT | $56 \times 56 \times 64$ | |
| CONVO3 | $56 \times 56 \times 128$ | 3×3 |
| ACTN | $56 \times 56 \times 128$ | |
| BATCHNORM | $56 \times 56 \times 128$ | |
| CONV4 | $56 \times 56 \times 128$ | 3×3 |
| ACTN | $56 \times 56 \times 128$ | |
| BATCHNORM | $56 \times 56 \times 128$ | |
| MAXPOOL | $28 \times 28 \times 128$ | 2×2 |
| DROPOUT | $28 \times 28 \times 128$ | |
| FULLY CONN | 512 | |
| ACTN | 512 | |
| BATCHNORM | 512 | |
| DROPOUT | 512 | |
| SOFTMAX | 2 | |

4.2 MaxPooling Layer

MaxPooling layer carryout a downsampling operation of the input map by calculating the maximum activation value in each feature map. The downsampling of MaxPooling is done to partly control overfitting and is formally written as:

$$y_j^l = f(\alpha_j^l \text{down}(y_i^{l-1} * M_{ij}^l) + a_j^l)$$

where α_j^l represent the multiplicative bias of every feature output map j that scale the output back to its initial range, $down(\cdot)$ can be substituted for either $avg(\cdot)$ or $max(\cdot)$ over an $n \times n$ window effectively scaling the input map by n times in every dimension.

Our model is trained using the Stochastic Gradients Descents (SGD) optimizer with an initial learning rate set to 0.001, batch size of 22 samples, momentum equals 0.9, regularization L2 equals 0.0005 to control overfitting and training loss. The SGD optimizer is given below as:

$$\alpha = \alpha - n \cdot \nabla_{\alpha} J(\alpha; x^{(i)}; y^{(i)})$$

where $\nabla_{\alpha} J$ is the gradient of the loss w.r.t α , n is the defined learning rate, α is the weight vector while x and y are the respective training sample and label.

4.3 Softmax Classifier

The Softmax classifier is used to process and classify the features that have been extracted from the convolutional stage. Softmax determine the probability of the extracted features and classify them into normal and abnormal classes defined as:

$$\sigma(q)_i = \frac{e^{q_i}}{\sum_{k=1}^k e^{q_k}}$$

where q mean the input vector to the output layer i that is depicted from the exponential element.

The structure of our model is presented in Fig. 2.

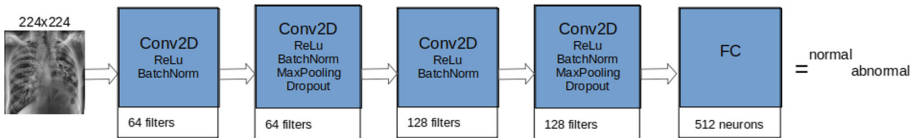


Fig. 2. Architecture of our ConvNet model.

5 Result

Training and validation of our model were performed on the Montgomery County (MC) TB specific dataset. The dataset which originally consists of 138 normal and abnormal samples were augmented to 5000 samples and split into 3,750 (75%) for training and 1,250 (25%) validation. The samples in the training set are indiscriminately selected i.e (shuffled) during the training to ensure features are extracted across all samples. Of course, the samples in the training set are

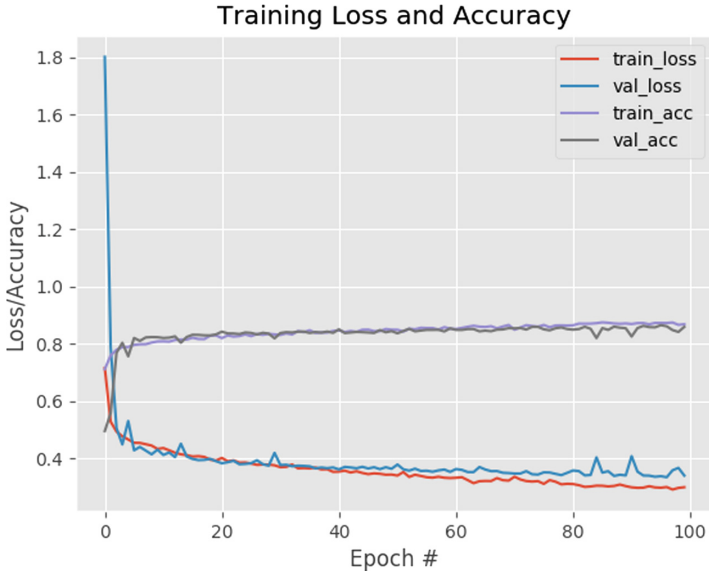


Fig. 3. Model accuracy and loss

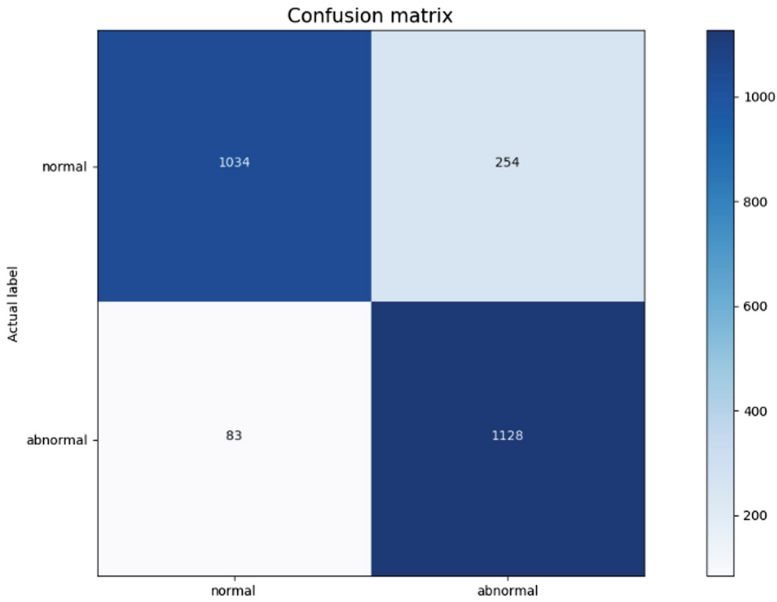


Fig. 4. Model confusion matrix

Table 3. Proposed model as compared

| Ref | Accuracy (%) |
|-----------------|--------------|
| 23 | 82.6, 84.7 |
| 12 | 83.7 |
| 11 | 80.4, 81.6 |
| 16 | 82 |
| 13 | 85 |
| 24 | 83.7 |
| Proposed | 87.1 |

entirely different from the samples contained in the validation set which are samples that the model has not seen before during model training. We use confusion matrix and accuracy as the metrics for evaluating the model. The 224×224 sample is fed into the model which was trained for 100 iterations with a batch size of 22 for SGD optimizer at an initial 0.001 learning rate. The model employs cross entropy loss function for weight update at every iteration and use Softmax function for processing and classifying the samples into normal and abnormal classes labeled 0 and 1. The proposed model achieved 87.1% validation accuracy. The Performance accuracy of the model and confusion matrix are presented in Figs. 3 and 4.

6 Conclusion

Presented in this paper is a model that aids early detection of Tuberculosis using CNN structure to automatically extract distinctive features from chest radiographs and classify them into normal and abnormal categories. We did not test other architectures in this research; instead, their performance accuracy is reported in Table 3. The histogram equalizer we applied to enhance the visibility of the data samples, which makes the extracted features more evident, is one of the contributing factors responsible for the improved performance. We will consider the Shenzhen, JSRT, KIT, and Indiana datasets in our future work while we continue to aim for optimal performance accuracy.

References

1. World Health Organization: Global tuberculosis report. WHO (2018)
2. World Health Organization: Global tuberculosis report. WHO (2019)
3. Zaman, K.: Tuberculosis: a global health problem. *J. Health Popul. Nutr.* **28**(2), 111 (2010)
4. Baral, S.C., Karki, D.K., Newell, J.N.: Causes of stigma and discrimination associated with tuberculosis in Nepal: a qualitative study. *BMC Public Health* **7**(1), 211 (2007)

5. World Health Organization: Tuberculosis prevalence surveys: a handbook. Report No.: WHO/HTM/TB/2010.17. World Health Organization, Geneva (2011)
6. World Health Organization: Chest radiography in tuberculosis detection: summary of current WHO recommendations and guidance on programmatic approaches (No. WHO/HTM/TB/2016.20). World Health Organization (2016)
7. Lakhani, P., Sundaram, B.: Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology* **284**(2), 574–582 (2017)
8. Pedrazzoli, D., Lalli, M., Boccia, D., Houben, R., Kranzer, K.: Can tuberculosis patients in resource-constrained settings afford chest radiography? *Eur. Respir. J.* **49**(3), 1601877 (2017)
9. Lönnroth, K., Migliori, G.B., Abubakar, I., et al.: Towards tuberculosis elimination: an action framework for low-incidence countries. *Eur. Respir. J.* **45**, 928–952 (2015)
10. Leibstein, J.M., Nel, A.L.: Detecting tuberculosis in chest radiographs using image processing techniques. University of Johannesburg (2006)
11. Rohilla, A., Hooda, R., Mittal, A.: TB detection in chest radiograph using deep learning architecture. In: *Proceeding of 5th International Conference on Emerging Trends in Engineering, Technology, Science and Management (ICETETSM-2017)*, pp. 136–147 (2017)
12. Hwang, S., Kim, H.E., Jeong, J., Kim, H.J.: A novel approach for tuberculosis screening based on deep convolutional neural networks. In: *Computer-Aided Diagnosis*, vol. 9785, p. 97852W. International Society for Optics and Photonics (2016)
13. Liu, C., et al.: TX-CNN: detecting tuberculosis in chest X-ray images using convolutional neural network. In: *IEEE International Conference on Image Processing (ICIP)*, pp. 2314–2318 (2017)
14. Konstantinos, A.: Testing for tuberculosis. *Aust. Prescr.* **33**(1), 12–18 (2010)
15. Srinivas, M., Roy, D., Mohan, C.K.: Discriminative feature extraction from X-ray images using deep convolutional neural networks. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 917–921 (2016)
16. Hooda, R., Sofat, S., Kaur, S., Mittal, A., Meriaudeau, F.: Deep-learning: a potential method for tuberculosis detection using chest radiography. In: *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 497–502 (2017)
17. Jaeger, S., Karargyris, A., Antani, S., Thoma, G.: Detecting tuberculosis in radiographs using combined lung masks. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4978–4981 (2012)
18. Karargyris, A., et al.: Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays. *Int. J. Comput. Assist. Radiol. Surg.* **11**(1), 99–106 (2015). <https://doi.org/10.1007/s11548-015-1242-x>
19. Van Ginneken, B., Katsuragawa, S., ter Haar Romeny, B.M., Doi, K., Viergever, M.A.: Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Trans. Med. Imaging* **21**(2), 139–149 (2002)
20. Shen, R., Cheng, I., Basu, A.: A hybrid knowledgeguided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs. *IEEE Trans. Biomed. Eng.* **57**(11), 2646–2656 (2010)
21. Hogeweg, L., Mol, C., de Jong, P.A., Dawson, R., Ayles, H., van Ginneken, B.: Fusion of local and global detection systems to detect tuberculosis in chest radiographs. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010*. LNCS, vol. 6363, pp. 650–657. Springer, Heidelberg (2010). <https://doi.org/10.1007/978-3-642-15711-0.81>

22. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *J. Big Data* **6**(1) (2019). Article number. 60. <https://doi.org/10.1186/s40537-019-0197-0>
23. National Center for Biomedical Communications. <http://lhncbc.nlm.nih.gov>