Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization

TAWSIFUR RAHMAN[1], **AMITH KHANDAKAR**[2], (Senior Member, IEEE),
MUHAMMAD ABDUL KADIR[1], **KHANDAKER REJAUL ISLAM**[3],
KHANDAKER F. ISLAM[2], **RASHID MAZHAR**[4,5], **TAHIR HAMID**[5,6],
MOHAMMAD TARIQUL ISLAM[7], (Senior Member, IEEE), **SAAD KASHEM**[8],
ZAID BIN MAHBUB[9], **MOHAMED ARSELENE AYARI**[10], AND
MUHAMMAD E. H. CHOWDHURY[2], (Senior Member, IEEE)

[1]Department of Biomedical Physics and Technology, University of Dhaka, Dhaka 1000, Bangladesh
[2]Department of Electrical Engineering, Qatar University, Doha 2713, Qatar
[3]Department of Orthodontics, Bangabandhu Sheikh Mujib Medical University, Dhaka 1000, Bangladesh
[4]Thoracic Surgery, Hamad General Hospital, Doha 3050, Qatar
[5]Department of Medicine, Weill Cornell Medicine-Qatar, Doha 24811, Qatar
[6]Cardiology, Hamad General Hospital, Doha 3050, Qatar
[7]Department of Electrical, Electronic, and Systems Engineering, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia
[8]Faculty of Robotics and Advanced Computing, Qatar Armed Forces-Academic Bridge Program, Qatar Foundation, Doha 24404, Qatar
[9]Department of Mathematics and Physics, North South University, Dhaka 1229, Bangladesh
[10]College of Engineering, Qatar University, Doha 2713, Qatar

Corresponding author: Mohamed Arselene Ayari (arslana@qu.edu.qa)

ABSTRACT Tuberculosis (TB) is a chronic lung disease that occurs due to bacterial infection and is one of the top 10 leading causes of death. Accurate and early detection of TB is very important, otherwise, it could be life-threatening. In this work, we have detected TB reliably from the chest X-ray images using image pre-processing, data augmentation, image segmentation, and deep-learning classification techniques. Several public databases were used to create a database of 3500 TB infected and 3500 normal chest X-ray images for this study. Nine different deep CNNs (ResNet18, ResNet50, ResNet101, ChexNet, InceptionV3, Vgg19, DenseNet201, SqueezeNet, and MobileNet) were used for transfer learning from their pre-trained initial weights and were trained, validated and tested for classifying TB and non-TB normal cases. Three different experiments were carried out in this work: segmentation of X-ray images using two different U-net models, classification using X-ray images and that using segmented lung images. The accuracy, precision, sensitivity, F1-score and specificity of best performing model, ChexNet in the detection of tuberculosis using X-ray images were 96.47%, 96.62%, 96.47%, 96.47%, and 96.51% respectively. However, classification using segmented lung images outperformed that with whole X-ray images; the accuracy, precision, sensitivity, F1-score and specificity of DenseNet201 were 98.6%, 98.57%, 98.56%, 98.56%, and 98.54% respectively for the segmented lung images. The paper also used a visualization technique to confirm that CNN learns dominantly from the segmented lung regions that resulted in higher detection accuracy. The proposed method with state-of-the-art performance can be useful in the computer-aided faster diagnosis of tuberculosis.

INDEX TERMS Tuberculosis detection, TB screening, deep learning, transfer learning, lungs segmentation, image processing.

I. INTRODUCTION

Tuberculosis (TB) is a communicable disease caused by a bacterium called Mycobacterium tuberculosis. It is the

leading cause of death from a single infectious disease [1]. Fortunately, this bacterial infectious disease can be well treated by antimicrobial drugs. Early diagnosis of tuberculosis and consequent administration of proper medication can cure this deadly disease [2]. Chest X-rays (CXR) are commonly used for detection and screening of pulmonary

The associate editor coordinating the review of this manuscript and approving it for publication was Shadi Alawneh.

tuberculosis [3], [4]. In clinical practice, chest radiographs are examined by experienced physicians for the detection of TB. However, this is time consuming and a subjective process. Subjective inconsistencies in disease diagnosis from radiograph is inevitable [5], [6]. Importantly, CXR images of tuberculosis are often misclassified to other diseases of similar radiologic patterns [7], [8], which may lead to wrong medication to the patients and thereby worsening the health condition. There is also a lack of trained radiologists in the low resource countries (LRC), especially in the rural areas. In this perspective, computer aided diagnosis (CAD) systems can play important role in the mass screening of pulmonary TB by analysing the chest X-ray images. The accessibility of large-scale labelled datasets and deep convolutional neural networks (CNNs) has led to huge success in image recognition. CNNs allow data-driven, highly representative, hierarchical image features to be learned from adequate training data, but obtaining data sets in the medical imaging domain as comprehensively annotated, as ImageNet remains a challenge [9]–[11]. It is also worth noting that the healthcare sector is entirely different from any other field being a high priority sector with customers willing to pay for highest quality of care and services. The healthcare sector has not fulfilled the aspirations of society, while the industry absorbs a large percentage of national budgets [12]. The medical experts examining medical data which substantially suffer from subjective differences and the quality of the images and by the fatigue caused due to heavy workload. Thus, application of machine learning in health care sector is gathering a lot of attention in recent times. Artificial intelligence (AI) based solutions have been proposed for many biomedical applications including brain tumor, lungs nodule, pneumonia, breast cancer detection, physiological monitoring and social sensing [13]–[19]. Among the deep machine learning (ML) techniques, convolutional neural networks (CNNs) have shown great promise in image classification and therefore widely adopted by the research community [20]–[22]. X-ray radiography is a low-cost imaging technique and there is an abundance of data available for training different machine learning models making deep learning techniques popular for the automatic detection of lung diseases from chest radiographs. Moreover, we have a team of medical doctors who are experts in interpreting chest X-ray images put us in a favourable condition for computer-aided diagnostic system development.

CNNs have been used in several recent studies for the detection of lungs diseases including pneumonia and tuberculosis by analysing chest X-ray images. In response to the COVID-19 pandemic situation in 2020, CNN based techniques have been used for the detection of the novel coronavirus infection form CXR images. Tahir *et al.* [23] classified different coronavirus families (SARS, MERS and COVID-19) using transfer learning of various pre-trained CNN models with sensitivity values greater than 90%. Chowdhury *et al.* [24] developed a trained model using chest X-ray dataset to distinguish COVID-19 pneumonia, viral

pneumonia and normal cases. Chhikara *et al.* [21] explored the possibility of detecting pneumonia from CXR images and evaluated the performance of some pre-trained models (Resnet, ImageNet, Xception, and Inception) applying preprocessing techniques like filtering and gamma correction. Abbas *et al.* [25] reported a modified transfer learning CNN termed as Decompose, Tranfer and Compose (DeTraC) that can deal with data imbalance in medical image classification. This CNN architecture was shown to have improved performance in detecting normal and abnormal x-rays with an accuracy of 99.8 %.

Several research groups used classical machine learning techniques for classifying TB and non-TB cases from CXR images [26]–[31]. The use of deep machine learning algorithms have been reported in the detection of tuberculosis by varying the parameters of deep-layered CNNs [32]–[37]. Concept of transfer learning in deep learning framework was used for the detection of tuberculosis utilizing pre-trained models and their ensembles [38]–[42]. Hooda *et al.* [32] presented a deep learning approach to classify CXR images into TB and non-TB categories with an accuracy of 82.09%. Evalgelista and Guedes [34] reported a computer-aided approach based on intelligent pattern recognition using CNNs for TB detection from chest X-ray images with an accuracy of 88.76%. Pasa *et al.* [35] proposed a deep network architecture optimized for the screening of tuberculosis with an accuracy of 86.82%. They also reported a tool for interactive visualization of TB cases. Nguyen *et al.* [36] evaluated the performance of a pre-trained model, DenseNet, to classify normal and tuberculosis images from Shenzhen (CHN) and Montgomery County (MC) databases [43] using fine-tuned model, and reported the Area Under the Curve (AUC) values of 0.94 and 0.82 respectively. Hernández *et al.* [37] proposed a method for the automatic classification of TB from X-Ray images using an ensemble of CNN models with an accuracy of 86%. Lopes and Valiati [39] used different pre-trained CNN models to classify the chest radiographs into TB positive and TB negative classes. The performance of the system was evaluated on two publicly available chest X-ray datasets (CHN and MC) and achieved an accuracy of 80%. Meraj *et al.* [40] used four different CNN models (VGG-16, VGG-19, RestNet50, and GoogLeNet) and explored the limits of accuracies for small-scale and large-scale CNN models in the classification of TB from chest X-rays. Ahsan *et al.* [41] proposed a generalized pre-trained CNN model for TB detection and achieved accuracies of 81.25% and 80% with and without the application of image augmentation respectively. Yadav *et al.* [42] reported the detection of tuberculosis using transfer learning technique, which showed an accuracy of 94.89%. Abbas *et al.* [25] proposed a CNN architecture based on a class decomposition approach to improve the performance of ImageNet pre-trained CNN models using transfer learning and achieved high accuracy in TB detection on Japanese Society of Radiological Technology (JSRT) database. It is worth mentioning that transfer learning methods have also been used to classify the

images of TB culture test. Chang *et al.* [22] used transfer learning technique on labelled tuberculosis culture images and achieved precision and sensitivity values of 99% and 98% respectively. However, classification of TB culture image requires specific samples from the patients and is not as robust compared to classification from chest X-rays which are readily available.

In clinical applications, an increase in accuracy of TB detection from chest radiographs with a robust and versatile method can make computer aided automatic diagnostic tools more reliable. The classification accuracy can be improved either by using different deep learning algorithms or by modifying the existing outperforming algorithms or combining several outperforming algorithms as an ensemble model. Typically, whole X-ray images were used for the detection of lung disorders using CNN. However, the X-ray images contains lungs as well as other regions of the thorax although the disease like TB is manifested in the lung region only. Thus, focusing on the lung region of the X-ray images during training and classification may significantly improve the performance of TB detection. To the best of our knowledge, no such work regarding the use of deep learning networks on segmented lungs for TB detection is reported. This paper focuses on the detection of TB using transfer learning based technique of CNNs on the original and segmented lungs in X-ray images. CNN based visualization techniques are also implemented to confirm that the deep networks perform classification using the main region of interest, i.e., lungs minimizing learning from irrelevant regions in chest X-rays.

Several important contributions were reported in this study. Firstly, two different U-net models were investigated for the segmentation of the chest X-ray images. Secondly, nine different pre-trained CNNs were applied for the detection of TB from the original and segmented lungs of X-ray images and their performances were analysed. Then, the performance of TB detection by the pre-trained networks using non-segmented X-ray images and segmented images were compared. Finally, state-of-the-art score class activation mapping (Score-CAM) based visualization technique was implemented and used to demonstrate the regions of X-ray images that contribute in the classification by the CNNs to confirm whether the segmented lungs X-ray images based classification is more reliable than that of whole X-ray images or not.

The rest of the paper is divided in the following sections: Section 2 summarizes different pre-trained networks for image classification, U-net models for lung segmentation and Score-CAM based visualization techniques. Section 3 describes dataset, pre-processing steps and methodology of this study, while Section 4 summarizes the results of the classification using whole X-ray images and segmented lung images and results are discussed and compared with some other recent studies. Finally, Section 5 concluded the study.

II. BACKGROUND
A. DEEP CONVOLUTIONAL NEURAL NETWORKS (CNNS) BASED TRANSFER LEARNING

As discussed earlier, deep CNNs have been popular due to their improved performance in image classification. The convolutional layers in the network along with filters help in extracting the spatial and temporal features in an image. Transfer learning can be useful in those applications of CNN where the dataset is not large. Recently, transfer learning has been successfully used in various field applications such as manufacturing, medical and baggage screening [44]–[46]. This removes the requirement of having large dataset and also reduces the long training period as is required by the deep learning algorithm when developed from scratch [47], [48].

Nine popular pre-trained deep learning CNNs such as ResNet18, ResNet50, ResNet101 [49], DenseNet201 [50], ChexNet [51], SqueezeNet [52], InceptionV3 [53], VGG19 [54] and MobileNetV2 [55] were used for TB detection. All of these networks apart from ChexNet were initially trained on ImageNet database. Residual Network (in short ResNet) was originally developed to solve vanishing gradient and degradation problem [49]. ResNet has several different variants: ResNet18, ResNet50, ResNet101 and ResNet152 based on the number of layers in the residual network. ResNet was successfully used in biomedical image classification for transfer learning. Typically, deep neural network layers learn low or high level features during training while ResNet learns residuals instead of features [56].

Dense Convolutional Network (in brief DenseNet) needs less number of parameters than a conventional CNN, as it does not learn redundant feature maps. The layers in DenseNet are very narrow, which add a lesser set of new feature-maps. DenseNet has four different common variants: DenseNet121, DenseNet169, DenseNet201 and DenseNet264. Each layer in DenseNet has direct access to the original input image and gradients from the loss function. Therefore, the computational cost significantly reduced, which makes DenseNet a better choice for image classification. ChexNet Pretrained model is a modified version of DenseNet121 and this network is specially trained on large number of chest X-ray images [51].

SqueezeNet and MobileNetv2 are very compact network compared to the other networks. The foundation of SqueezeNet network is a fire module, which consists of Squeeze Layer and Expand layer. The Squeeze layer has only 1×1 filters, which is feeding to an Expand layer, which has a mixture of 1×1 and 3×3 convolution filters [52]. VGG [54] addresses a very important aspect of CNNs, which is depth. The convolutional layers in VGG network use a very small receptive field. There are 1×1 convolution filters which act as a linear transformation of the input, which is followed by a rectified linear unit (ReLU) layer. The convolution stride is fixed to 1 pixel so that the spatial resolution is preserved after convolution. VGG has different variants: VGG16, and VGG19.

MobileNet structure is built on depth-wise separable convolutions except for the first layer which is a full convolution. All layers are followed by a batch normalization and ReLU nonlinearity with the exception of the final fully connected layer which has no nonlinearity and feeds into a Softmax layer for classification. A final average pooling reduces the spatial resolution to 1 before the fully connected layer. Counting depth-wise and pointwise convolutions as separate layers, MobileNet has 28 layers. Inception modules are used in Convolutional Neural Networks to allow for more efficient computation and deeper networks through a dimensionality reduction with stacked 1×1 convolutions. The modules were designed to solve the problem of computational expense, as well as over-fitting, among other issues.

B. SEGMENTATION

There are several variants of segmentation models based on U-nets are available in the literature. Two different variants of called original U-Net [57] and Modified U-Net [58] were investigated in this work to utilize best performing one. Figure 1 shows the architecture of original U-Net and Modified U-Net. The original U-net consists of a contracting path and an expanding path. The contracting path consists of the repeated application of two 3×3 convolutions (unpadded convolutions), each followed by a ReLU and a 2×2 max pooling operation with stride 2 for down sampling. The expanding path consists of an up sampling of the feature map followed by a 2×2 convolution (''up-convolution'') that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3×3 convolutions, each followed by a ReLU. Total 23 convolutional layers are used in the network. Modified U-Net also consists of a contracting path and an expanding path as U-Net. The contracting path includes four steps. Each step consists of two convolutional 3×3 filters followed by a 2×2 max pooling function and ReLU. The U-Net might learn redundant features in successive convolution layers. However, modified U-Net uses densely connective convolutions to mitigate this problem. Each step in the expanding path starts by performing an up-sampling function over the output of the previous layer. In the modified U-Net, the corresponding feature maps in the contracting path are cropped and copied to the expanding path. These feature maps are then concatenated with the output of the up-sampling function. Instead of a simple concatenation in the skip connection of U-Net, Modified Unet employs bidirectional Convolutional Long Short Term Memory (BConvLSTM) to combine the feature maps extracted from the corresponding contracting path and the previous expanding up-convolutional layer in a non-linear way.

C. VISUALIZATION TECHNIQUES

Increasing interest on the internal mechanisms of the CNNs and the reasoning behind a network making specific decisions have led to the developments of visualization techniques. The visualization techniques help in better visual representation

for interpreting the decision-making process of CNNs. These also increase the model's transparency by visualizing the logic behind the inference that can be interpreted in a way easily understandable to human, thereby increasing the confidence on the outcomes of the neural networks. Amongst the various visualization techniques like SmoothGrad [59], Grad-CAM [60], Grad-CAM++ [61], and Score-CAM [62]. Recently proposed Score-CAM was used in this work due to its promising performance in comparison to other techniques. Score-CAM gets rid of the dependence on gradients by obtaining the weight of each activation map through its forward passing score on target class, the final result is obtained by a linear combination of weights and activation maps. A sample image visualization with Score-CAM is shown in Figure 2, where the heat map indicates that the lungs regions dominantly contributed in the decision making in CNN. This can be helpful to understand how the network is taking its decision and also to improve the confidence of the end-user when it can be confirmed that all the time the network is taking decisions using the lungs of the chest X-rays.

In order to further confirm the deep layers can produce discriminating features between the TB and Normal patients, we have used t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization technique. It is a variant of Stochastic Neighbor Embedding and performs better in visualizing high dimensional data into two dimensional map [63]. This technique helps in a graphical analysis of deep learned features and have been very popularly used in many recent works [64]–[67].

III. METHODOLOGY

The overall methodology of this study is illustrated using Figure 3. Two different databases were created for this study. One was for lung segmentation while the other one was for TB classification. Three major experiments were carried out in this study. Firstly, two different U-Net models were investigated to identify the suitable network for segmenting lung regions of the X-ray images. Secondly, original chest X-ray images were used for TB classification using nine different pre-trained networks and evaluate classification reliability using Score-CAM technique. Thirdly, segmented lungs of X-ray images were used for TB classification using same networks and evaluated their performance using Score-CAM technique. Finally, the t-SNE technique was implemented on a python platform and dimensions, maximum iterations, perplexity- effective number of neighbors etc. parameters were modified from default values in order to confirm the performance of the best network [68].

A. DATASETS DESCRIPTION

1) LUNG SEGMENTATION

In this work, Kaggle Chest X-ray images and corresponding lung mask dataset [69] were used for training the lung segmentation models, where 704 X-ray images and their corresponding ground truth lung masks are available. All masks
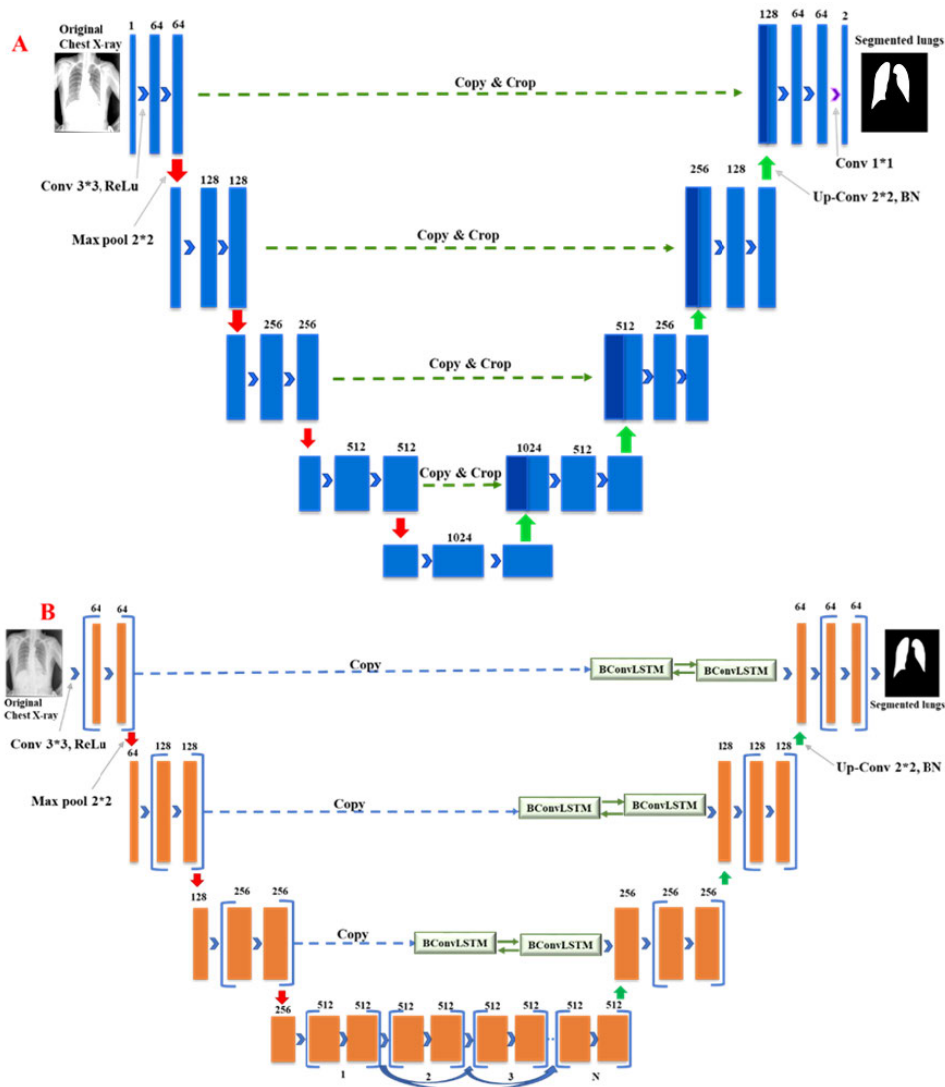
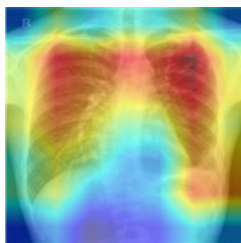**FIGURE 1.** Architecture of A) original U-Net and B) modified U-Net.

FIGURE 2. Score-CAM heat map on a chest X-ray image showing that different regions of the image were used in decision making by the CNN.

were annotated by expert radiologists; sample X-ray images and masks are shown in Figure 4. There are 360 normal X-ray images and 344 abnormal (infected lung) X-ray images available in the dataset. Therefore, U-Net networks were trained with both normal and abnormal images.

2) TB CLASSIFICATION

Four publicly accessible databases were used for TB classification problem. These are NLM dataset, Belarus dataset, NIAID TB dataset and RSNA dataset:

NLM dataset: National Library of Medicine (NLM) in U.S. [25] has made two lung X-ray datasets publicly available: the Montgomery and Shenzhen datasets. The Montgomery County (MC) and the Shenzhen, China (CHN) databases are comprised of 138 and 667 posterior-anterior (PA) chest X-ray images respectively. The resolution of the images of MC database was either 4, 020 × 4.892 or 4, 892 × 4.020 pixels whereas that for CHN database was variable but around 3000 × 3000 pixels. In the MC database, out of 138 chest X-ray images, 58 images were taken from different TB patients and 80 images were from normal subjects. In the CHN database, out of 662 chest X-ray images, 336 images were taken from different TB patients
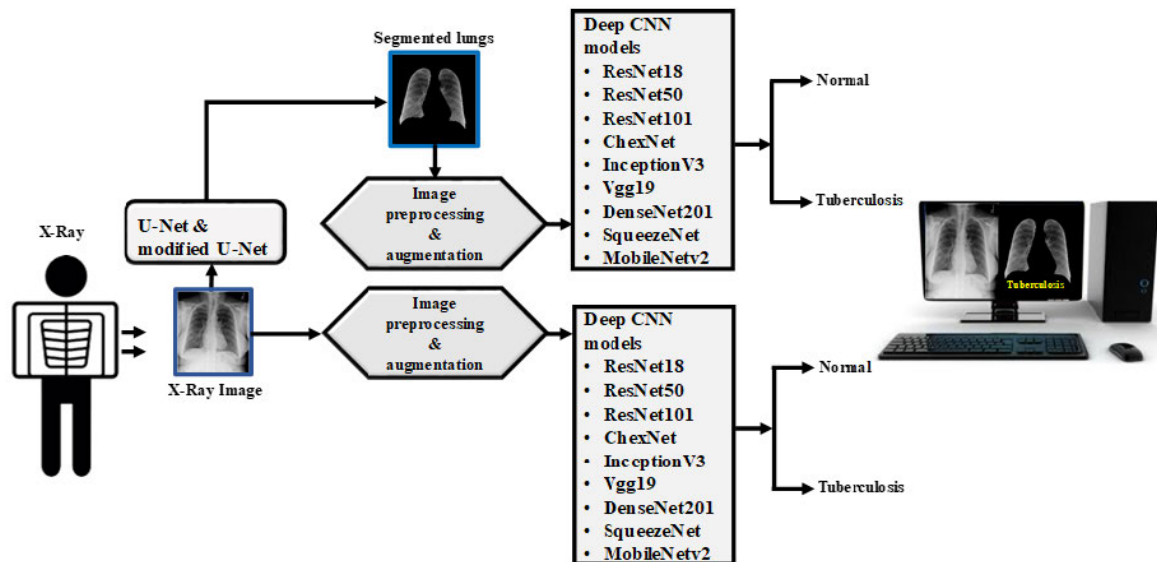
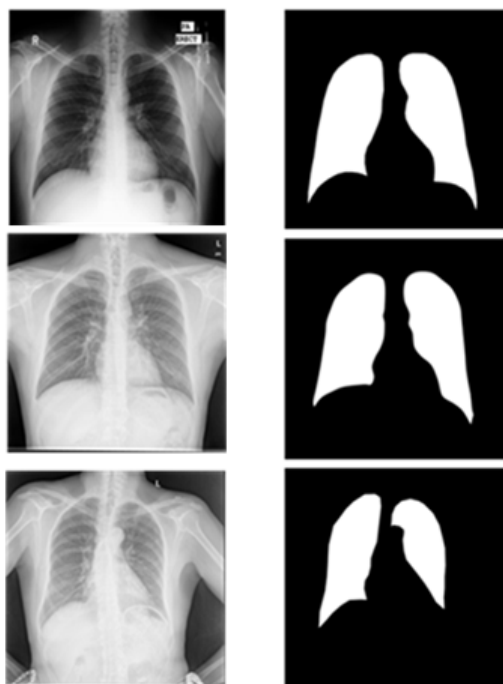**FIGURE 3.** Overview of the complete system.

FIGURE 4. Example of X-ray images and corresponding ground truth lung masks from Kaggle dataset.

and 324 images were from normal subjects. Therefore, in this NLM database, there are 406 normal and 394 TB infected X-ray images.

Belarus dataset: Belarus Set [70] was collected for a drug resistance study initiated by the National Institute of Allergy and Infectious Diseases, Ministry of Health, Republic of Belarus. The dataset contains 306 CXR images of 169 patients. Chest radiographs were taken using the Kodak Point-of-Care 260 system and the resolution of the images

was 2248 × 2248 pixels. All the images of this database are TB infected images.

NIAID TB dataset: NIAID TB portal program dataset [71], which contains about 3000 TB positive CXR images from about 3087 cases. All images were collected from seven different countries and all images are in Portable Network Graphics (PNG) format. In this study, we have used 2800 TB positive CXR images out of 3000 images. 200 poor quality images were discarded from this database.

RSNA CXR dataset: RSNA pneumonia detection challenge dataset [72], which is comprised of about 30,000 chest X-ray images, where 10,000 images are normal and others are abnormal and lung opacity images. All images are in Digital Imaging and Communications in Medicine (DICOM) format. To create a normal database of 3,500 chest X-ray images for this study, 3,094 normal images were taken from this database and rest of the 406 normal images were taken from the NLM database. However, the number of TB infected images by combining NLM and Belarus dataset was 700 and from NIAID TB dataset was 2800. In total, there were 3500 TB infected and 3500 normal X-ray images were used in this study. Samples of the X-ray images is shown in Figure 5.

B. PREPROCESSING

The size of the input images for different CNNs were different and therefore the datasets were preprocessed to resize the X-Ray images. In segmentation problem, for original U-Net and modified U-Net, the images were resized to 256 × 256 pixels. In classification problem, for InceptionV3 the images were resized to 227 × 227 pixels whereas for ResNet, DenseNet, ChexNet, VGG, MobileNetV2 and SqueezeNet, the images were resized to 224 × 224 pixels. All images were normalized using Z-score normalization using image database mean and standard deviation.

TABLE 1. Details of Training, Validation and Test set for U-Net segmentation models.

Dataset	No. of CXR images & masks	Train set/fold	Validation set/fold	Test set/fold
Kaggle lung x-ray & masks dataset	704	451	112	141

TABLE 2. Details of Training, Validation and Test set for classification problem.

Database	Types	Total No. of X-ray images/ class	Training without & with image augmentation		
			Training set/fold	Validation /fold	Test image/ fold
NLM, Belarus, NIAID and RSNA	Normal	3500	2240	560	700
	Tuberculosis	3500	2240	560	700
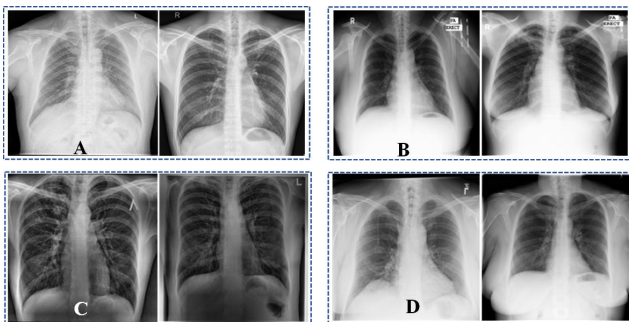
FIGURE 5. Example of CXR images from different datasets. (A) CHN dataset, (B) MC dataset, (C) Belarus dataset and (D) RSNA dataset.

C. EXPERIMENTS

1) LUNG SEGMENTATION

Original U-Net and modified U-Net were used separately on Kaggle CXR images and lung mask dataset for lung segmentation. Out of 704 CXR images and lung masks, 80% images and masks were used for training and 20% for testing as summarized in Table 1. Five-fold cross validation was used for training, validation and testing the entire dataset.

The networks were implemented using PyTorch library with Python 3.7 on Intel® Xeon® CPU E5-2697 v4 @ 2,30GHz and 64 GB RAM, with a 16 GB NVIDIA GeForce GTX 1080 GPU. Both the U-Net models were trained using Stochastic Gradient Descent (SGD) with momentum optimizer with learning rate, $\alpha = 10^{-3}$, dropout rate = 0.2, momentum update, $\beta = 0.9$, mini-batch size of 32 images with 50 back propagation epochs; early stopping criterion of 5 maximum epochs when no improvement in validation loss is seen.

2) TB CLASSIFICATION

Nine different CNN models were trained, validated and tested separately using non-segmented and segmented X-ray images for the classification of TB and non-TB normal images to investigate if data segmentation can improve classification

accuracy. The complete image set were divided into 80% training and 20% testing sub-sets for five-fold cross validation and 20% of training data were used for validation. For example, 80% (2800) of 3500 normal X-ray images were used for training and 20% (700) images were used for testing. However, 20% (560) of 2800 training images were used for validation and therefore, remaining 2240 images were used for training a fold. Table 2 shows the number of training, validation and test images used in the two experiments of non-segmented and segmented lungs images.

All nine CNNs were implemented using PyTorch library with Python 3.7 on Intel® Xeon® CPU E5-2697 v4 @ 2,30GHzand 64 GB RAM, with a 16 GB NVIDIA GeForce GTX 1080 GPU. Three comparatively shallow networks (MobileNetv2, SqueezeNet and ResNet18) and six deep networks (Inceptionv3, ResNet 50, ResNet101, CheXNet, VGG19 and DenseNet201) were evaluated in this study to investigate whether shallow or deep networks are suitable for this application. Three different variants of ResNet were used to compare specifically the impact of shallow and deep networks with similar structure. Performance difference due to initially trained on different image classes other than X-ray images were compared with CheXNet, which is a 121-layer DenseNet variant and the only network pre-trained on X-ray images. Nine pre-trained CNN models were trained using same training parameters and stopping criteria as mentioned earlier. However, only 15 back propagation epochs were used for classification problem. Five-fold cross-validation results were averaged to produce the final receiver operating characteristic (ROC) curve, confusion matrix, and evaluation matrices. Using image augmentation and having a validation image set, helps in avoiding overfitting of the models [73]. Table 3 shows comparative performance of both U-Net models in terms of the variable parameters such as Optimizer, Loss Function and Batch Size for image segmentation. Other segmentation parameters such as number of epochs and learning rate were programmed to change in case no improvement of performance was observed.

TABLE 3. Comparative performance of original U-Net and modified U-Net.

Batch Size	Loss Function	Optimizer	Network	Test loss	Test accuracy	IoU	Dice
4	Dice Loss	SGD	U-Net	0.039	98.03	92.4	96.0
8	Dice Loss	SGD	U-Net	0.277	97.46	90.5	94.94
16	Dice Loss	SGD	U-Net	0.0921	97.59	90.88	95.13
4	BCE Loss	SGD	U-Net	0.21	97.51	90.52	94.95
8	BCE Loss	SGD	U-Net	0.183	98.03	92.2	95.7
16	BCE Loss	SGD	U-Net	0.895	97.96	92.08	95.75
4	Dice Loss	Adam	U-Net	0.0385	98.11	92.72	96.14
8	**Dice Loss**	**Adam**	**U-Net**	**0.0385**	**98.14**	**92.82**	**96.19**
16	Dice Loss	Adam	U-Net	0.0407	98.12	92.8	96.19
4	BCE Loss	Adam	U-Net	0.051	98.11	92.65	96.09
8	BCE Loss	Adam	U-Net	0.3241	84.55	79.46	88.26
16	BCE Loss	Adam	U-Net	0.3352	84.15	78.9	87.86
4	Dice Loss	SGD	Modified U-Net	0.043	97.88	91.8	95.6
8	Dice Loss	SGD	Modified U-Net	0.0415	98.1	92.71	96.08
16	Dice Loss	SGD	Modified U-Net	0.042	98.02	92.7	96.01
4	BCE Loss	SGD	Modified U-Net	0.0789	97.87	91.88	95.55
8	BCE Loss	SGD	Modified U-Net	0.247	97.88	91.85	95.67
16	BCE Loss	SGD	Modified U-Net	0.0664	97.98	92,.25	95.88
4	Dice Loss	Adam	Modified U-Net	0.1099	84.75	80.02	88.47
8	Dice Loss	Adam	Modified U-Net	0.1116	84.14	79.68	88.21
16	Dice Loss	Adam	Modified U-Net	0.1191	84.09	78.75	87.75
4	BCE Loss	Adam	Modified U-Net	0.3185	84.79	79.73	88.4
8	BCE Loss	Adam	Modified U-Net	0.3322	84.78	79.4	88.22
16	BCE Loss	Adam	Modified U-Net	0.3176	84.88	79.9	88.48

TABLE 4. Details of training parameter for segmentation and classification.

Training parameter		
	Segmentation model	**Classification model**
batch size	8	32
learning rate	0.001	0.001
epochs	20	15
epochs patience	3	3
stopping criteria	5	5
optimizer	Adam	Stochastic gradient descent (SGD)

D. PERFORMANCE MATRIX

1) LUNG SEGMENTATION

The performance of different networks in image segmentation for the testing dataset was evaluated after the completion of training and validation phase and was compared using four performance metrics: loss, accuracy, IoU, Dice. The equations used to calculate accuracy, Intersection-Over-Union (IoU) or Jaccard Index and Dice coefficient (or F1-score) are

shown in equation (1-3).

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

$$IoU(Jaccard index) = \frac{(TP)}{(TP + FN + FP)} \quad (2)$$

$$Dice\ Coefficient(F1 - score) = \frac{(2 * TP)}{(2 * TP + FN + FP)} \quad (3)$$

2) TB CLASSIFICATION

The performance of different CNNs for testing dataset was evaluated after the completion of training and validation phase and was compared using six performance metrics: accuracy, sensitivity or recall, specificity, precision, area under curve (AUC), F1 score. The matrices were calculated using the following equations (4-8):

$$Accuracy = \frac{(TP + TN)}{(TP + FN) + (FP + TN)} \quad (4)$$

$$Sensitivity = \frac{(TP)}{(TP + FN)} \quad (5)$$

$$Specificity = \frac{(TN)}{(FP + TN)} \quad (6)$$

$$Precision = \frac{(TP)}{(TP + FP)} \quad (7)$$

$$F1Score = \frac{(2 * TP)}{(2 * TP + FN + FP)} \quad (8)$$

Here, true positive (TP), true negative (TN), false positive (FP) and false negative (FN) were used to denote number of tuberculosis images identified as tuberculosis, number of normal images identified as normal, number of normal images incorrectly identified as tuberculosis images and number of tuberculosis images incorrectly identified as normal, respectively.

In addition, the networks can be compared in terms of processing time for a test image δt_e)(and also the training time per epoch (δt_t). The processing time for a test image (δt_e) is the time taken by a network to classify an image (I) and the training time per epoch is the time taken by a network to train one epoch represented in equation 9 and 10 respectively.

$$\delta t_e = t_2 - t_1 \quad (9)$$

$$\delta t_t = t_4 - t_3 \quad (10)$$

where t_1 *and* t_2 are the start and end time for a network to classify an image I, respectively and t_3 *and* t_4 are the start and end time of training an epoch by a network. All the time is measured in seconds.

IV. RESULTS AND DISCUSSIONS

A. LUNG SEGMENTATION

Original U-Net and modified U-Net networks were trained, validated and evaluated on the test data for the segmentation of lungs of X-ray images. The best performing segmentation parameters are stated in Table 4 and can also be seen in the highlighted row of Table 3. Figure 6 shows some example test X-ray images, corresponding ground truth masks and segmented lung images generated by the two trained U-Net models for kaggle dataset. It can be noted that the original U-Net outperformed modified U-Net in the segmentation of lung regions on CXR images quantitatively and qualitatively.
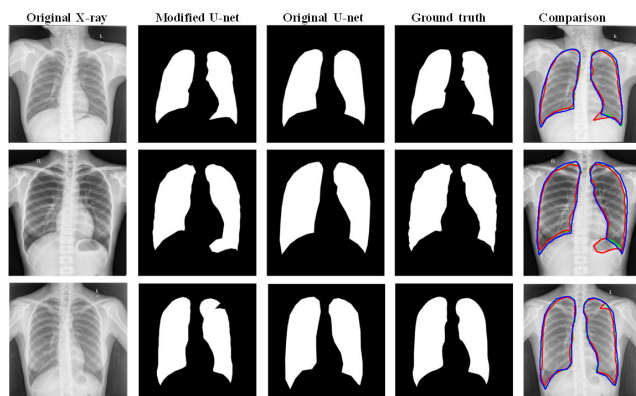

FIGURE 6. Sample test X-ray images, segmented lungs using two U-net models and ground truth were compared.

The better performing original U-Net model was then used to segment the classification database (3500 normal and
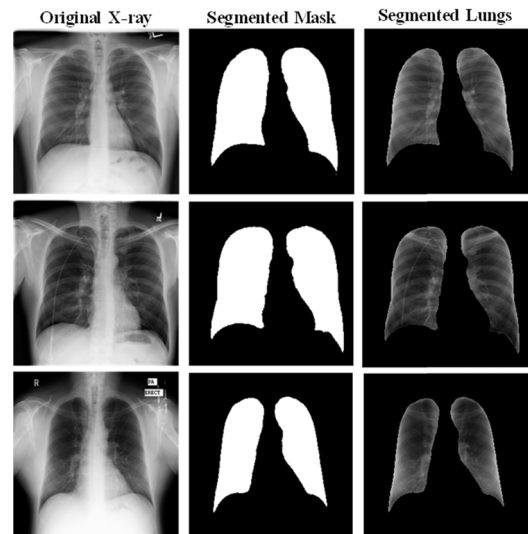

FIGURE 7. Samples X-ray images from classification database (left), generated masks by the trained original U-Net model (middle) and corresponding segmented lung (right).

3500 TB images), which was used for classification of TB and non-TB normal cases. It is important to see on a completely unseen image-set with TB infection and normal images how well the trained segmentation model works. It can be seen from Figure 7 that the original U-net model trained on Kaggle chest X-ray dataset can segment the lung areas of the X-ray images of the classification database very reliably. However, the quantitative evaluation on the classification dataset is not possible as there is no ground truth masks available for this database and therefore, qualitative evaluation was done to confirm that each X-ray image was segmented correctly.

B. TB CLASSIFICATION

As mentioned earlier, there are two different experiments (using non-segmented and segmented lungs X-ray images) were conducted for the classification of TB and normal (non-TB) cases. The comparative performance for different CNNs for the binary classification is shown in Table 5. It is apparent from Table 5 that all the evaluated pre-trained models perform very well in classifying TB and normal images in this two-class problem. Among the networks trained with X-ray images without segmentation, CheXNet is performing better for classifying the X-ray images. Even though CheXNet is shallower than DenseNet201, it was originally trained on X-ray images provides it additional benefits in classifying X-ray images and it is showing better performance than DenseNet201. It is not necessary that deeper network will perform better rather CheXNet is a very good example of transfer learning and it outperforms other networks for this problem. Similar performance was observed by the authors in the COVID-19 classification problem [24]. CheXNet is also fast in classify an image, evident from the processing time for testing an image (δt_e) being the smallest amongst all the other competing networks. The training time taken per epoch (δt_t) is also amongst the average compared with

TABLE 5. Comparative performance of different models for TB classification using with and without segmented data.

| Scheme			Weighted Average						
	Model	Overall Accuracy	Precision	Sensitivity	F1-score	Specificity	Processing time per test image (δt_e)	Training time per epoch (δt_t)
Without Segmentation	ResNet18	93.85	94.08	93.85	93.84	93.91	0.2	39.24
	ResNet50	93.11	93.4	93.11	93.09	93.16	0.35	42.7
	ResNet101	94.55	94.74	94.55	94.54	94.59	0.68	49.6
	ChexNet	**96.47**	**96.62**	**96.47**	**96.47**	**96.51**	**0.4**	**43.85**
	InceptionV3	95.72	95.92	95.73	95.73	95.92	1.02	75.59
	Vgg19	95.8	95.95	95.8	95.8	95.85	1.51	62.2
	DenseNet201	95.07	95.27	95.07	95.07	95.12	0.58	93.6
	SqueezeNet	94.18	94.31	94.18	94.17	94.21	0.19	38
	MobileNet	94.33	94.65	94.33	94.32	94.39	0.13	39.42
With Segmentation	ResNet18	96.84	97.14	96.48	96.96	97.42	0.24	32.43
	ResNet50	97.07	97.34	97.07	97.14	97.36	0.4	33.45
	ResNet101	97.96	98	97.96	97.96	97.93	0.75	43.45
	ChexNet	98.14	98.15	98.14	98.14	98.13	0.78	38.29
	InceptionV3	98.54	98.57	98.54	98.54	98.52	1.15	70.38
	Vgg19	97.91	97.95	97.92	97.92	97.89	1.7	59.5
	DenseNet201	**98.6**	**98.57**	**98.56**	**98.56**	**98.54**	**0.8**	**88.67**
	SqueezeNet	96.58	97.08	96.75	96.66	97.32	0.22	34.2
	MobileNet	96.9	97.29	96.68	96.86	97.62	0.15	33.23
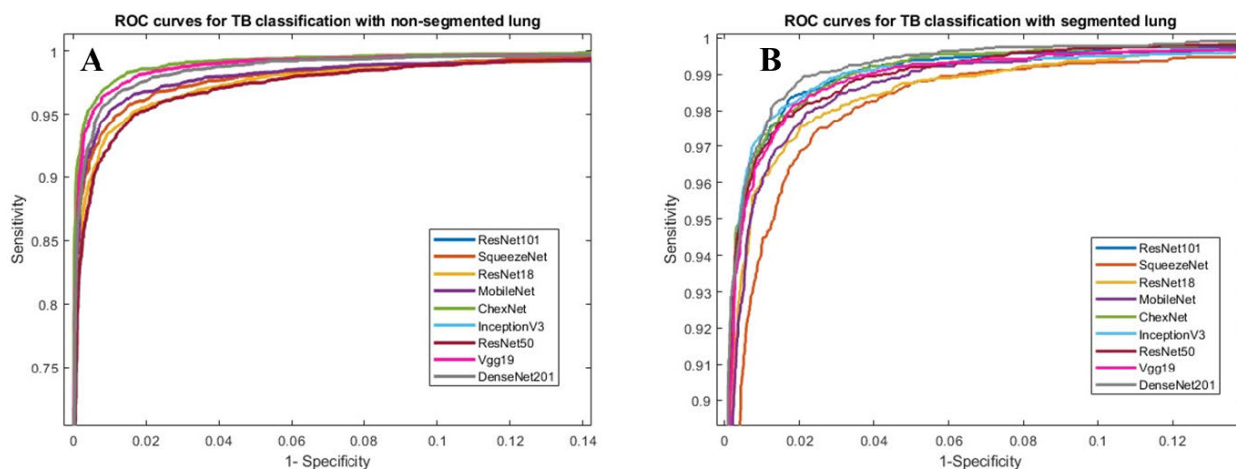
FIGURE 8. Comparison of the ROC curves for Normal, and Tuberculosis classification using CNN based models for non-segmented (A) and segmented (B) CXR images.

other networks which can also be attributed to the number of layers in this network being lesser than deeper networks such as InceptionV3, Vgg19, DenseNet201 and being deeper than ResNet18, and ResNet50. Even though CheXNet has more layers than ResNet101, however, faster than ResNet101, which could be linked to the number of network parameters.

Figure 8(A) clearly shows that the ROC curves for CheXNet is the best when considering non-segmented lungs.

On the other hand, segmented lung chest X-ray images provided a clear performance boost for all of the tested networks. This reflect the fact that shallow or deep, all CNNs can distinguish TB and normal lungs with very high reliability
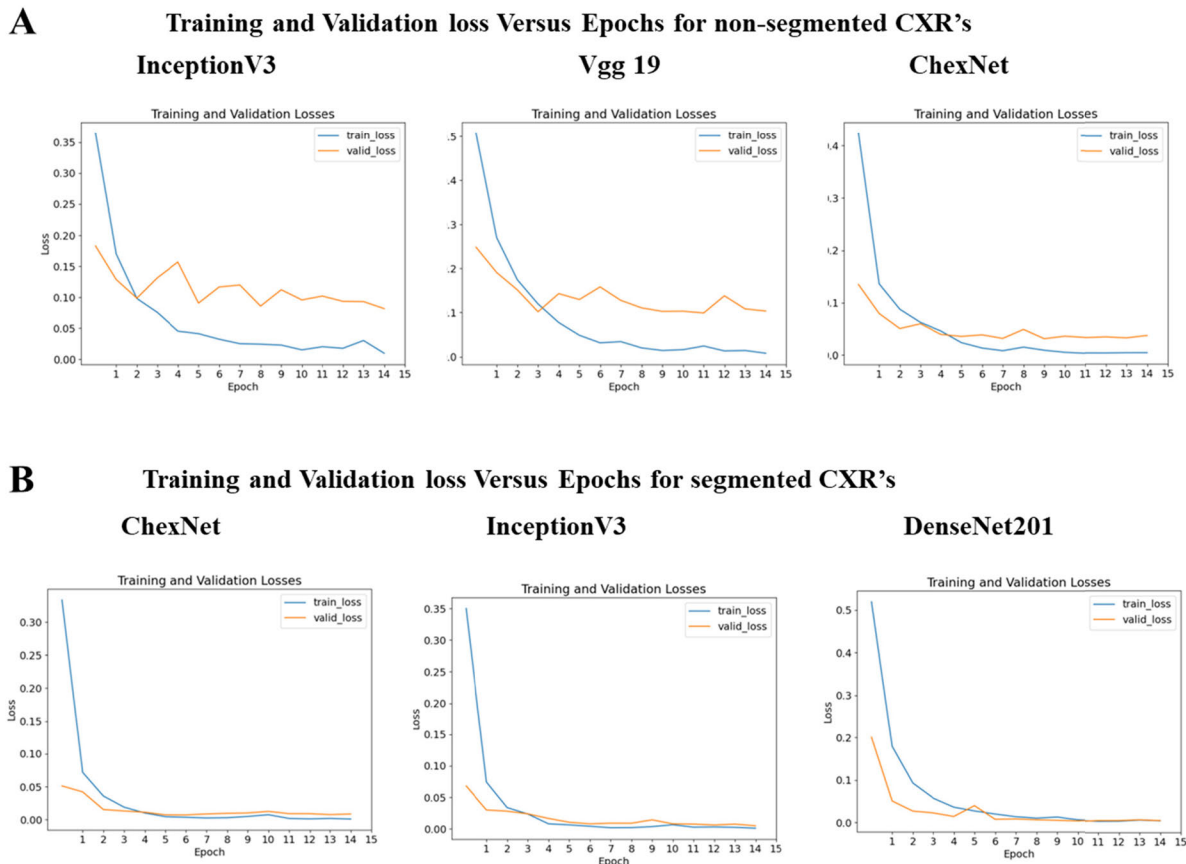
A **Training and Validation loss Versus Epochs for non-segmented CXR's**

InceptionV3 Vgg 19 ChexNet

B **Training and Validation loss Versus Epochs for segmented CXR's**

ChexNet InceptionV3 DenseNet201

FIGURE 9. Training and Validation Losses versus Epoch for non-segmented (A) and segmented (B) CXR Images.

when only lung images are used as input to the CNNs. This is apparent from the ROC curves of Figure 8 as well where a comparable performance from all the networks were shown. Although all the CNNs provide better performance in classification, DenseNet201 showed the most outstanding performance in classifying TB and normal images. Even though performance-wise DenseNet201 was the best performer, however, it was comparatively slow in training and takes more time to process a test image being the deepest network in the group. It is also interesting to see that δt_e increases and δt_t decreases for all the networks when dealing with segmented lung images compared to when working with non-segmented image. Figure 9 shows the training and validation loss versus epochs for the three best performing networks for segmented and non-segmented lung X-rays. It can also be seen that the networks reach and stabilizes with the lowest loss earlier when dealing with the segmented lungs CXR.

In summary, CheXNet and DenseNet201 are producing the highest classification accuracies of 96.47% and 98.6% for non-segmented and segmentation images respectively. DenseNet201 is performing well on the segmented lungs, which reflects that the deeper network can classify more accurately for segmented lung X-ray images. It is evident that segmentation improves overall performance of classification

significantly. However, as this problem is a binary problem and the lungs' regions of TB images are significantly different than that of normal images, all the tested networks performed well.

Figure 10 shows the confusion matrix for outperforming ChexNet model without segmented X-ray images and DenseNet201 model with segmented lungs X-ray images. It can be noticed that even with the best performing network, 115 out of 3500 TB images were miss-classified as normal and 132 out of 3500 normal X-ray images were miss-classified as TB image when the non-segmented X-ray images were used as input to the classifier. On the other hand, only 39 out of 3500 TB images were miss-classified as normal and 59 normal images were miss-classified as TB when the segmented lungs were used as input to the classifier. This is clearly an outstanding performance from any computer aided classifier and this can significantly help in the fast screening of TB by the radiologist immediately after acquiring the X-ray images.

C. VISUALIZATION USING SCORE-CAM

As mentioned earlier, it is important to see where network is learning for the relevant area of the X-ray images or it is learning from anywhere and any non-relevant information for classification. Therefore, Score-CAM based heat maps
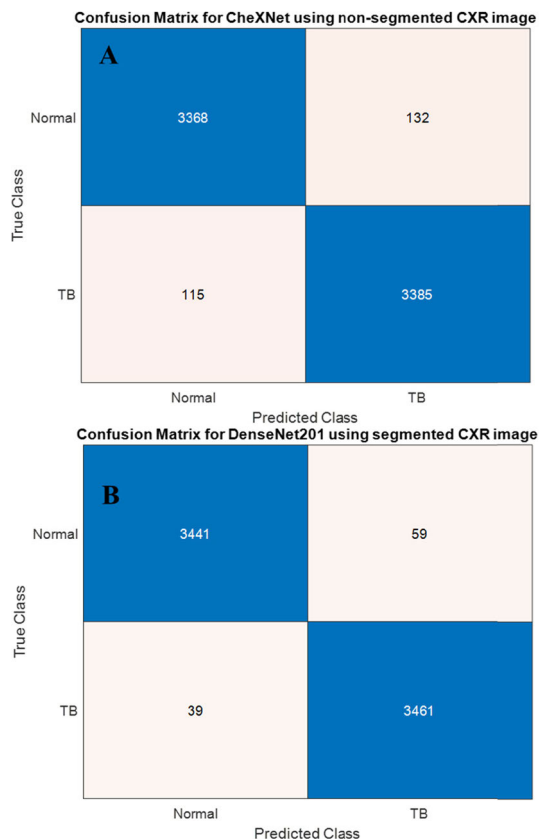
FIGURE 10. Confusion matrix for Normal and Tuberculosis (TB) classification for ChexNet model without segmented X-ray (A), and DenseNet201 model with segmented X-ray (B).

FIGURE 11. Score-CAM visualization of correctly classified TB infected chest X-ray: Original X-ray (left), Score-CAM heat map on original X-ray (middle) and Score-CAM heat map on segmented lungs (right).

FIGURE 12. Feature Space obtained using the best trained Densenet201 network and visualized using t-SNE visualization.

were generated for original (non-segmented) X-ray images and segmented X-ray images. Figure 11 shows the original X-ray samples along with the heat maps on non-segmented and segmented lung. In each of the non-segmented images, CNN is learning from the regions other than the lungs and the areas which are mostly contributing to take decision are not part of the lungs always or most of the cases. Therefore, even though the performance of the CNNs is quite good in classification of TB and normal images, the reliability of these networks would be criticized and several researchers have recently criticized CNNs for learning from non-relevant areas of the image [74]. However, classification using segmented lungs can overcome this drawback and it is evident from Figure 11 that the heat maps for segmented lungs indicate that the dominantly contributing region in the decision making of CNN are within the lung. As TB changes opacities of the lung regions on the CXR, thus learning by the deep networks with segmented lung images provided higher classification accuracies. Since network is now only learning from lung areas, it can learn differences of normal and TB infected lung images only and therefore, it can distinguish them with very high accuracy. We have further confirmed the performance of the best performing network with t-SNE visualization in Figure 12, using the 3500 TB and 3500 Normal segmented CXR images and the best trained Denset201 network.
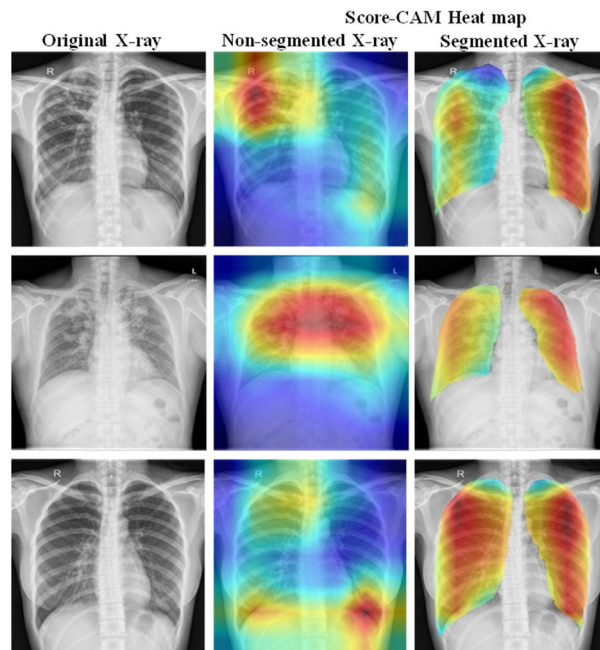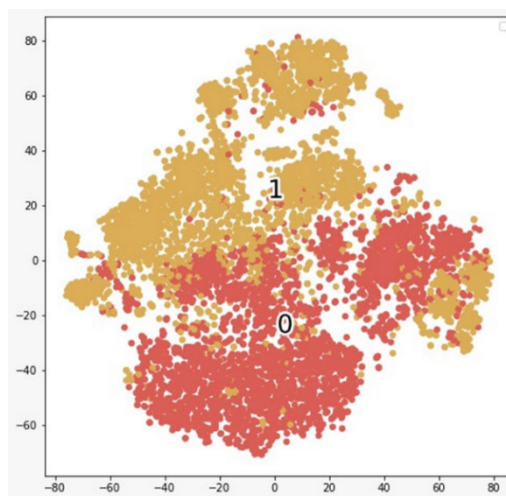
The graphical representation can be seen in Figure 12 which shows the clear distinction between TB and Normal classes represented by 1 and 0 respectively.

As mentioned earlier, out 3500 normal X-ray images, 59 was miss-classified to TB infected image while 39 out of 3500 TB images were miss-classified to normal by the CNN model trained with segmented lungs. It is therefore, important to investigate these four miss-classified images whether these are early stage of TB and therefore, CNN consider them normal or there is any potential reason of this miss-classification. Figure 13 shows that in some miss-classified images CNN took decision from the lower edge of the lungs and only from a smaller area of lungs and that portion of

TABLE 6. Comparison of findings of this study with other recent similar works.

Author	Year	Number of TB images	Method	Database	Evaluation Matrix
Hrudya et al. [26]	2015	138	Support Vector Machine (SVM)	MC	Not stated
Jaime et al. [27]	2016	392	Statistical Analysis	392 records collected from Cape Town in South Africa	AUC - 0.84, Specificity - 49% and Sensitivity 95 %
Rahul et al. [32]	2017	805	CNN	MC and CHN	Accuracy 82.09 %
Anuj et al. [38]	2017	805	CNN Transfer Learning	MC and CHN	Accuracy > 80 %
Lopes et al. [39]	2017	805	CNN Transfer Learning	MC and CHN	Accuracy 84.7% and AUC - 0.926
Abbas et al. [75]	2018	138	Knowledge transferred via Alexnet	MC	AUC – 99.8
Lucas et al. [34]	2018	1052	CNN and two ensembles	JSRT,MC and CHN	Accuracy 88.76 %
Ojasvi et al. [42]	2018	805	Transfer learning (ResNet)	MC and CHN	Accuracy 94.89 %
Niharika et al. [28]	2019	805	Support Vector Machine	MC and CHN	AUC - 0.96 and specificity - 100%
Pasa et al. [35]	2019	1111	Optimized CNN	MC, CHN and Belarus Dataset	AUC - 0.811 for MC, 0.9 for CHN and 0.925 for combined
Syeda et al. [40]		805	Transfer Learning	MC and CHN	AUC-0.85
Mostofa et al. [41]	2019	805	Transfer Learning (VGG 16 model)	MC and CHN	Accuracy 80% and 81.25% without and with augmentation
Quang et al. [36]	2019	805	Tuning of DenseNet model	MC and CHN	AUC-0.94 for CHN and 0.82 for MC
Alfonso et al. [37]	2019	805	3 pre-trained CNNs	MC and CHN	Accuracy 86%
Abbas et al. [25]	2020	247	DeTraC: Class decomposition	JSRT	Accuracy 99.8%
This paper	**2020**	**3500**	**Nine pre-trained CNN models with lungs segmentation**	**NLM (MC and CHN), Belarus, NIAID TB portal & RSNA**	**Accuracy 98.6%**
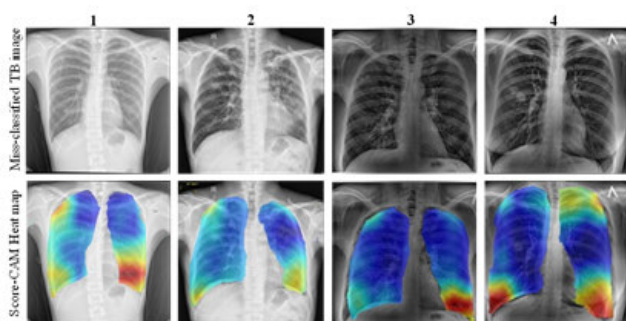
FIGURE 13. Miss-classified TB infected images and their corresponding Score-CAM heat map.

the lungs was normal for all of the images. First case of TB miss-classified image was very similar to normal image and radiologist was classified them as normal and therefore,

it may be very early stage of TB as it was labelled as TB infected image in the original dataset. However, second to fourth TB infected images are from mild to moderate TB patients and CNN miss-classified them as it is learning from wrong area of the lungs. Therefore, if the segmented lungs can be further segmented into patches which can be used as input to CNN model, which might further enhance the performance. This is the future direction of investigation of this work.

This state-of-the-art performance of our proposed method was compared with the recently published works in the same problem domain. Table 6 summarizes the comparison of the results presented in this paper to that of others for the detection of tuberculosis from chest X-ray images. In few studies [43], [75], the detection accuracies have been reported to be 99.8% using databases consisting of small number

of images. However, in our study we used larger datasets than others and found consistent results. We also used image segmentation techniques and evaluated classification performance using nine different CNN models that makes our method more robust and versatile with 98.6% accuracy. Moreover, the performance of our model is justified using Score-CAM based visualization technique to emphasize the importance of segmentation of X-ray images for CNN based classification tasks.

V. CONCLUSION

This work presents a transfer learning approach with deep Convolutional Neural Networks for the automatic detection of tuberculosis from the chest radiographs. The performance of nine different CNN models were evaluated for the classification of TB and normal CXR images. ChexNet model outperforms other deep CNN models for the datasets without image segmentation whereas DenseNet201 outperforms for the segmented lungs. The classification accuracy, precision and recall for the detection of TB were found to be 96.47%, 96.62%, and 96.47% without segmentation and 98.6%, 98.57%, and 98.56% with segmentation respectively. It was also shown that image segmentation can significantly improve classification accuracy. The Score-CAM visualization output confirms that lung segmentation helps in taking decisions from the lung region unlike the original x-rays where decision can be taken based on features outside the lung region. Therefore, segmentation of lungs is very crucial for computer aided diagnosis using radiographs. This state-of-the-art performance can be a very useful and fast diagnostic tool, which can save significant number of people who died every year due to delayed or improper diagnosis.

AUTHORS CONTRIBUTION

Experiments were designed by TR and MEHC. Experiments were performed by TR, AK, KFI and KRI. Results were analyzed by KRI, RM, TH, MEHC, MTI, ZBM and MAK. All the authors were involved in the interpretation of data and paper writing and revision of the article.

ACKNOWLEDGMENT

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

SUPPLEMENTARY MATERIALS

A chest X-ray database of 3500 Tuberculosis patients' image and 3500 normal images were released. This database was created from the 4 publicly available databases, which are referenced in the database. https://www.kaggle.com/tawsifurrahman/tuberculosis-tb-chest-xray-dataset

REFERENCES

[1] *Global Tuberculosis Report 2019*, WHO, Geneva, Switzerland, 2019.
[2] S. K. Sharma and A. Mohan, "Tuberculosis: From an incurable scourge to a curable disease-journey over a millennium," *Indian J. Med. Res.*, vol. 137, no. 3, p. 455, 2013.
[3] C. Silverman, "An appraisal of the contribution of mass radiography in the discovery of pulmonary tuberculosis," *Amer. Rev. Tuberculosis*, vol. 60, no. 4, pp. 466–482, 1949.
[4] A. H. van't Hoog, H. K. Meme, K. F. Laserson, J. A. Agaya, B. G. Muchiri, W. A. Githui, L. O. Odeny, B. J. Marston, and M. W. Borgdorff, "Screening strategies for tuberculosis prevalence surveys: The value of chest radiography and symptoms," *PLoS ONE*, vol. 7, no. 7, Jul. 2012, Art. no. e38691.
[5] A. P. Brady, "Error and discrepancy in radiology: Inevitable or avoidable?" *Insights Imag.*, vol. 8, no. 1, pp. 171–182, Feb. 2017.
[6] A. J. Degnan, E. H. Ghobadi, P. Hardy, E. Krupinski, E. P. Scali, L. Stratchko, A. Ulano, E. Walker, A. P. Wasnik, and W. F. Auffermann, "Perceptual and interpretive error in diagnostic radiology—Causes and potential solutions," *Academic Radiol.*, vol. 26, no. 6, pp. 833–845, Jun. 2019.
[7] M. van Cleeff, L. Kivihya-Ndugga, H. Meme, J. Odhiambo, and P. Klatser, "The role and performance of chest X-ray for the diagnosis of tuberculosis: A cost-effectiveness analysis in Nairobi, Kenya," *BMC Infectious Diseases*, vol. 5, no. 1, p. 111, Dec. 2005.
[8] S. Graham, K. D. Gupta, J. R. Hidvegi, R. Hanson, J. Kosiuk, K. A. Zahrani, and D. Menzies, "Chest radiograph abnormalities associated with tuberculosis: Reproducibility and yield of active cases," *Int. J. Tuberculosis Lung Disease*, vol. 6, no. 2, pp. 137–142, 2002.
[9] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
[10] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, May 2016.
[11] H. Ravishankar *et al.*, "Understanding the mechanisms of deep transfer learning for medical images," in *Deep Learning and Data Labeling for Medical Applications. DLMIA, LABELS* (Lecture Notes in Computer Science), vol. 10008, G. Carneiro *et al.*, Eds. Cham, Switzerland: Springer, 2016, doi: 10.1007/978-3-319-46976-8_20.
[12] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Classification in BioApps* (Lecture Notes in Computational Vision and Biomechanics), vol. 26, N. Dey, A. Ashour, and S. Borra, Eds. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-319-65981-7_12.
[13] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nature Rev. Cancer*, vol. 18, no. 8, pp. 500–510, 2018, doi: 10.1038/s41568-018-0016-5.
[14] M. E. H. Chowdhury, K. Alzoubi, A. Khandakar, R. Khallifa, R. Abouhasera, S. Koubaa, R. Ahmed, and M. A. Hasan, "Wearable real-time heart attack detection and warning system to reduce road accidents," *Sensors*, vol. 19, no. 12, p. 2780, Jun. 2019.
[15] M. E. H. Chowdhury, A. Khandakar, K. Alzoubi, S. Mansoor, A. M. Tahir, M. B. I. Reaz, and N. Al-Emadi, "Real-time smart-digital stethoscope system for heart diseases monitoring," *Sensors*, vol. 19, no. 12, p. 2781, Jun. 2019.
[16] K. Kallianos, J. Mongan, S. Antani, T. Henry, A. Taylor, J. Abuya, and M. Kohli, "How far have we come? Artificial intelligence for chest radiograph interpretation," *Clin. Radiol.*, vol. 74, no. 5, pp. 338–345, May 2019.
[17] T. Rahman, M. E. H. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, and S. Kashem, "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," *Appl. Sci.*, vol. 10, no. 9, p. 3233, May 2020.
[18] M. H. Chowdhury, M. N. I. Shuzan, M. E. H. Chowdhury, Z. B. Mahbub, M. M. Uddin, A. Khandakar, and M. B. I. Reaz, "Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques," *Sensors*, vol. 20, no. 11, p. 3127, Jun. 2020.
[19] A. Al Amin, S. Parvin, M. Kadir, T. Tahmid, S. K. Alam, and K. S.-E. Rabbani, "Classification of breast tumour using electrical impedance and machine learning techniques," *Physiol. Meas.*, vol. 35, no. 6, p. 965, 2014.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[21] P. Chhikara, P. Singh, P. Gupta, and T. Bhatia, "Deep convolutional neural network with transfer learning for detecting pneumonia on chest X-rays," in *Advances in Bioinformatics, Multimedia, and Electronics Circuits and Signals* (Advances in Intelligent Systems and Computing), vol. 1064, L. Jain, M. Virvou, V. Piuri, and V. Balas, Eds. Singapore: Springer, 2020, doi: 10.1007/978-981-15-0339-9_13.

[22] R.-I. Chang, Y.-H. Chiu, and J.-W. Lin, "Two-stage classification of tuberculosis culture diagnosis using convolutional neural network with transfer learning," *J. Supercomput.*, vol. 76, pp. 8641–8656, Jan. 2020.

[23] A. Tahir, Y. Qiblawey, A. Khandakar, T. Rahman, U. Khurshid, F. Musharavati, M. T. Islam, S. Kiranyaz, and M. E. H. Chowdhury, "Coronavirus: Comparing COVID-19, SARS and MERS in the eyes of AI," 2020, *arXiv:2005.11524*. [Online]. Available: http://arxiv.org/abs/2005.11524

[24] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al-Emadi, M. B. I. Reaz, and T. I. Islam, "Can AI help in screening viral and COVID-19 pneumonia?" 2020, *arXiv:2003.13145*. [Online]. Available: http://arxiv.org/abs/2003.13145

[25] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "DeTrac: Transfer learning of class decomposed medical images in convolutional neural networks," *IEEE Access*, vol. 8, pp. 74901–74913, 2020.

[26] H. Das and A. Nath, "An efficient detection of tuberculosis from chest X-rays," *Int. J. Adv. Res. Comput. Sci. Manage. Studies*, vol. 3, no. 5, pp. 149–154, May 2015.

[27] J. Melendez, C. I. Sánchez, R. H. H. M. Philipsen, P. Maduskar, R. Dawson, G. Theron, K. Dheda, and B. van Ginneken, "An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information," *Sci. Rep.*, vol. 6, no. 1, p. 25265, Jul. 2016.

[28] N. Singh and S. Hamde, "Tuberculosis detection using shape and texture features of chest X-rays," in *Innovations in Electronics and Communication Engineering* (Lecture Notes in Networks and Systems), vol. 65, H. Saini, R. Singh, G. Kumar, G. Rather, and K. Santhi, Eds. Singapore: Springer, 2019, doi: 10.1007/978-981-13-3765-9_5.

[29] B. van Ginneken, S. Katsuragawa, B. M. ter Haar Romeny, K. Doi, and M. A. Viergever, "Automatic detection of abnormalities in chest radiographs using local texture analysis," *IEEE Trans. Med. Imag.*, vol. 21, no. 2, pp. 139–149, 2nd Quart., 2002.

[30] S. Jaeger, A. Karargyris, S. Antani, and G. Thoma, "Detecting tuberculosis in radiographs using combined lung masks," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 4978–4981.

[31] J. Melendez, C. I. Sánchez, R. H. Philipsen, P. Maduskar, and B. van Ginneken, "Multiple-instance learning for computer-aided detection of tuberculosis," *Proc. SPIE*, vol. 9035, Mar. 2014, Art. no. 90351J.

[32] R. Hooda, S. Sofat, S. Kaur, A. Mittal, and F. Meriaudeau, "Deep-learning: A potential method for tuberculosis detection using chest radiography," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Kuching, Malaysia, Sep. 2017, pp. 497–502.

[33] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, Aug. 2017.

[34] L. G. C. Evalgelista and E. B. Guedes, "Computer-aided tuberculosis detection from chest X-ray images with convolutional neural networks," in *Proc. Anais do XV Encontro Nacional de Inteligência Artif. e Computacional (ENIAC)*, Oct. 2018, pp. 518–527.

[35] F. Pasa, V. Golkov, F. Pfeiffer, D. Cremers, and D. Pfeiffer, "Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Dec. 2019.

[36] Q. H. Nguyen, B. P. Nguyen, S. D. Dao, B. Unnikrishnan, R. Dhingra, S. R. Ravichandran, S. Satpathy, P. N. Raja, and M. C. H. Chua, "Deep learning models for tuberculosis detection from chest X-ray images," in *Proc. 26th Int. Conf. Telecommun. (ICT)*, Apr. 2019, pp. 381–386.

[37] A. Hernández, Á. Panizo, and D. Camacho, "An ensemble algorithm based on deep learning for tuberculosis classification," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.*, Manchester, U.K., 2019, pp. 145–154.

[38] A. Rohilla, R. Hooda, and A. Mittal, "Tb detection in chest radiograph using deep learning architecture," in *Proc. ICETETSM*, Aug. 2017, vol. 6, no. 8, pp. 1073–1085.

[39] U. K. Lopes and J. F. Valiati, "Pre-trained convolutional neural networks as feature extractors for tuberculosis detection," *Comput. Biol. Med.*, vol. 89, pp. 135–143, Oct. 2017.

[40] S. S. Meraj, R. Yaakob, A. Azman, S. N. Rum, A. Shahrel, A. Nazri, and N. F. Zakaria, "Detection of pulmonary tuberculosis manifestation in chest X-rays using different convolutional neural network (CNN) models," *Int. J. Eng. Adv. Technol. (IJEAT)*, vol. 9, no. 1, pp. 2270–2275, Oct. 2019.

[41] M. Ahsan, R. Gomes, and A. Denton, "Application of a convolutional neural network using transfer learning for tuberculosis detection," in *Proc. IEEE Int. Conf. Electro Inf. Technol. (EIT)*, May 2019, pp. 427–433.

[42] O. Yadav, K. Passi, and C. K. Jain, "Using deep learning to classify X-ray images of potential tuberculosis patients," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 2368–2375.

[43] S. Jaeger, S. Candemir, S. Antani, Y.-X. Wang, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quant. Imag. Med. Surg.*, vol. 4, no. 6, p. 475, 2014.

[44] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, "Multisource transfer learning with convolutional neural networks for lung pattern analysis," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 76–84, Jan. 2017.

[45] H. Yang, S. Mei, K. Song, B. Tao, and Z. Yin, "Transfer-learning-based online Mura defect classification," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 1, pp. 116–123, Feb. 2018.

[46] S. Akçay, M. E. Kundegorski, M. Devereux, and T. P. Breckon, "Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2016, pp. 1057–1061.

[47] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[48] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[49] *ResNet, AlexNet, VGGNet, Inception: Understanding Various Architectures of Convolutional Networks*. Accessed: Jul. 5, 2020. [Online]. Available: https://cv-tricks.com/cnn/understand-resnet-alexnet-vgg-inception/

[50] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[51] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-Rays with deep learning," 2017, *arXiv:1711.05225*. [Online]. Available: http://arxiv.org/abs/1711.05225

[52] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: http://arxiv.org/abs/1602.07360

[53] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[54] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[55] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[56] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proc. IEEE Int. Symp. Circuits Syst.*, Jun. 2010, pp. 253–256.

[57] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[58] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-Net with densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 406–415.

[59] D. Smilkov, N. Thorat, B. Kim, F. Viégas, and M. Wattenberg, "SmoothGrad: Removing noise by adding noise," 2017, *arXiv:1706.03825*. [Online]. Available: http://arxiv.org/abs/1706.03825

[60] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[61] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 839–847.

[62] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-CAM: Score-weighted visual explanations for convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 24–25.

[63] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[64] P. Perera and V. M. Patel, "Learning deep features for one-class classification," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5450–5463, Nov. 2019.

[65] A. D. Baxevanis, G. D. Bader, and D. S. Wishart, *Bioinformatics*. Hoboken, NJ, USA: Wiley, 2020.

[66] M. A. Myszczynska, P. N. Ojamies, A. M. B. Lacoste, D. Neil, A. Saffari, R. Mead, G. M. Hautbergue, J. D. Holbrook, and L. Ferraiuolo, "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases," *Nature Rev. Neurol.*, vol. 16, no. 8, pp. 440–456, Aug. 2020.

[67] J. Wang, C. Jiang, H. Zhang, Y. Ren, K.-C. Chen, and L. Hanzo, "Thirty years of machine learning: The road to Pareto-optimal wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1472–1514, 3rd Quart., 2020.

[68] S. Jaju. (2017). *Comprehensive Guide on t-SNE Algorithm With Implementation in R & Python*. Accessed: Sep. 1, 2020. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/01/t-sne-implementation-r-python

[69] Pandey. *Chest Xray Masks and Labels*. Accessed: Jun. 9, 2020. [Online]. Available: https://www.kaggle.com/nikhilpandey360/chest-xray-masks-and-labels

[70] B. P. Health. (2020). *Belarus Tuberculosis Portal*. Accessed: Jun. 9, 2020. [Online]. Available: http://tuberculosis.by/

[71] A. Rosenthal *et al. The TB Portals: An Open-Access, Web-Based Platform for Global Drug-Resistant-Tuberculosis Data Sharing and Analysis*. Accessed: Sep. 1, 2020. [Online]. Available: https://data.tbportals.niaid.nih.gov/

[72] kaggle. *RSNA Pneumonia Detection Challenge*. Accessed: Jun. 9, 2020. [Online]. Available: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data

[73] E. D. Science. *Overfitting in Machine Learning: What It Is and How to Prevent It*. Accessed: Jul. 7, 2020. [Online]. Available: https://elitedatascience.com/overfitting-in-machine-learning

[74] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.

[75] A. Abbas and M. M. Abdelsamea, "Learning transformations for automated classification of manifestation of tuberculosis using convolutional neural network," in *Proc. 13th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2018, pp. 122–126.

• • •