



## Research Article

## Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images

Vinayak Sharma<sup>1</sup>, Nillmani<sup>2,\*</sup>, Sachin Kumar Gupta<sup>1</sup>, Kaushal Kumar Shukla<sup>3</sup><sup>1</sup> School of Electronics and Communication Engineering, Shri Mata Vaishno Devi University Jammu and Kashmir, India<sup>2</sup> School of Biomedical Engineering, Indian Institute of Technology, Banaras Hindu University, Varanasi, India<sup>3</sup> Department of Computer Science and Engineering, Indian Institute of Technology, Banaras Hindu University, Varanasi, India

## ARTICLE INFO

## Keywords:

Tuberculosis  
Artificial intelligence  
Deep learning  
Segmentation  
Classification  
Chest X-ray

## ABSTRACT

**Objective** Tuberculosis (TB) is among the most frequent causes of infectious-disease-related mortality. Despite being treatable by antibiotics, tuberculosis often goes misdiagnosed and untreated, especially in rural and low-resource areas. Chest X-rays are frequently used to aid diagnosis; however, this presents additional challenges because of the possibility of abnormal radiological appearance and a lack of radiologists in areas where the infection is most prevalent. Implementing deep-learning-based imaging techniques for computer-aided diagnosis has the potential to enable accurate diagnoses and lessen the burden on medical specialists. In the present work, we aimed to develop deep-learning-based segmentation and classification models for accurate and precise detection of tuberculosis in chest X-ray images, with visualization of infection using gradient-weighted class activation mapping (Grad-CAM) heatmaps.

**Methods** First, we trained the UNet segmentation model using 704 chest X-ray radiographs taken from the Montgomery County and Shenzhen Hospital datasets. Next, we implemented the trained UNet model on 1,400 tuberculosis and control chest X-ray scans to segment the lung region. The images were taken from the National Institute of Allergy and Infectious Diseases (NIAID) TB portal program dataset. Then, we applied the deep learning Xception model to classify the segmented lung region into tuberculosis and normal classes. We further investigated the visualization capabilities of the model using Grad-CAM to view tuberculosis abnormalities in chest X-rays and discuss them from radiological perspectives.

**Results** For segmentation by the UNet model, we achieved accuracy, Jaccard index, Dice coefficient, and area under the curve (AUC) values of 96.35%, 90.38%, 94.88%, and 0.99, respectively. For classification by the Xception model, we achieved classification accuracy, precision, recall, F1-score, and AUC values of 99.29%, 99.30%, 99.29%, 99.29%, and 0.999, respectively. The Grad-CAM heatmap images from the tuberculosis class showed similar heatmap patterns, where lesions were primarily present in the upper part of the lungs.

**Conclusion** The findings may verify our system's efficacy and superiority to clinician precision in tuberculosis diagnosis using chest X-rays and raise the possibility of a valuable setup, particularly in environments with a scarcity of radiological expertise.

## 1. Introduction

Tuberculosis (TB) is among the most common communicable diseases caused by a bacterial infection, namely *Mycobacterium tuberculosis* [1]. TB is also the most common reason for infectious-disease-related mortality. In 2021, almost 10.6 million individuals worldwide became ill with TB and 1.6 million of them died, making it the second most common infectious killer after COVID-19 [1]. All age groups and genders from each nation are vulnerable, with 6 million, 3.4 million, and 1.2 million men, women, and children, respectively, reported to be ill with

TB in 2021 [1–2]. TB can be cured with early diagnosis and subsequent delivery of the right medicines [3]. Chest X-rays (CXR) are frequently used for identification of and screening for pulmonary TB. Chest radiographs are evaluated for the presence of TB by skilled doctors in clinical practice. However, this is a subjective and time-consuming process, and diagnoses inevitably have subjective discrepancies. In addition, patients with TB are frequently misdiagnosed based on CXRs as having other diseases with similar radiologic patterns, which can cause them to receive incorrect treatment and their health to deteriorate [4]. There is also a shortage of radiologists with training in resource-deficient nations,

\* Corresponding author: Nillmani, School of Biomedical Engineering, Indian Institute of Technology, Banaras Hindu University, Varanasi, India (Email: [nillmani.rs.bme17@itbhu.ac.in](mailto:nillmani.rs.bme17@itbhu.ac.in)).

<https://doi.org/10.1016/j.imed.2023.06.001>

Received 29 March 2023; Received in revised form 7 June 2023; Accepted 25 June 2023

2667-1026/© 2023 The Author(s). Published by Elsevier B.V. on behalf of Chinese Medical Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

particularly in rural regions. To overcome these limitations, computer-aided diagnosis (CAD) systems may be used to perform mass screening for TB, owing to great improvements in the performance of deep convolutional neural networks (CNNs) and the availability of extensive labeled datasets of radiological images [5–6]. Examination of medical data by medical specialists is significantly hampered by subjective judgments, the quality of the radiological images, and the exhaustion caused by a demanding workload [7]. Hence, the use of machine learning in the healthcare industry is currently a topic of much interest [8–9]. Many problems in healthcare, including detection of brain tumors [10–13], diagnosis of cardiovascular disease [14–17], detection of pneumonia [18–22], identification of lung nodules [23–24], diagnosis of breast cancer [25–27], and physiological monitoring [28], have been addressed using artificial intelligence (AI)-based solutions.

Deep learning (DL) approaches have been used with increasing success for the fast, accurate, and automatic detection of lung illnesses using chest radiological images [29]. CXR is preferable to computed tomography scans and magnetic resonance imaging for this purpose as it is a low-cost, readily available, low-radiation-dose imaging technology [30–32]. Moreover, many doctors skilled in deciphering CXR scans have used their expertise in the creation of DL-based CAD tools.

Inbuilt DL-CAD tools can be made more dependable in clinical applications if the accuracy of TB diagnosis using CXRs can be improved through the use of a robust and adaptable technique. Optimization of DL networks, modification of already effective techniques, and combination of several effective algorithms can all increase classification accuracy [33]. Traditionally, CNNs were applied to entire X-ray images to identify lung diseases, that is, to CXRs showing other parts of the thorax as well as the lungs. However, as TB only manifests in the lungs, concentrating on the lung area of CXRs during model formulation can greatly enhance the performance of TB diagnosis systems [34]. Various researchers have tried to develop or implement new and existing DL techniques to detect TB automatically using CXR scans [35–36]. However, there is not yet any highly accurate and robust system that can be generalized for practical applications.

This study focused on the development of segmentation-based classification approaches that could be highly accurate, robust, and generalizable for the detection of TB in practical applications. The motive for the segmentation-based approach is to limit learning from irrelevant sections of the CXR and improve accuracy.

## 2. Methods

Figure 1 illustrates the overall methodology adopted in this study. First, we selected a well-established CXR dataset containing CXR images and corresponding masks for segmentation model training. Next, we modified and optimized a highly successful UNet segmentation network for training. Following the model's training, performance was evaluated using diverse evaluation metrics, namely accuracy, Dice, Jaccard, and area under the curve (AUC). Further, the segmentation model was applied to TB and normal CXR image datasets to segment the lung region and remove the background. Next, we optimized and applied a highly efficient Xception DL convolutional neural network (CNN) to classify the segmented lung images. We evaluated the model's performance using various metrics including accuracy, precision, recall, F1-score, and AUC, as well as receiver operating characteristic (ROC) curves. Finally, we generated gradient-weighted class activation mapping (Grad-CAM) heatmaps of the images to visualize the patterns and locations of the lesions present in the lungs.

### 2.1. Data collection and patient demographics

For the segmentation part of the experiment, we used a combination of the Montgomery County dataset and the Shenzhen Hospital dataset. The datasets were taken from the open-source Kaggle “Chest X-ray masks and labels dataset” [37]. The combined dataset contained

704 CXRs and their corresponding masks; 494 randomly selected images (70%) were used for training of the model, 140 images (20%) were used for validation, and 70 images (10%) were used for testing of the model. The experiment involved five-fold cross-validation with each image used at least once for training and testing.

For the classification part of the experiment, we used the “National Institute of Allergy and Infectious Diseases (NIAID) TB portal program dataset.” This dataset was taken from the open-source Kaggle “Tuberculosis chest X-ray images dataset” [38] and contained 700 TB and 3,500 control CXR images, of which we used all 700 TB images and 700 control CXR images. For the classification problem, 1,120 randomly selected images (80%) were used for training, 140 images (10%) were used for validation, and the remaining 140 images (10%) were used for testing of the model.

### 2.2. Network architecture and optimization

UNet, which was first developed by Olaf Ronneberger in 2015 [39], is one of the most popular and efficient models for diverse radiological image segmentation in medical imaging. UNet has been proven to be reliable in several studies and has been used by numerous researchers. Therefore, in the present study, we used the UNet model to segment CXR images. Figure 2 shows the detailed architecture of the UNet network. As shown in the diagram, the UNet network has two phases: the encoder and the decoder. The encoder phase has several convolutional layers followed by ReLU and MaxPooling. The decoder phase contains up-convolution followed by depth concatenation, softmax, and MaxPooling. The image provided as input to the encoder is  $224 \times 224$  pixels in size. The encoder extracts the features of images in the form of numerical values between 0 and 255, which are later normalized to be between 0 and 1. The image size is reduced during transfer to the successive convolution after each MaxPooling layer. A bridge channel connects the encoder to the decoder and transfers the image's features to the decoder. The decoder phase decodes the numerical features into the form of an image and recreates the images, converting the pixel values  $< 0.5$  to 0 and  $\geq 0.5$  to 1. The encoder and decoder phases are also connected through skip connections that are used to transfer image features from an encoder to a decoder directly; this prevents the network from losing image information through loss from neural nodes. The final output images from the decoder have the binary pixel values 0 or 1, where 0 represents black and 1 represents white. This results in the generation of final black and white masks, which are used to generate the region of interest, i.e., the lung section from the CXRs.

For the segmentation problem, we employed a five-fold cross-validation protocol. As the segmentation dataset contained only 704 CXR images, the cross-validation protocol was highly beneficial as it allowed us to use each image for training at least once; this helped to maximize the utilization of the dataset. We split the dataset using 60:20:20 ratios to train, validate, and test the model. The UNet model was trained for 100 epochs in each fold, with the learning rate and dropout set to 0.001 and 0.25, respectively. The batch size for training and validation was maintained at four images per batch. The loss function implemented was cross-entropy, denoted  $L_{CE}$  and derived as in Equation (1).

$$L_{CE} = [(y_i \times \log a_i) + (1 - y_i) \times \log(1 - a_i)] \quad (1)$$

Here,  $i$  is the image,  $y_i$  is the GT (Ground truth) label 1,  $(1 - y_i)$  is GT label 0, and  $a_i$  is the Softmax classifier probability.

After segmentation to classify the lung regions, we applied the Xception DL model. Figure 3 illustrates the detailed architecture of the Xception neural network, which was first developed by Chollet in 2017 [40]. As shown in the diagram, the Xception model consists of three phases: entry flow, middle flow, and exit flow. CXR images with a resolution of  $224 \times 224$  pixels are given as input to the entry flow. Each flow has several neural nodes, containing convolution layers followed by ReLU and

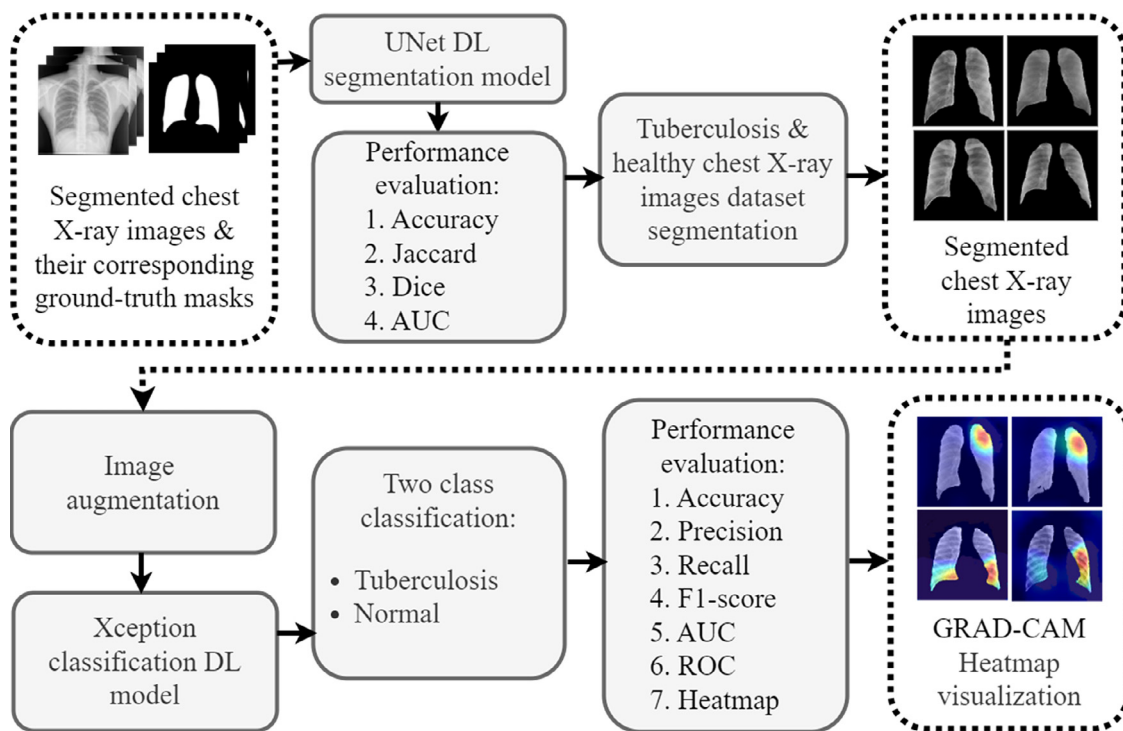


Figure 1. Schematic diagram representing overall methodology.

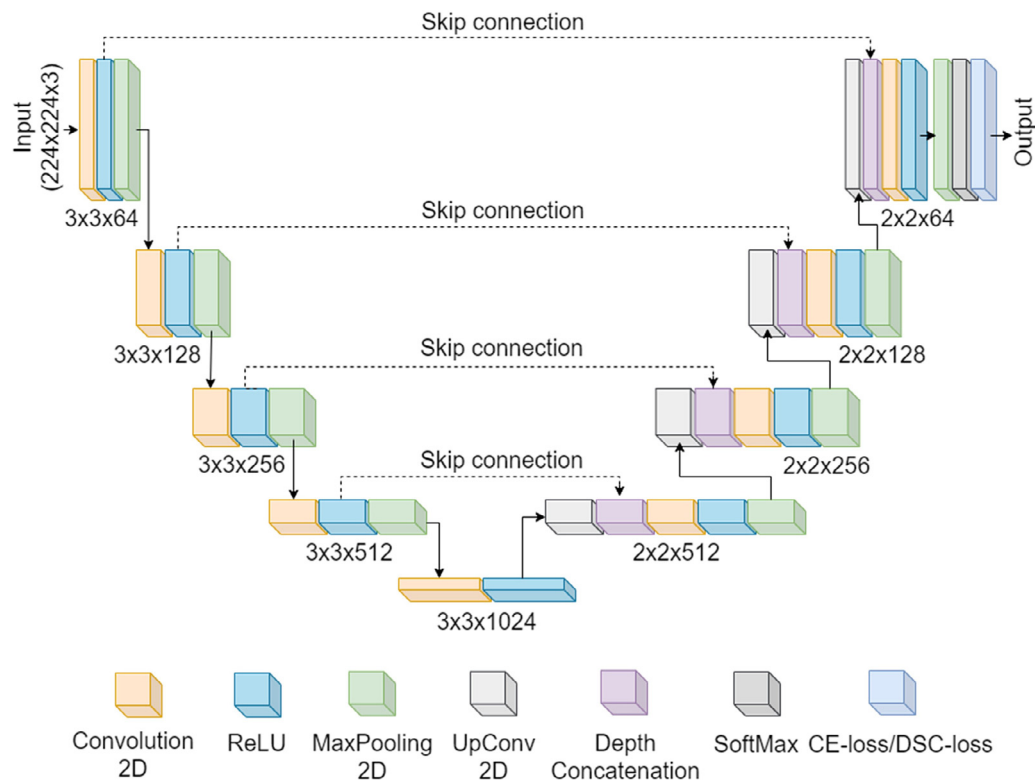


Figure 2. The architecture of the UNet segmentation network.

pooling layers. The nodes are also directly connected by a few convolutions, which simultaneously transfer features to the next neural nodes without any loss. The output from the entry flow is used as input to the middle flow, and the output from the middle flow is used as input to the exit flow. The concluding classifier layer of the exit flow is a logistic regression layer, which classifies the images based on the features ex-

tracted by the neural nodes and predicts the class for each image as its output.

For the classification problem, 100 epochs were used to train the Xception network. A learning rate of 0.001 was maintained, and the dropout rate was 0.25. The training and validation steps were performed with eight images per batch, and the best-performing epoch with the

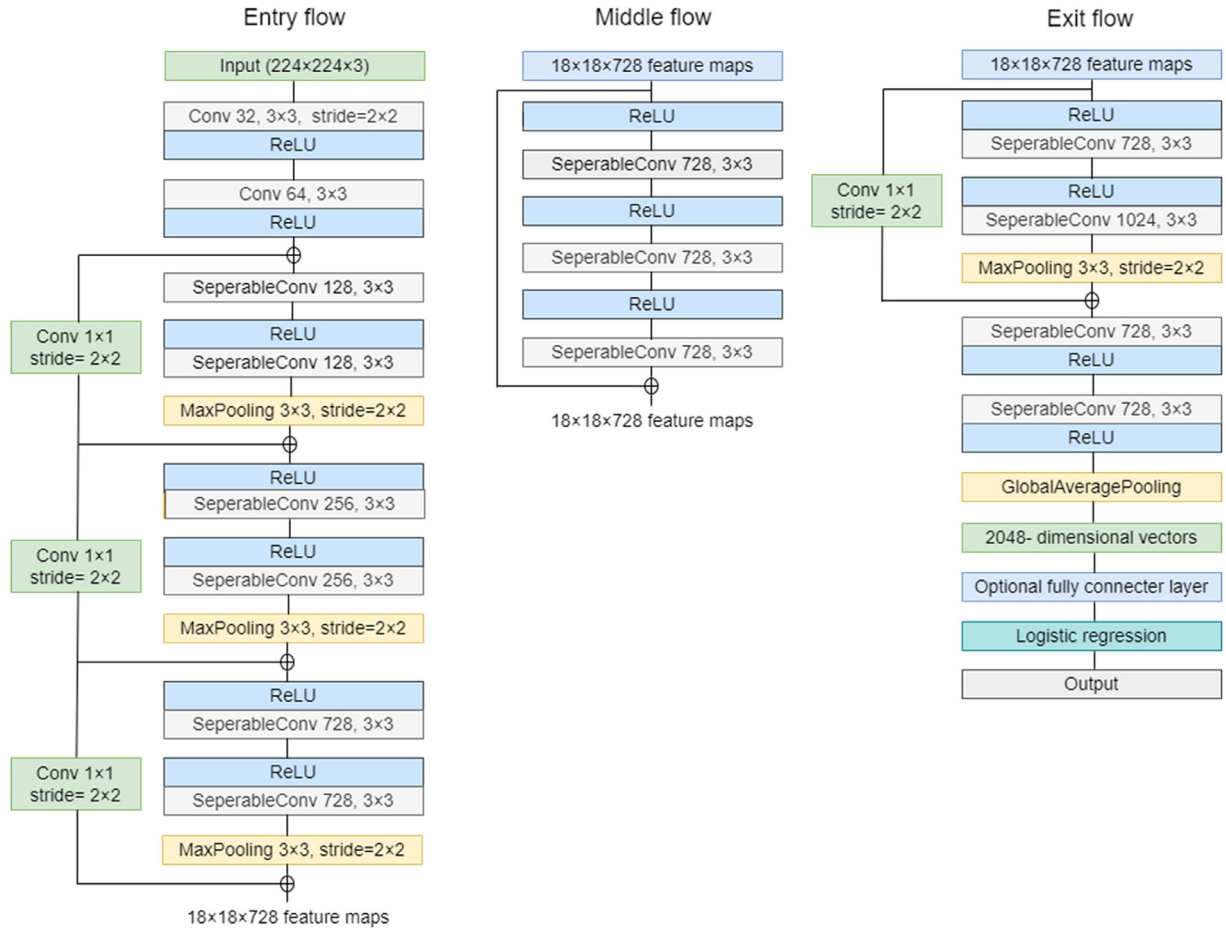


Figure 3. The architecture of the Xception classification network.

highest validation accuracy was saved as the final model during the training. The loss function was categorical cross-entropy, denoted  $L_{CCE}$  and derived as in Equation (2).

$$L_{CCE} = \frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C 1_{y_i \in C_c} \log a_{model}(y_i \in C_c) \quad (2)$$

Here,  $N$  is the total number of images,  $C$  is the number of categories or classes, and  $1_{y_i \in C_c}$  denotes the  $i$ th observation of the  $c$ th category.

The experiments were carried out using Python 3.9 on a workstation with an Intel Core i7 8th-generation processor, 16 GB of RAM, and an 8 GB NVIDIA Quadro P4000 GPU.

### 2.3. Performance evaluation metrics

Performance assessment is a crucial part of analyzing the competence of a DL-CAD system. Various different metrics can be used for the assessment of the network's performance. In the present work, we used accuracy, precision, recall, and F1-score as metrics, calculated as in Equations (3) and (4):

$$Accuracy = \frac{True\ Positives + True\ negatives}{Total\ number\ of\ cases} \quad (3)$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (4)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (5)$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (6)$$

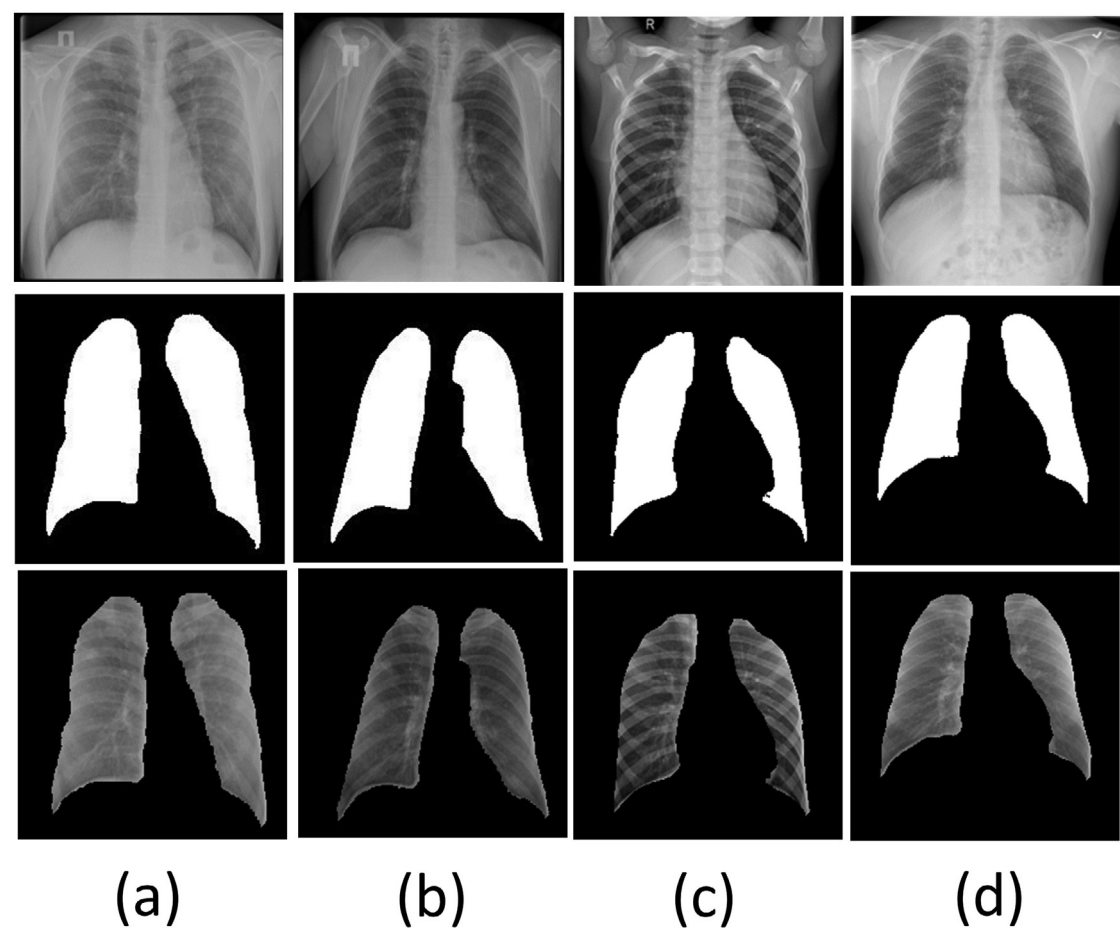
### 2.4. Gradient-weighted class activation mapping (Grad-CAM) visualization

The Grad-CAM method uses gradients for the classification score and the final convolutional feature map to pinpoint the areas of an input image that have the greatest influence on the classification score. The final score depends mainly on the data in the locations where this gradient is large. The Grad-CAM function produces a significance map by calculating the derivative of the reduction layer's output for a certain class with respect to a convolutional feature [41], automatically choosing appropriate layers to compute the significance map for classification tasks. Grad-CAM heatmaps are very advantageous for analysis of medical images as they represent the most significant regions of the images with respect to infection. The heatmap displays the areas of lesion or infection within the image, as well as showing the region from which the network has extracted the most significant features used to determine the classification of the image. Generally, such heatmaps show similar patterns in similar infection types. In our system using the gradients of the target, i.e., TB in the Xception-based classification model, Grad-CAM generates a coarse localization map and displays significant locations as heatmap scans [20].

## 3. Results

The segmentation UNet model achieved accuracy, Jaccard, Dice, and AUC values of 96.35%, 90.38%, 94.88%, and 0.99, respectively. Figure 4 shows a sample of images segmented by the UNet model. The top row shows the original CXR scans, the middle row shows the masks generated by the model, and the bottom row shows the final segmented lung images. Columns (a) and (b) represent TB-infected images, and





**Figure 4.** Segmentation results. (a,b) TB images; (c,d) normal images. Top row, original chest CXR images; middle row, UNet-generated masks; bottom row, final segmented lung regions.

**Table 1** Performance metrics for classification by the Xception model

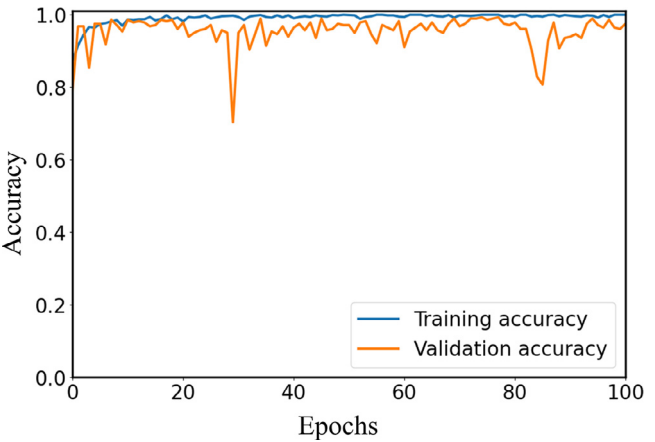
Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
Xception	99.29	99.30	98.29	99.29	0.999

columns (c) and (d) represent normal images. The border of the lung region and the visualization confirm that the model performed outstandingly when segmenting CXRs.

Table 1 presents the performance metrics for classification by the Xception model.

Learning curves are a frequently used analytical technique in machine learning. The curves obtained from learning on training and validation datasets can be used to evaluate an algorithm’s performance. After the completion of each cycle or epoch during training, the models are assessed or validated using the validation dataset. Finally, the measured performance is plotted to give learning curves, which can be examined to assess the performance and learning-related issues of the network. Figure 5 shows the training and validation accuracy curves for the Xception model, demonstrating stable and improving learning for successive epochs and maximum training and validation accuracy values of 100% and 99.29%, respectively. Figure 6 shows the training and validation loss of the Xception model, with curves representing the loss reduction with successive epochs The stable and significant output demonstrates the excellence and reliability of the model’s performance.

Figure 7 shows the confusion matrix for classification by the DL Xception model, representing the actual and predicted classification of images. In the test set, all 70 control images and 69 of 70 TB images were correctly classified, with just one image misclassified. The confusion matrix verifies the model’s efficacy and precise classification results.



**Figure 5.** Training and validation accuracy curves for the Xception model.

ROC curves illustrate a model’s performance by comparing the true positive and false positive rates. AUC values near 1.0 indicate superior model performance, whereas values near 0 indicate poor performance. ROC curves are generated using the prediction probabilities and the actual classes for each image. Figure 8 shows the ROC curves and AUC values for classification with the Xception model. The AUC values for the normal and TB classes were 0.99 and 1.0, respectively. The micro and macro averages of AUC were both 1.0. These AUC values, which were close to the maximum, indicate the model’s precision and superior performance.

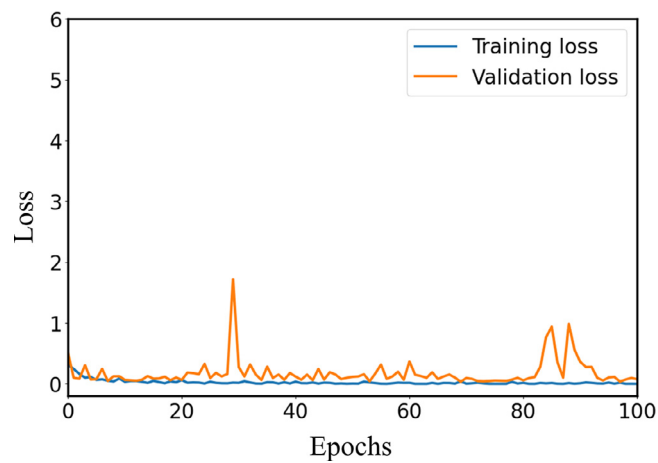


Figure 6. Training and validation loss curves for the Xception model.

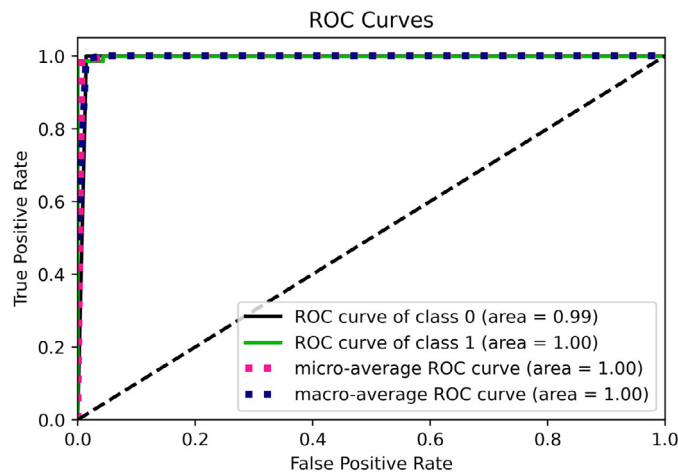


Figure 8. ROC curves for the Xception model.

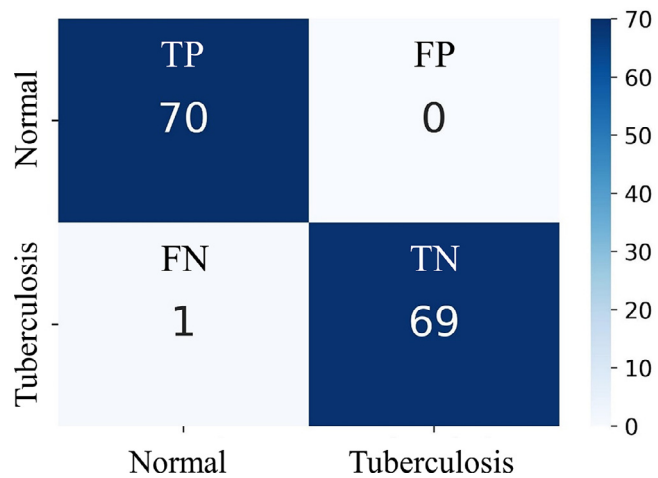


Figure 7. Confusion matrix for classification by the Xception model. TP: true positive; FP: false positive; FN: false negative; TN: true negative.

Figures 9 and 10 show heatmaps for samples of CXR images from the TB and control classes, respectively, taken from the “NIAID TB portal program dataset” [38]. In each figure, row (a) shows the original CXR images from each class, whereas row (b) shows the corresponding segmented lung regions from the CXRs. Row (c) represents the final GRAD-CAM heatmap images for each class. The heatmap images for the TB class showed similar heatmap patterns for each image, with lesions primarily present in the upper parts of the lungs. By contrast, the heatmaps for normal images showed completely different patterns from the TB images and similar patterns among images of the same class.

4. Discussion

Automated CXR image analysis techniques have developed intensely over the past 10 years in response to the limitations of knowledge and accessibility that hinder early TB detection [42]. DL using CNNs serves as a foundation for medical imaging and AI-embedded CAD solutions [43],

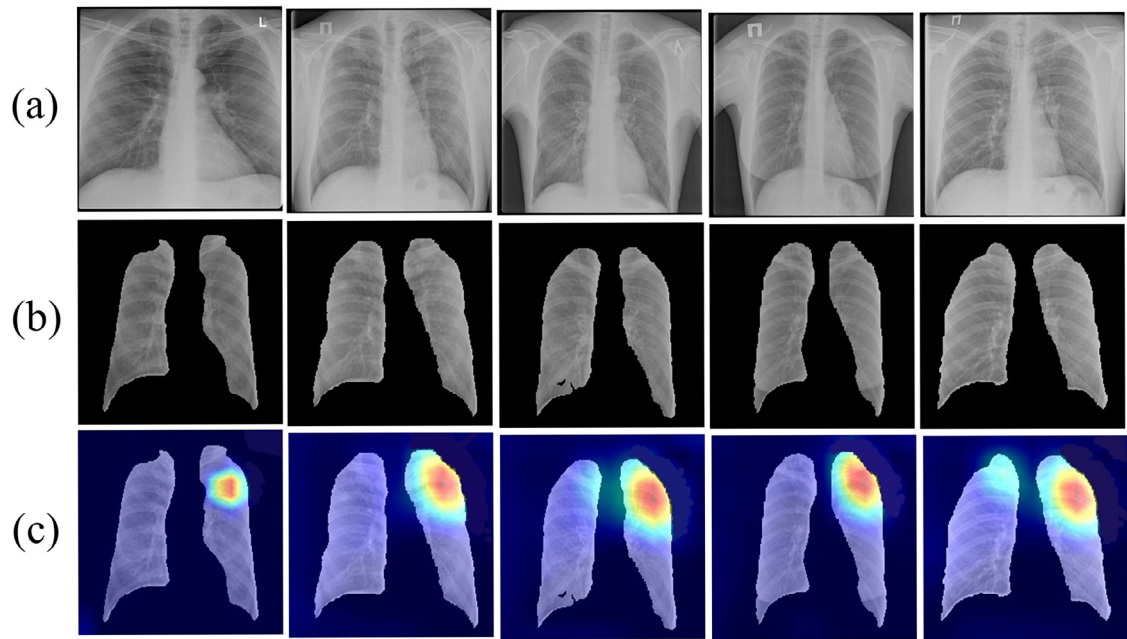
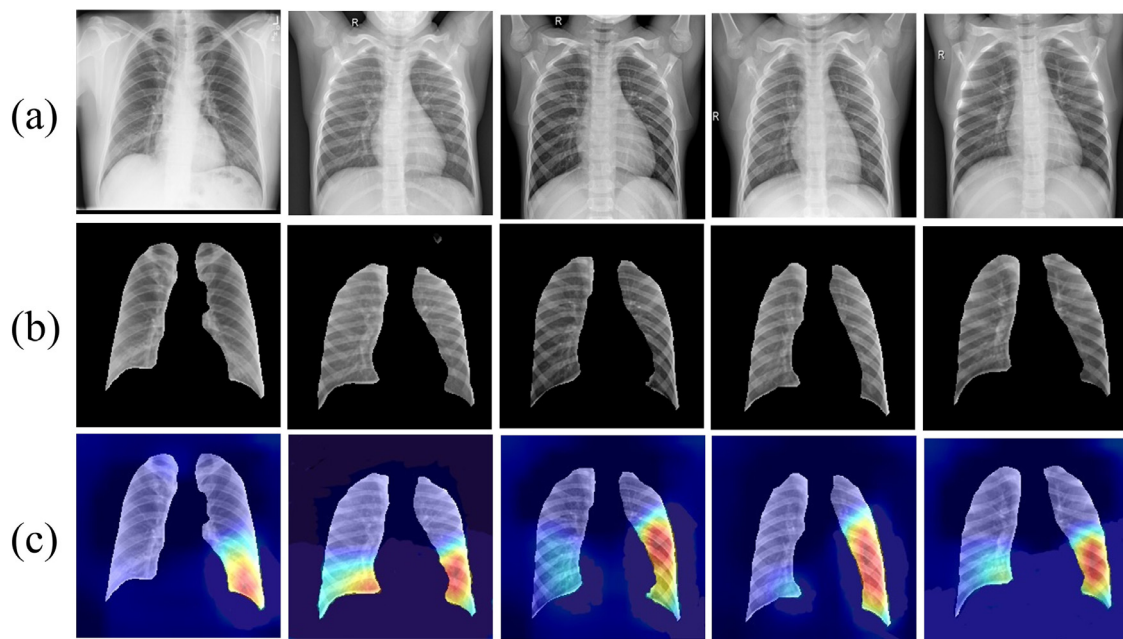


Figure 9. GRAD-CAM heatmap visualizations of TB-infected CXR images. (a) Original CXR images; (b) Corresponding segmented lung images; (c) Corresponding heatmaps.



**Figure 10.** GRAD-CAM heatmap visualizations of control CXR images. (a) Original CXR images; (b) Corresponding segmented lung images; (c) Corresponding heatmaps.

in which clinical decision-making is intended to be complemented by AI-guided solutions (such as DL techniques). Such imaging systems have the potential to lessen the burden on radiologists by prioritizing cases [44]. Conventional machine-learning techniques employ hand-crafted features that are restricted to a particular issue or dataset because feature extraction is an expert-based process. Deep features, on the other hand, can be extracted automatically without any explicit parameter adjustment. For low-resource countries in particular, DL-facilitated systems for TB detection are a possible answer to physicians' problems and enable the fast, automatic, and precise TB diagnosis that is crucial to achieving global control of the disease [45].

Various researchers have developed or implement novel and existing methods to achieve this goal and provide reliable solutions. Table 2 presents a comparison of the proposed method against previously implemented methods for TB detection. Ahsan et al. [46] applied the VGG16 DL classification network to a combination of two popular datasets: the Montgomery County dataset, which contains 58 TB and 80 healthy CXR images; and the Shenzhen Hospital dataset, which contains 336 TB and 326 normal images. They used a total of 800 images (394 TB-infected and 406 control CXR images) and split their dataset using a 3:1 train/test ratio, achieving an accuracy of 81.25%. Instead of using binary classification, i.e., TB versus healthy, Xie et al. [47] focused on the detection of TB-infected areas or lesions using a region-based faster RCNN (Region-based Convolutional Neural Network) method. They employed a combination of three datasets: the Montgomery County dataset; the Shenzhen Hospital dataset; and the First Affiliated Hospitals Dataset from Xian Jiao Tong University, Shanxi, China, which contains a total of 5,344 images with 2,962 TB and 2,382 control images. They used a total of 6,144 images, comprising 3,356 TB and 2,788 normal images, and opted for a 80:20 train/test split. They achieved a maximum accuracy of 92.60% and an AUC of 0.98.

Sahlol et al. [48] implemented a new hybrid network, Mobile Net-AEO, in which artificial ecosystem-based optimization (AEO) works as a feature selector, keeping the relevant features and discarding the redundant features that are generated by the CNN. They used a combination of two datasets, the Shenzhen Hospital dataset and the Mendeley Dataset (UK), which contains a total of 6,421 (3,883 TB and 2,538 normal) CXR images. Their study used a full dataset of 7,083 images (4,219 TB and 2,864 normal), which was split into training/test sets

with a 80:20 ratios, and demonstrated an accuracy of 94.10%. Rahman et al. [49] employed nine different transfer-learning-based deep neural networks, namely VGG19, ResNet18, ResNet50, ResNet101, MobileNet, DenseNet201, SqueezeNet, and ChexNet, to classify CXR images into normal and TB classes. Before classification, they segmented the lung region using a modified UNet network. They used a combination of four datasets: Montgomery, Shenzhen, Belarus TB Dataset (Belarus), and Radiological Society of North America Dataset (RSNA). The Belarus dataset contained 422 CXR images from TB-infected patients, whereas the RSNA dataset contained a total of 26,684 images (17,833 TB and 8,851 control images). This gave a total of 27,906 images, of which 18,649 were TB and 9,257 were normal images. Using a five-fold cross-validation approach, they achieved the maximum accuracy of 98.60% with DenseNet201. Guo et al. [50] employed an ensemble method averaging six deep neural networks, namely VGG16, VGG19, ResNet50, ResNet101, Inception V3, and ResNet34. Using the Shenzhen dataset with ten-fold cross-validation approach, they achieved an accuracy of 94.50% in their experiments. Abideen et al. [51] implemented a B-CNN (Bayesian-based CNN) method for the classification of CXR images into TB and normal classes. They used a combination of the Montgomery and Shenzhen datasets, giving a total of 800 CXR images with 394 TB and 406 normal images, which they split into training/test sets with an 80/20 ratio. They achieved an accuracy of 96.42%. Ayaz et al. [52] used an ensemble model that combined hand-crafted features and DL features, with a Gabor filter to extract hand-crafted features and transfer-learning-based pre-trained networks to extract deep features. The pre-trained DL networks were Inceptionv3, MobileNet, Xception, ResNet50, and InceptionResnetv2. They used a combination of Montgomery and Shenzhen datasets that contained a total of 800 CXR images (394 TB and 406 normal images) and implemented a ten-fold cross-validation approach for training. They achieved an accuracy of 93.47%, with an AUC of 0.97. Rajaraman et al. [53] used a bone-suppression-based ResNet-BS model to detects and remove occluding bony structures from CXRs, with a combination of the Montgomery and Shenzhen datasets. Their findings confirmed the better performance of DL classification models on bone-suppressed CXR images, with an accuracy of 92.30% and an AUC of 0.96. Acharya et al. [54] developed a normalization-free CNN and used it for classification of images into TB and normal classes. They used a combination of different datasets: the TBX11K (Tuberculosis X-ray)

**Table 2** Comparison of the proposed method against previous state-of-the-art methods

References	Methods	Dataset (image size, n)	Total number of images used (TB + normal)	Accuracy (%)	AUC
Ahsan et al. [46]	VGG16	Montgomery (138) + Shenzhen (662)	800 (394 + 406)	81.25	-
Xie et al. [47]	Faster RCNN	Montgomery (138) + Shenzhen (662) + First Affiliated Hospitals Dataset (5, 344)	6,144 (3,356 + 2,788)	92.60	0.98
Sahlol et al. [48]	Mobile Net-AEO	Shenzhen (662) + Mendeley dataset (UK) (6421)	7,083 (4,219 + 2,864)	94.10	-
Rahman et al. [49]	CheXNet	Montgomery (138) + Shenzhen (662) + Belarus TB Dataset (Belarus) (422) + Radiological Society of North America Dataset (RSNA) (26,684)	27,906 (18,649 + 9,257)	98.60	-
Guo et al. [50]	Ensemble (ResNet34, ResNet50, ResNet101, VGG16, VGG19, and Inception V3)	Shenzhen (662)	662 (336 + 326)	94.50	-
Abideen et al. [51]	B-CNN	Montgomery (138) + Shenzhen (662)	800 (394 + 406)	96.42	-
Ayaz et al. [52]	Ensemble (MobileNet, Inceptionv3, ResNet50, InceptionResnetv2, and Xception)	Montgomery (138) + Shenzhen (662)	800 (394 + 406)	93.47	0.97
Rajaraman et al. [53]	ResNet-BS	Montgomery (138) + Shenzhen (662)	800 (394 + 406)	92.30	0.96
Acharya et al. [54]	NFNet (normalization-free network)	TBX11K -Tuberculosis X-ray (800) + Shenzhen (662) + Montgomery (138) + National Institute of TB and respiratory diseases, New Delhi (88) + NLM, Belarus, NIAID, SNA (7,000)	8,688 (4,936 + 3,752)	96.91	0.993
Zhou et al. [55]	UNet & ResNet	Private (4,856)	4,856 (2,736 + 2,120)	94.8	0.998
Present work	UNet & Xception	NIAID TB portal program dataset (1,400)	1,400 (700 + 700)	99.29	0.999

dataset, Shenzhen dataset, Montgomery dataset, Belarus, NIAID, SNA dataset, and the X-ray images dataset from the National Institute of Tuberculosis and Respiratory Diseases, New Delhi. This yielded a total of 8,688 CXR images with 4,936 TB and 3,752 normal images for their experiments. They demonstrated a classification accuracy of 96.91% with an AUC of 0.993. Zhou et al. [55] used a private dataset compiled from five different sources containing 4,856 CXR images (2,736 TB and 2,120 normal images). They used the UNet model to segment CXR images and achieved a Dice value of 0.958. After segmentation, they used the ResNet DL model to classify the segmented lung images into TB and normal classes. They achieved a classification accuracy of 94.8% with an AUC of 0.998.

The proposed method implements segmentation and classification models to detect TB in CXR images. First, the UNet segmentation model is used to segment CXR images and select the region of interest, i.e., lung sections, removing the background. Thereafter, the Xception DL classification model is used to classify the lung sections into TB and normal classes. Here, we used the NIAID TB portal program dataset from the National Institute of Allergy and Infectious Diseases, which comprises a total of 1,400 images with 700 TB infected and 700 control images. We implemented a five-fold cross-validation process for segmentation and used an 80:20 split pattern for the training/test ratio. We achieved a classification accuracy of 99.29% with an AUC of 0.999. Our model outperformed all comparable methods in terms of performance and accuracy. We also used a substantial number of CXR images, ensuring that our system is highly reliable, precise, and robust for practical applications as a support system for radiologists or as a possible setup in resource-deficient provinces.

Our system is highly cost-effective compared to with traditional diagnostic methods. Biochemical tests and analyses often require expensive chemicals and machines and highly trained professionals, including radiologists. Our AI system requires only X-ray images and a simple computer system and can be run by personnel with an average level of training. Our approach is also very fast, generating multiple results in a few seconds. By contrast, a biochemical analysis can take more than a day for a single test.

Despite several advantages, the AI system has some limitations; for instance, the variation between X-ray machines and their output quality may affect the results of any AI system. This limitation can be overcome by training on larger databases and diverse image types [56]; however, retraining on large databases will require high-performing GPUs or Supercomputer frameworks that may incur higher costs and require more time. The noise present in X-ray images may also affect the system's output. However, this could be overcome by implementing various denoising methods or color-normalization techniques [57]. A significant concern for AI-based detection systems is institutional approval for medical use. Even after extensive development and testing, AI-based detection always requires approval for use as a primary diagnostic method. However, AI systems may readily be used as a supporting method [58].

TB is among the most common infectious diseases, causing extensive deaths each year. It frequently goes undiagnosed, especially in rural regions, owing to limited resources and a lack of radiologists. The use of AI-embedded CAD techniques has the potential to provide accurate diagnoses and reduce the workload of medical practitioners. In this work, we present DL-based segmentation and classification models for automated, fast, and precise detection of TB in CXR images. First, we applied the UNet segmentation model to 704 CXR images and their corresponding masks and achieved accuracy, Jaccard, Dice, and AUC values of 96.35%, 90.38%, 94.88%, and 0.99, respectively. The UNet model was further used for TB and control CXR image segmentation. Thereafter, the segmented lung images were classified into binary classes (TB and normal) using the DL Xception model. We achieved classification accuracy, precision, recall, F1-score, and AUC values of 99.29%, 99.30%, 99.29%, 99.29%, and 0.999, respectively. Our system's high accuracy and robustness establish its effectiveness and superiority for practical applications in TB diagnosis using chest radiographs, with the possibility of providing services in environments with a scarcity of radiological expertise.

In the next phase of the work, we will potentially retrain our system in an extensive data framework or on more recent, diversified datasets. Gathering more datasets from various machines is desirable as it enables the system's performance to be evaluated on a broader range of datasets. Pruning and stochastic imaging are examples of superior training meth-



ods that could improve the system's performance and reduce storage requirements [59]. We will also use workstations and more sophisticated GPUs to enhance the output and shorten the learning time. Deep architectural feature classification could also be attempted using more contemporary techniques such as tree seed algorithm-optimized artificial neural networks [60], or a hybrid pipeline that combines AlexNet and BiLSTM could be employed as part of a bidirectional long short-term memories (BiLSTM) layer.

## Conflicts of interest statement

The authors declare that there are no conflicts of interest.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Author contributions

All authors have reviewed and agreed to the published version of the manuscript. Writing the first draft of the manuscript: Vinayak Sharma. Conceptualization, literature retrieval, data collection: Vinayak Sharma, Nillmani. Manuscript revision: Vinayak Sharma, Nillmani, Sachin Kumar Gupta, Kaushal Kumar Shukla. Supervision: Sachin Kumar Gupta, Kaushal Kumar Shukla.

## References

- [1] World Health Organization. Tuberculosis. Available from <https://www.who.int/news-room/fact-sheets/detail/tuberculosis>. (Access on 17 March, 2023).
- [2] Estill J, Xun Y, Wu S, et al. Tuberculosis screening among children and adolescents in China: insights from a mathematical model. *Intell Med* 2022. doi:10.1016/j.imed.2022.09.001.
- [3] Yilancioglu K, Cokol M. Design of high-order antibiotic combinations against M. tuberculosis by ranking and exclusion. *Sci Rep* 2019;9(1):11876. doi:10.1038/s41598-019-48410-y.
- [4] Ito K. Limits of chest X-ray investigation in the diagnosis of recurrent pulmonary tuberculosis. *Kekkaku* 2005;80(7):521–6.
- [5] Hesamian MH, Jia W, He X, et al. Deep learning techniques for medical image segmentation: achievements and challenges. *J Digit Imaging* 2019;32(4):582–96. doi:10.1007/s10278-019-00227-x.
- [6] Pal D, Reddy PB, Roy S. Attention UW-Net: a fully connected model for automatic segmentation and annotation of chest X-ray. *Comput Biol Med* 2022;150:106083. doi:10.1016/j.combiomed.2022.106083.
- [7] Altaf F, Islam SMS, Akhtar N, et al. Going deep in medical image analysis: concepts, methods, challenges, and future directions. *IEEE Access* 2019;7:99540–72. doi:10.1109/access.2019.2929365.
- [8] Esteve A, Chou K, Yeung S, et al. Deep learning-enabled medical computer vision. *NPJ Digit Med* 2021;4(1):5. doi:10.1038/s41746-020-00376-2.
- [9] Pandey A, Singh SK, Udmale SS, et al. Epileptic seizure classification using battle royale search and rescue optimization-based deep LSTM. *IEEE J Biomed Health Inform* 2022;26(11):5494–505. doi:10.1109/jbhi.2022.3203454.
- [10] Tripathi S, Verma A, Sharma N. Automatic segmentation of brain tumour in MR images using an enhanced deep learning approach. *Comput Methods Biomech Biomed Engin* 2020;9(2):121–30. doi:10.1080/21681163.2020.1818628.
- [11] Tripathi S, Sharma N. Computer-based segmentation of cancerous tissues in biomedical images using enhanced deep learning model. *IETE Tech Rev* 2021;39(5):1208–22. doi:10.1080/02564602.2021.1994044.
- [12] Tandel GS, Tiwari A, Kakde OG. Performance optimisation of deep learning models using majority voting algorithm for brain tumour classification. *Comput Biol Med* 2021;135:104564. doi:10.1016/j.combiomed.2021.104564.
- [13] Maji D, Sigedra P, Singh M. Attention Res-UNet with guided decoder for semantic segmentation of brain tumors. *Biomed Signal Process Control* 2022;71:103077. doi:10.1016/j.bspc.2021.103077.
- [14] Jain PK, Dubey A, Saba L, et al. Attention-based UNet deep learning model for plaque segmentation in carotid ultrasound for stroke risk stratification: an artificial intelligence paradigm. *J Cardiovasc Dev Dis* 2022;9(10):326. doi:10.3390/jcdd9100326.
- [15] Jain PK, Sharma N, Giannopoulos AA, et al. Hybrid deep learning segmentation models for atherosclerotic plaque in internal carotid artery B-mode ultrasound. *Comput Biol Med* 2021;136:104721. doi:10.1016/j.combiomed.2021.104721.
- [16] Jain PK, Sharma N, Kalra MK, et al. Far wall plaque segmentation and area measurement in common and internal carotid artery ultrasound using U-series architectures: an unseen artificial intelligence paradigm for stroke risk assessment. *Comput Biol Med* 2022;149:106017. doi:10.1016/j.combiomed.2022.106017.
- [17] Jain PK, Sharma N, Saba L, et al. Unseen artificial intelligence-deep learning paradigm for segmentation of low atherosclerotic plaque in carotid ultrasound: a Multicenter Cardiovascular Study. *Diagnostics* 2021;11(12):2257. doi:10.3390/diagnostics11122257.
- [18] Singla M, Ghosh D, Shukla KK. pin<sup>+</sup>-TSVM: a robust transductive support vector machine and its application to the detection of COVID-19 infected patients. *Neural Process Lett* 2021;53(6):3981–4010. doi:10.1007/s11063-021-10578-8.
- [19] Nillmani, Jain PK, Sharma N, et al. Four types of multiclass frameworks for pneumonia classification and its validation in X-ray scans using seven types of deep learning artificial intelligence models. *Diagnostics* 2022;12(3):652. doi:10.3390/diagnostics12030652.
- [20] Nillmani, Sharma N, Saba L, et al. Segmentation-based classification deep learning model embedded with explainable AI for COVID-19 detection in chest X-ray scans. *Diagnostics* 2022;12(9):2132. doi:10.3390/diagnostics12092132.
- [21] Bhattacharyya A, Bhaik D, Kumar S, et al. A deep learning based approach for automatic detection of COVID-19 cases using chest X-ray images: a comprehensive study. *Biomed Signal Process Control* 2022;71:103182. doi:10.1016/j.bspc.2021.103182.
- [22] Nayak SR, Nayak DR, Sinha U, et al. Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: a comprehensive study. *Biomed Signal Process Control* 2021;64:102365. doi:10.1016/j.bspc.2020.102365.
- [23] Gu Y, Chi J, Liu J, et al. A survey of computer-aided diagnosis of lung nodules from CT scans using deep learning. *Comput Biol Med* 2021;137:104806. doi:10.1016/j.combiomed.2021.104806.
- [24] Halder A, Dey D, Sadhu AK. Lung nodule detection from feature engineering to deep learning in thoracic CT images: a comprehensive review. *J Digit Imaging* 2020;33(3):655–77. doi:10.1007/s10278-020-00320-6.
- [25] Debelee TG, Schwenker F, Ibenhal A, et al. Survey of deep learning in breast cancer image analysis. *Evol Syst* 2019;11(1):143–63. doi:10.1007/s12530-019-09297-2.
- [26] Bai J, Posner R, Wang T, et al. Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: a review. *Med Image Anal* 2021;71:102049. doi:10.1016/j.media.2021.102049.
- [27] Roy S, Whitehead TD, Li S, et al. Co-clinical FDG-PET radiomic signature in predicting response to neoadjuvant chemotherapy in triple-negative breast cancer. *Eur J Nucl Med Mol Imaging* 2021;49(2):550–62. doi:10.1007/s00259-021-05489-8.
- [28] Villarreal M, Chaichulee S, Jorge J, et al. Non-contact physiological monitoring of preterm infants in the Neonatal Intensive Care Unit. *NPJ Digit Med* 2019;2:128. doi:10.1038/s41746-019-0199-5.
- [29] Liu F, Tang J, Ma J, et al. The application of artificial intelligence to chest medical image analysis. *Intell Med* 2021;1(3):104–17. doi:10.1016/j.imed.2021.06.004.
- [30] Pillay T, Andronikou S, Zar HJ. Chest imaging in paediatric pulmonary TB. *Paediatr Respir Rev* 2020;36:65–72. doi:10.1016/j.prrv.2020.10.002.
- [31] Singh M, Venkatesh V, Verma A, et al. Segmentation of MRI data using multi-objective antlion based improved fuzzy c-means. *Biocybern Biomed Eng* 2020;40(3):1250–66. doi:10.1016/j.bbe.2020.07.001.
- [32] Singh M, Verma A, Sharma N. Multi-objective noise estimator for the applications of de-noising and segmentation of MRI data. *Biomed Signal Process Control* 2018;46:249–59. doi:10.1016/j.bspc.2018.07.012.
- [33] Vo DM, Nguyen NQ, Lee SW. Classification of breast cancer histology images using incremental boosting convolution networks. *Inf Sci* 2019;482:123–38. doi:10.1016/j.ins.2018.12.089.
- [34] Shi F, Wang J, Shi J, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. *IEEE Rev Biomed Eng* 2020;14:4–15.
- [35] Santosh K, Allu S, Rajaraman S, et al. Advances in deep learning for tuberculosis screening using chest X-rays: the last 5 years review. *J Med Syst* 2022;46(11). doi:10.1007/s10916-022-01870-8.
- [36] Singh M, Pujar GV, Kumar SA, et al. Evolution of machine learning in tuberculosis diagnosis: a review of deep learning-based medical applications. *Electronics* 2022;11(17):2634. doi:10.3390/electronics11172634.
- [37] Pandey N. Chest X-ray masks and labels. Available from <https://www.kaggle.com/datasets/nikhilpandey360/chest-xray-masks-and-labels>. (Access on 10 January, 2023).
- [38] Rahman T. Tuberculosis (TB) chest X-ray database. Available from <https://www.kaggle.com/datasets/tawfufurrahman/tuberculosis-tb-chest-xray-dataset>.
- [39] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *Lecture Notes in Computer Science*; 2015. p. 234–41. doi:10.1007/978-3-319-24574-4\_28.
- [40] Chollet F. Xception: deep learning with depthwise separable convolutions. In: *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA; 2017. doi:10.1109/CVPR.2017.195.
- [41] Selvaraju RR, Cogswell M, Das A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization. In: *Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)*; 2017. doi:10.1109/ICCV.2017.74.
- [42] Chen X, Wang X, Zhang K, et al. Recent advances and clinical applications of deep learning in medical image analysis. *Med Image Anal* 2022;79:102444. doi:10.1016/j.media.2022.102444.
- [43] Chandra TB, Verma K, Singh BK, et al. Automatic detection of tuberculosis related abnormalities in chest X-ray images using hierarchical feature extraction scheme. *Expert Syst Appl* 2020;158:113514. doi:10.1016/j.eswa.2020.113514.
- [44] Chartrand G, Cheng PM, Vorontsov E, et al. Deep learning: a primer for radiologists. *Radiographics* 2017;37(7):2113–31. doi:10.1148/rg.2017170077.
- [45] Hwang EJ, Park CM. Clinical implementation of deep learning in thoracic radiology: potential applications and challenges. *Korean J Radiol* 2020;21(5):511–25. doi:10.3348/kjr.2019.0821.
- [46] Ahsan M, Gomes R, Denton A. Application of a convolutional neural network using transfer learning for tuberculosis detection. In: *Proceedings of 2019 IEEE International Conference on Electro Information Technology (EIT)*; 2019. doi:10.1109/eit.2019.8833768.

- [47] Xie Y, Wu Z, Han X, et al. Computer-aided system for the detection of multicategory pulmonary tuberculosis in radiographs. *J Healthc Eng* 2020;2020:9205082. doi:[10.1155/2020/9205082](https://doi.org/10.1155/2020/9205082).
- [48] Sahlol AT, Abd Elaziz M, Tariq Jamal A, et al. A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features. *Symmetry* 2020;12(7):1146. doi:[10.3390/sym12071146](https://doi.org/10.3390/sym12071146).
- [49] Rahman T, Khandakar A, Kadir MA, et al. Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. *IEEE Access* 2020;8:191586–601. doi:[10.1109/access.2020.3031384](https://doi.org/10.1109/access.2020.3031384).
- [50] Guo R, Passi K, Jain CK. Tuberculosis diagnostics and localization in chest X-rays via deep learning models. *Front Artif Intell* 2020;3:583427. doi:[10.3389/frai.2020.583427](https://doi.org/10.3389/frai.2020.583427).
- [51] Ul Abideen Z, Ghafoor M, Munir K, et al. Uncertainty assisted robust tuberculosis identification with Bayesian convolutional neural networks. *IEEE Access* 2020;8:22812–25. doi:[10.1109/ACCESS.2020.2970023](https://doi.org/10.1109/ACCESS.2020.2970023).
- [52] Ayaz M, Shaukat F, Raja G. Ensemble learning based automatic detection of tuberculosis in chest X-ray images using hybrid feature descriptors. *Phys Eng Sci Med* 2021;44(1):183–94. doi:[10.1007/s13246-020-00966-0](https://doi.org/10.1007/s13246-020-00966-0).
- [53] Rajaraman S, Zamzmi G, Folio L, et al. Chest X-ray bone suppression for improving classification of tuberculosis-consistent findings. *Diagnostics* 2021;11(5):840. doi:[10.3390/diagnostics11050840](https://doi.org/10.3390/diagnostics11050840).
- [54] Acharya V, Dhiman G, Prakasha K, et al. AI-assisted tuberculosis detection and classification from chest X-rays using a deep learning normalization-free network model. *Comput Intel Neurosc* 2022;2022:1–19. doi:[10.1155/2022/2399428](https://doi.org/10.1155/2022/2399428).
- [55] Zhou W, Cheng G, Zhang Z, et al. Deep learning-based pulmonary tuberculosis automated detection on chest radiography: large-scale independent testing. *Quant Imaging Med Surg* 2022;12(4):2344–55. doi:[10.21037/qims-21-676](https://doi.org/10.21037/qims-21-676).
- [56] Kazemzadeh S, Yu J, Jamshy S, et al. Deep learning detection of active pulmonary tuberculosis at chest radiography matched the clinical performance of radiologists. *Radiology* 2023;306(1):124–37. doi:[10.1148/radiol.212213](https://doi.org/10.1148/radiol.212213).
- [57] Thanh D, Surya P, Hieu LM. A review on CT and X-ray images denoising methods. *Informatica* 2019;43(2):151–9. doi:[10.31449/inf.v43i2.2179](https://doi.org/10.31449/inf.v43i2.2179).
- [58] Kempt H, Nagel SK. Responsibility, second opinions and peer-disagreement: ethical and epistemological challenges of using AI in clinical diagnostic contexts. *J Med Ethics* 2021;48(4):222–9. doi:[10.1136/medethics-2021-107440](https://doi.org/10.1136/medethics-2021-107440).
- [59] Agarwal M, Agarwal S, Saba L, et al. Eight pruning deep learning models for low storage and high-speed COVID-19 computed tomography lung segmentation and heatmap-based lesion localization: a multicenter study using COVLIAS 2.0. *Comput Biol Med* 2022;146:105571. doi:[10.1016/j.combiomed.2022.105571](https://doi.org/10.1016/j.combiomed.2022.105571).
- [60] Cinar AC. Training feed-forward multi-layer perceptron artificial neural networks with a tree-seed algorithm. *Arab J Sci Eng* 2020;45(12):10915–38. doi:[10.1007/s13369-020](https://doi.org/10.1007/s13369-020).