International Journal of Medical Engineering and Informatics

ISSN online: 1755-0661 - ISSN print: 1755-0653
https://www.inderscience.com/ijmei

Layer-based deep net models for automated classification of pulmonary tuberculosis from chest radiographs

Sushil Ghildiyal, Saibal Manna, N. Ruban

DOI: 10.1504/IJMEI.2021.10043722

Article History:

Received:	14 September 2020
Last revised:	01 February 2021
Accepted:	17 February 2021
Published online:	30 November 2022

Layer-based deep net models for automated classification of pulmonary tuberculosis from chest radiographs

Sushil Ghildiyal*

School of Electrical Engineering,
Vellore Institute of Technology,
Vellore, Tamil Nadu, India
Email: info.sushil123@gmail.com
*Corresponding author

Saibal Manna

Department of Electrical Engineering,
School of Electrical Engineering,
NIT,
Jamshedpur, Jharkhand, India
Email: mannasaibal1994@gmail.com

N. Ruban

School of Electrical Engineering,
Vellore Institute of Technology,
Vellore, Tamil Nadu, India
Email: nruban@vit.ac.in

Abstract: Tuberculosis (TB) is a highly infectious bacterial disease. However, it can affect any body part, but is majorly a lung infection; which is potentially fatal and contagious. Like most of the serious health issues, the recovery rate of a symptomatic TB patient completely depends on the early detection and treatment. Deep learning algorithms-based computer aided diagnosis (CAD) system, can provide aid in early detection of the disease. In this regard, a method to detect infection of tuberculosis, which uses deep learning network to classify CXR images as normal or abnormal, is presented. Convolutional neural network (CNN), visual geometry group (VGG16) and high-resolution network (HRNet) models are used and their performance has been compared based on the validation loss and validation accuracy. The HRNet provides 89.7% accuracy with comparatively less loss among the proposed algorithms. The models are also deployed in android application for active clinical trials.

Keywords: tuberculosis; deep neural network; convolutional neural; CNN; VGG16; high-resolution network; HRNet.

Reference to this paper should be made as follows: Ghildiyal, S., Manna, S. and Ruban, N. (2023) 'Layer-based deep net models for automated classification of pulmonary tuberculosis from chest radiographs', *Int. J. Medical Engineering and Informatics*, Vol. 15, No. 1, pp.58–69.

Biographical notes: Sushil Ghildiyal received his MTech in Control and Automation from the Vellore Institute of Technology, India in 2019. He is currently pursuing his PhD in the Department of Computer Science and Engineering from the IIT Ropar, India. He is involved in research in the areas of classification of breast cancer detection using deep learning and optimisation of generative adversarial learning.

Saibal Manna received his MTech in Control and Automation from the Vellore Institute of Technology, India in 2019. He is currently pursuing his PhD in the Electrical Engineering Department from the NIT Jamshedpur, India. His research interests include control system application, renewable energy, and deep learning.

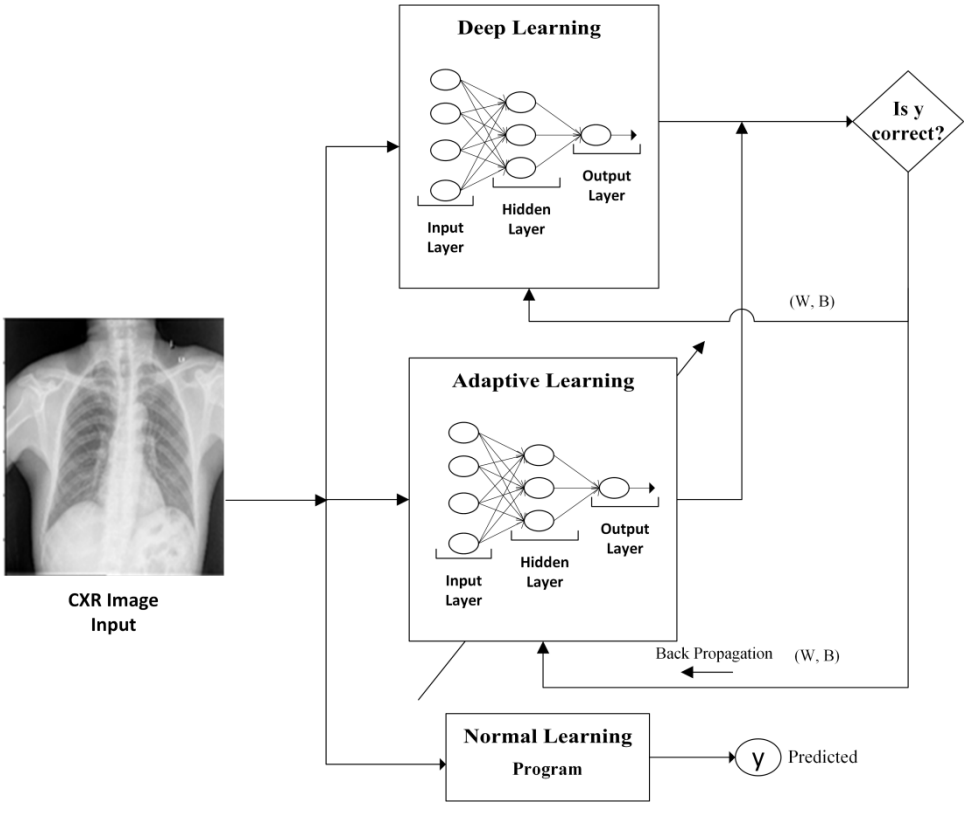
N. Ruban received his PhD in Electrical Engineering from the Vellore Institute of Technology, India. He is currently working as an Associate Professor with the School of Electrical Engineering. His current research interests include biomedical engineering, digital signal processing and machine learning.

1 Introduction

The bacteria of tuberculosis (TB) affects lungs adversely, it can be cured and prevented from spreading. If TB is identified at early stages and treated adds a great value in cost reduction of treatment and thus helps in retaining health of patient (WHO, 2018). The severity of the problem can be seen from the initiative taken by World Health Organization (WHO) as EndTB, a strategy which focuses on reducing mortality rate by 95% and incidence of this disease by 90% till the end of 2035 (Bar et al., 2015). Each year a million of people die due to TB across the world. Globally, TB is majorly the prominent cause of death. In 2017, as per the WHO report nearly 10 million people diagnosed with TB disease in which an estimated death of 1.3 million people was recorded. The disease mostly shows its presence in the regions of Africa and South-East Asia. The disease is highly infectious and is cause by a bacteria called Bacillus Mycobacterium TB. The pulmonary TB (PTB) mainly affects lungs although it is present in all part of body. The PTB is a communicable disease and symptoms are seen near the clavicles region. Based on various factors the PTB shows variety of pathological patterns in the lungs. PTB can be cured if, detected in early stages.

Magnetic resonance imaging, computed tomography and chest radiography are some of the radiological procedures used for detecting TB. After detection a confirmatory test is conducted on patient namely sputum cytology or sputum culture for conforming the presence of TB. Although many imaging modalities are available, the traditional chest radiographs (CXRs) are used widely, because they have lower radiation dose, low in cost and is accepted widely. The CXR is recommended by WHO for screening on large scale of patients so as to reduce normal cases subjected to costly tests. As a result, large number of patients and lack of skilled radiologist, there is higher chance that CXRs are prone to human errors, thus may result in cases undiagnosed. Hence, to avoid this problem, researchers are developing computer-aided diagnosis (CAD) a system that detects TB automatically from provided CXRs. This technology can be a boon for developing and underdeveloped countries, where medical resources are lean for large number of patients.

Figure 1 This shows various learning approach like deep learning, adaptive learning and normal learning to gain knowledge on given dataset



Deep learning is a subset of artificial intelligence (AI) that mimics the human brain working in interpreting data and forming patterns for decision-making. It enables computer methods that comprises of various layers of processing to learn data representations with multiple abstraction levels. These approaches have adequately upgraded the state-of-the-art in object detection, speech recognition, object recognition, drug discovery and genomics whereas in Normal learning or traditional way of learning is to create a program that takes an input and as per the code, it generates an output.

Adaptive learning systems need the minimum human effort and when new information is encountered, they only need human feedback when it matters most and constantly broaden their knowledge. Recently, adaptive learning places a vital role in adaptation of hyper parameters in deep learning. Figure 1 represents all the learning approaches for a given dataset. One thing that can speed up the learning of the model is to gradually reduce the learning rate over time and this adaptation of learning rate is called adaptive learning.

$$\alpha = \frac{1}{1 + \left(\text{decay}_{\text{rate}} * \text{epoch}_{\text{num}}\right)} \alpha_0$$

1 epoch = 1 pass through data

Using this adaptation rule, with the function of the epoch_num the learning rate (α) slowly decreases.

2 Literature review

TB can be detected in chest X-rays by observing the appearance of a particular pattern's in images. The earliest suggestion for identification of TB using CAD system is, where a multi-scale filter bank is used for the extraction of features from lung images (Van Ginneken et al., 2002). A weighted nearest neighbour method is used for classification and leave-one-out cross-validation (LOOCV) is used for validation. The scheme is applied to two private datasets and obtained area under the curve (AUC) 0.986 and 0.82 respectively. The authors provide a TB identification method that integrates a pixel-level textural abnormality analysis with other methods. They received AUC between 0.67 and 0.86 (Hogeweg et al., 2010). The semi-automated technique is suggested where decision tree classifier is used. All experiments were carried out in a limited private dataset with 94.9% accuracy (total images used were 95 in which 50 belong to normal and 45 to abnormal classes) (Tan et al., 2012).

The authors suggested a model for TB detection in CXR images. This model consists of a series of modules which must follow a series of steps in order to perform the classification. It begins from the CXR dataset with a wavelet denoising pre-processing module, guided by Gist and pyramid histogram of oriented gradients (PHOG) feature extraction module. The corresponding features are then chosen according to the chi-square distribution. Ultimately, support vector machine (SVM) classifier is used for model construction (Chauhan et al., 2014). Convolutional neural network (CNN) is broadly used in image recognition, object detection, image analysis and face recognition (Manna et al., 2020). AlexNet network (Krizhevsky et al., 2012) with a large input layer, followed by a new convolution and a new max-pooling layer is used for TB identification. This proposal involves the developing of a custom network, modification of the existing CNN to the particular TB problem identification and recalculation of existing weights (Hwang et al., 2016). A pre-trained ImageNet is used to identify CXR images as having signs of TB or as good, using a pre-trained CNN model in different ways (Abbas and Abdelsamea, 2018). A shallow CNN approach is used to classify lung image patches of computed tomography (CT) images into five separate classes with three fully connected and only one convolution layer (Li et al., 2014).

TB identification technique is proposed in which long Gabor mask, Intensity mask and lung model mask are used for lung segmentation. The pathological patterns in the CXR are then detected using various shape and texture descriptors. The histogram bins are used to present the distribution of each descriptor, and the each histogram bin value is considered as a function for each descriptor. SVM is used to identify the CXR images into healthy and unhealthy classes (Jaeger et al., 2012). In comparison to patch-based CNNs, the author used the entire lung image with different scales to feed a CNN into six different classes for lung images classification to cope with the shortcomings of conventional patch-based algorithms in differentiating between healthy and unhealthy lungs (Gao et al., 2018a).

A multi-resolution approach to grey-scale and rotation invariant texture classification based on local binary patterns (Ahonen et al., 2004) is proposed (Ojala et al., 2002). A patch-based CNN model was used as a feature extractor to recognise abnormalities in CT

image lung nodules, where two pooling and two convolution layers were used to create multi-scale features with different sizes of the nodule patches (Shen et al., 2015). A patch-based CNN model is used for lung pattern classification (Anthimopoulos et al., 2016). An automated approach is suggested in which two different feature sets are used, namely content-based image retrieval (CBIR)-based characters and object-based detection characters. SVM is eventually used as the classifier to identify CXR as healthy and unhealthy. Tests are collected using three datasets, two for training and one is used for system checking. In terms of AUC, the efficiency of CBIR feature vectors and object detection has been measured as 0.90 and 0.87 (Jaeger et al., 2014b). A pre-trained CNN is used as feature extractor for TB detection (Lopes and Valiati, 2017). Gao et al. (2018b) investigates the application of CT pulmonary images to identify and classify TB at five stages of severity, to track treatment effectiveness.

As per the literature survey, some of the papers carried their experiment with a small amount data of CXR images in which they obtained a good accuracy, but in that case, the over-fitting could happen. Moreover, some paper experimented with large dataset but in that either the accuracy was less and/or computation cost was increased as well as most of the methods for advanced diagnosis of the infection is still facing cost constraints for wider adaptability. In this paper, we compared different models and deployed the model in android application for clinical uses.

3 Experiments

For the recognition of TB, the task is measured to be a binary classification with the last layer given the probability that the CXR image has TB or not. The main objective is to generate a setup that can be deployable in an application and can be widespread in a much wider sense. Moreover, we also need to focus at the architecture, which is ideal, the hyper-parameters range and features inside the architectures that are unfavourable for the performance of the certain task and the choice of cost function so that it can be used for the diversity of diseases and at last, the model can be explicable for medical use. In these investigations, keeping all the other constraints such as primary stopping, learning rate and the train, validation splitting to be constant we just varied the model among conventional architectures: CNN, VGG16, and HRNet.

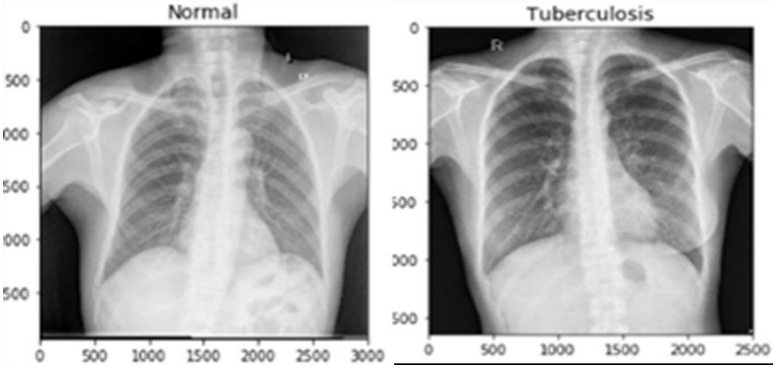
3.1 Dataset

On a sample of 800 CXR images taken from the two public datasets, namely the Shenzhen set and montgomery set (Jaeger et al., 2014a), the algorithm's performance is validated. Both the sets are managed by National Library of Medicine (NLM), USA and are freely accessible. The total number of 138 CXR images (80 normal and 58 TB) were collected under the TB control program in the country of Montgomery comprises the Montgomery collection and Shenzhen set composed of 662 CXR images (326 normal and 336 TB). The detailed statistics of the data is described in Table 1. The dataset is spitted into two groups training and validation. Training set is used to train the deep learning models. Lastly, performance evaluation was done by the unseen images in a validation set. Figure 2 shows the random representation of the CXR images with and without TB.

Table 1 Detailed statistics of the dataset

Dataset	*Shenzhen data*	*Montgomery data*
Total CXR images	662	138
CXR normal images	326	80
CXR TB images	336	58
Image type	Greyscale	Greyscale
Image view	Frontal	Frontal
Size (MB)	3,584	585
Image format	PNG	PNG
Resolution	Not specific resolution	4,892 × 4,020 / 4,020 × 4.892

Figure 2 This shows the random images from the dataset with and without TB



3.2 Data augmentation

To speed up the training part all images in the dataset have been compressed into 96 × 96 pixels. Up-gradation of the dataset is performed using transformation and deformation techniques. The process of data augmentation involves, transition, flipping, cropping, random small noise, rotation with diverse angles and colour jittering. This process results a total of 2,040 CXR images with (96 × 96) pixels. Therefore, for training the models a total of 2,040 images were taken into account in which 1,035 images were normal CXR images and 1,005 were abnormal or TB images. For validating the model, the total numbers of images were 120 (61 normal and 59 TB). A number of training and validations run up to 100 epochs in CNN.

3.3 CNN architecture and training

The architecture of classification model, i.e., CNN is consisting of two stages. The performance of the deep learning model and quantity of images for training is increased using data pre-processing and data augmentation techniques and the classification action of the model is analysed by adjusting features or weights using different learning techniques such as deep, fine tuning and shallow. The CNN has been trained to predict the presence (1,035 images) or absence (1,005 images) of the TB in GPU mode using

tensorflow framework. For training and validation section the dataset was split into 85% and 15% respectively. We used convolution neural network (CNN) which composed of three layers of convolution, max-pooling, kernel sizes and fully connected layers. Here we used kernel size of 3 × 3, max-pool size of 2 × 2 and the loss function used was binary cross entropy. The equation (1) represents the loss function of the CNN model.

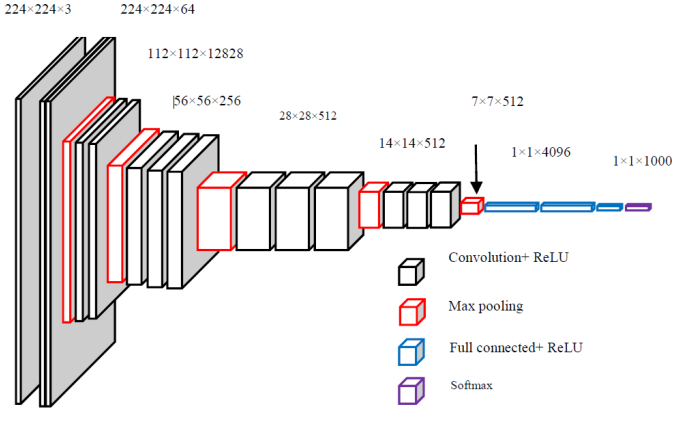
$$L = \frac{1}{m}\sum_{i=1}^{m}\left(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\right) \tag{1}$$

where L denotes loss function, m denotes no. of instances, $\hat{y}_i$ represents estimated value and y represents actual value.

3.4 VGG16 model

VGG16 in Figure 3 illustrate the architecture and layers involved. It is an implementation from the CNN. This model obtained 92.7% best five test accuracy in ImageNet, which is 14 million image dataset with 1,000 classes. The most incredible thing about VGG16 is that it concentrated on having convolution layer of 3 × 3 filter with step 1 and only used the same max-pool and padding layer of 2 × 2 filter of step 2. It is consistent in the entire system following this arrangement of convolutions and max-pool layers. It has two fully connected layers in the end and then a soft-max for output. The 16 in VGG16 corresponds to a weight of 16 layers. This is a huge network and has approximately 138 million parameter. This model is implemented in wide deep learning image classification problems.

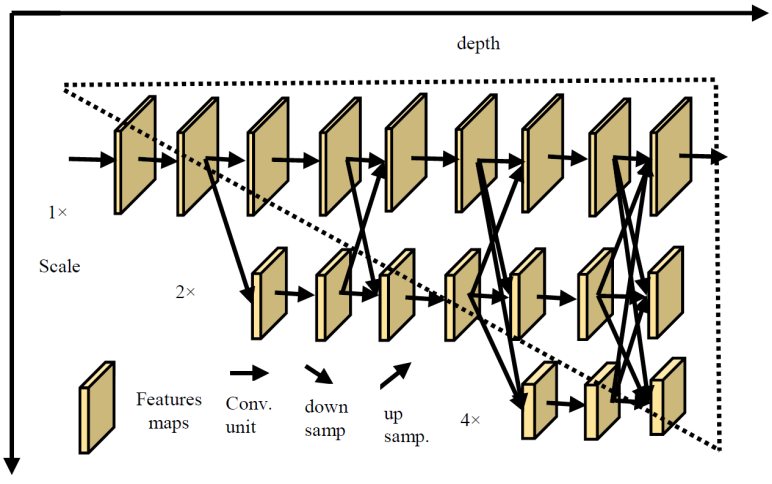
Figure 3 Illustrate VGG-16 architecture and its layers (see online version for colours)



3.5 HRNet model

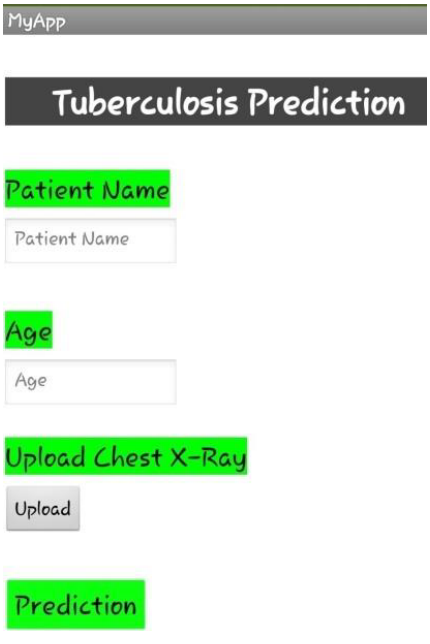
HRNet is developed at Microsoft and has signified state of art presentation in the areas of semantic segmentation, image classification, facial detection, object detection and pose estimation (Sun et al., 2019).

Figure 4 Illustrate the architecture of HRNet (see online version for colours)



Notes: It consists of parallel high to low resolution sub-networks with repeated information exchange across multi-resolution sub-networks. The vertical and horizontal directions correspond to the depth of the network and the scale of the feature maps.

Figure 5 Android app for TB prediction (see online version for colours)



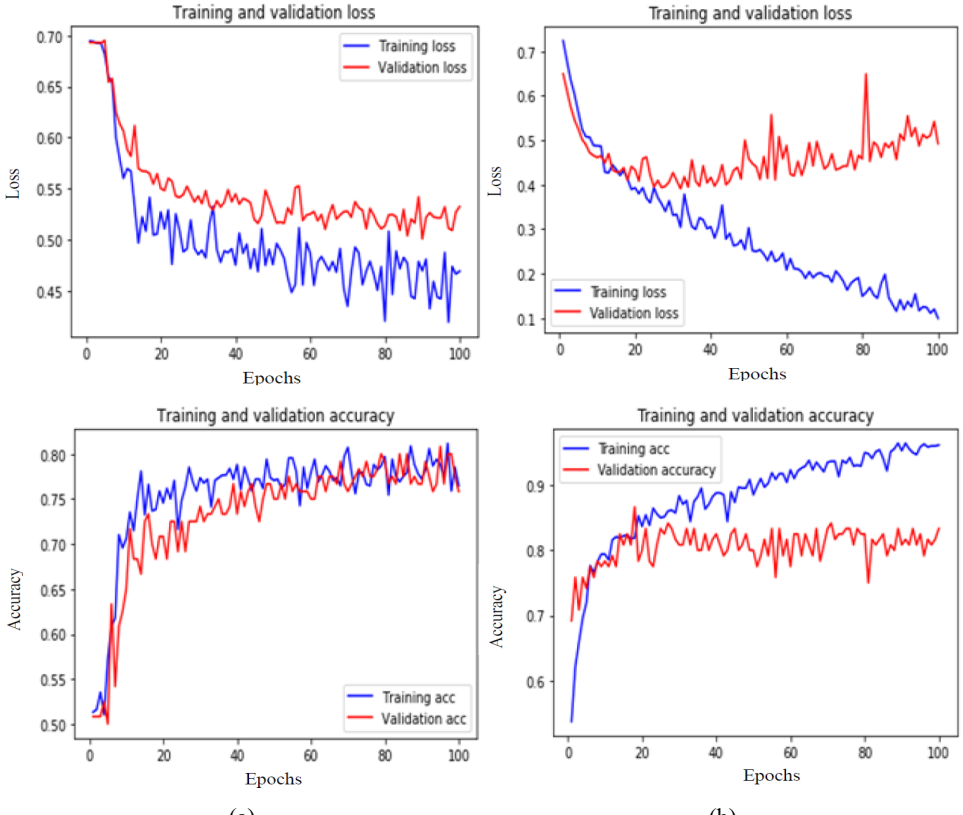
HRNet that will hold images at high resolution throughout the entire process. Here first step from a high-resolution subnetwork, add up high-low resolution sub-networks sequentially in order to create further phases and link up the multi-resolution

sub-networks in parallel. Multi-scale fusions executed over and over during the entire process by transfers of information through parallel multi-resolution sub-networks.

The proposed network estimates crucial points for high-resolution output. Figure 4 shows the resultant network of high resolution network (HRNet) model. Compared with existing widely-used networks, HRNet has advantage as it bonds high to low resolution subnetworks in parallel instead of in series as done in most prevailing solutions. Hence, to maintain the high resolution instead of improving the resolution through a low to high process, and therefore the predicted heat map is potentially spatially more precise.

An android app was also developed using MIT app inventor platform. Figure 5 illustrating the application developed for TB prediction using chest X-ray. In this app the user or doctor need to upload personal credentials and then user has to upload the chest X-ray image after a while it will predict whether the patient is suffering from TB or not.

Figure 6 These graphs shows the training loss vs. validation accuracy and training accuracy vs. validation accuracy, (a) using CNN model (b) VGG16 model (see online version for colours)



4 Results

Table 2 shows the results, i.e., validation loss and validation accuracy (in percent) of the deep learning models used for the detection of TB for clinical use. Figure 6 illustrates the graph of validation loss and accuracy loss with respect to the number of epochs in CNN and VGG16 model. From Figure 6(a) the CNN model is converging around 20 to 40 epochs and it is becoming stable thereafter. We get a similar convergence for VGG16 at 20 epochs the loss function is slightly oscillating between 0.4 to 0.5 in the range of 20 to 100 epochs, the convergence pattern remains unchanged even after 100 epochs as shown in Figure 6(b). Figures 6(c) and 6(d) shows the training and validation accuracy of the CNN and VGG16 models. The accuracy of the training and validation models are almost same in the range of 80% approximately for CNN model. The drop in the accuracy is due to the lack in accurate interpretation of the data. But the training accuracy of VGG16 model is more than 90% whereas for the validation it reached 83% which is better than CNN model, since it used pre-trained weights for its training.

Table 2 Validation loss and accuracy of deep learning models

Deep learning model	*Validation loss (%)*	*Validation accuracy (%)*
CNN	52.1	80.3
VGG16	49.2	83.3
HRNet	40.1	89.7

5 Conclusions

In this paper, it discusses the impact of learning diverse information levels based on information in order to classify the lungs region as normal or TB in two classes within the CXR. The algorithms are trained and validated with two public datasets, namely the Shenzhen set and Montgomery set. Both the sets are managed by National Library of Medicine (NLM), USA and are freely accessible. The results clearly shows that HRNet gave the best accuracy of 89.73% on validation dataset, whereas 83.33% and 80.33% with VGG16 and CNN algorithms respectively. When CNN is trained from scratch with 2,040 images it does not interpret accurate data and lead to a bit low classification accuracy as well as to get the better prediction the number of epochs were increased which results in high computational cost. The major limitation for the CNN model, is the data size, since the data size is insufficient for the accurate classification, we tried with other models for better classification. HRNet and VGG16 we have pre trained weights so the accuracy of the system increases with less computational cost which results in better prediction.

References

Abbas, A. and Abdelsamea, M. (2018) 'Learning Transformations for automated classification of manifestation of tuberculosis using convolutional neural network', *IEEE*.

Ahonen, T., Hadid, A. and Pietik€ainen, M. (2004) 'Face recognition with local binary patterns', in *European Conference on Computer Vision*, Springer, pp.469–481.

Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A. and Mougiakakou, S. (2016) 'Lung pattern classification for interstitial lung diseases using a deep convolutional neural network', *IEEE Transactions on Medical Imaging*, Vol. 35, No. 5, pp.1207–1216.

Bar, Y., Diamant, I., Wolf, L. and Greenspan, H. (2015) 'Deep learning with non-medical training used for chest pathology identification', in *SPIE Medical Imaging, International Society for Optics and Photonics*, pp.94140V–94140V.

Bergamo, A., Torresani, L. and Fitzgibbon, A.W. (2011) 'PiCoDes: learning a compact code for novel-category recognition', in *Advances in Neural Information Processing Systems (NIPS)*, December, Vol. 1, No. 5, p.6.

Chauhan, A., Chauhan, D. and Rout, C. (2014) 'Role of gist and PHOG features in computer-aided diagnosis of tuberculosis without segmentation', *PloS One*, Vol. 9, No. 11, p.e112980.

Gao, M., Bagci, U., Lu, L., Wu, A., Buty, M., Shin, H.C. and Mollura, D.J. (2018a) 'Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks', *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Vol. 6, No. 1, pp.1–6.

Gao, X.W., James-Reynolds, C. and Currie, E. (2018b) 'Analysis of tuberculosis severity levels from CT pulmonary images based on enhanced residual deep learning architecture', *Neurocomputing* [online] https://doi.org/10.1016/j.neucom.2018.12.086.

Hogeweg, L., Mol, C., de Jong, P.A., Dawson, R., Ayles, H. and van Ginneken, B. (2010) 'Fusion of local and global detection systems to detect tuberculosis in chest radiographs', in *Medical Image Computing and Computer-Assisted Intervention – Miccai*, Springer, pp.650–657.

Hwang, S., Kim, H-E., Jeong, J. and Kim, H-J. (2016) 'A novel approach for tuberculosis screening based on deep convolutional neural networks', in *SPIE Medical Imaging, International Society for Optics and Photonics*, pp.97852W–97852W.

Jaeger, S., Candemir, S., Antani, S., Wáng, Y.X.J., Lu, P.X. and Thoma, G. (2014a) 'Two public chest X-ray datasets for computer-aided screening of pulmonary diseases', *Quantitative Imaging in Medicine and Surgery*, Vol. 4, No. 6, p.475.

Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Xue, Z., Palaniappan, K., Singh, R.K., Antani, S. et al. (2014b) 'Automatic tuberculosis screening using chest radiographs', *IEEE Transactions on Medical Imaging*, Vol. 33, No. 2, pp.233–245.

Jaeger, S., Karargyris, A., Antani, S. and Thoma, G. (2012) 'Detecting tuberculosis in radiographs using combined lung masks', in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, pp.4978–4981.

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) 'ImageNet classification with deep convolutional neural networks', in *Advances in Neural Information Processing Systems*, pp.1097–1105.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D.D. et al. (2014) 'Medical image classification with convolutional neural network', *13th International Conference on Control Automation Robotics & Vision*, pp.844–848.

Lopes, U.K. and Valiati, J.F. (2017) 'Pre-trained convolutional neural networks as feature extractors for tuberculosis detection', *Computers in Biology and Medicine*, Vol. 89, pp.135–143.

Manna, S., Ghildiyal, S. and Bhimani, K. (2020) 'Face recognition from video using deep learning', *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, IEEE.

Ojala, T., Pietikainen, M. and Maenpaa, T. (2002) 'Multiresolution gray-scale and rotation invariant texture classification with local binary patterns', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp.971–987.

Shen, W., Zhou, M., Yang, F., Yang, C. and Tian, J. (2015) 'Multi-scale convolutional neural networks for lung nodule classification', in *International Conference on Information Processing in Medical Imaging*, Springer, pp.588–599.

Sun, K., Xiao, B., Liu, D. and Wang, J. (2019) 'Deep high-resolution representation learning for human pose estimation', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June, pp.5686–5696.

Tan, J.H., Acharya, U.R., Tan, C., Abraham, K.T. and Lim, C.M. (2012) 'Computer-assisted diagnosis of tuberculosis: a first order statistical approach to chest radiograph', *J. Med. Syst*., Vol. 36, No. 5, pp.2751–2759.

Van Ginneken, B., Katsuragawa, S., ter Haar Romeny, B.M., Doi, K. and Viergever, M.A. (2002) 'Automatic detection of abnormalities in chest radiographs using local texture analysis', *IEEE Transactions on Medical Imaging*, Vol. 21, No. 2, pp.139–149.

World Health Organization (WHO) (2018) *Global Tuberculosis Report*.