University of Sindh Journal of Information and Communication Technology (USJICT)
Volume 7, Issue 1
ISSN-E: 2523-1235, ISSN-P: 2521-5582
Website: http://sujo.usindh.edu.pk/index.php/USJICT/

Automated Pulmonary Tuberculosis Screening Based on Deep Convolutional Neural Networks Using Chest Radiographs

Amrat Anmol[1], Mahwish Ilyas[1], Muhammad Bilal[2,3,*], Hikmat Ullah Khan[4], Muhammad Ramzan[5]

[1] Department of Computer Science, University of Sargodha, Sargodha 40100, Pakistan
[2] Department of Computing and Information Systems, School of Engineering and Technology, Sunway University, Petaling Jaya 47500, Selangor, Malaysia
[3] Department of Pharmaceutical Outcomes and Policy, Malachowsky Hall for Data Science and Information Technology, University of Florida, Gainesville 32610, Florida, USA
[4] Department of Information Technology, University of Sargodha, Sargodha 40100, Pakistan
[5] Department of Software Engineering, University of Sargodha, Sargodha 40100, Pakistan
amratanmol@gmail.com; mahwishilyas@gmail.com; muhammad.bilal@ufl.edu; hikmat.khan@uos.edu.pk; muhammad.ramzan@uos.ed.pk

Abstract: Pulmonary Tuberculosis (PTB) is a highly dangerous illness that can have severe consequences if left untreated. Researchers and physicians are taking more interest in using chest radiography to automate the investigation of pulmonary tuberculosis. For early detection and treatment, computer-aided diagnosis (CAD) systems are being used to automate the diagnosis of TB. Previous literature shows that many deep learning-based approaches (DL) have been introduced for the classification of pulmonary tuberculosis using chest radiographs. In this study, we introduced a modified convolutional neural network (CNN) model comprised of 21 layers with different architectural layers. This study aims to classify images into TB and Normal. The presented methodology is trained and evaluated on the dataset created by combining multiple publicly available datasets. Results of the proposed modified CNN model have been compared with the seven individual pre-trained state-of-art models, including VGG16, VGG19, ResNet50, InceptionV3, Xception, MobileNetV2, and EfficientNetB7. The comparison demonstrates that the modified CNN model outperformed pre-trained models by achieving an accuracy of 0.9081, precision of 0.9317, recall of 0.9323, F1 score of 0.9307, and AUC of 0.9474.

Keywords: Tuberculosis; Classification; CNN, Deep learning; Pre-trained; Chest Radiographs.

I. INTRODUCTION

Pulmonary Tuberculosis (PTB) is an infectious disease that is a major reason of illness and the deadliest disease in the world. TB has the highest rate of death from a single infectious agent until the virus, say, the coronavirus (COVID-19) outbreak, ranking higher than HIV/AIDS. Tuberculosis is a treatable and preventable disease. In the face of the COVID-19 outbreak, TB fatalities have increased for the first time in over a decade due to the limited access to TB diagnosis and treatment. In 2020, about half of the people diagnosed with tuberculosis (TB) did not receive treatment and were not reported [1]. Several lung investigation techniques are available, but chest radiographs are a common screening tool for the diagnosis of pulmonary problems [2]. It is chosen over other imaging techniques like computed tomography (CT) because it is less expensive, more generally available, and has a high measure for pulmonary tuberculosis. Chest X-rays (CXRs) are widely used by experienced physicians [3].

Mass screening is prevalent in underdeveloped countries because these nations have a significant risk of tuberculosis due to excess numbers of people, poverty, and starvation. Typically, Chest radiographs collected from screening places were examined by radiologists. Unfortunately, experienced radiologists are unavailable compared to the demand for suspected patients. The performance of radiologists varies significantly due to the differences in expertise and fatigue effects. CAD systems have been created to help radiologists reduce inter-observer and intra-observer discrepancies. Sometimes, radiologists face difficulty diagnosing a patient with lung infection because several diseases have similar signs and symptoms [4]. The image processing approach will be beneficial to detect the tiny details that are missed by the radiologist during the X-ray examination [5]. The notion of CAD diagnosis for chest X-rays has existed for the last 50 years. CAD systems have traditionally performed jobs by using machine learning approaches.

DL has recently demonstrated superhuman performance on several image-based categorization tasks [6]. The advances in recognizing objects in natural images have revived interest in using DL to analyze medical images. The amazing effectiveness of CNNs for the classification of medical images has been defined in several recent studies [7]. In addition, CNN has been used in the past to develop

CAD systems for disease detection [8]. Several CNNs have been used to classify chest radiographs [9]. Pre-trained models are also widely used for image classification problems and have achieved outstanding performance on extremely complex datasets [10]. Transfer learning is an optimization method in the neural network. It will help to enhance the efficiency of the already trained model [11]. In the medical field, the pulmonary tuberculosis datasets are limited, which makes the existing diagnostic systems less effective. Computer-aided diagnosis CAD systems require many chest radiographs CXRs. Mostly, previous research was done using small datasets and combined datasets formed from publicly available or private radiographs. In this study, we used another combined dataset, which has been collected from four publicly available datasets, namely Shenzhen, Montgomery, JSRT, and Belarus. The included size of the dataset is also different from previous reports.

We have studied deep neural networks that include several hidden layers and have proved very successful in image classification. We propose a Modified CNN model that will automatically detect TB manifest using a huge number of CXR at a time. In comparison to the costly, time-consuming, and inaccurate due to a lack of professional radiologists and inconvenient method of manual CXR examination, the suggested model will alleviate the inconvenience of patients waiting in line for days to be examined, provide guaranteed improved diagnosis, performance, accuracy, and eventually reduce screening costs. This paper also evaluated the performance of seven different pre-trained CNN models with fine-tuning for the classification of TB and normal CXR images. For this purpose, we have used VGG16 [12], VGG19, ResNet50 [13], InceptionV3 [14], Xception [15], MobileNetV2 [16] and EfficientNetB7 [17] among the available alternatives.

The rest of the paper is divided into the following sections. Section 2 provides the previous research on deep learning-based tuberculosis detection. In Section 3, a brief description of the datasets, preprocessing techniques, proposed methodology, and pre-trained architectures is discussed. Section 4 discusses classification results using chest radiographs and compares them with other models. Finally, the conclusion is drawn in Section 5.

II. Literature Review

Automated detection of chest radiographs has received increasing interest. Different deep convolutional networks have been used to detect pulmonary tuberculosis. Lakhani et al. [18] implemented two DCNN classifiers, AlexNet and Google-Net, and presented their ensemble network. Four experiments were performed on four HIPAA-compliant datasets using the pre-trained network without augmentation, pre-trained network with augmentation, untrained network without augmentation, and untrained network with augmentation. Results show that the ensemble network gives outstanding performance, achieving 99.9% of AUC. Hooda et al. [19] presented a hybrid of three convolutional architectures, namely AlexNet, Google-Net, and ResNet. Architecture is applied to combined images of the Shenzhen dataset, Montgomery, Belarus, and JSRT dataset, which achieve accuracy and AUC of 88.24% and 0.93, respectively. This study also claims ensemble architecture successfully achieved higher performance than other techniques, which were reported in their literature. Mustapha et al. [20] presented a model based on deep CNN to efficiently classify Tb and non-Tb images from the Montgomery X-rays dataset. The technique of augmentation was applied to increase the size of the dataset from 138 samples to 5000 samples, and some preprocessing techniques, histogram equalization, noise removal, horizontal, vertical flip, random zoom, and rotation were used. The proposed model achieves 87.1% validation accuracy.

Verma et al. [21] proposed a novel framework based on a neural network, which efficiently classified different pulmonary infections, namely pulmonary tuberculosis, viral pneumonia, and bacterial pneumonia. The proposed architecture achieves 99.01% high accuracy. Hwang et al. [22] developed a DL-based algorithm for the automatic detection of pulmonary tuberculosis. This study performed the image classification of TB and localization of lesion abnormalities. The proposed model has 27 layers with 12 connections applied on six independent datasets. The algorithm achieved 0.977–1.000 performance and 0.973–1.000 localization performance. Sangheum et al. [23] developed another tuberculosis screening system based on the well-known deep CNN architecture AlexNet. One extra convolutional layer is added as feature extraction. Three radiograph datasets, namely KIT, Montgomery, and Shenzhen, were included in the study. Results were generated by performing two experiments with and without transfer learning. Results are 0.96, 0.93, 0.88 in terms of AUC, and 0.902, 0.905, 0.903 in terms of accuracy for three independent datasets respectively obtained, which demonstrates that the proposed model gives high performance by using transfer learning.

Rehman et al. [24] evaluated nine different pre-trained models for tuberculosis classification and segmentation. Evaluation has been performed by conducting three experiments: segmentation of X-rays, classification of X-ray images, and segmented images. Comparison results demonstrate that ChexNet outperformed X-ray images by 96.47% in terms of accuracy. Further, Dens Net shows 98.6% higher accuracy than other models for segmented images. Munadi et al. [25] evaluated three image enhancement techniques on the Shenzhen dataset. By employing EfficientNet-B4 for detection. The model achieved 0.899 and 0.948 accuracy and AUC with unsharp masking. Meanwhile, the High-Frequency Emphasis Filtering model obtained 0.8908 accuracy and 0.931 AUC. Hooda et al. [26] proposed a CNN model with seven convolutional layers and three fully connected ones. Publicly available datasets, Shenzhen and Montgomery, were included in the study. Images were resized to the 224×224 size. Results are compared based on the different optimizers, namely, Adam, momentum, and SGD, which show that the proposed model achieved 82.09% higher accuracy with Adam than with other optimizers.

For the diagnosis of pulmonary abnormalities, Ashan et al. [27] applied the VGG16 model to the Shenzhen and Montgomery datasets by using the concept of pre-trained learning. Two experiments have been performed, the model applied to and without augmented images. For data augmentation, random rotation, cropping, flipping, brightness, contrast, zooming, and bilinear interpolation were used. By comparing results, it indicated that VGG16 provides 81.25% high accuracy on augmented images. Hooda et al. [28] presented a proposed model based on CNN architecture that has ten blocks that are divided into four sets. An ensemble network is also used for their experiment. In the designed ensemble network, the block-based proposed architecture and two other standard architectures, AlexNet and ResNet, have been used. Four datasets were involved. The technique of data augmentation with histogram equalization has been adopted. Block-based architecture attained 0.876 and 0.93, and ensemble architecture achieved 0.90 and 0.96 in terms of accuracy and AUC. Liu et al. [29] designed a simple CNN network for pulmonary tuberculosis segmentation and classification. The proposed system has two regions: segmentation and feature extraction. ResNet-152, Inception v4, and Inception-ResNet v2 worked separately as a classification backbone. Three datasets were part of the study, namely, Shenzhen and Montgomery, and the other is a private labeled dataset. In three different experiments, ResNet152 attained a high accuracy of 0.923 on Shenzhen and 0.89 on the Montgomery dataset. Abbas and Abdelsamea [30] evaluated a pre-trained AlexNet model on the Montgomery dataset. Transformation techniques have been adopted that help to expand the small-size dataset. Fine-tuned based CNN model achieved 0.998 of AUC. Results have been compared with various networks, which concluded that fine-tuned networks performed better than other deep-tuned networks and shallow-tuned networks.

Heo et al. [31] presented a deep-learning-based tuberculosis diagnoses system that also considered demographic information, namely age, height, weight, and gender. They were able to attain an AUC of 0.92. They concluded that demographic information can enhance the accuracy of TB detection. Imam et al. [32] proposed a modified inception model using the 'Mixed 8' layer and evaluated various architectures, namely, VGG16 and VGG19, by adding extra layers. The Modified proposed model achieved 0.91 validation accuracy on combined datasets of Shenzhen and Montgomery dataset. The model for the remaining datasets, including Shenzhen and Montgomery, has not performed well.

Al-Timemy et al. [33] proposed a new methodology for the detection of Five class datasets of x-rays, including COVID-19, viral, and bacterial pneumonia, TB, and healthy cases, also for three class datasets including COVID-19 TB, and healthy cases. The methodology consists of a pipeline of feature extraction with 14 pre-trained models, which are integrated with traditional machine learning methods and an ensemble of subspace discriminants for classification. Results show that the pipeline with ResNet50 as deep feature extraction produced the best detection accuracy of 91.6± 2.6% for five class datasets and 98.6±1.4% accuracy for three class datasets. Luajan-Garca et al. [34] proposed a NanoChest-net for the classification using chest radiographs, such as tuberculosis and COVID-19. The proposed model with fewer parameters achieved the results only on the pneumonia and COVID-NORMAL datasets. It has given AUC up to 99.2%. The model for the remaining datasets, including Shenzhen and Montgomery, has not performed well.

III. Materials and Methods

In this section, we have included a summary of the datasets chosen, as well as the preprocessing approaches used. We also discuss the proposed methodology and the pre-trained architectures in detail, and we provide a short overview of the evaluation metrics used. The overall method is illustrated using

.

A. Datasets Description

In the medical field, there is a shortage of publicly accessible datasets for the classification of pulmonary tuberculosis. Many studies have used small datasets such as the Shenzhen and Montgomery datasets. They consist of 662 and 138 images. This study used four standard publicly accessible datasets to create a combined dataset. Sample CXR images of the combined dataset are shown in Figure. 1. A brief description of these datasets is as follows:

1) Shenzhen dataset

This dataset has been generated by Shenzhen No.3 Hospital in Shenzhen, Guangdong, China [35]. It contains 662 chest X-rays, consisting of 326 normal cases and 336 abnormal cases, and some additional clinical information. Images are in PNG format and approximately 3K × 3K pixels in size. All images of the Shenzhen dataset have been included in our combined dataset.

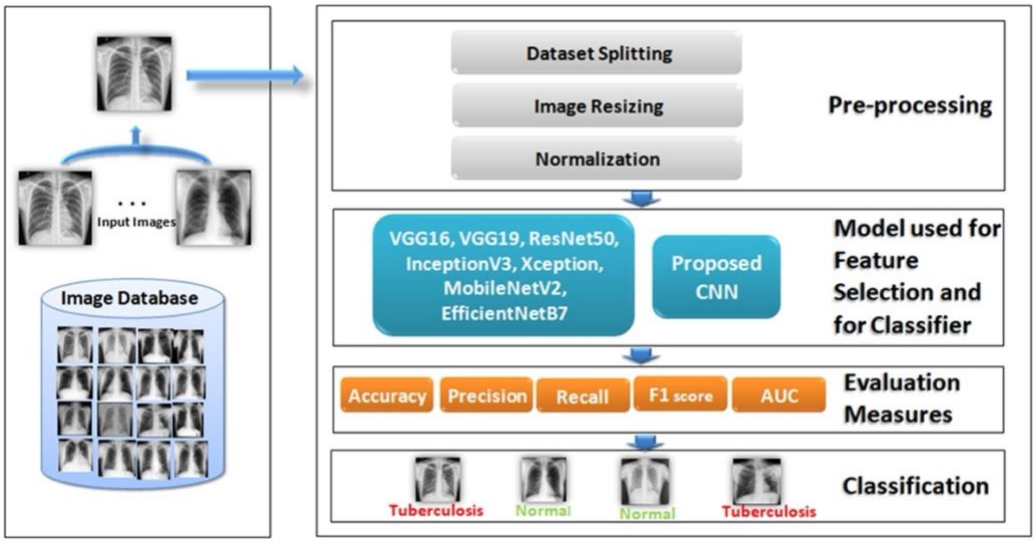

Figure 1. The proposed deep learning framework for pulmonary Tuberculosis classification

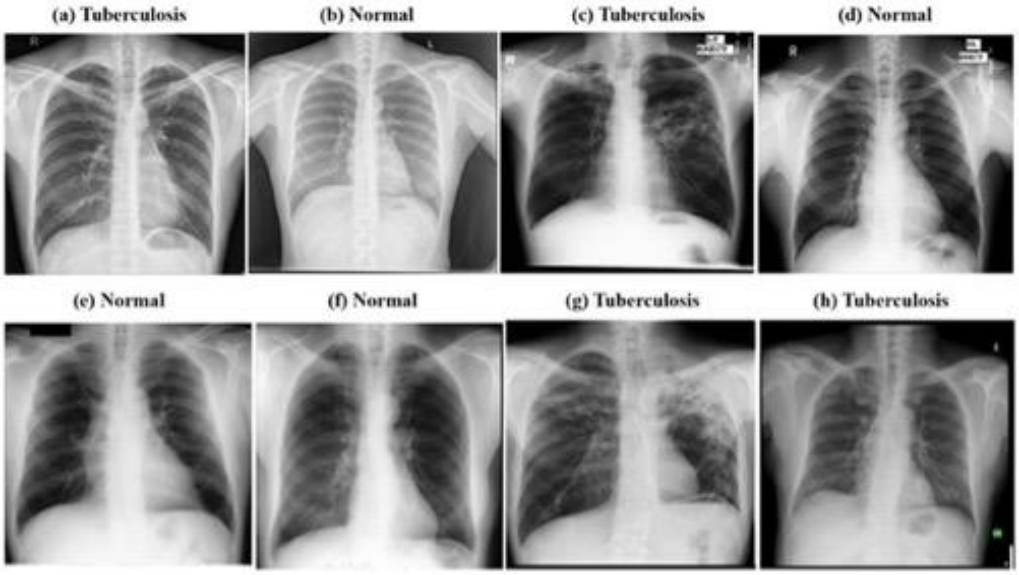

Figure. 1. Sample CXR images of the combined dataset (a) Shenzhen (Tuberculosis), (b) Shenzhen (Normal) (c) Montgomery (Tuberculosis), (d) Montgomery (Normal), (e) JSRT (Normal), (f) JSRT (Normal), (g) Belarus (Tuberculosis), (h) Belarus (Tuberculosis)

2) Montgomery dataset

This dataset has been generated by the tuberculosis control program of the Department of Health and Human Services of Montgomery County, MD, USA [35]. It contains images of 138 chest X-rays, consisting of 80 normal and 58 abnormal cases. Our combined dataset includes some additional clinical information. The images are in PNG format. All images of the Montgomery dataset have been included in our combined dataset.

1) JSRT dataset

This dataset has been generated by the Japanese Society of Radiological Technology (JSRT) and the Japanese Radiological Society [36]. The dataset contains 247 CXRs, consisting of 154 nodule images and 93 normal CXRs, and includes some additional clinical information like age, gender, and location of nodule. The X-rays have an image resolution of 2,048×2,048 pixels. Only 93 normal CXRs have been included in our combined dataset.

2) Belarus dataset

For a drug resistance study, this dataset was generated by the National Institute of Allergy and Infectious Diseases, Ministry of Health, the Republic of Belarus [37]. It contains 420 TB CXRs and CT and includes some additional clinical information. The CXRs have an image resolution of 2,248×2,248 pixels. We have included 306 TB CXRs for our combined dataset.

B. Distribution of Dataset

The combined dataset comprises images of 1199 CXRs, 700 of which have TB manifestations and 499 of which are normal. **Error! Reference source not found.** shows the original distribution of each dataset for both classes, TB and Normal. Of 1199 images, 840 are randomly chosen for training, and the remaining 359 are used for testing. The testing set is utilized to determine the final performance of the model.

TABLE I. DETAIL OF COMBINED DATASET.

Datasets	TB	Normal	Total
Shenzhen	336	326	662
Montgomery	58	80	138
JSRT	0	93	93
Belarus	306	0	306
Total	700	499	1199

C. Preprocessing

In our combined dataset, chest radiograph images belong to different pixel intensities and different dimensions. Therefore, the dataset was preprocessed to resize the CXR to 224×224 pixels. Bilinear interpolation was applied on both training and testing sets because the target size and loaded image size were different from each other. Further, the normalization technique was also used to change the range of pixel intensity value by dividing images by 255.

D. Proposed Architecture

Previous studies have proposed various CNN models to solve the problems. In this study, we proposed a modified CNN architecture consisting of five convolution blocks, with the max-pooling layer and dropout layer in each. The architectural design of the proposed model is shown in **Error! Reference source not found.**.

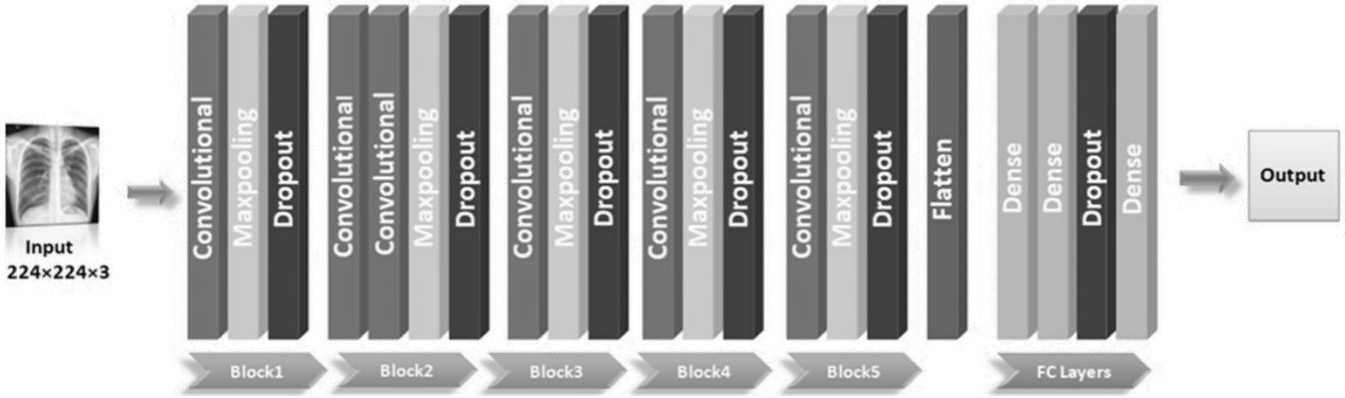

Figure 3. Proposed modified CNN architecture.

The images of size 224×224×3 was taken as input to the network. Each layer takes input for the next layer from its previous layer. The output of each layer works as an input for the next layer with a feature map. The first block consists of one convolution layer with 32 filters and padding, as well as kernel size 3. The second region consists of two convolution layers with 64 numbers of filters, padding, and a kernel size of 3. The third block consists of one layer with 128 filters and the same padding, with kernel size 3. The fourth and fifth regions comprised of one layer with 256 and 512 filters, respectively, and padding as well as kernel size 3. All convolution layers are activated using the Relu function. Each block has one max-pooling and a dropout layer with a 0.2 probability. The Max pooling layer facilitates lessening the spatial size of the features and assists in avoiding over-fitting. Then, the output of the block was flattened into a one-dimensional. A flatten layer is used to transition to the fully connected layer. In this network, we have used three dense or fully connected layers. After two fully connected layers, we used a dropout layer with a 0.5 probability. It's a basic regularization approach for preventing the overfitting of the layers. First, two layers are activated with the activation function Relu, and the last layer is activated with the sigmoid. The input and output size of each layer is shown in Table 2.

Adam optimizer is employed as an optimization approach, with an initial lr of 0.001. Cross-entropy is the loss function used to modify the weights at each epoch. The batch size of 32 images has been used for the training set rather than provided all at once. The model is trained over 70 epochs, each has 27 steps. Hyper-parameters are used carefully to optimize performance. All these configurations aid in the prevention of over-fitting. Hyper-parameters for the proposed model are shown in Table 3.

TABLE 3. DETAIL OF HYPER-PARAMETERS FOR THE PROPOSED MODEL.

Hyper-parameters for Proposed Model	
Batch size	32
Epochs	70
Learning Rate	0.001
Optimizer	Adam
Loss function	Binary cross-entropy

TABLE 2. DETAIL OF PROPOSED ARCHITECTURE.

	Layer Name	Input Size	Filters	Padding	Kernel size	Output Size
Block1	Convolution Layer	224×224×3	32	same	3×3	224×224×32
	Max-pooling Layer	224×224×32	-	-	-	112×112×32
	Dropout Layer	112×112×32	-	-	-	112×112×32
Block2	Convolution Layer	112×112×32	64	same	3×3	112×112×64
	Convolution Layer	112×112×64	64	same	3×3	112×112×64
	Max-pooling Layer	112×112×64	-	-	-	56×56×64
	Dropout Layer	56×56×64	-	-	-	56×56×64
Block3	Convolution Layer	56×56×64	128	same	3×3	56×56×128
	Max-pooling Layer	56×56×128	-	-	-	28×28×128
	Dropout Layer	28×28×128	-	-	-	28×28×128
Block4	Convolution Layer	28×28×128	256	same	3×3	28×28×256
	Max-pooling Layer	28×28×256	-	-	-	14×14×256
	Dropout Layer	14×14×256	-	-	-	14×14×256
Block5	Convolution Layer	14×14×256	512	same	3×3	14×14×512
	Max-pooling Layer	14×14×512	-	-	-	7×7×512
	Dropout Layer	7×7×512	-	-	-	7×7×512
	Flatten	7×7×512	-	-	-	25088
FC Layers	Dense	25088	1024	-	-	1024
	Dense	1024	256	-	-	512
	Dropout Layer	512	-	-	-	512
	Dense	512	1	-	-	1

TABLE 4. DETAIL OF HYPER-PARAMETERS FOR THE PRE-TRAINED MODELS.

Pre-trained Models	Fine-tuned at	Batch size	Learning Rate	Momentum	Optimizer	Modified FC layer	Dropout rate
VGG16	'block5_pool'	32	0.0001	0.9	SGD	512	0.5
VGG19	'block5_pool'	32	0.0001	0.9	SGD	512	0.5
ResNet50	'conv5_block3_out'	32	0.0001	0.9	SGD	1024	0.2
InceptionV3	'mixed7'	32	0.0001	0.9	SGD	1024	0.5
Xception	'block14_sepconv2'	32	0.0001	0.9	SGD	1024	0.5
MobileNetV2	'block 16_expand'	32	0.0001	0.9	SGD	512	0.5
EfficientNetB7	'top_conv'	32	0.0001	0.9	SGD	1024	0.5

In addition, we have also used seven pre-trained convolutional neural networks (CNN) that have performed well on the ImageNet dataset. As mentioned earlier, these state-of-art models including VGG16,12 VGG19, ResNet50,13 InceptionV3,14 Xception,15 MobileNetV2,16 and EfficientNetB717. The characteristics of medical data differ from the data on which these pre-trained models were trained, so the approach of fine-tuning has been adopted to utilize the concept of transfer learning [38]. Fine-tuning these architectures indicates re-training by replacing the fully connected and modified layers. Initially, freeze the entire convolutional base by setting the trainable flag of the model to false to prevent the weights from being updated while training. Fine-tuning has been recognized as one of the successful transfer learning techniques. Hyperparameter values are optimized individually for all architectures. Hyper-parameters for the selected pre-trained models are shown in Detail of hyper-parameters for the pre-trained models.

1) VGG16 and VGG19

Here, we have used two variants of VGG, including VGG16 (with 16 deep layers) and VGG19 (with 19 deep layers). Resized (224×224 pixels) images are provided to the pre-trained networks. For both networks, the SGD carried out an optimization algorithm with lr of 0.00001 and a momentum of 0.9. We used the VGG as the base model up to the 'block5_pool' layer. Then, the output of the 'block5_pool' layer is flattened. After that, 512 units of fully connected layer with Relu function have been placed. Between two dense layers, the dropout layer is used with a probability of 0.5 to avoid overfitting.

2) ResNet50

In this study, we used ResNet50, which consists of 50 layers. Resized (224×224 pixels) images were provided to the pre-trained network. SGD is an optimization approach with an lr of 0.00001 and a momentum of 0.9. We used ResNet50 as the base model up to the 'conv5_block3_out' layer, and then the output layer was flattened. After that, 1024 units of fully connected layer function were added.

Between two dense layers, the dropout layer is used with a probability of 0.2.

3) InceptionV3

Here, resized (224×224 pixels) images have been provided to the pre-trained network. SGD is used as an optimization algorithm with lr of 0.00001 and a momentum of 0.9. We used Inceptionv3 as the base model up to the 'mixed7' layer, and then the output layer was flattened. After that, 1024 units of fully connected layer function were added. Between two dense layers, the dropout layer is employed with a probability of 0.5 to neglect the overfitting.

4) Xception

In this study, an input size of 224×224 pixels is used. SGD is used as an optimization approach with lr of 0.00001 and a momentum of 0.9. We used the Xception as a base model up to the 'block14_sepconv2' layer, and then the output of that layer was flattened. After that, 1024 units of fully connected layer function were added. Between two dense layers, the dropout layer is employed with a probability of 0.5.

5) MobileNetV2

In this study, an input size of 224×224 pixels is used. SGD is used as an optimization algorithm with lr of 0.00001 and a momentum of 0.9. We used the MobileNetV2 as a base model up to the 'block_16_expand' layer, then flattened the output layer. Afterward, 512 units of fully connected layer function were used. Between two dense layers, the dropout layer is used with a probability of 0.5.

6) EfficientNetB7

In this study, an input size of 224×224 pixels is used. SGD is used as an optimization algorithm with lr of 0.00001 and a momentum of 0.9. We used the efficientNetB7 as a base model up to the 'top_conv' layer, and then the output of the layer was flattened. After that, 1024 units of fully connected layer of 1024 units function were used. Between two dense layers, the dropout layer is used with a probability of 0.5.

E. Evaluation Metrics

The following measures are used to compute the classification results: Accuracy, Precision, Recall, F-measure, and AUC. Sometimes accuracy is not a precise measure when you are working with an imbalanced dataset [39]. AUC is a metric that combines performance across all classification thresholds. It is a better measure than accuracy [40]. AUC is classification-threshold-invariant. It assesses the accuracy of the predictions regardless of the classification threshold used. Therefore, precision, recall, F1-Score, and AUC are always appropriate metrics to measure the performance of a classifier.

IV. Results and Discussion

In this section, we will present the complete performance analysis of the imbalanced dataset for the classification of pulmonary tuberculosis. As mentioned earlier, all experiments were carried out using a combined dataset. First, we experimented using a 21-layer-based modified CNN model. Second, we have evaluated the results of the seven pre-trained state-of-the-art models using the concept of fine-tuning.

The testing performance of all models is determined using the same number of the testing set. A summary of prediction performance analysis using the test set is shown in **Error! Not a valid bookmark self-reference.**. Out of 225 TB X-ray images, miss-classified as Normal images counted as 17 images, while 16 out of 134 were miss-classified as Normal X-ray images showing TB-infected images by the proposed CNN architecture.

TABLE 5. SUMMARY OF PREDICTION RESULTS USING THE TEST SET.

Models	TP	FN	TN	FP
VGG16	203	22	107	27
VGG19	208	17	97	37
InceptionV3	190	35	106	28
ResNet50	205	20	115	19
Xception	200	25	98	36
MobileNetV2	195	30	111	23
EfficientNetB7	201	24	109	25
Proposed	208	17	118	16

In the case of the VGG16 model, 203 images were correctly predicted as TB-infected, 22 images were falsely rejected as normal images, and 27 images were falsely classified as TB-infected. The VGG19 architecture misclassified 17 TB images as Normal images, and 37 out of 134 Normal X-ray images were miss-classified as TB-infected images, which is a significantly high rate of wrongly positive prediction compared to others.

In the case of ResNet50, 20 images were falsely rejected as normal images, and 19 images were falsely classified as TB-infected. The predictions of InceptionV3 show that 35 TB images were misclassified as Normal images, which is a significantly higher rate of wrongly negative prediction than others. In the case of the Xception model, 25 TB images were miss-classified as Normal images, and 36 Normal images were predicted as TB images. In the case of MobileNetV2, 30 TB x-ray images and 23 Normal x-ray images were wrongly classified. The prediction results of efficientNetB7 show that 201 images were correctly predicted as TB-infected, and 24 images were falsely rejected as normal images. In contrast, 25 images were falsely classified as TB-infected. By comparing the prediction results, the proposed architecture achieved the best performance with a high rate of true-positive and true-negative and a low rate of false-positive and false-negative.

A. Comparative Analysis

In **Error! Reference source not found.**, we also compare our proposed model with seven pre-trained models using different evaluation metrics. The classification accuracy of the proposed models, VGG16, VGG19, ResNet50, Inceptionv3, Xception, MobileNetv2, and EfficientNetB7, is 0.9081, 0.8635, 0.8496, 0.8914, 0.8245, 0.8301, 0.8524, and 0.8635, respectively.

As shown in **Error! Reference source not found.**, the training and testing performance of all the models are plotted in terms of evaluation metrics that depict the testing

accuracy, precision, recall, f1score, and AUC of the proposed model higher than those of other pre-trained models. Although the datasets used in this study are the same as those used in the study of Hooda et al. [19], their ensemble method achieved 0.93 AUC on the combined dataset. However, our model performs better than their method in terms of AUC, which is 0.94.

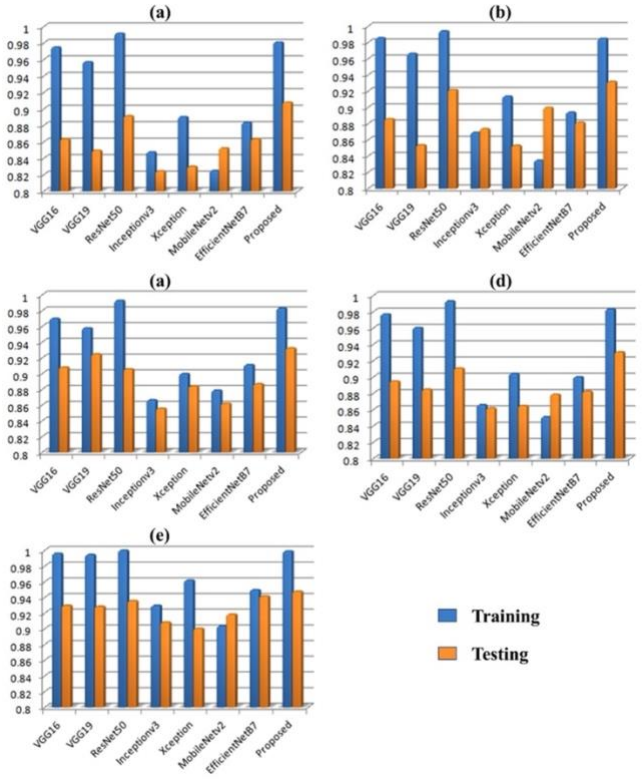

Figure 2. The final performance of training and testing in terms of evaluation metrics is as follows: (a) accuracy, (b) precision, (c) recall, (d) F1 score, and (e) AUC.

TABLE 6: COMPARISON OF RESULTS

Models	Total epoch Time in (seconds)	Evaluation Time in (seconds)
VGG16	473.7	82.8
VGG19	398.3	26.6
InceptionV3	233.6	24.5
ResNet50	357.2	24.2
Xception	365.0	25.4
MobileNetV2	**173.0**	24.9
EfficientNetB7	795.4	31.1
Proposed	256.2	**22.8**

In addition, the integration of real-time health monitoring systems using the Internet of Things (IoT) [41] and the fusion of artificial intelligence and human computer interaction [42] in healthcare will enhance patient care and outcomes.

TABLE shows the computational findings, which demonstrate that MobileNetV2 takes less computational time as a training network and that the proposed model consumes less computational time for testing images. The proposed model, as shown in Table 7, is comparatively computationally efficient compared to other pre-trained models such as VGG16, VGG19, RestNet50, and EfficientNetB7.

V. CONCLUSION

This paper presents an automated pulmonary tuberculosis screening based on deep convolutional neural network architecture. The proposed architecture supports 21 layers. Moreover, the results measure of seven different pre-trained models were also evaluated by adopting the concept of fine-tuning. In this study, a combined dataset is formed by using four standard publicly accessible datasets, including Shenzhen, Montgomery, JSRT, and Belarus, to solve the small dataset problem. The classification accuracy, precision, recall, F1 Score, and AUC for the proposed architecture were found to be 0.9081, 0.9317, 0.9323, 0.9307, and 0.9474, respectively. Results demonstrate that our proposed architecture performed better than pre-trained state-of-the-art in all selected metrics. The proposed architecture has proved to be more efficient by burning less computational power. In future work, we will consider the preparation of a new combined dataset by using TB images with other pulmonary disease images. Moreover, we will focus on the modification of hyperparameters used to build deep learning-based models. In addition, the integration of real-time health monitoring systems using the Internet of Things (IoT) [41] and the fusion of artificial intelligence and human computer interaction [42] in healthcare will enhance patient care and outcomes.

TABLE 7. PROCESSING TIME OF PROPOSED MODEL AND PRE-TRAINED MODELS.

Models	Subsets	Accuracy	Precision	Recall	F1 Score	AUC
VGG16	Training	0.9750	0.9856	0.9698	0.9770	0.9959
	Testing	0.8635	0.8859	0.9080	0.8949	0.9295
VGG19	Training	0.9571	0.9664	0.9578	0.9606	0.9943
	Testing	0.8496	0.8537	0.9246	0.8849	0.9283
InceptionV3	Training	0.8476	0.8691	0.8664	0.8659	0.9295
	Testing	0.8245	0.8736	0.8556	0.8619	0.9081
ResNet50	Training	0.9917	0.9941	0.9926	0.9931	0.9998
	Testing	0.8914	0.9217	0.9058	0.9108	0.9355
Xception	Training	0.8905	0.9137	0.8997	0.9042	0.9616
	Testing	0.8301	0.8531	0.8839	0.8648	0.9002
MobileNetV2	Training	0.8250	0.8347	0.8786	0.8513	0.9035
	Testing	0.8524	0.8998	0.8619	0.8788	0.9182
EfficientNetB7	Training	0.8833	0.8940	0.9112	0.9001	0.9495
	Testing	0.8635	0.8815	0.8868	0.8823	0.9415
Proposed	Training	0.9810	0.9849	0.9829	0.9832	0.9988
	Testing	0.9081	0.9317	0.9323	0.9307	0.9474

REFERENCES

[1] Organization W.H. 2020. "*Global tuberculosis report 2020: executive summary*."

[2] Konstantinos A. 2010. Testing for tuberculosis. *Aust. Prescr.* **33**: 12–18.

[3] Degnan A.J., E.H. Ghobadi, P. Hardy, *et al.* 2019. Perceptual and Interpretive Error in Diagnostic Radiology—Causes and Potential Solutions. *Acad. Radiol.* **26**: 833–845.

[4] Sangani S.G., S. B S, S. V Grover, *et al.* 2014. Spectrum of Findings in Pulmonary Infections on Ct / Hrct Scan of the Chest. *J. Evol. Med.*

Dent. Sci. **3**: 10927–10985.

[5] Wang X., Y. Peng, L. Lu, *et al.* 2017. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017* **2017-Janua**: 3462–3471.

[6] Krizhevsky A., I. Sutskever & G.E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**: 84–90.

[7] Li Q., W. Cai, X. Wang, *et al.* 2014. Medical image classification with convolutional neural network. *2014 13th Int. Conf. Control Autom. Robot. Vision, ICARCV 2014* **2014**: 844–848.

[8] Shariaty F. & M. Mousavi. 2019. Application of CAD systems for the automatic detection of lung nodules. *Informatics Med. Unlocked* **15**: 100173.

[9] Çallı E., E. Sogancioglu, B. van Ginneken, *et al.* 2021. Deep learning for chest X-ray analysis: A survey. *Med. Image Anal.* **72**: 102125.

[10] Han X., Z. Zhang, N. Ding, *et al.* 2021. Pre-trained models: Past, present and future. *AI Open* **2**: 225–250.

[11] Torrey L., J.S.-H. of research on machine learning & undefined 2010. Transfer learning. *igi-global.com*.

[12] Simonyan K. & A. Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.* 1–14.

[13] He K., X. Zhang, S. Ren, *et al.* 2016. Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2016-Decem**: 770–778.

[14] Szegedy C., V. Vanhoucke, S. Ioffe, *et al.* 2016. Rethinking the Inception Architecture for Computer Vision. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2016-Decem**: 2818–2826.

[15] Chollet F. 2017. Xception: Deep learning with depthwise separable convolutions. *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017* **2017-Janua**: 1800–1807.

[16] Sandler M., A. Howard, M. Zhu, *et al.* 2018. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 4510–4520.

[17] Tan M. & Q. V. Le. 2019. EfficientNet: Rethinking model scaling for convolutional neural networks. *36th Int. Conf. Mach. Learn. ICML 2019* **2019-June**: 10691–10700.

[18] Lakhani P. & B. Sundaram. 2017. Deep learning at chest radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology* **000**: 61–98.

[19] Hooda R., A. Mittal & S. Sofat. 2019. Automated TB classification using ensemble of deep architectures. *Multimed. Tools Appl.* **78**: 31515–31532.

[20] Oloko-Oba M. & S. Viriri. 2020. Diagnosing tuberculosis using deep convolutional neural network. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **12119 LNCS**: 151–161.

[21] Verma D., C. Bose, N. Tufchi, *et al.* 2020. An efficient framework for identification of Tuberculosis and Pneumonia in chest X-ray images using Neural Network. *Procedia Comput. Sci.* **171**: 217–224.

[22] Hwang E.J., S. Park, K.N. Jin, *et al.* 2019. Development and Validation of a Deep Learning-based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiographs. *Clin. Infect. Dis.* **69**: 739–747.

[23] Hwang S., H.-E. Kim, J. Jeong, *et al.* 2016. A novel approach for tuberculosis screening based on deep convolutional neural networks. *Med. Imaging 2016 Comput. Diagnosis* **9785**: 97852W.

[24] Rahman T., A. Khandakar, M.A. Kadir, *et al.* 2020. Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. *IEEE Access* **8**: 191586–191601.

[25] Munadi K., K. Muchtar, N. Maulina, *et al.* 2020. Image Enhancement for Tuberculosis Detection Using Deep Learning. *IEEE Access* **8**: 217897–217907.

[26] Hooda R., S. Sofat, S. Kaur, *et al.* 2017. Deep-learning: A potential method for tuberculosis detection using chest radiography. *Proc. 2017 IEEE Int. Conf. Signal Image Process. Appl. ICSIPA 2017* 497–502.

[27] Ahsan M., R. Gomes & A. Denton. 2019. Application of a convolutional neural network using transfer learning for tuberculosis detection. *IEEE Int. Conf. Electro Inf. Technol.* **2019-May**: 427–433.

[28] Hooda R., A. Mittal & S. Sofat. 2020. A Novel Ensemble Method for PTB Classification in CXRs. *Wirel. Pers. Commun.* **112**: 809–826.

[29] Liu J., Y. Liu, C. Wang, *et al.* 2018. "*An original neural network for pulmonary tuberculosis diagnosis in radiographs*." Springer International Publishing.

[30] Abbas A. & M.M. Abdelsamea. 2019. Learning Transformations for Automated Classification of Manifestation of Tuberculosis using Convolutional Neural Network. *Proc. - 2018 13th Int. Conf. Comput. Eng. Syst. ICCES 2018* 122–126.

[31] Heo S.J., Y. Kim, S. Yun, *et al.* 2019. Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers' health examination data. *Int. J. Environ. Res. Public Health* **16**:.

[32] Imam O.T., M. Haque, C. Shahnaz, *et al.* 2020. Detection of Tuberculosis from Chest X-Ray Images Based on Modified Inception Deep Neural Network Model. *Proc. 2020 IEEE Int. Women Eng. Conf. Electr. Comput. Eng. WIECON-ECE 2020* 360–363.

[33] Al-Timemy A.H., R.N. Khushaba, Z.M. Mosa, *et al.* 2021. An Efficient Mixture of Deep and Machine Learning Models for COVID-19 and Tuberculosis Detection Using X-Ray Images in Resource Limited Settings. *Stud. Syst. Decis. Control* **358**: 77–100.

[34] Luján-García J.E., Y. Villuendas-Rey, I. López-Yáñez, *et al.* 2021. Nanochest-net: A simple convolutional network for radiological studies classification. *Diagnostics* **11**:.

[35] Jaeger S., S. Candemir, S. Antani, *et al.* 2014. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant. Imaging Med. Surg.* **4**: 475–7.

[36] Shiraishi J., S. Katsuragawa, J. Ikezoe, *et al.* 2000. Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *Am. J. Roentgenol.* **174**: 71–74.

[37] Rosenthal A., A. Gabrielian, E. Engle, *et al.* 2017. The TB portals: An open-access, web-based platform for global drug-resistant-tuberculosis data sharing and analysis. *J. Clin. Microbiol.* **55**: 3267–3282.

[38] Hussain M., J.J. Bird & D.R. Faria. 2019. A study on CNN transfer learning for image classification. *Adv. Intell. Syst. Comput.* **840**: 191–202.

[39] Weng C.G. & J. Poon. 2008. A new evaluation measure for imbalanced datasets. *Conf. Res. Pract. Inf. Technol. Ser.* **87**: 27–32.

[40] Huang J. & C.X. Ling. 2005. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.* **17**: 299–310.

[41] Mahar, A.A., Kumar, D. and Nisar, K., 2021. Real-Time Health Monitoring System using IoT for Comatose Patients. *University of Sindh Journal of Information and Communication Technology* , **5(3)**, 145-151.

[42] Muhammad, A.H. and Faisal, A. 2022. Integration of Artificial Intelligence and Human Computer Interaction in Healthcare. *University of Sindh Journal of Information and Communication Technology* , **6(3)**, 101-107.