

# Detection of Pulmonary Tuberculosis Manifestation in Chest X-Rays Using Different Convolutional Neural Network (CNN) Models

Syeda Shaizadi Meraj, Razali Yaakob, Azreen Azman, Siti Nurulain Mohd Rum, Azree Shahrel Ahmad Nazri, Nor Fadhlin Zakaria

**Abstract:** Tuberculosis (TB) is airborne infectious disease which has claimed many lives than any other infectious disease. Chest X-rays (CXRs) are often used in recognizing TB manifestation site in chest. Lately, CXRs are taken in digital formats, which has made a huge impact in rapid diagnosis using automated systems in medical field. In our current work, four simple Convolutional Neural Networks (CNN) models such as VGG-16, VGG-19, ResNet50, and GoogLeNet are implemented in identification of TB manifested CXRs. Two public TB image datasets were utilized to conduct this research. This study was carried out to explore the limit of accuracies and AUCs acquired by simple and small-scale CNN with complex and large-scale CNN models. The results achieved from this work are compared with results of two previous studies. The results indicate that our proposed VGG-16 model has gained highest score overall compared to the models from other two previous studies.

**Keywords:** Tuberculosis, Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNN), Deep Learning (DL)

## I. INTRODUCTION

TB is an infectious disease. It is caused by bacteria, Mycobacterium Tuberculosis (MTB). It is the ninth prominent disease for high mortality rate worldwide, after HIV/AIDS. In 2017, an estimated 10 million people fell ill with TB; 1.6 million succumbed to death from TB, around 0.5 million people developed Multi Drug-Resistant TB (WHO, 2018).

TB is a contagious disease. It spreads through air, when an infected person sneezes or coughs. The TB bacteria which affects the lungs is known as pulmonary tuberculosis. The bacteria can also affect different parts of the body such as brain, spine, kidney or lymph nodes etc (CDC, 2017). When the bacteria infects different parts of the body, it is known as Extra pulmonary tuberculosis (EPTB).

**Revised Manuscript Received on September 22, 2019.**

Syeda Shaizadi Meraj, Faculty of Computer Science and Information Technology

Razali Yaakob, Faculty of Computer Science and Information Technology

Azreen Azman, Faculty of Computer Science and Information Technology

Siti Nurulain Mohd Rum, Faculty of Computer Science and Information Technology

Azree Shahrel Ahmad Nazri, Faculty of Computer Science and Information Technology

Nor Fadhlin Zakaria, Department of Medicine, Medical and Health Science Faculty, Universiti Putra Malaysia, 43400, Serdang, Selangor

Conventional diagnosis is one where the patient visits the doctor after prolonged cough or illness. Then the doctor examines for the signs and symptoms of the disease. Later the blood sample is sent for biopsy, which might take a couple of weeks. The patients might be asked to get a chest x-ray, and there might be a high probability of patient being infected with TB. Therefore, the conventional procedure of TB diagnosis, requires quite some time. Hence, to overcome this, a rapid automated diagnosing method can be used to help doctors (Meraj et al, 2019).

## Artificial Intelligence (AI)

AI Can Be Defined As, Network That Replicates The Structure And Function Of The Human Intelligence. The Most Elementary Unit Of ANN Is A “Perceptron”, Which Is A Single Artificial Neuron With A Feedforward Neural Network (Engelbrecht, 2007). It Does Not Have Any Hidden Layers.

When A Layer Of Neurons Are Connected To Other Layer Of Neurons Through Weights And Biases In Each Layer, Then It Is Called As Multi-Layer Perceptron (MLP) Also Known As Feedforward Neural Networks (Haykin, 2009). The MLP typically consists of an input layer and an output layer with one or more hidden layers. The hidden layers can be any number of arbitrary layers.

The concept of MLP is applied in Deep Learning (Dara et al., 2018). Deep Learning (DL) is a sub-field of Machine Learning (ML) algorithms, that uses Artificial Neural Networks (ANNs). DL is based on the concept of ANNs. The DL architecture cascades multiple layers of nonlinear processing units for feature extraction and transformation (Deng et al., 2014). The starting layer is input layer and the ending layer is output layer similar to ANNs. Varying on the number of layers given in-between input and output layers, it automatically extracts feature without human intervention, unlike other traditional ML algorithms. Currently, DL has many applications in diverse domains such as, automatic speech and image recognition (Simonyan & Zisserman, 2014), visual art processing, natural language processing, bioinformatics (Chicco et al., 2014); (Min et al., 2017), drug discovery and toxicology.

Convolutional Neural Networks (CNN) or ConvNets is a class of DL, which are inspired by biological visual cortex. CNNs are feed forward ANNs, as the information passes right through the network of the model. They have become

popular in analysing visual imagery.

Currently, they have become the most preferred technique in Biomedical field for the identification and classification of radiological images (Litjens et al., 2017). CNNs are comprised of neurons with learnable weights and biases. It may consist of few layers to several layers. However, the common layers are convolutional layer, pooling layer followed by Fully Connected (FC) layer (Lee et al., 2017). In addition to the mentioned layers other specialized layers can be included or omitted.

Transfer learning is a method to reuse pre-trained model on a new task. The challenges faced in medical field, it is rather difficult to obtain well annotated and large dataset of medical images. Studies have yielded impressive results by using transfer learning on pre-trained ImageNet model (Tajbakhsh et al., 2016); (Liu et al., 2017). In the current study, using keras high level API, different CNN models are implemented. It automatically downloads the weights that have been trained on ImageNet, for the image classification.

## Previous Studies

The emergence of sophisticated equipment in the field of radiology, has immensely helped in Computer Aided Diagnosis (CAD). The Digital Radiographic (DR) picture is directly taken in Joint Photographic Experts Group (JPEG) or Digital Imaging and Communications in Medicine (DICOM) format and are used in diagnosis and manifestation of disease inside the body. These types of images are easy to access and analyse in large quantity. Few

studies that have been carried out using CNN methodology for the detection of TB are discussed here.

Table 1 shows several works of CNN on TB detection. Most of the authors have used their own private TB dataset obtained from hospitals. However, in this work, we will compare with two previous works, which are (Lopes & Valiati, 2017) since they have used two publicly available TB dataset, Montgomery County (MC) and Shenzhen (SH) CXR sets (Jaeger et al., 2014).

Based on Table 1, Lopes & Valiati, (2017) have used both MC and SH datasets, executed on three CNN models such as GoogLeNet (Szegedy et al., 2014), ResNet (He et al., 2016) and VggNet (in their case, they have used Vgg19) and an ensemble of them. Meanwhile, Islam et al., (2017) have used six models viz., AlexNet, Vgg16, Vgg19, ResNet50, ResNet101, ResNet152 and an ensemble of them.

## Dataset

The U.S National Library of Medicine (NLM) has compiled three datasets of chest radiography available publicly to help research community, which are Montgomery County (MC), Shenzhen (SH) and Indiana University chest X-Ray dataset ((Jaeger et al., 2014). We have used two datasets namely, MC and SH data sets in our study. The MC dataset consists of 138 CXRs, of which 80 are normal cases and 58 are TB cases. The SH dataset consists of 326 normal and 336 abnormal CXRs (Jaeger et al., 2014).

**Table. 1 Comparison of evaluation from various CXRs and CNN methods**

Author and Year	Dataset	Features/Parameters	Method	Training algorithm	Classification accuracy (%)	Sensitivity (%)	Specificity (%)
Hwang et al., 2016	Private Dataset, MC and SH Datasets	CXRs	CNN	Modified AlexNet	90.00	—	—
Lakhani and Sundaram, 2017	MC, SH, Thomas Jefferson University Hospital and Belarus Tuberculosis Portal	CXRs	CNN*	Ensemble	96.00	—	—
				Radiologist augmented	98.70	—	—
Lopes and Valiati, 2017	MC and SH Datasets,	CXRs	CNN*	Ensemble	84.60	—	—
Islam et al., 2017	SH Dataset	CXRs	CNN*	Ensemble	90.00	88.00	92.00
Alcantara et al., 2017	5,000 CXRs from Partners in Health at Peru and various other image DB's	CXRs	CNN	Binary classification	89.60	—	—
				Multi-class classification	62.07	—	—

## II. MATERIALS AND METHODS

The CNN models were executed on windows 7 operating system, with 500 GB hard disk, 8GB RAM and no GPU



support. Keras API was used to build the models with TensorFlow backend. Python programming language, implemented in Jupyter notebook (Kluyver et al., 2016).

### Pre-Processing

MC and SH datasets comprise of total 138 and 662 PNG images respectively, with their descriptions in a corresponding text file. For training 0.75% of the images were allocated and the rest 0.25% were used for validating. Only image augmentation was carried out with flips,

rotations, zooming. No other pre-processing techniques were carried out.

### Model architecture

We implemented four simple small individual CNN models, such as VGG-16, VGG-19, ResNet50 and GoogLeNet models. The structure of all four standard CNN architecture is modified in the same way, as shown in the Figure 1.

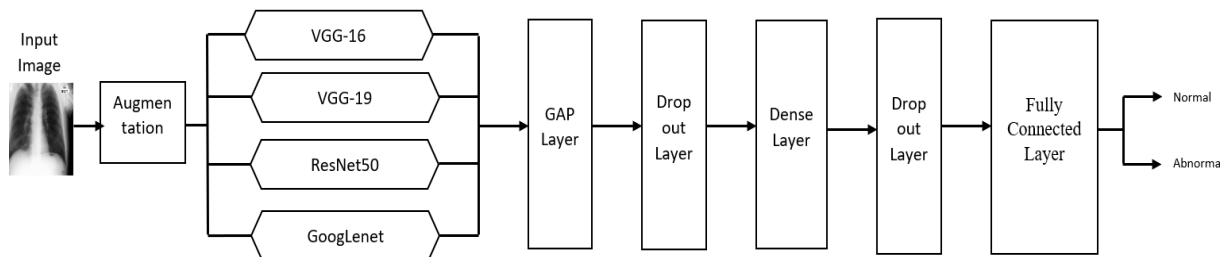


Fig. 1 Shows the architecture of the CNN models

The pre-trained weights were automatically downloaded, when the models were instantiated. All input image are resized to 224\*224, which are default size for the models as mentioned in keras documentation. The CXR images are labels with '0' for normal cases and '1' for TB positive. Next, we have added Global Average Pooling layer (GAP) for the dimensionality reduction of the previous layer. As the layer is flattened after the GAP, a dropout of probability 0.5 is introduced. To reduce overfitting, dropout switches off the influence of probability of the number of neurons given. Subsequently, FC layer (Dense layer) which is a type Neural Network. It receives single vector as an input to classify the image input as normal or abnormal. Non-linearity is induced in CNN by Sigmoid activation function. The model's learning rate was set to 0.0001 with Adam optimizer. Number of epochs is specified as 10 with 35 epochs per step. The reason to define a small epoch size is due to time and computation cost of our system. Binary crossentropy was used as loss function, as it is a binary classification problem. Early stopping was implemented to stop training, as the improvement in model stopped. The architecture of VGG-16 can be visualised, by using model. Summary function in keras to view the CNN architecture in detail.

### Evaluation Metrics

The evaluation of the models metrics was carried out in four measures viz., accuracy, area under Receiver Operating Characteristic (ROC) curve (AUC), Sensitivity and Specificity. However, in this work, we will show only accuracy and AUC. Accuracy is measured as number of correctly classified patient cases to the total number of patient cases.

A ROC curve is a graphical representation of competence of a classification model. It plots two parameters on the graph, which are True Positive Rate (TPR)/sensitivity and False Positive Rate (FPR)/specificity (Tharwat, 2018).

## III. RESULTS AND DISCUSSION

We have computed the accuracy and AUC, on our four models i.e VGG-16, VGG-19, ResNet50 and GoogLeNet on both SH and MC datasets.

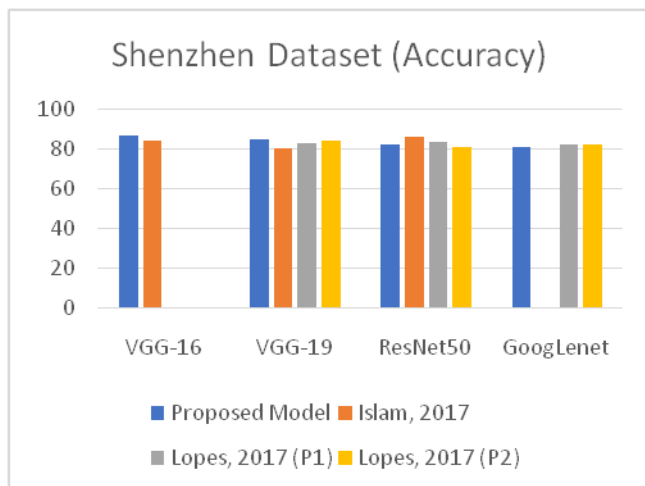
Islam et al., (2017), has implemented six models as mentioned earlier. However, we only compare with three models, namely VGG-16, VGG-19, and ResNet50. They have not mentioned the implementation of their methodology. They have conducted their study only on SH dataset.

On the other hand, (Lopes & Valiati, 2017) has utilized both SH and MC datasets. They have employed two approaches, i.e proposal 1(P1) and proposal 2(P2) in their work. They have executed VGG-19, ResNet50 and GoogLeNet models in their research. In order to compare our study with theirs, we have employed models similar to them but in our study, we use the standard CNN. In this work, we do not compare ensemble models from both studies.

Table 2 summarises the comparison of the accuracies scored by our four proposed models with studies by (Islam et al., 2017) and (Lopes & Valiati, 2017) on SH dataset respectively. Our proposed VGG-16 and VGG-19 models has scored highest accuracies overall compared to the studies with (Islam et al., 2017) and (Lopes & Valiati, 2017). Islam et al., (2017) and (Lopes & Valiati, 2017) have achieved highest accuracies on ResNet50 and GoogLeNet models respectively.

**Table. 2** Comparison of accuracies SH Dataset

Shenzhen Dataset (Accuracy)				
Model	Proposed Model	Islam et al., 2017	Lopes et al., 2017 (P1)	Lopes et al., 2017 (P2)
VGG-16	86.74	84	-	-
VGG-19	84.33	80	82.8	83.7
ResNet50	81.92	86	83.4	80.8
GoogLeNet	80.72	-	82.2	82.3

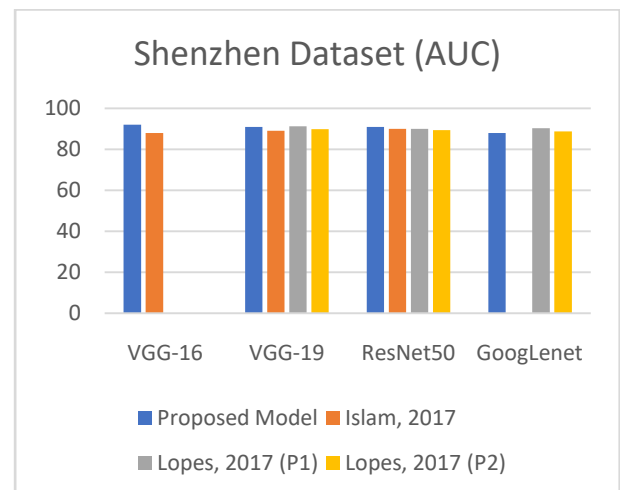


**Fig. 2** Graph showing accuracies obtained on SH dataset

Figure 2 illustrates the details of the Table 2 in a bar graph. As, there are no results to compare from (Lopes & Valiati, 2017) study on VGG-16 model and also by (Islam et al., 2017) on GoogLeNet model respectively. Hence, they are not shown in the graph.

**Table. 3** Comparison of AUCs on SH Dataset

Shenzhen Dataset (AUC)				
Model	Proposed Model	Islam et al., 2017	Lopes et al., 2017 (P1)	Lopes et al., 2017 (P2)
VGG-16	92.0	88	-	-
VGG-19	91.0	89	91.2	89.9
ResNet50	91.0	90	90	89.4
GoogLeNet	88.0	-	90.3	88.7

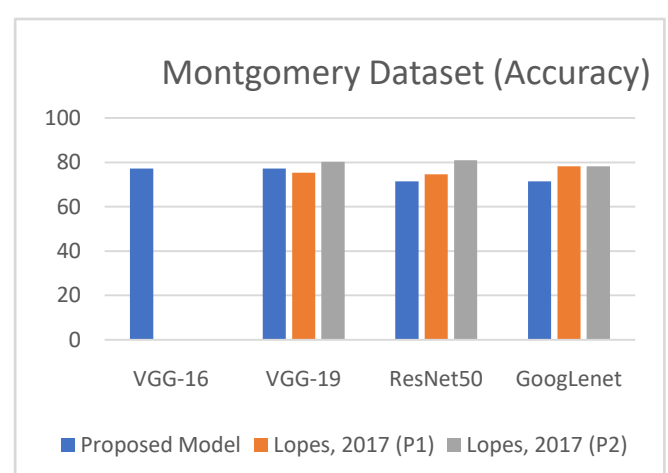


**Fig. 3** Graph showing AUC scores on SH dataset

Table 3 shows the details about the AUCs scored on the SH dataset. Our VGG-16 model has scored the highest overall. Both our VGG-16 and ResNet50 have same AUC scores respectively. There is a menial difference, in our score and the score of (Lopes & Valiati, 2017) VGG-19 (P1) model. Lopes & Valiati, (2017) has better AUC results in both P1 and P2 respectively. Figure 3 visualizes the details of the Table3. The graph is not plotted for the models that do not have scores.

**Table. 4** Comparison of accuracies on MC Dataset

Montgomery Dataset (Accuracy)			
Model	Proposed Model	Lopes et al., 2017 (P1)	Lopes et al., 2017 (P2)
VGG-16	77.14	-	-
VGG-19	77.14	75.3	80.3
ResNet50	71.42	74.6	81
GoogLeNet	71.42	78.2	78.2



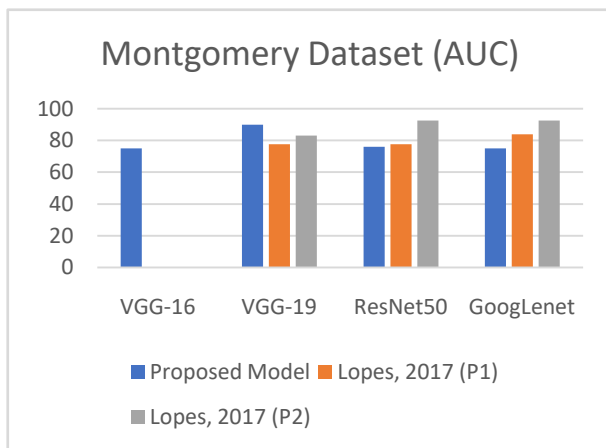
**Fig. 4** Graph showing accuracies scored on MC dataset



Table 4 shows the comparison on accuracies on MC dataset. Our models have achieved 77.14 AUC on both VGG-16 and VGG-19 models respectively. Also, scored 71.42 on both ResNet50 and GoogLeNet models respectively. However, our overall results obtained on MC dataset are inferior compared to (Lopes & Valiati, 2017) (P1 and P2) respectively. Figure 4 highlights the gist of Table 4.

**Table. 5 Comparison of AUCs on MC Dataset**

Montgomery Dataset (AUC)			
Model	Proposed Model	Lopes et al., 2017 (P1)	Lopes et al., 2017 (P2)
VGG-16	75.0	-	-
VGG-19	90.0	77.7	83.1
ResNet50	76.0	77.6	92.6
GoogLeNet	75.0	83.8	92.5



**Fig. 5 Graph showing AUCs on MC dataset**

Table 5 describes the comparisons of AUCs on MC dataset. Only our model VGG-19 has scored the highest AUC overall. However, (Lopes & Valiati, 2017) (P2) has attained comparatively superior results that other models. Figure 5 graphically represent the details of Table 5.

#### IV. CONCLUSION

In conclusion, we have compared four CNN models with the previous two studies. As, they have used the same publicly available dataset. It was easier to compare the results, which is not the case with other studies, as they have used private datasets. VGG-16 model on SH dataset has achieved good performance results. However, due to imbalance and limited dataset, the performance of MC was inferior in comparison with SH dataset. The reason to conduct this study was to investigate, if we would be able to achieve competent results despite low computational power and small model architecture. The future works will include collecting TB patient data from specified local TB hospitals. Design a new framework that can predict better and accurate results. Also, compare between TB X-ray scan and TB CT scans of same patient. Visualizing the effect of segmentation on both 2D and 3D images.

#### ACKNOWLEDGMENT

This work is funded by Universiti Putra Malaysia under Geran Putra (GP/2018/9596100).

#### REFERENCES

- Alcantara, M. F., Cao, Y., Liu, C., Liu, B., Brunette, M., Zhang, N., Sun, T., et al. (2017). Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor communities in Peru. *Smart Health*, 1, 66–76. Elsevier.
- CDC.(2017).Centers for Disease Control and Prevention. Retrieved from [www.cdc.gov/tb/publications/faqs/qa\\_introduction.htm](http://www.cdc.gov/tb/publications/faqs/qa_introduction.htm)
- Chicco, D., Sadowski, P. & Baldi, P. (2014). Deep autoencoder neural networks for gene ontology annotation predictions. *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics* (pp. 533–540).
- Dara, S., Tumma, P., Eluri, N. R. & Kancharla, G. R. (2018). Feature Extraction In Medical Images by Using Deep Learning Approach. *International Journal of Pure and Applied Mathematics*, 120(6), 305–312.
- Deng, L., Yu, D. & others. (2014). Deep learning: methods and applications. *Foundations and Trends in Signal Processing*, 7(3-4), 197–387. Now Publishers, Inc.
- Engelbrecht, A. P. (2007). *Computational intelligence: an introduction*. John Wiley & Sons.
- Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Pearson education Upper Saddle River.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
- Hwang, S., Kim, H.-E., Jeong, J. & Kim, H.-J. (2016). A novel approach for tuberculosis screening based on deep convolutional neural networks. *Medical Imaging 2016: Computer-Aided Diagnosis* (Vol. 9785, p. 97852W).
- Islam, M. T., Aowal, M. A., Minhaz, A. T. & Ashraf, K. (2017). Abnormality Detection and Localization in Chest X-Rays using Deep Convolutional Neural Networks. *arXiv preprint arXiv:1705.09850*.
- Jaeger, S., Candemir, S., Antani, S., Wang, Y.-X. J., Lu, P.-X. & Thoma, G. (2014). Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative imaging in medicine and surgery*, 4(6), 475. AME Publications.
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., Kelley, K., et al. (2016). *Jupyter Notebooks—a publishing format for reproducible computational workflows*. *ELPUB* (pp. 87–90).
- Lakhani, P. & Sundaram, B. (2017). Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*, 162326. Radiological Society of North America.
- Lee, J.-G., Jun, S., Cho, Y.-W., Lee, H., Kim, G. B., Seo, J. B. & Kim, N. (2017). Deep learning in medical imaging: general overview. *Korean journal of radiology*, 18(4), 570–584.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., et al. (2017). A survey on deep learning in medical image analysis. *arXiv preprint arXiv:1702.05747*.
- Liu, C., Cao, Y., Alcantara, M., Liu, B., Brunette, M., Peinado, J. & Curioso, W. (2017). TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network. *Image Processing (ICIP), 2017 IEEE International Conference on* (pp. 2314–2318).
- Lopes, U. & Valiati, J. (2017). Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Computers in Biology and Medicine*, 89, 135–143. Elsevier.
- Meraj, S. S., Yaakob, R., Azman, A., Rum, S. N. M. & Nazri, A. S. A. (2019). Artificial Intelligence in Diagnosing Tuberculosis: A Review. *International Journal on Advanced Science, Engineering and Information Technology*, 9(1), 81–91.
- Min, S., Lee, B. & Yoon, S. (2017). *Deep learning in bioinformatics. Briefings in bioinformatics*, 18(5), 851–869. Oxford University Press.
- Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

## Detection of Pulmonary Tuberculosis Manifestation In Chest X-Rays Using Different Convolutional Neural Network (CNN) Models

21. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., et al. (2014). Going Deeper with Convolutions. CoRR, abs/1409.4842.
22. Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B. & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging, 35(5), 1299–1312. IEEE.
23. Tharwat, A. (2018). Classification assessment methods. Applied Computing and Informatics. Elsevier. WHO. (2018). Global tuberculosis report 2018.

