

Original paper

Deep learning-based automatic detection of tuberculosis disease in chest X-ray images

Eman Showkatian^{1,B,C,E,F}, Mohammad Salehi^{1,B,C,E}, Hamed Ghaffari^{1,A,C,D,E}, Reza Reiazi^{1,2,A,D,G}, Nahid Sadighi^{3,D,F}

¹Department of Medical Physics, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

²Department of Radiation Physics, Division of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

³Advanced Diagnostic & Interventional Radiology Research Center (ADIR), Tehran University of Medical Sciences (TUMS), Tehran, Iran

Abstract

Purpose: To train a convolutional neural network (CNN) model from scratch to automatically detect tuberculosis (TB) from chest X-ray (CXR) images and compare its performance with transfer learning based technique of different pre-trained CNNs.

Material and methods: We used two publicly available datasets of postero-anterior chest radiographs, which are from Montgomery County, Maryland, and Shenzhen, China. A CNN (ConvNet) from scratch was trained to automatically detect TB on chest radiographs. Also, a CNN-based transfer learning approach using five different pre-trained models, including Inception_v3, Xception, ResNet50, VGG19, and VGG16 was utilized for classifying TB and normal cases from CXR images. The performance of models for testing datasets was evaluated using five performances metrics, including accuracy, sensitivity/recall, precision, area under curve (AUC), and F1-score.

Results: All proposed models provided an acceptable accuracy for two-class classification. Our proposed CNN architecture (i.e., ConvNet) achieved 88.0% precision, 87.0% sensitivity, 87.0% F1-score, 87.0% accuracy, and AUC of 87.0%, which was slightly less than the pre-trained models. Among all models, Exception, ResNet50, and VGG16 provided the highest classification performance of automated TB classification with precision, sensitivity, F1-score, and AUC of 91.0%, and 90.0% accuracy.

Conclusions: Our study presents a transfer learning approach with deep CNNs to automatically classify TB and normal cases from the chest radiographs. The classification accuracy, precision, sensitivity, and F1-score for the detection of TB were found to be more than 87.0% for all models used in the study. Exception, ResNet50, and VGG16 models outperformed other deep CNN models for the datasets with image augmentation methods.

Key words: tuberculosis, machine learning, deep learning, transfer learning.

Introduction

Tuberculosis (TB) is a lung disease that occurs owing to bacterial infection [1]. TB is caused by a type of bacteria called the bacillus *Mycobacterium tuberculosis*. This airborne communicable disease is considered as the one of the top 10 leading causes of death worldwide [1]. In contrast to other infectious diseases, early diagnosis of TB is relatively diffi-

cult, and therefore several diagnostic tests will be required to detect TB [2]. Hence, World Health Organization (WHO) recommends broad use of systematic screening strategy to provide early diagnosis of TB. Although chest X-ray (CXR) has low specificity and remarkable intra- and inter-observer variations, WHO recommends postero-anterior chest radiography as one of the primary methods for detection and screening of TB owing to its relatively high sensitivity [3,4]. As stated above, however, there are significant intra- and

Correspondence address:

Reza Reiazi, Department of Radiation Physics, Division of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA, e-mail: rreiazi@mdanderson.org

Authors' contribution:

A Study design · B Data collection · C Statistical analysis · D Data interpretation · E Manuscript preparation · F Literature search · G Funds collection

inter-observer variations in the interpretation of CXR, resulting in misdiagnosis of TB [5,6]. Moreover, CXR interpretation is a time-consuming and a subjective process. Also, there is similarity between radiologic patterns of TB and other lung diseases that can lead to misdiagnosis [6]. To this end, in recent years, computer aided diagnosis (CAD) systems have emerged to automatically detect TB from chest radiography [7]. CAD solutions apply image segmentation, texture and shape feature extraction, and classification procedure for diagnosis of pulmonary TB from chest radiography [8].

Over the last decade, artificial intelligence (AI) based solutions have been applied for automated detection of breast cancer and brain tumor, physiological monitoring, etc. [9-11]. Machine learning as a branch of AI enables the computers learn without human intervention (self-learn), automatically without being explicitly programmed [12-14]. In other words, machine learning can discover pattern in data (e.g., images). Deep learning is a subfield of machine learning which uses multiple layers to extract higher-level features from the raw input. In recent years, deep learning algorithms are considered as the state of the art for image classification. Among deep learning approaches, deep convolutional neural networks (DCNNs) have emerged as an attractive technique for image classification [12-14]. Recently, several studies have used CNNs for automated detection of lung diseases such as pneumonia from CXR [11]. In deep learning framework, concept transfer learning is utilized to identify TB using pre-trained models and their ensembles [15]. Herein, we aimed to train a CNN model from scratch to automatically detect TB from CXR images and compare its performance with transfer learning based technique of five different pre-trained CNNs.

Material and methods

Dataset

In the current study, two publicly available datasets of postero-anterior chest radiographs were used, which are from Montgomery County, Maryland, and Shenzhen, China (Table 1) [16]. Figure 1 shows sample cases of normal and TB of both datasets. Both datasets are publicly available here: <http://archive.nlm.nih.gov/repos/chestImages.php>.

China Set – The Shenzhen set

It was collected by the National Library of Medicine, Maryland, USA in cooperation with Shenzhen No. 3 Peo-

ple's Hospital, Guangdong Medical College, Shenzhen, China. The radiographs were captured using a Philips DR Digital Diagnose system. This dataset consists of 662 CXR images, of which 326 are normal cases and 336 are infected with TB along with radiologist reading.

Montgomery County X-ray Set

The radiographs of this dataset were acquired by The National Library of Medicine in collaboration with The Department of Health and Human Services of Montgomery County, MD, USA. This dataset has a total of 138 postero-anterior chest radiographs, of which 58 CXR images are TB and 80 are normal CXR images. Radiology readings also are available for this dataset.

Data preprocessing and augmentation

In the present study, the CXR images were resized to 256×256 pixels because the size of input images was different. Then, data augmentation techniques were applied. It has been reported that using data augmentation can result in improving classification accuracy of deep learning algorithms. Furthermore, data augmentation can significantly increase the samples in the datasets for training models. Herein, we performed image augmentation with a rotation range of 10, a width shift range of 0.1, a height shift range of 0.1, a zoom range 0.1, and horizontal/vertical flipping. Figure 2 shows the samples of image augmentation. After data augmentation, 2040 images were belonged to 2 categories for training phase (49.2% TB and 50.8% normal). 120 images (49% TB and 51% normal) for validation phase and 120 images (49% TB and 51% normal) for testing phase were belonged.

Proposed model

In this work, we trained a convolutional neural network (CNN or ConvNet) from scratch that the term "Conv" indicates the convolution layer. :

Convolutional layer

Convolutional layers are considered as the core building block used in CNN. CNNs use convolution in place of general matrix multiplication. Convolutional layers consist of a series of filters also known convolutional kernels. The main task of convolutional layer is to extract certain features

Table 1. Summary of tuberculosis (TB) chest X-ray datasets

Dataset	No. of healthy cases	No. of TB cases	File type	Bit depth	Radiology system	Resolution	Average age (years)	Sex
Shenzhen, China	80	58	PNG	8 bit	DR	3000 x 3000	33.4 ± 14.0	66.4% (men) and 33.6% (women)
Montgomery County, MD, USA	326	336	PNG	8 bit	CR	4020 x 4892	33.1 ± 18.1	44.2% (men) and 55.8% (women)

TB – tuberculosis, PNG – portable network graphic, DR – digital radiography, CR – computed radiography

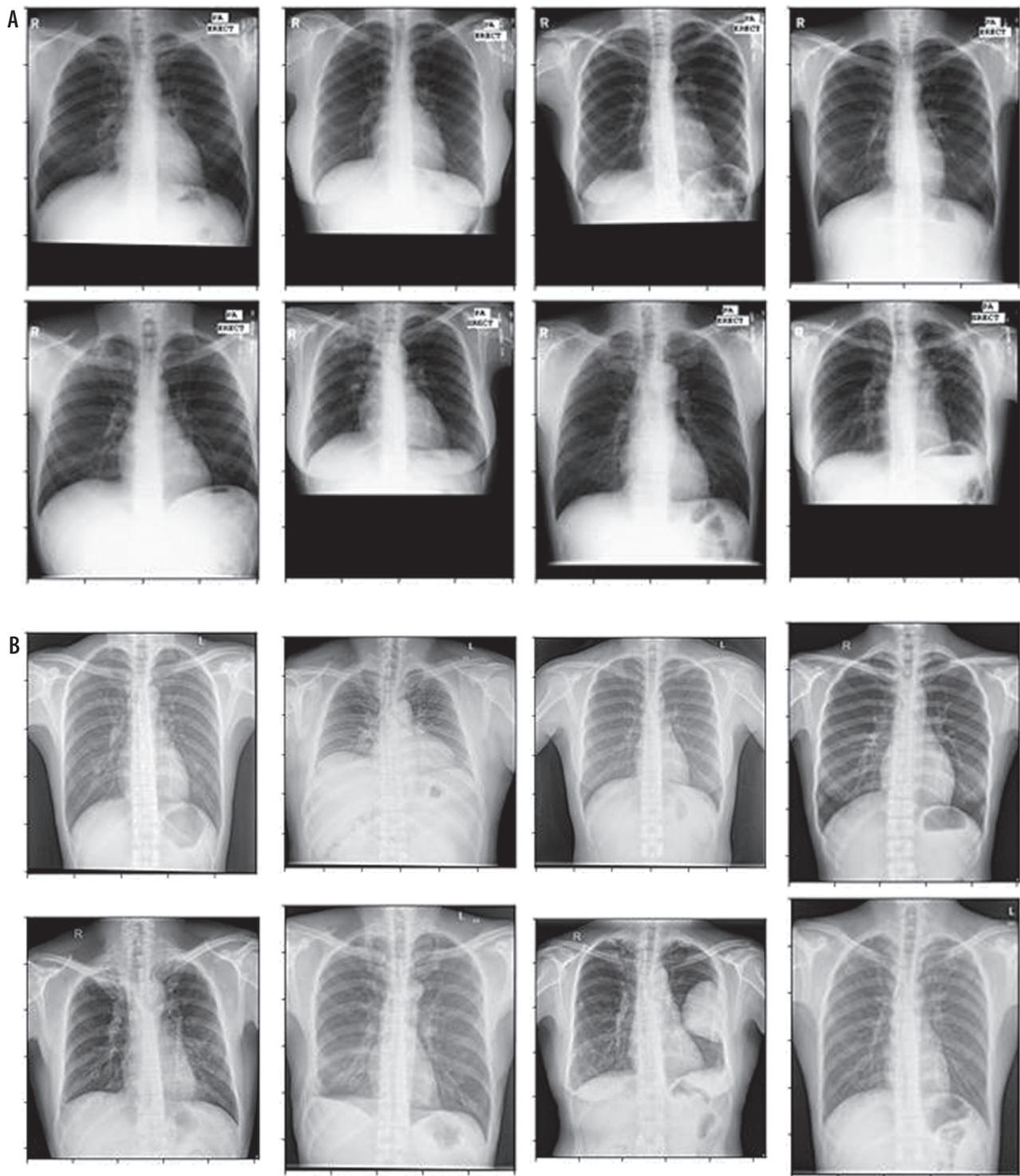


Figure 1. Samples of chest X-ray images from the Shenzhen, China dataset (A) and Montgomery County, MD dataset (B) with corresponding labels

from an input image using kernels. Usually, Rectified Linear Unit (ReLU) is used as activation layer for deep learning.

Pooling layer

In CNN, after convolutional layer, an optional pooling or down sampling layer is applied to decrease the spatial size of the input data, resulting in reducing the number of parameters in the network. Max Pooling is known as the most

common pooling technique. Beside, average pooling and L2-norm pooling are other pooling techniques.

Fully connected layer

Fully connected layers connect each neuron in previous layer to every neuron in next layer. The output from the final Pooling or convolutional layer that is flattened is the input to the fully connected layer. Softmax and Sup-

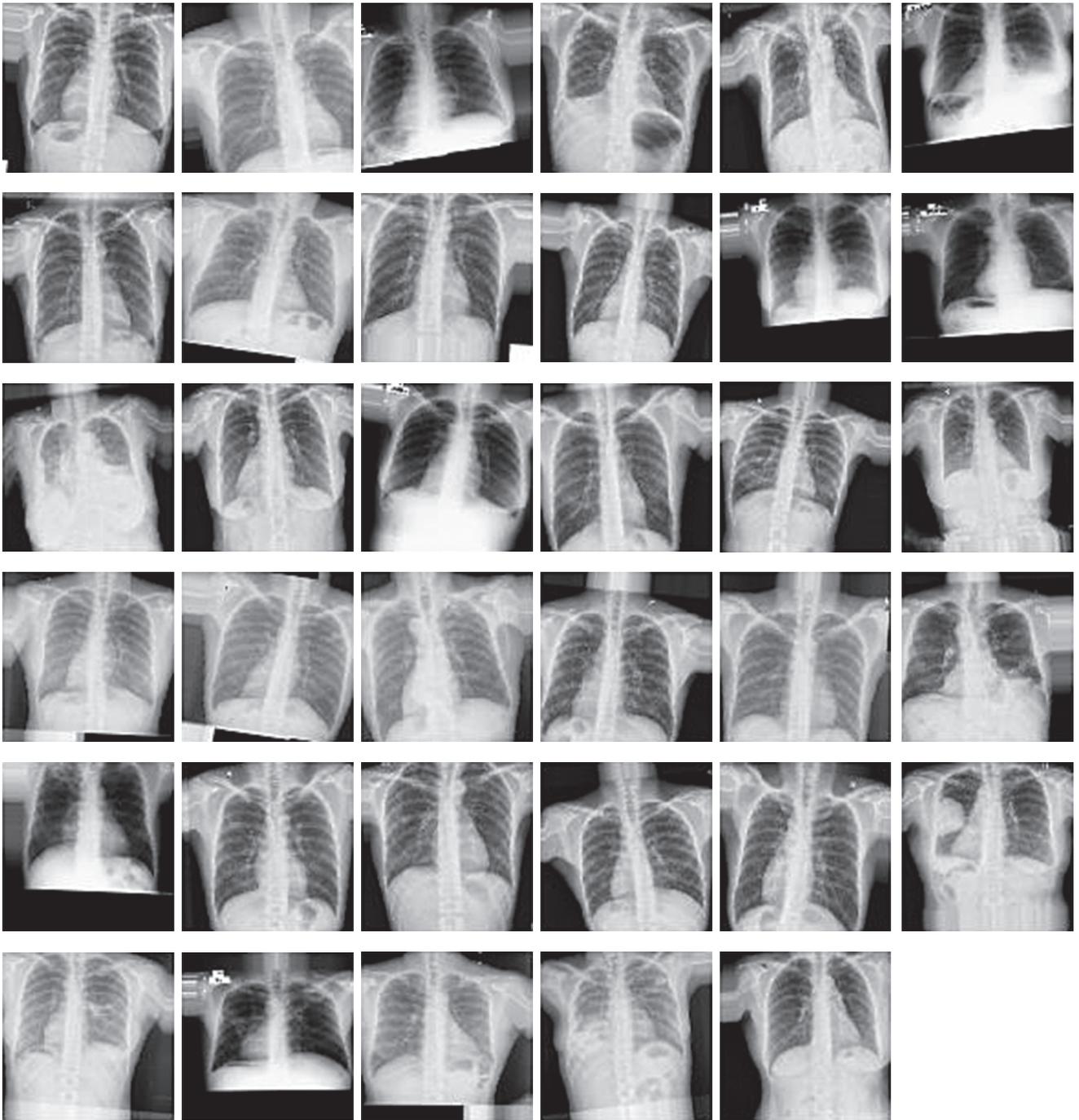


Figure 2. The samples of image augmentation

port Vector Machines (SVM) are the two most important classifiers used in CNN.

Figure 3 illustrates training process of our proposed CNN model. The size of input image was 96×96 pixels. Our model consists of three blocks. In the first block, three convolutional layers were used, each of which has 32 filters, each with 3×3 dimensions, and a ReLU activation function. Then, we used a Max Pooling layer with a pool size (2, 2) to perform down-sampling. As stated earlier, this factor was used to reduce the size of the images, which reduces the model parameters while preserving

the image information. The second and third blocks are the same. But in the second block, 64 filters with size 3×3 and in the third block, 128 filters with size 3×3 have been used. A flatten layer was used to convert the tensor extracted from the third block into a one-dimensional vector. Next, a dense layer with 256 nodes and a ReLU activation function was used. Then, we utilized a dropout with a threshold of 0.03 to improve generalization. Finally, we used the last dense layer for reducing the vector of height 64 to a vector of 2 elements. The output of the classification model in this problem was two-class classification.

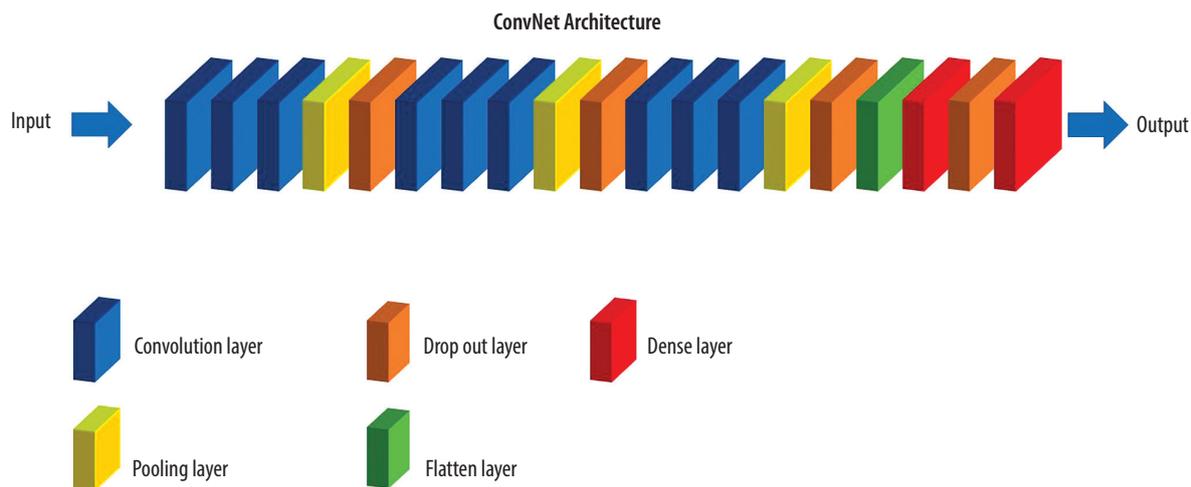


Figure 3. The illustration diagram of the CNN architecture proposed in this study

Pre-trained transfer models

Herein, the datasets containing TB CXR images were used. For the training process, the image augmentation techniques were used following dataset pre-processing and splitting. The data augmentation results in preventing the risk of over-fitting.

A CNN-based model was used to detect TB from CXR images. In the current work, a CNN-based transfer learning approach using five different pre-trained models, including Inception_v3, Xception, ResNet50, VGG19, and VGG16 was utilized for the classification of CXR images to normal and TB (binary classification). The overall architecture of the proposed CNN models consists of feature extractors and a classifier (sigmoid activation function). In other words, the architecture of all pre-trained transfer models is similar and contains 4 following parts: (1) convolution layer, (2) pooling layer, (3) flattening layer, and (4) fully connected layer. Herein, the fully connected layer comprises following layers: (1) flatten, (2) dense with a size of 256 units, (3) dropout with a threshold of 0.2, and (4) a last dense with the two-element softmax activation to predict the normal and TB cases (binary classification).

Training phase

We trained the proposed models using the categorical cross-entropy loss function and Adam optimizer to re-

duce the dimension of extracted features. For training, the following solver parameters were applied: learning rate of 0.00001, batch size of 50, and epoch value of 200. As mentioned above, data augmentation methods have been used to increase training efficiency and prevent over-fitting. In the present study, we used Python library for training, validation, and testing of different algorithms. The training of CNNs was implemented using a standard PC with a GeForce GTX 8 GB NVIDIA and 32 GB RAM. The performance of binary classification of the proposed models was evaluated using the Holdout method.

Evaluation criteria

The performance of different proposed models for testing dataset was evaluated and compared using five performance metrics, including accuracy, sensitivity/recall, Precision, Area under curve (AUC), and F1 score, as shown in equation (1-4).

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{1}$$

$$\text{Recall} = \frac{(TP)}{(TP + FN)} \tag{2}$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{3}$$

$$\text{F1-score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{4}$$

Table 2. Model performance on the test set

Models	Precision	Recall	F1-score	Accuracy	AUC	Validation loss
ConvNet	0.88	0.87	0.87	0.87	0.87	0.39
Exception	0.91	0.91	0.91	0.90	0.91	0.38
Inception_V3	0.88	0.88	0.88	0.88	0.88	0.50
ResNet50	0.91	0.91	0.91	0.90	0.91	0.41
VGG16	0.91	0.91	0.91	0.90	0.91	0.26
VGG19	0.90	0.90	0.90	0.89	0.90	0.53

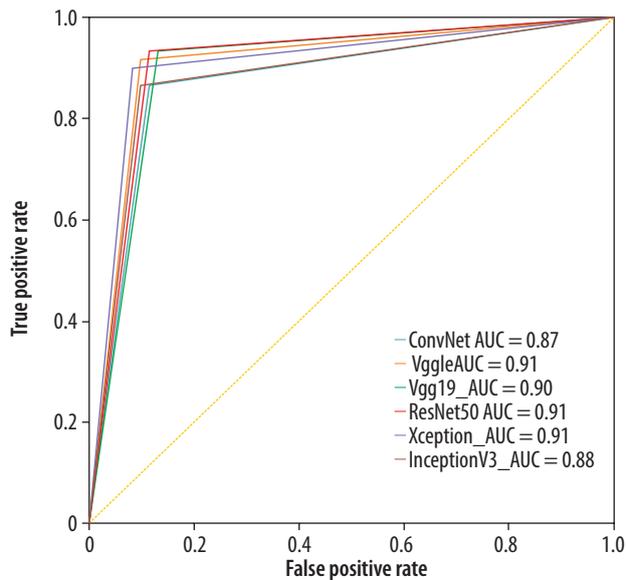


Figure 4. Receiver operation characteristic (ROC) curves of all convolutional neural network (CNN) models on the dataset

where: TP – true positive, FP – false positive, TN – true negative, and FN – false negative.

Results and Discussion

Table 2 outlines the performance metrics of each proposed model for detection TB on the test dataset. As shown in Table 2, all CNNs provided an acceptable accuracy for two-class classification. Our proposed CNN architecture (*i.e.*, ConvNet) achieved 88.0% precision, 87.0% sensitivity, 87.0% F1-score, 87.0% accuracy, and AUC of 87.0%, which is slightly less than the pre-trained models. Among all models, Exception, ResNet50, and VGG16 provided the highest classification performance of automated TB classification with precision, sensitivity, F1-score, and AUC of 91.0%, and 90.0% accuracy, as observable in Table 2. Loss for validation dataset for each proposed model is outlined in Table 2. The closer the value of the validation loss is to zero, the more efficient the model. Previously, several groups of researchers have applied classical machine learning techniques to classify TB and non-TB cases from CXR images [17,18]. Using deep learning algorithms have been also reported in the detection of TB.

The idea of transfer learning in deep learning framework is the detection of TB utilizing pre-trained models and their ensembles. In a study, Hooda *et al.* have used a deep learning approach to classify chest radiographs into TB and non-TB categories with an accuracy of 82.09% [19]. Also, Evalgelista *et al.* have reported accuracy of 88.76% for automated detection of TB using CNNs [20]. In another study, Nguyen *et al.* have evaluated the performance of a pre-trained DenseNet model to classify TB and normal cases from CXR images from Shenzhen, China and Montgomery County databases using fine-tuned model, and found the AUC values of 0.94 and 0.82, respectively [21]. Ahsan *et al.* have utilized a generalized pre-trained CNN model for classifying TB. Their model provided accuracies of 81.25% and 80% with and without using image augmentation, respectively [22]. Yadav *et al.* reported 94.89% accuracy for TB detection using transfer learning technique [23]. As outlined in Table 2, our data are in line with the aforementioned studies.

Figure 4 illustrates the receiver-operating characteristic (ROC) curves of all CNN models on the test set. The ROC curve indicates the trade-off between recall and specificity [24]. The AUC is considered as an effective way to indicate the accuracy of ROC produced by each proposed model.

Conclusions

Our study presents a transfer learning approach with deep CNNs to automatically classify TB and normal cases from the chest radiographs. The performance of six different CNN models was evaluated for the detection TB CXR images. Exception, ResNet50, and VGG16 models outperformed other deep CNN models for the datasets with image augmentation methods. The classification accuracy, precision, sensitivity, and F1-score for the detection of TB were found to be more than 87.0% for all models used in the study. Using bigger dataset can increase the accuracy and robustness of our proposed models, which this point needs to be addressed in future works.

Conflict of interest

The authors report no conflict of interest.

References

- World Health Organization. Global tuberculosis report 2020. https://www.who.int/publications/i/item/9789240013131_eng.pdf (Accessed: 13.12.2020).
- Sathitratanaheewin S, Sunanta P, Pongpirul K. Deep learning for automated classification of tuberculosis-related chest X-Ray: dataset distribution shift limits diagnostic performance generalizability. *Heliyon* 2020; 6: e04614.
- Silverman C. An appraisal of the contribution of mass radiography in the discovery of pulmonary tuberculosis. *Am Rev Tuberc* 1949; 60: 466-482.
- van't Hoog AH, Meme HK, Laserson KF, *et al.* Screening strategies for tuberculosis prevalence surveys: the value of chest radiography and symptoms. *PLoS One* 2012; 7: e38691.
- Ahmad Khan F, Pande T, Tessema B, *et al.* Computer-aided reading of tuberculosis chest radiography: moving the research agenda forward to inform policy. *Eur Respir J* 2017; 50: 1700953.
- Graham S, Das GK, Hidvegi RJ, *et al.* Chest radiograph abnormalities associated with tuberculosis: reproducibility and yield of active cases. *Int J Tuberc Lung Dis* 2002; 6: 137-142.

7. Hooda R, Mittal A, Sofat S. Segmentation of lung fields from chest radiographs-a radiomic feature-based approach. *Biomed Eng Lett* 2018; 9: 109-117.
8. Jaeger S, Karagyris A, Candemir S, et al. Automatic tuberculosis screening using chest radiographs. *IEEE Trans Med Imaging* 2014; 33: 233-245.
9. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer* 2018; 18: 500-510.
10. Rahman T, Chowdhury ME, Khandakar A, et al. Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection using Chest X-ray. *Appl Sci* 2020; 10: 3233.
11. Mohammadi R, Salehi M, Ghaffari H, et al. Transfer Learning-Based Automatic Detection of Coronavirus Disease 2019 (COVID-19) from Chest X-ray Images. *J Biomed Phys Eng* 2020; 10: 559-568.
12. Choy G, Khalilzadeh O, Michalski M, et al. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology* 2018; 288: 318-328.
13. Do S, Song KD, Chung JW. Basics of Deep Learning: A Radiologist's Guide to Understanding Published Radiology Articles on Deep Learning. *Korean J Radiol* 2020; 21: 33-41.
14. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *Radiographics* 2017; 37: 505-515.
15. Lopes UK, Valiati JF. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Comput Biol Med* 2017; 89: 135-143.
16. Jaeger S, Candemir S, Antani S, et al. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imaging Med Surg* 2014; 4: 475-477.
17. Melendez J, Sánchez CI, Philipsen RHHM, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Sci Rep* 2016; 6: 25265.
18. Ginneken BV, Katsuragawa S, Romeny BMtH, et al. Automatic detection of abnormalities in chest radiographs using local texture analysis. *IEEE Trans Med Imaging* 2002; 21: 139-149.
19. Hooda R, Sofat S, Kaur S, et al. Deep-learning: a potential method for tuberculosis detection using chest radiography. 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)2017. p. 497-502.
20. Evalgelista L, Guedes E. Computer-Aided Tuberculosis Detection from Chest X-Ray Images with Convolutional Neural Networks 2018.
21. Nguyen QH, Nguyen BP, Dao SD, et al. Deep Learning Models for Tuberculosis Detection from Chest X-ray Images. 2019 26th International Conference on Telecommunications (ICT) 2019. p. 381-385.
22. Ahsan M, Gomes R, Denton A. Application of a Convolutional Neural Network using transfer learning for tuberculosis detection. 2019 IEEE International Conference on Electro Information Technology (EIT) 2019. p. 427-433.
23. Yadav O, Passi K, Jain CK. Using Deep Learning to Classify X-ray Images of Potential Tuberculosis Patients. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2018. p. 2368-2375.
24. Streiner DL, Cairney J. What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry/La Revue canadienne de psychiatrie* 2007; 52: 121-128.