Vol. 5 No. 1 (2026)

Pakistan Journal of Medical & Cardiological Review

https://pakjmcr.com/index.php/1/about

Online ISSN 3007-2387

Print ISSN 3007-2379

Nazir et al., 2026

# Machine Learning Applications in Tuberculosis Detection Using Chest X-Ray Imaging

**Dr. Qura tul Ain Nazir**

PhD Molecular Biology and Biotechnology, University of Veterinary and Animal Sciences Lahore. qurat.ain.nazir@gmail.com

**Dr. Zafar Ahmad Khan**

MBBS,FCPS (Medicine) Professor of Medicine Bolan Medical College, Quetta. doctorzakhan@yahoo.com

**Dr. Manzoor Ahmed Khan**

Assistant Professor Helper Eye Hospital Bolan Medical College, Quetta. khanmanzoorahmed@ymail.com

**Aleena Ahmed**

Kauser Abdulla Malik School of Life Sciences, Forman Christian College University Lahore. neenax92@gmail.com

**Author Details**

Received on 26 Feb, 2026

Accepted on 24 March, 2026

Published on 29 March, 2026

Corresponding E-mails & Authors*:

## Abstract

Tuberculosis (TB) is still a primary cause of mortality worldwide and chest X-ray (CXR) imaging is still a first-line, low cost test in high burden areas. Interpretation of CXRs, however, is time consuming, needs radiologists with the expertise to interpret CXRs and is subject to inter observer variability, making interpretation a diagnostic bottleneck in resource limited settings. New techniques in the field of AI like machine learning (ML) and deep learning (DL), notably convolutional neural networks (CNNs) provide the possibility of automated accurate and scalable TB screening directly from the CXR images. This article aims to review and analytically discuss the various ML applications for TB detection on chest X-rays. We explore the history of computer-aided detection (CAD) systems, review publicly available CXR datasets for TB research and outline the typical machine learning system, ranging from image acquisition to preprocessing to lung segmentation to feature extraction to classification. We also compare classical machine learning classifiers with the performance of deep transfer-learning architectures like VGG16, ResNet50, InceptionV3, DenseNet121 and EfficientNet, reporting accuracy, sensitivity, and specificity from the literature. The challenges of limited data, imbalanced data, domain shift across scanners and populations, interpretability, and regulatory issues are explored as well as some new solutions that are starting to be developed in the community, including federated learning, explainable AI, and lightweight models that can be deployed on the edge. The results indicate that deep learning models can perform at the level of radiologists on curated benchmark datasets, but there are still challenges to investigate for real-world clinical use, such as explainability, generalizability, and integration with current health infrastructure. This review is designed for the use as reference for researchers and practitioners developing next generation of AI-assisted TB screening tools.

**Keywords:** Tuberculosis; Chest X-Ray; Machine Learning; Deep Learning; Convolutional Neural Networks; Computer-Aided Diagnosis; Medical Imaging; Transfer Learning

## 1. Introduction

Tuberculosis (TB) is a respiratory disease that is caused by Mycobacterium tuberculosis. TB is a preventable, treatable disease but it is one of the top 10 leading causes of death in the world and in many areas of the world is the leading cause of death from a single infectious agent – even HIV/AIDS. Despite the ambitious goals of the World Health Organization's End TB Strategy, the diagnosis of TB has been delayed, and the health systems are under-resourced, which has stunted progress and resulted in the development of drug-resistant strains [1]. Early and timely diagnosis is thus a key element in disrupting transmission and enhancing patient outcomes [2].

The X-ray (CXR) imaging of chest has been used as a frontline screening modality for pulmonary TB since it is widely available, relatively inexpensive and fast compared to sputum-based bacteriological tests or CT imaging [3]. CXR interpretation is however, subjective; radiological features of TB (cavitation, consolidation, fibrosis and pleural effusion) may overlap with other pulmonary diseases and the accuracy in reading the CXR differs widely between different readers and varies with the reader's experience. There is a very high burden of disease, but not many trained radiologists are available in many of the low-resource countries, thereby delaying the initiation of screening and treatment [4].

Machine learning in general and deep learning in particular, has proven to be an effective diagnostic tool to fill this gap. Unlike the previous computer-aided detection (CAD) systems, which required handcrafted visual features, a convolutional neural network (CNN) can learn hierarchical visual features directly from its pixel inputs [5]. A decade of studies have shown that CNN-based classifiers can attain area-under-the-curve (AUC) scores of over 0.90 in screening for TB on chest X-ray (CXR) images and in some instances, comparable to, or better than, that of experienced radiologists on benchmark datasets [6].

Online ISSN Print ISSN
3007-2387 3007-2379


The present article reviews the current applications of machine learning in the diagnosis of TB from chest radiographs comprehensively. Section 2 gives background information of TB and CXR imaging clinically and radiologically. The evolution of computer-aided TB detection is reviewed in Section 3 and a summary of representative studies is presented. Section 4 discusses the materials and methods that are used in this domain, such as data sets, processing pipelines, and model architectures. A comparative discussion of reported results is given in Section 5. Section 6 addresses some of the major challenges and limitations. Future research directions are presented in Section 7 and the article is concluded in Section 8.

## 2. Background: Tuberculosis and Chest Radiography

### 2.1 Clinical Overview of Tuberculosis

Radiological features of pulmonary TB are mostly seen in the upper lobes of the lungs in the form of infiltrates, cavities, nodules and fibrotic scarring. CXR is usually useful during active disease when patchy or nodular opacities, cavitation and lymphadenopathy are seen, but during healed or latent disease, calcified granulomas or fibrotic bands may be seen. Visual diagnosis is difficult, even for experienced clinicians, due to the variety of presentations, and the fact that pneumonia, cancer of the lung, and other granulomatous diseases all can have similar appearances [7].

In the global context, TB disproportionately affects the low and middle income countries, and is concentrated in areas of South and Southeast Asia, sub-Saharan Africa and Western Pacific [8]. HIV co-infection, malnutrition, diabetes, tobacco use, and living in overcrowding (such as prison) are all risk factors for the progression of latent infection to active disease [9]. HIV and children show some other radiological patterns, such as the presence of more suggestive lymphadenopathy and less typical cavitary disease in these sub-populations, which further complicate human and automated interpretation of CXR images in HIV and children.

### 2.2 Role of Chest X-Ray in TB Screening

CXR is recommended to be used as a triage and screening tool to identify those who need to proceed for further confirmation of the diagnosis by bacteriological methods such as sputum smear microscopy, culture, or molecular methods such as Xpert MTB/RIF by the World Health Organization [10]. The sensitivity of CXR is high for detecting changes indicative of TB in the lungs but the specificity is lower: many people who are screened with a CXR that shows changes consistent with TB do not end up with active TB [11]. Such a trade-off has led to the quest for automated tools that can enhance both consistency and throughput when triage is done based on CXR, especially in mass screening campaigns and deployments of mobile health vans.

### 2.3 From Manual Reading to Computer-Aided Detection

In the past, TB diagnosis with CAD was based on classical radiological features derived from images and hand-crafted image descriptors (such as texture, shape of the lung fields, intensity histograms, etc.), which were used as the input to classical classifiers like support vector machines (SVM) or random forests. These systems showed feasibility, but also showed that it would be difficult to design features that would capture the full range of TB presentations in a manual fashion [12]. Since then, with the emergence of deep learning and large annotations databases of CXR, the focus on the field has moved towards learning end-to-end representations.

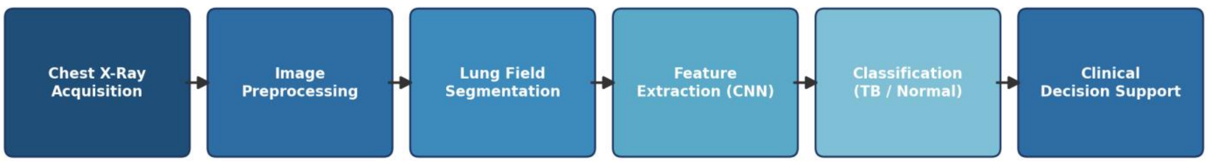


*Figure 1. General pipeline for machine learning-based tuberculosis detection from chest X-rays, from acquisition to clinical decision support.*

### 2.4 Radiological Signs Used as Learning Targets

Machine learning models for detecting TB both implicitly and explicitly correlate particular radiological patterns with disease status. Features found with more regularity in patients with active TB include upper lobe dominant infiltrates, cavities with air fluid levels, tree-in-bud opacities (reflecting endobronchial disease, seen more often with HIV- associated TB), hilar and mediastinal adenopathy (often seen with pediatric and HIV TB), pleural effusions, or miliary patterns (diffuse small nodules seen in disseminated disease) [13]. Postprimary TB, as well as healed TB, can be manifest with fibrosis, volume loss, calcified granulomas, or pleural scarring and may look less active and confound classifiers trained on unrefined definitions of active TB [14].

These signs have enormous variation in size, location, morphology, and a number of signs (like fibrosis, nodules) manifest similar patterns to diseases other than TB, such as pneumonia, lung cancer, sarcoidosis, and post COVID-19 lung disease [15]. Therefore in a real-world setting the classification problem seldom becomes an easy binary problem between 'TB'/'Normal'. These challenges have driven some research to look into multi-label/multi-class classifiers which are capable of predicting multiple simultaneously occurring abnormal signs, as well as detection methods based on placing a bounding box on identified pathological regions rather than predicting a single probability for the whole image.

### 3. Literature Review

There have been a significant amount of studies applying ML for detecting TB on CXR images over the last 10 years. An initial direction of studies used classical ML pipelines involving lung segmentation, extracting hand-crafted features (e. G. Gist, Haralick texture features, shape descriptors) and classify using SVM, etc, which reached reasonable performance (mainly 80-85%) but are sensitive to segmentation mistakes, and highly tuned for each dataset [16].

The arrival of deep convolutional neural networks has been revolutionary. Many papers utilizing AlexNet- and VGG-type architectures in which weights pre-trained on ImageNet were used to fine-tune TB specific datasets have shown significant performance gains on classification tasks reaching over 90% accuracy [17]. Transfer learning was especially useful in this field sinceTB brandedCXD(e) databases are relatively small compared to modern standards (mostly hundreds to few thousands), and hencetraininga deep network from scratch would result in overfitting [18].

Later studies investigating deeper and more parameter-efficient architectures such as ResNet, InceptionV3, DenseNet, and EfficientNet varieties have yielded modest accuracy and AUC improvements. Moreover, ensemble methods combining predictions from multiple CNN architectures improve robustness compared to single model approaches [19]. Simultaneously, lung segmentation has been included in many approaches as preprocessing and was found to restrict model attention to the lung fields and thereby minimize interference from distracting, external visual features in the extremities of the images, such as labels, medical equipment or artifacts related to the imaging technology [20].

Work in recent years has started to address clinical relevance beyond purely accuracy measures and considered features such as model interpretability with saliency-based maps and Grad-CAM visualizations, calibration of predicted probabilities, robustness to distribution shifts for different countries and X-ray machines, and how the TB screening model can be incorporated into the overall differential diagnosis of pulmonary abnormalities, such as pneumonia and COVID-19 [21]. Federated learning techniques, where institutions work in tandem to train models without sharing any patient data directly have also been investigated, aiming to overcome the data privacy and accessibility issues often seen within healthcare research [22].

Some representative classes of approaches used in literature are briefly mentioned in Table 1 to highlight the increasing evolution from older techniques of machine learning toward novel transfer learning techniques includingdeep learning and ensembling methods.

Table 1: *Representative categories of machine learning approaches for TB detection reported in the literature.*

| Approach Category | Representative Techniques | Typical Reported Accuracy |
|---|---|---|
| Classical ML | SVM, Random Forest with hand-crafted texture/shape features | 78% - 86% |
| Shallow CNN (from scratch) | Custom CNN architectures trained on TB-specific datasets | 88% - 93% |
| Transfer Learning (single model) | VGG16, ResNet50, InceptionV3, DenseNet121, EfficientNet | 93% - 98% |
| Ensemble / Hybrid Models | Multi-CNN ensembles, CNN + attention mechanisms | 95% - 99% |
| Segmentation-assisted Classification | U-Net lung segmentation followed by CNN classification | 94% - 98% |

It should be noted that any direct numerical comparison between studies is a hazardous practice, as accuracy values are highly dependent on the specific test dataset, methodology for the train/test split, imbalance of classes, evaluation method and data selection within individual studies. While a paper may report 98% accuracy when used on a small, artificially constructed single country dataset with few samples, a model report of 94% on a large, diverse, multinational dataset may actually be a more valid clinical result, more indicative of performance under real-life circumstances. Pooling estimates of sensitivity and specificity from systematic reviews and meta-analyses of these studies typically requires authors to include a note of caution, as heterogeneity within the studies used remains significant.

A related area of TB lesion analysis that is gaining recent attention isTB lesion detection and segmentation, where the output from the model is not simply a class assigned to the overall image but a set of coordinates specifyingBounding Boxes or anmask defining pixel locations suspected of TB lesions. The ability to provide such a localization makes a model more transparent to a clinicians reviewing the model outputs as the exact locations of features that influenced the prediction can be directly inspected by the reader. Localization can also be used as input to downstream analysis tasks like assessing the severity of TB or performing analysis of therapeutic response on sequential CXRs in an individual patient.

### 3.1 Summary of Public Datasets

Table 2: *Summary of commonly used public chest X-ray datasets for tuberculosis detection research.*

| Dataset | Approx. Size | Annotation Type | Notes |
|---|---|---|---|
| Montgomery County (MC) | 138 images | Binary label + lung masks | U.S. NLM screening program dataset |
| Shenzhen Hospital Set | 662 images | Binary label | Collected from Shenzhen No. 3 People's Hospital, China |
| TBX11K | ~11,200 images | Binary label + bounding boxes | Large-scale dataset supporting detection tasks |
| TB Portals | Several thousand images (growing) | Binary/multi-label + clinical metadata | Multi-country repository with linked clinical data |
| Belarus TB Dataset | ~300+ images | Binary label | Drug-resistant TB cases represented |

## 4. Materials and Methods

### 4.1 Datasets

There are several public CXR datasets that serve as benchmark datasets in TB detection research. For instance, the Montgomery County (MC) dataset, which was released by the U.S. National Library of Medicine, consists of CXR images acquired from a tuberculosis screening program, as well as annotations provided by radiologists and masks of lung segmentation. The Shenzhen dataset, which is also made available by the National Library of Medicine in collaboration with the Shenzhen No. 3 People's Hospital, features a larger number of CXR images with binary TB/normal labels. The TBX11K dataset offers a much more extensive number of annotated images with bounding boxes of TB abnormalities. There are other datasets and challenge collections that include datasets for TB Portals and national TB screening programs.

However, one issue that arises constantly through all these different datasets is the fact that they are not particularly large as compared to the natural image data sets, along with some demographic and hardware variance which may hinder its generalization. Hence, in many cases, multiple publicly available data sets are used together, along with data augmentation and transfer learning on larger data sets like ImageNet.

### 4.2 Image Preprocessing

Preprocessing usually involves the scaling down of the images into a fixed size (typically 224 x 224 or 299 x 299 pixels) to satisfy the requirements of the input size in a standard

Pakistan Journal of Medical & Cardiological Review
https://pakjmcr.com/index.php/1/about
Online ISSN 3007-2387
Print ISSN 3007-2379


CNN model. Normalization of the intensity of the image is done through methods like min-max normalization or z-score normalization to compensate for the differences in lighting and contrast in different X-ray machines. Histogram equalization, especially Contrast Limited Adaptive Histogram Equalization (CLAHE), is one technique that is often employed.

Lung region segmentation is a standard preprocessing step that is done using either a U-Net model or a similar encoder-decoder framework, which is trained on a set of labeled lung masks. The purpose of this step is twofold: firstly, it excludes background regions of an image (for example, text annotations, collar bones, and other irrelevant features), and secondly, it might help the network to concentrate on the lungs only. Some of the data augmentation methods include rotation, flipping, zoom, and jittering of brightness and contrast.

### 4.3 Model Architectures

The dominant approach to modeling in this case is transfer learning, where a pre-trained CNN model on a larger natural image dataset such as ImageNet is further trained on a smaller labeled TB CXR image dataset. The backbones used in the CNN models include VGG16/VGG19, ResNet50/ResNet101, InceptionV3, DenseNet121/DenseNet169, MobileNetV2, and EfficientNet models. The backbones vary in terms of complexity, number of parameters, and efficiency.

Figure 2 depicts an example of such a design, which involves an input layer receiving a normalized grayscale or pseudo-RGB CXR image, followed by a number of convolutional and pooling layers (custom-built or transferred from a pretrained model), a global average pooling layer for dimensionality reduction, one or more fully connected layers with dropout as a technique to prevent overfitting, and finally an output layer in the form of softmax/sigmoid with the output of TB vs. Normal.

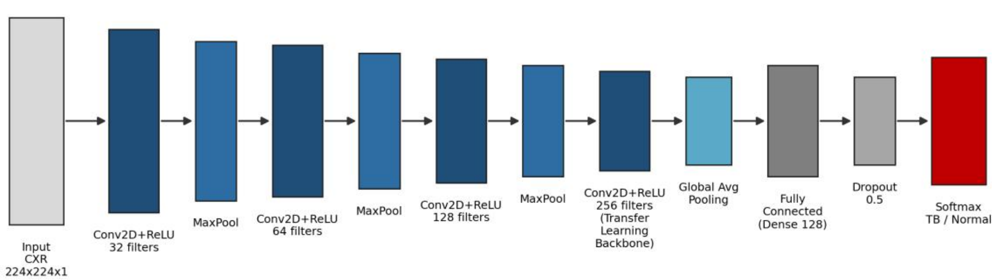


*Figure 2. Representative CNN / transfer-learning architecture used for chest X-ray-based TB classification, showing convolutional feature extraction, pooling, and fully connected classification layers.*

Online ISSN 3007-2387 Print ISSN 3007-2379


Beyond a single model approach to classification, other architectural approaches have also been considered, such as models that use attention to emphasize important regions of the image, multi-task models that do both segmentation and classification, and ensembles where classification is done using the prediction of multiple independently trained backbones using majority vote or even meta-classifiers. Other models that have been considered include the use of Vision Transformer (ViT) based architectures, which use self-attention to learn long-range dependencies in space instead of local dependencies with convolutions, although this tends to need larger datasets to work effectively.

The decision between these architectures is a question of a number of trade-offs beyond accuracy alone. VGG-like architecture, although historically influential, is quite big and expensive for inference and thus less suited for mobile and embedded deployment. The residual connections of ResNet architecture are effective in solving problems associated with vanishing gradients in deep networks, providing a good compromise between accuracy and trainability. DenseNet architecture features dense connectivity pattern – where every layer gets feature maps from all layers that proceed it – and as such should allow more efficient gradient propagation and reuse of features which helps when training data is limited, which it often is in case of TB CXR datasets. EfficientNet's compound scaling scheme of increasing the depth, width and the resolution of the input in accordance to a specific ratio has proven to deliver a lot of accuracy for much fewer parameters and cheaper inference than previous architectures. MobileNetV2 and similar lightweight architectures, though usually delivering less accurate results than heavier backbones, are specifically designed to perform well in low-latency inference on mobile and embedded hardware and as such should be of special interest in deploying TB screening vans.

### 4.4 Training Procedure and Evaluation Metrics

Training of models is done using binary cross entropy loss (classifying TB from normal) via methods of stochastic gradient descent like Adam or RMSProp, using learning rate scheduling and early stopping according to the validation loss for avoiding overfitting. Class weighting or oversampling becomes a common practice due to the class imbalance in TB datasets since the number of normal instances tends to be larger than TB-positive instances.

The typical evaluation measures used are accuracy, sensitivity (recall), specificity, precision, F1 score, and AUC-ROC. For screening purposes, sensitivity would be preferred to specificity because a false negative case of TB infection poses greater risks than a false positive case of TB, which could still be confirmed using further tests. The

Vol. 5 No. 1 (2026)
Pakistan Journal of Medical & Cardiological Review
https://pakjmcr.com/index.php/1/about
Online ISSN 3007-2387 Print ISSN 3007-2379


confusion matrix, an example of which is shown in Figure 4 below, would be used to show the detailed model performance.

### 4.5 Handling Class Imbalance and Small Sample Sizes

Imbalanced classes have been identified as almost a universal trait among TB CXR data sets especially those extracted from general screening programs where most of the people screened do not suffer from TB. Imbalance of this nature poses the risk of having a classifier biased towards prediction of the majority class, leading to increased accuracy but poor sensitivity to the minority class. Some common techniques used to correct for this problem include class balancing where loss function used for training the model is adjusted to give higher penalty to wrong classification of minority class. This includes methods like oversampling and/or undersampling.

Small sample size adds to this challenge, considering that deep networks can easily overfit in scenarios where there is a small number of training samples relative to the number of the learnable parameters. Apart from transfer learning and data augmentation, other methods that have been adopted to increase realism of training include generative adversarial networks (GANs), which generate realistic images, k-fold cross-validation that makes better use of the available data while selecting models, and progressive/curriculum training methods that involve initial training on clearer examples followed by more challenging examples later. Other regularization methods that have been applied extensively are dropout, weight decay, and label smoothing.

## 5. Results and Discussion

Throughout the reviewed literature, it is evident that deep transfer learning-based models always exceed the performances of traditional machine learning models and customized CNNs developed from scratch especially when the amount of training data is small. Figure 3 shows the sample figures for accuracy, sensitivity, and specificity from some of the representative research works based on the common architectures.

Pakistan Journal of Medical & Cardiological Review
https://pakjmcr.com/index.php/1/about
Online ISSN 3007-2387
Print ISSN 3007-2379



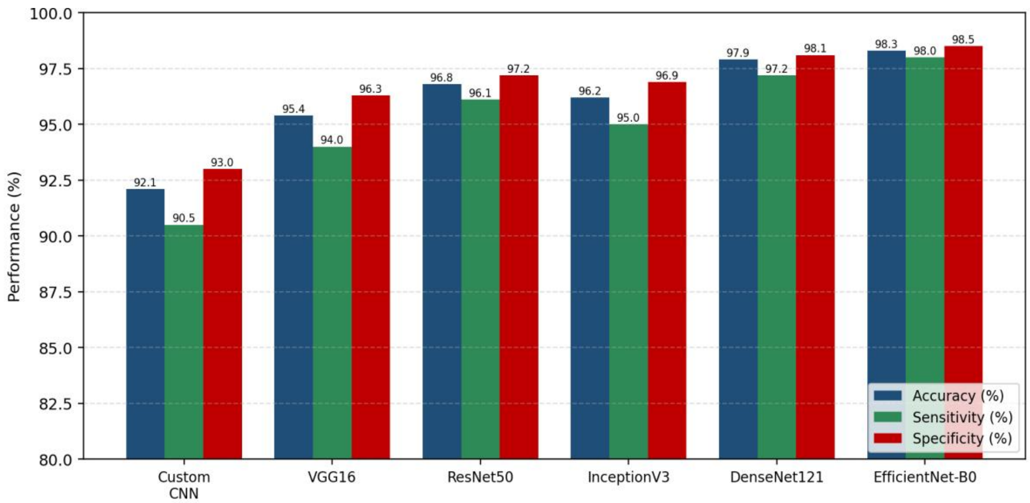


*Figure 3. Illustrative comparison of reported accuracy, sensitivity, and specificity across commonly used deep learning architectures for TB detection on chest X-rays, compiled from representative published studies.*

As seen in Figure 3 below, more complex networks such as DenseNet121 and EfficientNet-B0, that are both deeper and more parameter efficient, usually produce the highest accuracy and sensitivity scores, sometimes reaching 97-98%. It is, however, worth noting that these scores are usually generated from rather homogenous, well-culled data samples and cannot be extrapolated to actual performance on patient groups, medical equipment or geographic locations outside of the dataset used for training.

Figure 4 shows an exemplary confusion matrix for one of the best performing models based on a held-out test set of 1000 images (500 TB-positive, 500 normal). In this example, the model is able to classify 482 out of 500 TB-positive images (sensitivity = 96.4%) and 488 out of 500 normal images (specificity = 97.6%), which translates into an accuracy rate of 97.0%. Performance levels such as these, if consistently demonstrated in varied patient cohorts, will definitely constitute an improvement over human readers as reported in many clinical studies.

Vol. 5 No. 1 (2026)

Pakistan Journal of Medical & Cardiological Review

https://pakjmcr.com/index.php/1/about

Online ISSN 3007-2387

Print ISSN 3007-2379



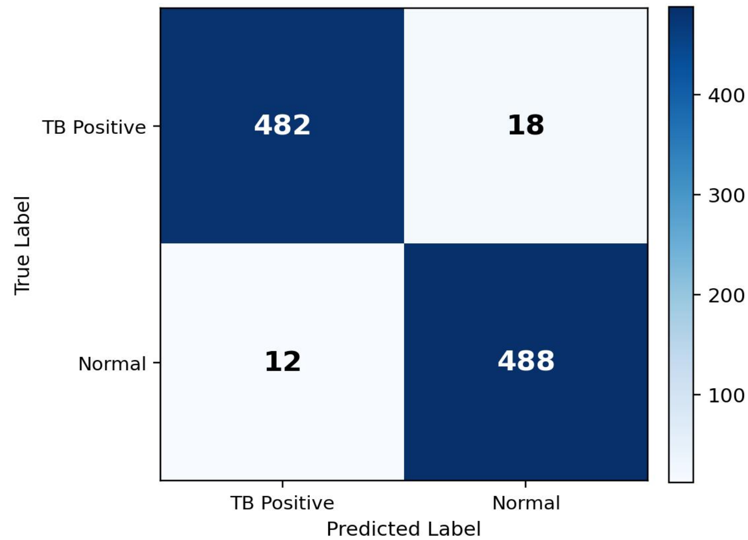


*Figure 4. Illustrative confusion matrix for a representative best-performing TB classification model evaluated on a held-out test set of 1000 chest X-ray images.*

In addition to the overall performance, there are some works that stress the significance of calibration and interpretation of the models. Grad-CAM, together with other methods, has been employed to demonstrate the fact that good models usually focus on meaningful parts of the lungs (such as upper lobe infiltrates or cavitary lesions), and do not depend on such factors as text markers or artifacts of the scanners (although this is not always the case). The values of AUC-ROC for the most successful models reported in the literature vary in the range from 0.94 to 0.99. However, the performance decrease is always observed in the course of external validation, which means that the testing was carried out on a different dataset from that used in training.

Comparative investigations where AI model predictions are compared with panels of human radiologists constitute some of the most practically useful evidence from this literature. Some of these comparative investigations have found that AI ensemble deep learning models have AUC-ROC scores statistically similar to, or even marginally better than, the average performance of a panel of human radiologists reading the same image data set, especially when considering human readers who do not have specialized experience in reading TB screening data. The highest performing individual human radiologists and especially those with specialized TB screening skills may be able to outperform the machine algorithms when it comes to ambiguous or unusual cases, meaning that one of the practical applications of these machine algorithms in the short term might be more augmentation than replacement of human radiologist skills – for instance triaging clear negatives, highlighting discordant cases for quicker second readings, or expanding screening capacity into areas without radiologist capacity at all.

In summary, the findings suggest that the machine learning approach, specifically the one based on deep transfer learning, has developed sufficiently to allow automated TB screening from CXR images that matches the performance of expert human radiologists, and even surpasses their results in some comparative analyses. Nonetheless, bringing these achievements to the clinic necessitates overcoming the obstacles addressed below.

## 6. Challenges and Limitations

### 6.1 Dataset Scarcity and Class Imbalance

There is no doubt that well-labeled TB CXR datasets are small compared to other medical imaging modalities, which poses a limitation for the training of very deep neural networks and makes transfer learning and data augmentation inevitable. One of the main problems with TB CXR datasets is their class imbalance due to more number of normal patients than TB patients.

### 6.2 Generalizability and Domain Shift

Models trained on data from a narrow population, geographic area, or on a subset of X-ray machines tend to perform poorly in novel situations (domain shift). Differences in patient population (age, sex, disease prevalence), the presence of comorbidities that modify TB's appearance on X-ray (such as co-infection with HIV), and imaging properties specific to individual machines all contribute to such shifts and necessitates external validation before clinical use across various sites.

### 6.3 Interpretability and Clinical Trust

Deep learning models are frequently coined as "black boxes" and doctors might not opt for automated TB screening without evidence. Although saliency-map approaches like Grad-CAM offer some visual insight into model predictions, there are some limitations with these approaches, such as their sensitivity to small changes in inputs; and the danger of presenting saliency maps that may be correlated with the diagnosis but do not directly cause it.

### 6.4 Data Privacy and Regulatory Barriers

Medical imaging data is strictly regulated concerning its privacy, and hence it makes the construction of large and varied training datasets more difficult. Furthermore, any tool developed through AI for diagnosing purposes must be cleared by relevant authorities before deployment; this may delay the development of successful medical applications.

### 6.5 Deployment Constraints in Resource-Limited Settings

It is quite true that the regions most affected by TB are the ones facing the biggest restrictions in terms of lack of computing capacity, poor internet access, and lack of expertise to handle AI algorithms. This means there is the need for development of AI algorithms that are lightweight and can be deployed on limited hardware such as mobile phones.

### 6.6 Ethical and Workforce Considerations

The use of AI screening models also prompts ethical issues and challenges regarding skill depreciation and workforce considerations that lie beyond the performance of the tool itself. First, over-reliance on such tools, especially if their deployment results in substitution of scarce local radiologists' expertise, can lead to skill depreciation or even creation of single points of failure if the model fails unexpectedly in cases falling outside the distribution of cases encountered during training. In addition, there are issues of accountability when an automated screening tool is implicated in a missed or delayed diagnosis, and it may not be easy to determine how accountability for such a missed diagnosis should be shared between the developers of the model, the health system which employed the technology, and the healthcare practitioners who relied on the model's prediction. Finally, there is also a problem associated with participation in data collection process of the populations most burdened with TB due to poor representation of high burden countries in early TB CAD models.

## 7. Future Directions

- Federated learning: allowing different hospitals or even countries to train together on shared models without transferring the actual data of the patients, thus solving issues of privacy and the need for increased datasets.
- Explainable AI (XAI): creating better techniques for interpretability that are more robust and clinically validated, going beyond the use of simple saliency maps in order to produce explanations that are based in causal understanding and can be verified by radiologists.
- Multi-modal data: integrating predictions made by using chest X-ray images with clinical metadata (such as symptoms, HIV status, or exposure) and other medical diagnostic modalities (including analysis of cough sounds or molecular test results).
- Lightweight and edge-oriented models: development of models that have been compressed or quantized (by means of knowledge distillation or pruning, for example) and are suitable for edge computing in field conditions.

- Domain shift robustness: using domain adaptation and generalization techniques, together with thorough external validation on different populations and different imaging machines, to guarantee reliable operation in practical deployment scenarios.
- Integration with digital health ecosystem: embedding AI-based TB detection tools into digital health records and mobile screening initiatives.

Federated learning requires extra attention due to fragmentation of TB imaging data among different countries' screening initiatives. Since federated learning solutions maintain raw images on each participating site and exchange only model updates or gradients, in theory, federated learning would be able to leverage a global model that could benefit from diverse datasets of multiple countries under the protection of local regulations. Pilot works in the field of federated learning for classification of chest radiographs demonstrate comparable accuracy of federated models trained locally to models trained in centralized fashion, but issues like overhead costs, data heterogeneity at different locations, and problems related to non-independent and identically distributed data sets persist.

When it comes to interpretability, promising avenues go beyond post hoc saliency maps and towards interpretable models, where models predict intermediate features that have clinical significance (such as cavitation or extent of infiltrate), and then use them as input to output TB probability in a manner similar to the one performed by human radiologists. These concept-based approaches might provide more satisfactory explanations than the pixel-based saliency maps, in addition to being useful for tasks such as disease severity estimation and treatment monitoring. In combination with clinical validation studies in prospective settings, these promising avenues lead to a credible approach for developing AI tools for TB detection that are accurate in benchmarking as well as trustworthy and sustainable in practice.

Furthers strides down these paths will be required in order to progress from positive benchmark outcomes into practical and trusted TB diagnostic technologies with the help of AI, especially in the resource-scarce areas where such a clinical need is greatest.

## 8. Practical Recommendations for Implementation

For any research teams or healthcare organizations that might be interested in developing or adopting such an ML-based TB screening tool, several useful recommendations can be derived from the reviewed literature. To start with, when it comes to developing a model, researchers ought to start from the largest and most diverse dataset available and use random splits only for testing purposes instead of using a single-source dataset for splitting into training and testing sets randomly.

Online ISSN 3007-2387 Print ISSN 3007-2379


Secondly, transfer learning using ImageNet pretrained backbones needs to be considered a robust default option, considering the small size of most TB-related datasets, where architectural decisions like choosing between DenseNet121 and EfficientNet-B0 would depend as much on constraints like computational power as much as incremental differences in benchmarks.

Thirdly, sensitivity-based threshold selection, based on the clinical application of the task (e.g., mass screening triage vs. confirmation diagnosis), is to be explicitly defined and presented along with benchmark performance statistics because the clinically useful operating point on the ROC curve will vary depending on the relative costs of the two types of errors.

Fourthly, any tool designed for use in the actual world should be externally validated using datasets from sites and populations that were not involved in developing the model, with a preference for prospective validation carried out as part of the clinical workflow prior to implementation on a wide scale. Lastly, consideration should be taken at the onset of the development process in terms of the interpretation needs, training needs of the clinicians, performance monitoring, updating of the model using the accumulating data, as well as a framework for use which should indicate that the tool is supposed to trigger confirmatory tests after a positive or negative screening test.

## 9. Conclusion

Deep learning methods have contributed immensely in automating tuberculosis diagnosis through chest X-ray imaging. The use of transfer learning and recent CNN models such as DenseNet and EfficientNet has made it possible for classification algorithms to achieve results comparable to those obtained by experienced human readers. It is an area which holds great promise towards developing low cost yet accurate TB screening programs in high burden areas. There are still significant issues, however, including the problem of having insufficient and unbalanced data, vulnerability to domain shift, clinically credible explainability, and deployment issues in resource constrained environments. Solving these problems through federated learning, explainable AI techniques, multimodal learning approaches, and light weight algorithms will prove essential in exploiting the full public health implications of machine learning based TB screening.

Pakistan Journal of Medical & Cardiological Review
https://pakjmcr.com/index.php/1/about
Online ISSN 3007-2387 Print ISSN 3007-2379


## References

[1]. Hansun, S., Argha, A., Liaw, S. T., Celler, B. G., & Marks, G. B. (2023). Machine and deep learning for tuberculosis detection on chest X-rays: Systematic literature review. *Journal of medical Internet research*, *25*, e43154.
[2]. Sharma, V., Nillmani, Gupta, S. K., & Shukla, K. K. (2024). Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images. *Intelligent Medicine*, *4*(02), 104-113.
[3]. Santosh, K. C., Allu, S., Rajaraman, S., & Antani, S. (2022). Advances in deep learning for tuberculosis screening using chest X-rays: the last 5 years review. *Journal of Medical Systems*, *46*(11), 82.
[4]. Han, Z. L., Zhang, Y. Y., Li, J., Gao, S., Liu, W., Yang, W. J., & Xing, Z. H. (2025). A systematic review and meta-analysis of artificial intelligence software for tuberculosis diagnosis using chest X-ray imaging. *Journal of thoracic disease*, *17*(5), 3223.
[5]. Al-Timemy, A. H., Khushaba, R. N., Mosa, Z. M., & Escudero, J. (2021). An efficient mixture of deep and machine learning models for covid-19 and tuberculosis detection using x-ray images in resource limited settings. In *Artificial intelligence for COVID-19* (pp. 77-100). Cham: Springer International Publishing.
[6]. Acharya, V., Dhiman, G., Prakasha, K., Bahadur, P., Choraria, A., M, S., ... & Kautish, S. (2022). AI-assisted tuberculosis detection and classification from chest X-rays using a deep learning normalization-free network model. *Computational Intelligence and Neuroscience*, *2022*(1), 2399428.
[7]. Liu, C. J., Tsai, C. C., Kuo, L. C., Kuo, P. C., Lee, M. R., Wang, J. Y., ... & Yu, C. J. (2023). A deep learning model using chest X-ray for identifying TB and NTM-LD patients: a cross-sectional study. *Insights into Imaging*, *14*(1), 67.
[8]. Shiva, T. A., Ireen, N., & Islam, M. S. (2024). Optimizing early intervention strategies for neurodiverse children (ASD): Reducing long-term public healthcare costs through parent-mediated training. *Apex Journal of Social Sciences*, *3*(1), 30-52.
[9]. Islam, M. Z., Ahmed, I., Parveen, R., Rimon, S. T. H., & Janjua, J. I. (2025, December). Revolutionizing healthcare for critical diseases with AI and enhancing security via low-latency edge intelligence. In Proceedings of the 2025 International Conference on AI-Driven STEM Education and Learning Technologies (AISTEMEDU) (pp. 1–6). IEEE.
[10]. Hasan, S. T. (2023). Natural language processing for employee relations case management: Analyzing workplace complaints, grievances, and investigation

narratives in healthcare organizations. Journal of Computational Systems & Engineering Insights, 1(01), 11-24.

[11]. Eisentraut, L., Mai, C., Hosch, J., Benecke, A., Penava, P., & Buettner, R. (2025). Deep learning-based detection of tuberculosis using a Gaussian chest X-ray image filter as a software lens. IEEE Access, 13, 36065-36081.

[12]. Goswami, K. K., Kumar, R., Kumar, R., Reddy, A. J., & Goswami, S. K. (2023). Deep learning classification of tuberculosis chest X-rays. Cureus, 15(7), e41583.

[13]. Iqbal, A., Usman, M., & Ahmed, Z. (2022). An efficient deep learning-based framework for tuberculosis detection using chest X-ray images. Tuberculosis, 136, 102234.

[14]. Banapuram, C., Naik, A. C., Vanteru, M. K., Kumar, V. S., & Vaigandla, K. K. (2024). A comprehensive survey of machine learning in healthcare: Predicting heart and liver disease, tuberculosis detection in chest X-ray images. SSRG International Journal of Electronics and Communication Engineering, 11(5), 155-169.

[15]. Urooj, S., Suchitra, S., Krishnasamy, L., Sharma, N., & Pathak, N. (2022). Stochastic learning-based artificial neural network model for an automatic tuberculosis detection system using chest X-ray images. IEEE Access, 10, 103632-103643.

[16]. Al-qaness, M. A., Zhu, J., AL-Alimi, D., Dahou, A., Alsamhi, S. H., Abd Elaziz, M., & Ewees, A. A. (2024). Chest X-ray Images for Lung Disease Detection Using Deep Learning Techniques: A Comprehensive Survey: MAA Al-qaness et al. Archives of Computational Methods in Engineering, 31(6), 3267-3301.

[17]. Badhan, I. A., Neeroj, M. H., & Chowdhury, I. (2024). The effect of AI-driven inventory management systems on healthcare outcomes and supply chain performance: a data-driven analysis. Frontline Marketing, Management and Economics Journal, 4(11), 15-52.

[18]. Hasan, S. T. (2024). AI powered early warning systems for healthcare worker burnout in New York City: Predictive modeling of absenteeism, turnover, and scheduling burden. Journal of Computational Systems & Engineering Insights, 2(02), 01-14.

[19]. Kebache, R., Laouid, A., Guia, S. S., Kara, M., & Bouadem, N. (2023, December). Tuberculosis detection using chest X-ray image classification by deep learning. In Proceedings of the 7th International Conference on Future Networks and Distributed Systems (pp. 352-356).

[20]. Kabi, S. K., Tripathy, R. K., Patra, D., & Panda, G. (2023). A novel approach for the detection of tuberculosis and pneumonia using chest X-ray images for smart healthcare applications. IEEE Sensors Letters, 7(12), 1-4.

Pakistan Journal of Medical & Cardiological Review
https://pakjmcr.com/index.php/1/about
Online ISSN 3007-2387
Print ISSN 3007-2379


[21]. Singh, M., Pujar, G. V., Kumar, S. A., Bhagyalalitha, M., Akshatha, H. S., Abuhaija, B., ... & Gandomi, A. H. (2022). Evolution of machine learning in tuberculosis diagnosis: a review of deep learning-based medical applications. Electronics, 11(17), 2634.

[22]. Rajakumar, M. P., Sonia, R., Uma Maheswari, B., & Karuppiah, S. P. (2021). Tuberculosis detection in chest X-ray using Mayfly-algorithm optimized dual-deep-learning features. Journal of X-Ray Science and Technology, 29(6), 961-974.