Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

MULTI-SOURCE CHEST X-RAY DATASETS FOR ACCURATE AND EXPLAINABLE TUBERCULOSIS (TB) DIAGNOSIS

Reeba Waris Ali[*1], Saira Khatoon[2], Sohaib Shazadi[3], Samar Abbas[4], Zaeem Nazir[5]

[*1,2,3,4,5]*Department of Computer Science, University of Narowal*

[*1]reebawarisali@gmail.com, [2]Saraashfaq81@gmail.com,[3]sohaibshahzadi02@gmail.com, [4]samarabbas0425@gmail.com, [5]zaeem.nazir@gmail.com

DOI: https://doi.org/10.5281/zenodo.19367284

Keywords
Gradient-weighted Class Activation Mapping (Grad-CAM), Explainable Artificial Intelligence (XAI), Deep Learning (DL), Multi-Source Data Integration, Chest X-ray (CXR) Imaging, and Tuberculosis (TB) Diagnosis

Article History
Received: 01 February 2026
Accepted: 17 March 2026
Published: 31 March 2026

Corresponding Author: *
Reeba Waris Ali

Abstract
Tuberculosis (TB) is another cause of death, which relates to infectious diseases in the world. Chest X-ray (CXR) is an inexpensive and convenient screening tool for TB. Nevertheless, in the majority of cases, X-rays need to be interpreted manually, making them slow, subjective, and subject to inter- and intra-observer variability. In addition, most existing automated TB detection approaches are trained on a single data source, which limits their ability to generalise to other populations. In response to this, the current research paper proposes the multi- source CXR fusion methodology that will be more reliable and open the TB diagnosis. The publicly available CXR datasets are merged with various clinical settings and geographical areas in order to maximize the data in the datasets and reduce bias in the datasets. The preprocessing steps are followed by the merging of datasets, all of which are intensive preprocessing steps, including image normalization, lung part segmentation, contrast intensification, and class balancing, and all the sources must be consistent and compatible to facilitate fusion in the end. The feature extraction and classification problem is solved using a deep-based model using a convolutional neural network (CNN) as a pre-trained model, where it is possible to learn the discriminative TB-related patterns on the joint data automatically. Gradient-weighted Class Activation Mapping (Grad- CAM) is an explainable Artificial Intelligence (XAI) technique that is applied to improve interpretability and facilitate clinical decision making. Such methods present graphically where the contribution of each part of the lungs in predicting the model is maximum, such that the clinicians may have a more detailed view of the diagnostic outcomes and may verify them. This approach is highly likely to be applied in real-life medical practice, e.g., the community with limited resources is where a TB screening is highly needed in a real-time context.

1. Introduction
Even with impressive gains in the field of drug development and global disease control programs, TB remains a significant global public health problem. TB is a pneumonia illness that is caused by Mycobacterium Tuberculosis and is transmitted by air. Such a pattern of transmission contributes greatly to the level of infections, especially in areas where there is low economic capacity, overpopulation, and underdeveloped healthcare facilities. The report of the World Health Organization (WHO) indicates that the annual number of new cases of TB in millions, and the impact of the disease is spread unequally among the low and middle-income countries as a result of late diagnosis and access to effective treatment facilities [1].

As the role of experienced health care professionals is significant in TB diagnosis, such issues as high workloads, an insufficient number of trained radiologists, variability on the side of the observer, and the subjectivity of CXR interpretation severely limit their practice. Failure to identify TB at an early stage or the incorrect diagnosis of the disease results in worse outcomes, including machine learning and mortality in patients. This fact highlights the growing need to have accurate diagnostic solutions, and at the same time, fast, scalable, and ubiquitous [2]. Conventional methods of diagnostics, e.g., microscopy of sputum smears and culture-based testing, are still restricted in terms of time and sensitivity, especially with early-stage or smear-negative TB.

Due to recent years' study, there has been a rapid development of Artificial Intelligence (AI) and Machine Learning (ML) that has created new opportunities in automated analysis of medical images that can offer alternative methods of diagnosis of TB to traditional ones. In this area, Deep Learning (DL), and specifically Convolutional Neural Networks (CNNs), have proved to be extremely successful in medical imaging operations, such as disease detection, image segmentation, and classification operations. CNN-based models can independently learn complex and noteworthy features in the chest X-rays to make the necessary discovery of TB-related abnormalities that might be missed in a manual analysis [3]. Several studies have documented the great diagnostic accuracy of deep learning models that are trained using CXR datasets in the presence of controlled experimental conditions. Such systems have high potential to assist radiologists in terms of decreasing diagnostic work and increasing the rate of detection of cases. Nevertheless, although their performance indicators are encouraging, there are a number of issues that restrict their use in practice in hospitals. The first and most obvious ones are poor generalization of the different datasets, image quality variability, and, most significantly, there is no transparency in model decision-making. This aspect of CNNs raises worrying questions on the part of clinicians about trust, accountability, and clinical reliability in situations involving high-stakes medical decision making [4].

To deal with these restrictions, XAI methods have been created to offer information about the way deep learning models reach their forecasts. To demonstrate the areas of the chest X-rays that moderate diagnostic scores, one can use techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM), saliency maps, and attention-based systems. Such approaches can emphasize clinically relevant pulmonary abnormalities encompassing infiltrates, cavities, nodules, and consolidations in the lumbar node of TB detection and therefore agree with available radiological knowledge [5].

The explanation of explainability in automated TB detection structures is a must-have element in the development of confidence in the clinician, enhancing transparency, and promoting activity within a realistic healthcare environment. Explainable models can also help medical professionals not only to check AI-generated predictions but to improve the clinical decision-making process, which presupposes interpretable visual evidence. Due to that, high-performing deep learning has been integrated.

This research is supposed to have the following objectives:

1. To combine multi-source chest X-ray results to make a sound TB diagnosis.
2. To minimise inter-dataset variability, employ standardized preprocessing and lung segmentation.
3. To add balance to classes and provide feature consistency in the combined data.
4. To train a deep learning model for accurate TB detection.
5. To give explainable and clinically interpretable model predictions.

This piece of work will allow us to contribute to the research, and this will also provide a practical solution that will not only help the patients but also the healthcare providers.

2. Related Work

Deep learning has presented unprecedented performance in medical image analysis, especially radiological image classification,

which is fully automated. In several studies, CXR images have been diagnosed using CNNs to identify pulmonary illnesses such as TB. The X-ray-based techniques are attractive, as they are common and cheap; they use the CNN algorithm on X-ray and CT images. Two-dimensionality, though, poses the disadvantage of being able to represent only the intricate nature of the lungs, and in most cases, the accuracy of the diagnosis of subtle or intermediate diseases is compromised [6]. Recent studies indicate that CT-based deep learning models achieve higher performance as compared to X-ray- based systems because of their capability to capture the volumetric and spatial features of lung aberrations. But CT is very expensive, subjects the patients to increased radiation, and is not accessible in many rural or low-resource regions with a high prevalence of tuberculosis. The current research primarily relies on synthetic pictures or actual CT data instead of modality transformation to be generically augmented.

The main concern of the early research on CAD was the detection of nodules in the lungs and not the disease. It was found out that automated systems could be useful in aiding radiologists, although they could not be used independently to diagnose. Van Ginneken also stated that decades of efforts had not resulted in completely reliable systems of automated interpreting of chest radiographs. Traditional machine learning methods were handcrafted features such as the texture, shape, symmetry, a histogram-based feature, and an edge, for TB-specific diagnosis. Such classifiers as SVM, MLP, Random Forests, and Bayesian networks followed the line. Order of accuracy S 97 and M 96. [7]. Deep CNN based on this transfer learning in ImageNet has been observed to be encouraging, especially in cases where medical datasets labelled are scarce. Discriminative features of TB could be extracted with models such as VGG, ResNet, DenseNet, and MobileNet on the basis of comparative studies. There are successfully implemented cases of the use of CNNs to recognize infectious diseases such as COVID-19, tuberculosis (TB), and other lung conditions using CXR images. Chest X-rays are very cheap and reasonably available, but with the limitation that its two dimensional nature limits the anatomical details that can be effectively used in diagnosing the condition. Computer tomography (CT) scans, on the other hand, provide more detailed 3D information and have been more successful in measuring and diagnosing the severity of tuberculosis. Through many studies, it has been proven that.

The CT-based deep learning models outperform the X-ray-based techniques as they are able to better display the architecture of the lungs and patterns of diseases. Enhanced diagnostic performance has been further achieved with the use of multi-modal learning methods, which integrate the data on many (many) varying imaging modalities. They obtain an accuracy of 86% [8]. Earlier research has shown that the accuracy of the classification can be increased through the integration of the CT data with other clinical data or through a high number of X-ray projections. Nevertheless, the research on the use of synthetic CT scans created based on the X-rays of the chest organs, in particular, in the detection and classification of TB, has not been conducted sufficiently.

The conventional diagnosis techniques, which are usually expensive, time-consuming, and require lab materials, are sputum microscopy, culture, and molecular tests. In recent years, AI, especially machine learning and deep learning, has proven to be a promising technology to make diagnosing and treating tuberculosis more efficient and accurate, often even reaching and surpassing the accuracy of trained radiologists. Past studies have demonstrated that deep learning models trained on CT images and CXR images can identify the presence of pulmonary TB with high accuracy, often that of a competent radiologist, without any doubt. Radiomics AI methods have proven to be highly effective in terms of detection and also differentiating tuberculosis against other chronic diseases of the lungs, such as lung cancer and non-TB mycobacterial infections. Liang et al. obtain DL and radiomics-based AI models, which often have similar radiological features [9]. However, the research just being published can be considered as sufficient evidence of the possible benefits of AI as an addition to the contemporary treatment

methods of TB and resistance prediction at its first manifestation.

DAI-based CAD systems have demonstrated tremendous potential over the last couple of years to assist radiologists by enhancing the accuracy of the diagnosis and reducing the occurrence of interpretation errors. A large gap in the development of TB detection models with high diagnostic accuracy, timely inference, and low computation cost still remains. DL ways, specifically CNNs, have emerged as the most common system in automated TB detection on CXRs.[10],. Such a discontinuity is material given the upscaling cases of TB undiagnosis, and the requirement of scalable webbing patterns in areas that have small coffers. Consequently, the present exploration has become feather-light effective deep literacy infrastructures, save slice- edge performance, and facilitated real- time deployment, which are being prodded as optimized models such as LightTBNet.

Recent research has used the DL algorithms more frequently to automate TB detection on CXR film to overcome these shortcomings. CNN and transfer literacy models EfficientNet and Inception have performed well in TB bracket tasks, where they are continuously achieving delicacy situations of above 90. The algorithms that CNN and Transfer Learning EfficientNet that they implemented in the course of their work [11]. An example of closely available datasets that have been significantly utilized are TBX- 11K, the Montgomery County dataset, and the Shenzhen Hospital CXR dataset. It has been observed that dataset emulsion significantly enhances bracket delicacy and recall in the discovery of tuberculosis when employed with a style of preprocessing such as CLAHE, normalization, and addition of data. Also, emulsion- based styles achieve stability with a wide range of imaging scripts.

Early and proper diagnosis is highly needed to decrease mortality and avoid the spread of the disease. CXR imaging is the most common mechanism of imaging used in the diagnosis of tuberculosis because of its low cost and availability. In order to overcome these problems, authors have examined deep learning, multi-source dataset fusion, and explainable artificial intelligence (XAI) algorithms to identify tuberculosis in patients automatically. Such models as CNN, ResNet, DenseNet, VGG, and others.

The ability to detect TB using the chest X-ray photographs has been documented through Inception, which is highly accurate, with the benefit of using feature fusion methods such as concatenation and attention-based fusion. ViT and hybrid CNN-Transformer models have also gained popularity [12] due to their ability to extract local and global image features and, therefore, have superior classifying capabilities when confronted by large, diverse TB datasets. At the same time, explainable AI (XAI) has become one of the prominent fields of research. The proliferation of the dangerous lung disease TB is best suppressed at the first stages so as to prevent further growth. Even though CXR imaging is also common in identifying TB, manual analysis is dependent on the radiologists and may be prone to errors, especially in localities where there is a deficit of medical staff. Deep learning and machine learning were created by researchers to boost diagnosis. The identification of TB using chest X-rays became popular under CNN. Pretrained CNNs such as VGG16, AlexNet, and DenseNet can automatically learn picture features and achieve high accuracy rates 98% accuracy rate [13], but this method lacks clarity as to why or why not a specific region of the lung is used to make a prediction. Explainable AI (XAI) techniques, like Class Activation Maps (CAM) and LiME, have been created to answer the question of how and why a particular part of the lung is so important in prediction. The following procedures improve the faith in the decision of models.

TB is often diagnosed based on images of CXR, but manual interpretation depends on radiologists and is not accurate. Subsequently, using deep-learning models to extract characteristics of the X-ray images automatically, and especially CNN, was more successful[14]. use CNN, Hybrid CNN + HOG, and pretrained models (VGG16, AlexNet). and associates. To support the development of deep and textural features, hybrid models that can combine CNN with Histogram of Oriented Gradients (HOG) were proposed. Although the

use of pretrained models such as VGG16 and AlexNet led to higher detection accuracy, they are black-box models that do not explain why they do so and fail to integrate features.

To solve these problems, recent studies have focused on automated TB detection and triage with the use of deep learning and artificial intelligence. The early deep learning methods were mostly based on CNNs trained on labelled CXR collections, such as the Shenzhen, Montgomery, and ChestXray collections. More recently, studies of transfer learning and self-supervised learning have been examined in order to better leverage large quantities of unlabelled medical images (theoretical study). ViT has been shown to have improved representation learning results than traditional CNNs when combined with self-supervised learning methods such as DINO-based distillation.

Following a study of research papers, published during the period of 2020-25, a research gap is identified that the majority of the studies on tuberculosis (TB) detection use the single-source datasets of chest X-rays, which restricts the ability of the model to be robust and generalized in other clinical settings. Though the accuracy of deep learning models is high, the models are combined with uniform preprocessing to minimize bias in the dataset. Moreover, imbalance in classes in the combined datasets is also not satisfactorily addressed, which influences its sensitivity to TB-positive cases. Most available models are also not provided with explainability as black-box systems, which makes them a barrier to clinical trust. Although there are explainable AI methods, they are not often used together with multi-source fused datasets within a single framework.

Table 1:

Author Name	Year	Algorithm	Accuracy	Data Fusion Technique	Ensemble Learning	Explainable AI
Tawsifur Rahman, Amith Khandakar, et al. [6]	2020	CNN for X-ray and CT images	99%	NO	NO	NO
Lewis, Mahmoodi, Zhou Elena et al. [7]	2021	CNNs, Deep Learning based CT models, Multi-modal Deep Learning models, Feature fusion / Multi-view learning models	86%	NO	NO	NO
Jignesh, Chowdary et al. [8]	2021	Traditional ML: SVM, MLP, Random Forest, Bayesian; Deep Learning: VGG, ResNet, DenseNet, MobileNet	S 97%, M 96%	NO	NO	NO
Liang et al. [9]	2022	Deep Learning, Radiomics-based AI models	95%	NO	NO	NO
Daniel Capellán-Martin et al. [10]	2023	Lightweight CNN (LightTBNet)	90%	NO	NO	NO
E. Mahamud, N. Fahad, M. Assaduzzaman et al. [11]	2024	CNN, Transfer Learning EfficientNet	99%	NO	NO	NO
A. Khan, Akram, Sharif, Y. Javed, and Muhammad et al. [12]	2024	NN (ResNet, DenseNet, VGG, Inception), Vision Transformer (ViT), CNN-Transformer Hybrid	99%	NO	NO	NO
Maheswari et al. [13]	2024	CNN, VGG16, AlexNet, DenseNet; Explainable AI: CAM, LIME	98%	NO	NO	YES
Ayalew et al. [14]	2025	CNN, Pretrained models (VGG16, AlexNet), Hybrid CNN + HOG	91%	NO	NO	NO
N. Patel et al. [15]	2025	CNN, Vision Transformer (ViT), Self-supervised learning (DINO-based)	*N/A (theoretical)*	NO	NO	NO
Our Work	**2026**	**CNN AND RepViT-CXR Algorithm**	**96%**	YES	YES	YES

The table below provides a comparative overview of the previous studies:

3. Methodology

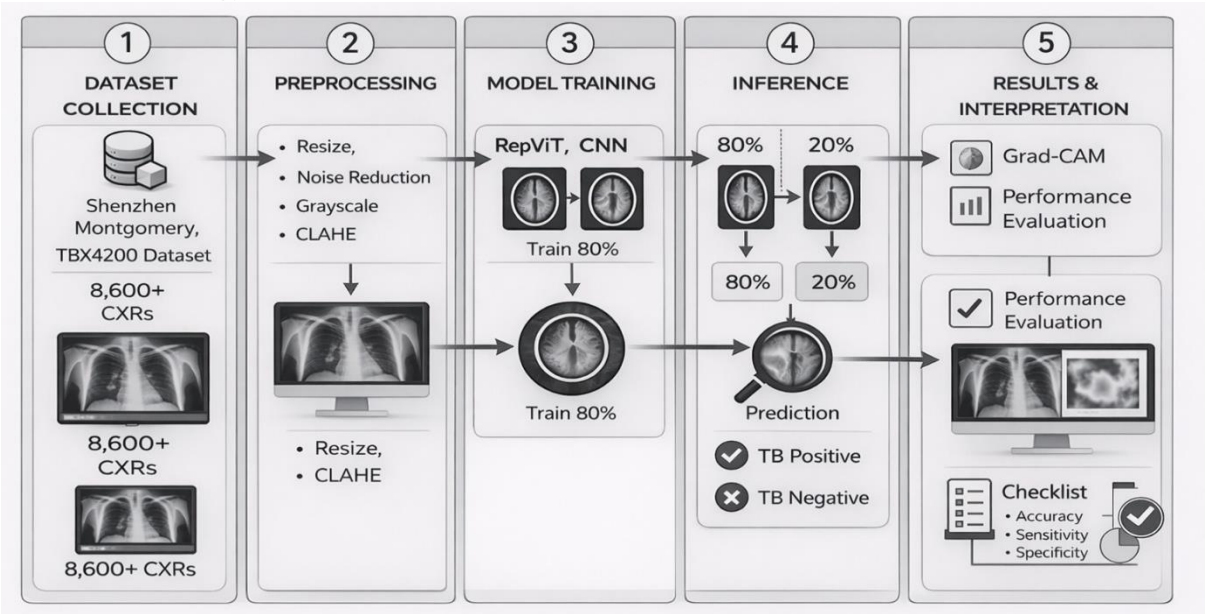


Fig.3.1: Proposed tuberculosis (TB) detection methodology pipeline illustrating dataset integration, preprocessing, model training, inference, and result interpretation in a structured workflow.

3.0.1 Overall System Pipeline

The figure is a descriptive framework of the proposed TB detection pipeline that can be divided into 5 consecutive steps. This will begin with the data sets being collected, and several chest X-ray (CXR) datasets will be pooled to enlarge the diversity of the data in the form of Shenzhen, Montgomery, and TBX4200. In preprocessing, image size and standardization, image removal of noise, color grayscale, and CLAHE image enhancers are ways to standardize images to improve the quality of the resulting image and its consistency. Subsequently, 80 percent of the data is then utilized to train deep learning models, i.e., CNN and RepViT, and they get to learn the discriminative features during the model training stage. This inference phase consists of the trained models making a classification on the remaining 20 percent of test data in order to arrive at predictions on the TB-positive and TB-negative cases. Finally, performance assessments and Grad-CAM visualization will also be part of the results and interpretation step, as they will provide a quantitative assessment of the model decisions and interpretability. Through this pipeline, there is an intricate, efficient, and clinically meaningful framework of automated detection of TB.

In addition, the provided pipeline is concerned with performance and interpretability, implying that the created system would not simply be accurate, but it needs to be a clinically reliable system as well. Data integration using multiple sources helps in offsetting data differences, in addition to interpolating the models in different populations. It also smooths input data using structured preprocessing, which results in less noise and better feature extraction. In addition to that, the explainable AI techniques, such as Grad-CAM, also act as a visual justification of the model decisions and, therefore, contribute to the rates of openness and confidence among the medical professionals. Overall, the end-to-end architecture of the pipeline offers scalability of the deployment and exhibits high potential for assisting the medical workers in diagnosing TB at an early and accurate stage.

3.1 Dataset

The presented research relies on diverse publicly available CXR collections and is aimed at creating an efficient and user-friendly method of diagnosing tuberculosis (TB). The information was obtained through the known and popular medical imaging repositories to ensure reliability, variety, and clinical significance. The datasets were analyzed separately, then merged together to analyze the size, the distributions of the classes, and consistency in labeling. This close analysis plays a very important role in medical imaging studies to identify potential labelling conflicts, remove dataset bias, and prevent unintended data leakage when training the model.

Kaggle Dataset 1: Tuberculosis Chest X-ray Images (Mendeley)

The initial data was acquired at Kaggle and was initially obtained from the Mendeley Data repository. This dataset is a selective dataset of chest X-rays.

images that were specifically dedicated to detecting tuberculosis.

- TB images: 2,494
- Normal images: 514
- Total images: 3,008

The dataset is rather large in the number of TB proportions, which makes class imbalance more evident with a greater proportion of TB-positive samples, which might also lead to bias when used separately.

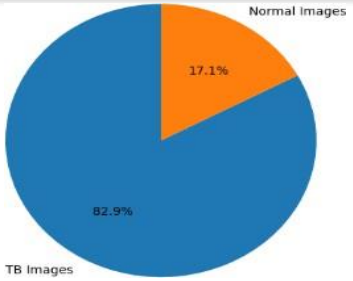


Fig. 3.2 Dataset 1 representation

Kaggle Dataset 2: Tuberculosis (TB) Chest X-ray Dataset (TB X-ray 4200)

The second dataset obtained in Kaggle is also known as the TB X-ray 4200 data set. This data is popularly utilized in the academic literature, and it includes a definite binary class.

- TB images: 700
- Normal images: 3,500
- Total images: 4,200

As opposed to the former dataset, this dataset is extremely concentrated on the Normal category, and there is a need to combine and balance the set of data with the aim of achieving impartial model training.

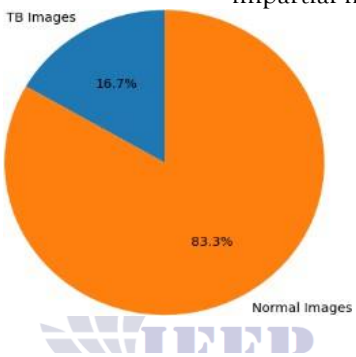


Fig. 3.3 Dataset 2 representation

Shenzhen Hospital X-ray Chest Dataset (NIH / NLM)

The dataset on Shenzhen Hospital Chest X-ray was obtained at the National Library of Medicine (NLM), National Institutes of Health (NIH). It is a high-authoritative and clinically-validated database comprising the chest X-rays of Shenzhen No.3 People's Hospital in China.

- TB images: 336
- Normal images: 326
- Total images: 662

This dataset is surprisingly small, but it can help make the datasets more diverse and enhance the generalization of the model.

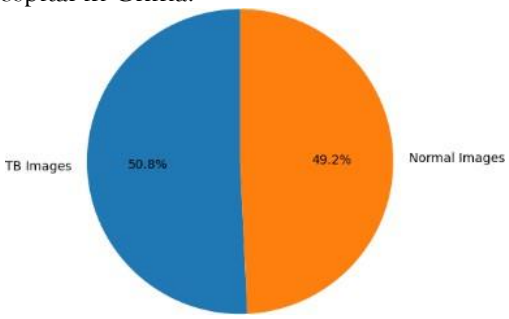


Fig. 3.4 Dataset 3 representation

3.1.1 Dataset Preprocessing

Class labels already exist in all three sets in the case of fusion. They, however, varied in their folder structure, names, and quality of pictures. A single pretreatment pipeline was applied to each dataset to standardize the inputs before

merging. Before dataset fusion, there was one common processing phase for all pictures to ensure consistency of the datasets, as well as to improve the learning of the model.

Resizing the Images

All the X-rays of the chest were going to be down-sampled by 224 pixels. The reason behind this resolution was as follows:

- As an input size, it is a standard size with pretrained CNN and Vision Transformer models. Balancing is achieved on anatomical details and computational efficiency.
- It allows drawing fair comparisons of different designs.

Greyscale Conversion

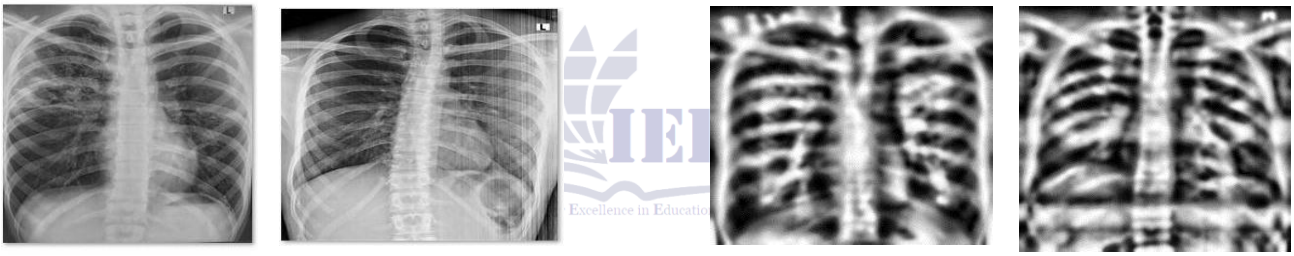
The X-ray chest images are greyscale in nature. Photos were all specially converted to single-channel greyscale to remove the colour channels that were not necessary and to reduce computing overhead. This step: keeps therapeutically significant intensity patterns, reduces the complexity of the model, and enhances convergence stability.

CLAHE or Contrast Limited

To increase local contrast and make lung features more visible, CLAHE was used. For chest X-rays, this method works especially well because it intensifies minor TB-related abnormalities such as cavities and infiltrates. Unlike global histogram equalization, it prevents noise from being over-amplified. It enhances deep learning models' feature extraction. [17], [18]

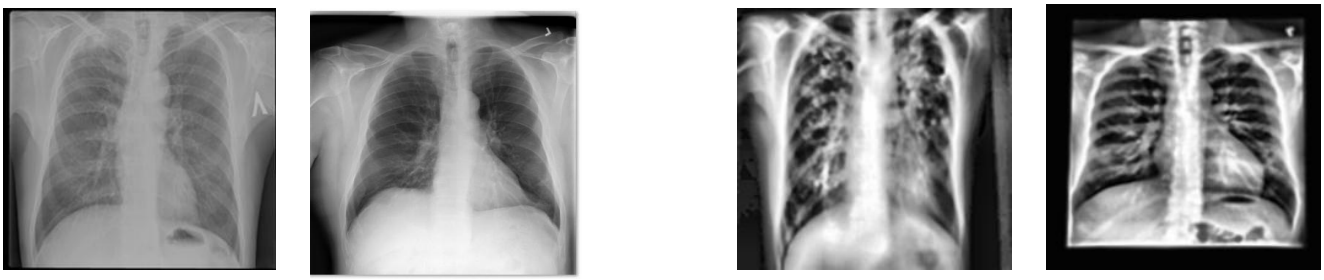
Kaggle Dataset 1: *Tuberculosis Chest X-rays Images* (Mendeley)

Before preprocessing **After preprocessing**



Kaggle Dataset 2: *Tuberculosis (TB) Chest X-ray Dataset* (TB X-ray 4200)

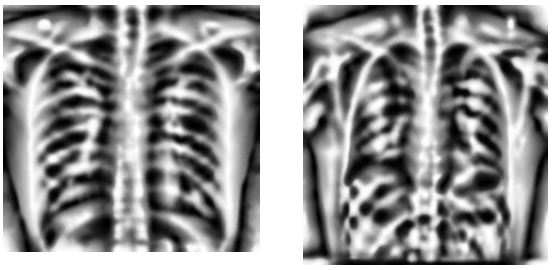
Before preprocessing **After preprocessing**



Shenzhen Hospital CXR Dataset (NIH / NLM)

Before preprocessing

After preprocessing



3.2 Strategy for Data Fusion

A data fusion method using a data-level fusion technique was used to merge all three datasets into a single one. The fusion that transpired between class labels into two categories: 1)Normal, 2)TB.

- TB images: 3523
- Normal images: 4340
- Class disparity: 817 pictures

This imbalance would skew the model towards the majority class (Normal), which would reduce the sensitivity of TB identification.

Class Balancing

A separate TB class was used to equalize class imbalance with 4309 photos equaling the distribution of the Normal class. Balancing (without causing a leakage of data across train and test sets) was accomplished after preprocessing, by means of controlled oversampling of TB pictures.

TB pictures: 4309, regular pictures: 4340, and 8649 photos. This balanced dataset improves the robustness, recall, and fairness of the trained models.

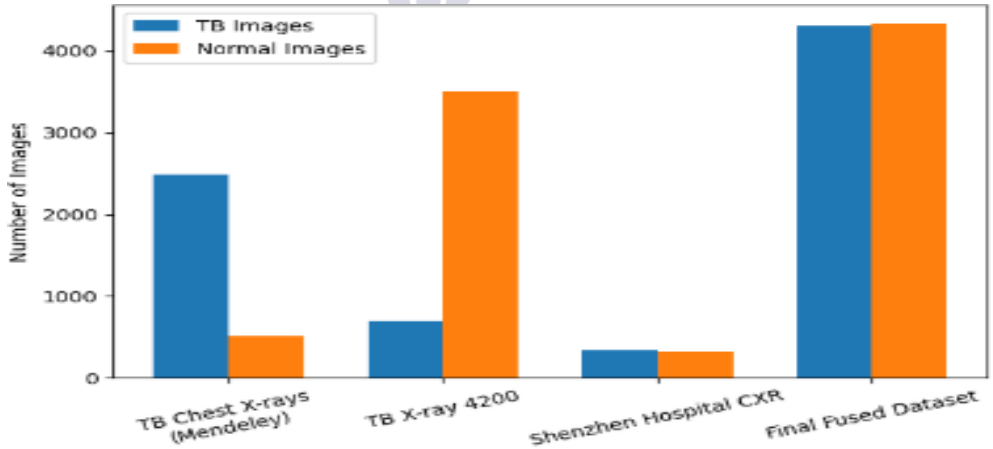


Figure 3.5: Graphical representation of the dataset

This bar chart shows values of TB (Tuberculosis) and the standard chest X-ray image in three source datasets and in their final fused dataset. TB Chest X-rays (Mendeley) provides the largest number of TB images of the individual sources, which is about 2,500, yet only about 500 of normal images, showing a severe class imbalance. The TB X-ray 4200 data is the opposite end, where there are about 700 TB images and about 3,500 normal images, which is once again a significant disparity. The Shenzhen Hospital CXR dataset is the least contributive data, both TB and normal images being approximately 300-350 each, so it is the most balanced one out of the three sources. On combining the three datasets into the Final Fused Dataset, both groups have a similar number of TB and normal chest X-rays of around 4,300 and 4,300, respectively. This balancing is vital in training unbiased machine learning models to detect TB, as equal class

representation will avoid the bias of the models towards the dominant one.

3.3 Splitting Datasets

To ensure balance of classes, the resulting fused dataset was stratified (80% - 20%), i.e., 80% of the data was used as the training set and 20% as the testing set: 20%.

3.4 Deep Learning Models

There were three deep learning architecture types that were examined in order to evaluate whether this detection technique can work with TB cases based on chest X-rays.

3.4.1 Tuberculosis Prediction Model CNN-Based

The proposed development is a deep learning algorithm based on a Convolutional Neural Network (CNN) to identify TB with a CXR image automatically. To detect TB in a CXR image, a CNN model is developed to effectively learn discriminative representations of spatial features in the image that classify the object into a normal cell as opposed to a TB-infested cell[15]. The model employs a custom featherlight CNN as opposed to a veritably deep one; it is computationally efficient and applicable in real-world clinical environments. The model was estimated and trained using fused TB data, robustness, and icing diversity in learning complaint patterns.

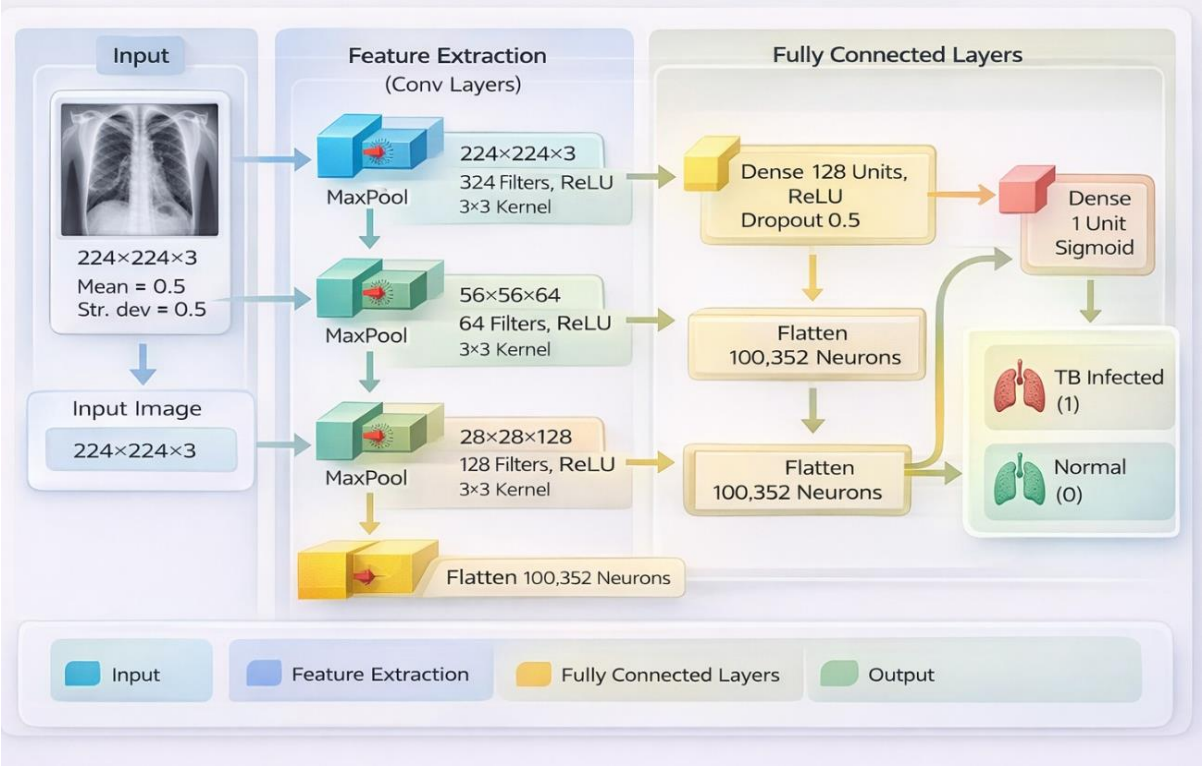


Figure 3.6: Architecture of CNN-based automated Tuberculosis detection on Chest X-ray images.

Working Medium of the CNN Model:

To manage the thickness of all the datasets, all CXR images were downsized to 224 x 224 pixels. The images were all turned into three-channel grayscale images as they were originally grayscale, so that they would match common convolutional operations. Image normalization- This will be done using a mean and standard deviation of 0.5, which will stabilize the change in grades throughout the training process and accelerate the convergence process. The point birth model has three convolutional layers, which represent higher-order position representations at a high rate. The first cast (3 → 32 channels) detects low position characteristics that are equal to edges, outlines, and texture variation of CXR images. The other caste (32 → 64 channels) gains a new collection of fundamental patterns as well as more themes, including the edges of the lungs and the ambiguous areas of TB lesions. The third subcaste (64 → 128 channels) is listening

to the complex spatial distribution, including cavitation zones, nodes, and unusual lung textures, which enables the model to identify the properties of tuberculosis positively.

3.4.2 RepVision Transformer (ViT) Model

In this study, the double bracket of CXR images applies the ViT deep learning approach to the study. The interpretation of images as patch sequences and the self-attention process to analyze the global contextual association allows the Vision Transformer to surpass the conventional methods. The model is particularly beneficial to medical picture analysis because significant clinical patterns can extend across many geographical regions. The model employs patch embedding, positional encoding, as well as an embedding of multiple layers of a piled head of motor encoders. Transfer learning is asserted to make the model through action-at-a-point mean of already well-trained weights, which permits successful convergence and improved performance when trained on a small medical data mass.

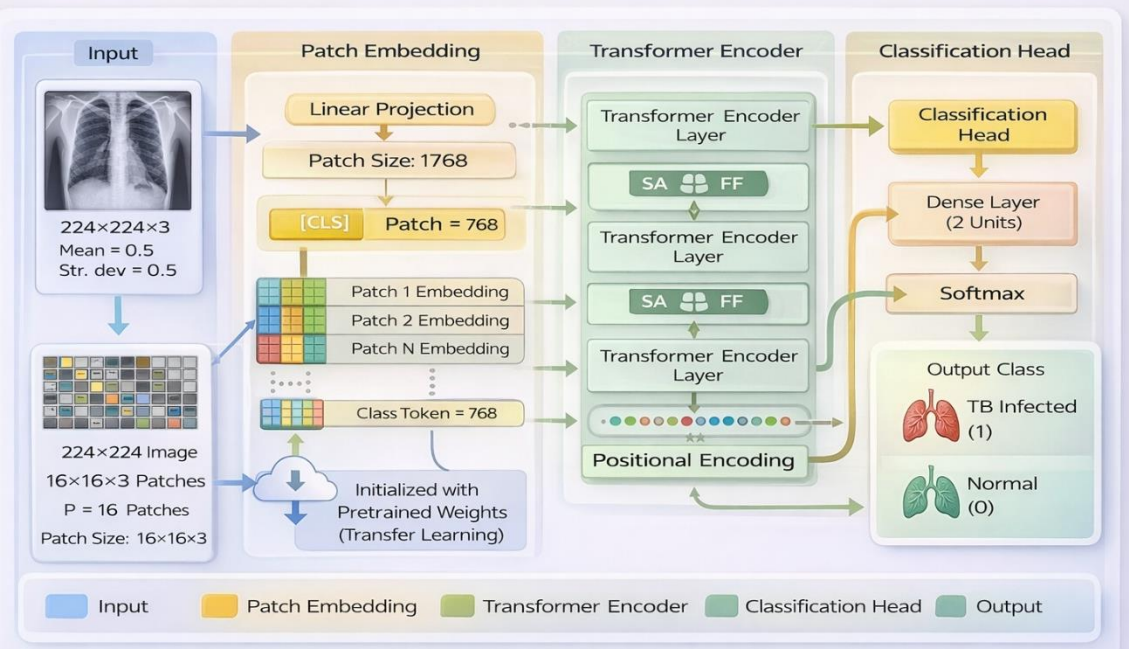


Figure 3.7: Vision transformer(ViT)-based model of tuberculosis classification with patch embeddings and self-attention.

Working Medium of the Vision Transformer Model

The input image is downsized (224 × 224) and homogenized. It is divided into fixed-size patches (say 16 squares), and each patch is smoothed and linearly projected to a patch embedding. Positional embeddings are added to conserve on the spatial. The sequence of patch embeddings undergoes multiple encodings of Global dependences with Multi-Head Tone-Attention (MHSA) and an improvement of points with a feed-forward network (FFN). Constant training is secured by residual connections and normalization of the subcaste. Residual connections. The information is summarized in a special (CLS) token, the final representation of which is input into a fully connected bracket head with softmax activation in order to give probabilities of the classes. The model is optimized on cross-entropy loss by gradient-based optimization (Adam). Finally, the ones that are not observed are given in the same channel, and the most probable is the prognosticated one.

It is an improved architecture of U-Net, and it adds in the encoder and decoder branch the residual blocks. The very consecutive blocks of convolutional layers are mixed with stochastic blocks in the subnet contracting type of the network, and as the space resolution reduction continues, at the same time, the channels in which the features are represented are also raised, and such is particularly handy in aiding the generation of the high degree semantic

representations. At the bottom level, the dropout is activated on the residual loophole module to get smaller and more discriminative representations, and other incentives to minimize biases of overgenerativity. The broad path, in its turn, spreads the spatial resolution similarly by a set of up-sampling algorithms partitioned by the reserved blocks. The synthesis of all these maps occurs at the various levels of the corresponding encoder and is connected with each other via skip connections to the output of the corresponding decoders and, hence, retains the integrity of small spatial patterns and context. Residual learning helps the gradient propagation process to be less tedious since it reduces the difficulty of vanishing gradients, enabling it to train harder network architectures. The last and the last one is the 1x1 convolution that is succeeded by a sigmoid activation, yielding us the final segmentation map.

3.4.3 Explainable AI in Ensemble Learning in the Detection of Tuberculosis

CBT: Inaccurate diagnosis of TB based on Chest X-ray (CXR) images is especially important when modern facilities have limited access to trained radiologists due to resource constraints. Although deep learning techniques, such as CNNs and ViTs, have demonstrated high predictive accuracy in automated TB detection, their tendency to produce black-box results makes their application in clinical settings challenging. Clinicians not only need correct predictions but also explanations that can be adopted and interpreted to provide details of the places in the lungs that result in a positive diagnosis. To overcome this difficulty, explainable artificial intelligence (XAI) systems have been incorporated in ensemble learning systems, which are both highly accurate and interpretable. In the work, ensemble learning was used by mixing a CNN and a ViT using the strengths of all of them to enhance their effectiveness. The CNN is good at capturing localized spatial characteristics, whereas the ViT provides a global context of lung patterns. The ensemble prediction is made by averaging the results of the two models, and this makes the prediction significantly more robust, and the uncertainty level is also decreased when compared to the case of single-model predictions. But it is not enough to get an effective prediction; it is also important to know what areas affected the model to give the decision, so that clinical reliance is possible. Grad-CAM was run on the CNN part of the ensemble to produce what it can interpret as it did (establishment of interpretable explanations). Grad-CAM uses the flow of the target class into the last convolutional layers, which is the gradient, to create a rough localization image of the significant areas of the input image. In the case of ViT, the attention maps of the preceding self-attention layers were obtained, and they were then handled to emphasize the compelling patches. Given that ViTs are not based on convolutional layers, attention-based visualization is another framework that can be used to comprehend the role of global features in the model's decision.

4. Results and Discussion

Google Colaboratory (Colab) will be taken as the key development and implementation environment in this paper. Google Colab is a free, web-based application that lets researchers write and run Python code on a high-performance local machine without using a high-performance web browser. It has gained immense popularity with respect to machine learning and deep learning research due to the fact that these offer very powerful computational resources, including GPUs and TPUs, which also help in achieving high speed in training and testing models.

This study was conducted using Google Colab to train and evaluate deep learning models in the detection of tuberculosis (TB) in chest x-ray (CXR) images. Other notable libraries on the site are TensorFlow, Keras, OpenCV, NumPy, and Matplotlib that can be used in image pre-processing, feature extraction, feature development, and model performance analysis. Besides, it could be easily applied to Google Drive and enables more efficient storing and access to data, which contributes to making it a highly cost-effective and user-friendly tool in terms of implementing the deep learning-based system of medical image classification.

The model was optimized using the Adam optimizer, and this algorithm of updating model parameters employs an adaptive learning rate mechanism to update model parameters in an effective and efficient manner based on the historical gradients. The learning rate of 0.0001 was selected to offer effective and stable training, as well as various advantages, including faster convergence, automatic changes in the parameters, and results of high quality in medical imaging data. This model took 10 epochs to be trained, and during training, the training loss gradually decreased, and the training accuracy also increased. These results demonstrate that the model utilized had successful feature learning and a successful convergence, which demonstrates that the suggested strategy can be efficient in the process of identifying patterns of tuberculosis with CXR pictures.

4.1 Statistical Measures

In the assessment of the proposed model, some classic classification measures were used, which were evaluated in accordance with the metrics of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN):

Accuracy (ACC): $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Misclassification Rate (MCR): $$MCR = \frac{FP + FN}{TP + TN + FP + FN} = 1 - Accuracy$$

Specificity (SPC): $$Specificity = \frac{TN}{TN + FP}$$

Recall / Sensitivity (SEN): $$Sensitivity = \frac{TP}{TP + FN}$$

Precision / Positive Predictive Value (PPV): $$= \frac{Precision}{TP + FP}$$

Negative Predictive Value (NPV): $$NPV = \frac{TN}{TN + FN}$$

False Positive Rate (FPR): $$FPR = \frac{FP}{FP + TN} = 1 - Specificity$$

False Negative Rate (FNR): $$FNR = \frac{FN}{FN + TP} = 1 - Sensitivity$$

Negative Likelihood Ratio (LR−): $$LR- = \frac{1 - Sensitivity}{Specificity}$$

F1-Score: $$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

These are the basic statistics tools applied to measure the efficiency of the suggested classification model. These measures are based on the elements of the confusion matrix, that is, True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Accuracy and misclassification rate give a general understanding of the correctness of the model, whereas sensitivity (recall) and specificity give an understanding of how the model is able to detect positive and negative cases correctly, respectively. The reliability of the predictions is also further measured by precision and negative predictive value. Moreover, misclassification tendencies may also be seen through such error metrics as false positive and false negative rates. The addition of the likelihood ratios (LR+ and LR−) allows greater interpretation about the context of the uses of the decision, and the F1-score is highly useful, unlike the probability, which offers a rough score measure of the precision and recall, especially in skewed datasets. A combination of all the above

measurements guarantees a thorough measure of the model's performance in classification.

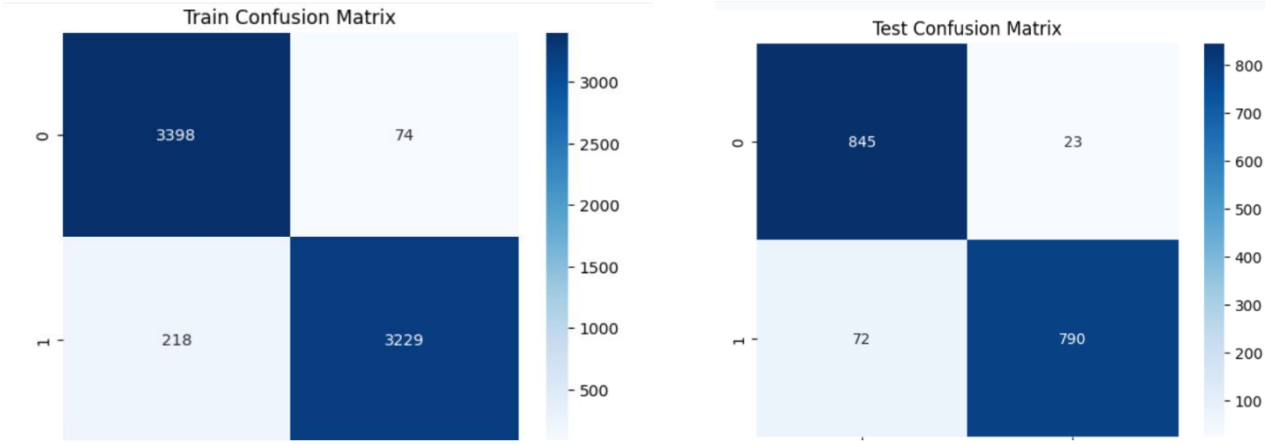


Figure 4.1: Confusion matrices of the proposed model on the training and testing sets. The matrices provide the summarization of the classification performance by the True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The misclassification rate is low, with a high level of correct classification, indicating effective learning, as shown by the training confusion matrix. In the same manner, the testing confusion matrix demonstrates steady testing on unseen information, attesting to the model's capability of generalisation and strength.

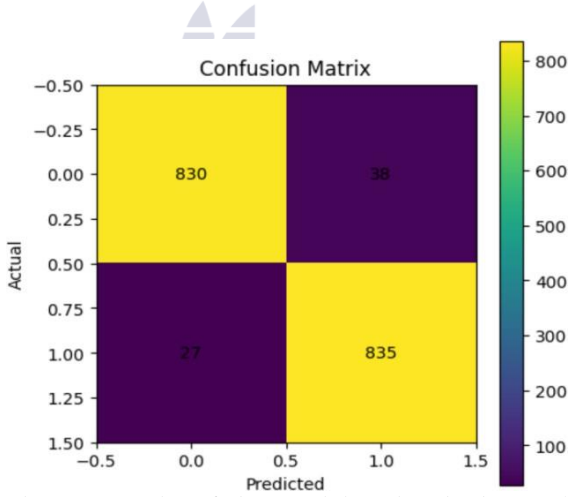


Figure 4.2: Matrix showing the test results of the model with a high number of correctly identified samples in both classes

The confusion matrix shows that the model is performing well on the whole, as it more correctly identifies samples in each of the classes. This implies that the model has learnt how to identify the differentiating trends, and it will be able to effectively make good predictions. The reality that the samples of misclassified incidents are not very many is also representative in the sense that the incident limits used during the training are understandable. The false positive and false negative values are minimal, and this implies that the model can recognize the smallest changes in image properties that are present in the trends that it receives during the training phase. The confusion matrix at the testing stage provides a closer view of the model in identifying the generalization capability. The majority of the test samples are categorized in the appropriate manner, and these results show the stability and high power of the trained model in the unknown data. Nonetheless, false

negatives exist several times, and they might be caused by a similar radiographic appearance or image quality. These fallacies note that medical interpretation of images is intrinsically complicated, and that recall is of particular concern when the performance is judged, particularly when the correct detection is essential to both operational and situational awareness.

Table 2:
Comparison of CNN and RepViT models on major performance metrics of classification, with the sacrifice of accuracy, precision, sensitivity, or specificity.

Measurement	CNN	RepViT
ACC	96.2	94.6
MCR	3.8	5.4
SPC	95.6	97.1
SEN	96.8	92.1
PPV	95.6	97.3
NPV	96.8	91.6
FPR	4.4	2.9
FNR	3.2	7.9
LR+	0.22	0.318
LR–	3.35	8.14
F1	0.957	0.946

The above table shows a comparison of the CNN model and RepViT model based on various evaluation metrics. The findings reveal that the CNN model has better predictive performance in terms of the overall accuracy (96.2) and misclassification rate (3.8) than RepViT, which has been lower. Another aspect in which CNN-based methods achieve a better score compared to RepViT is sensitivity (recall) and negative predictive value (NPV), since CNN can recognize positive cases and real negatives better. Contrastingly, RepViT has a better specificity (97.1%), accuracy (PPV = 97.3%), meaning that it is more qualified to correctly detect negative cases and decrease false positive estimates. Moreover, RepViT has a lower false positive rate (FPR), which is pertinent in the context of the application where false alarms are perilous. Nonetheless, it has a more significant false negative rate (FNR), which implies a compromise between false negatives.

The values of F1-score between the two models are comparable, but CNN is slightly higher than RepViT, which means that it is more balanced in its precision and recall. On the whole, it can be stated that both CNN and RepViT are effective patterns to compare because their performances are more balanced and accurate, whereas RepViT is better in terms of specificity and precision, and each model can be used in a specific application based on the priority of evaluation measures.

5. Explainable AI in Ensemble Learning

Explainable AI inclusion into ensemble learning has a twofold advantage of both accuracy and. Automated TB Detection interpretability. The suggested framework allows clinicians to visualize the areas that lead to model predictions, hence boosting trust and making decisions easier. The explanations between both models were combined to form an ultimate heatmap about the ensemble. decision-making process. Two methods were

discussed: a naive average between the single maps and a combination of them with weights applied, and they were determined by the performance of each model. validation data. The last heatmap was superimposed on the initial chest X-ray images, which indicated the existence of areas of the lungs that are most suggestive of TB infection. The spots marked in red in the heatmap represent regions with a high value of positive influence on the prediction of TB, and cooler regions imply non-relevant or normal areas.

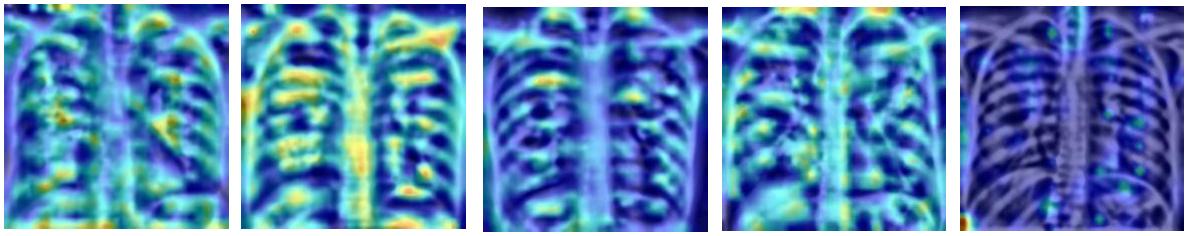


Figure 5.1: Grad-CAM visualization of Ensemble-based visualization of TB-relevant regions using chest X-ray images. The resulting heatmaps are created as a result of combining the explanations of several models, either by averaging or weighted approaches, with red areas implying high contribution to positive TB predictions.

6. Future Directions

Despite the high performance of the presented framework in the field of tuberculosis (TB) detection, there are some directions to consider in an attempt to increase its efficiency and applicability to the clinic. To begin with, future researchers can include bigger and more diversified multi-institutional data in their research to enhance model generalization and minimize possible bias in datasets. Introduction of other imaging techniques, including CT scans, might be useful as a complement to diagnostic evidence and enhance general detection rates. Second, it can explore new advanced deep learning architectures and hybrid models that help to maximize performance without compromising the computational efficiency, especially when deployed in resource-limited settings. Pruning, quantization, and knowledge distillation are also model optimization techniques that can be considered to allow real-time inference on edge devices. Moreover, the incorporation of clinical metadata (i. e., history, symptoms, and laboratory findings of the patient) and imaging data might strengthen the robustness and reliability of the diagnostic system. Regarding interpretability, further research can involve the addition of better explainable AI (XAI) methods to get more information and enhance the level of trust in clinicians.

Lastly, to analyse the practical utility of the proposed system, its reliability, and ethical implications, possible validation in real-world clinical practice and interaction with healthcare would be necessary before the system is rolled out on a large scale.

7. Conclusion

This research paper is an effective and explainable deep learning model of tuberculosis (TB) detection on the basis of training many publicly available datasets of chest X-rays (CXR). The proposed approach will be capable of alleviating inter-dataset disparity and bias because it will consolidate the datasets into a single one and apply the uniform means of preprocessing, including scaling of pictures, grayscale, CLAHE optimization, and class balance. It has been demonstrated that a CNN-based model gives superior diagnostic outcomes of 96.2, and the RepViT-based model also generates competitive outcomes with an accuracy of 94.6.

The explainable artificial intelligence (XAI) techniques (in particular Grad-CAM) were applied to detect the clinically important areas in the lung images, thereby enhancing model transparency and interpretability because deep learning models are black-box in nature. Taken collectively, the findings indicate that the multi-source datasets applied together with explainable deep learning may increase the robustness, generalisation, and clinical

reliability of TB detectors significantly. Such an approach exhibits a great potential of being applicable in real-life clinical practice and more so in resource-strained settings where speedy, quality, and interpretable screening tools will be critical.

8. Acknowledgments

The authors also recognized the accessibility of the clinical imaging data and the computer resources that enabled the successful conduction of the study.

REFERENCES

World Health Organization, "Global Tuberculosis Report 2023," Geneva, Switzerland, 2023.

K. D. A. Toman, "Tuberculosis case detection and diagnostic delays," *Int. J. Tuberc. Lung Dis.*, vol. 24, no. 4, pp. 345–352, 2020.

A. Rajpurkar *et al.*, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 6527–6537, 2017.

G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42,
pp. 60–88, 2017.

R. R. Selvaraju *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, 2020.

R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in Proc. IEEE Int.Conf. Computer Vision (ICCV), Venice, Italy, 2017, pp. 618–626.

M. M. Rahman, M. A. Moni, R. Islam, and M. R. Islam, "Deep learning–based classification of tuberculosis using chest X-ray images with segmentation and transfer learning," IEEE Access, vol. 8, pp. 98,123–98,135, 2020.

A. Lewis, E. Mahmoodi, Y. Zhou, M. Coffee, and E. Sizikova, "Improving tuberculosis (TB) prediction using synthetically generated computed tomography (CT) images," in Proc. IEEE/CVF Int. Conf. Computer Vision Workshops (ICCVW), Montreal, QC, Canada, 2021, pp. 3265–3273, doi: 10.1109/ICCVW54120.2021.00365.

G. J. Chowdary, G. Suganya, M. Premalatha, and K. Karunamurthy, "Class dependency based learning using Bi-LSTM coupled with the transfer learning of VGG16 for the diagnosis of tuberculosis from chest X-rays," arXiv preprint, arXiv:2108.04329, 2021.

W. Liang, S. Liang, G. Wang, J. Sun, J. Liu, C. Wang, H. Duan, J. Meng, and J. Li, "Artificial intelligence in pulmonary tuberculosis: A review of deep learning and radiomics applications," Frontiers in Medicine, vol. 9, Art. no. 935080, Jul. 2022, doi: 10.3389/fmed. 2022.935080.

A. Alhudhaif, M. Alsubaie, M. A. Khan, and S. Kadry, "Multi-Source Chest X-ray Dataset Fusion for Accurate and Explainable Tuberculosis Diagnosis," arXiv preprint arXiv:2309.02140, Sep. 2023.

E. Mahamud, N. Fahad, M. Assaduzzaman, et al., "A deep convolutional neural network–based framework for lung disease classification using chest X-ray images," Decision Analytics Journal, vol. 12,
p. 100499, 2024.

M. A. Khan, T. Akram, M. Sharif, M. Y. Javed, and N. Muhammad, "Tuberculosis detection from chest X-ray image modalities based on transformer and convolutional neural network," Computers in Biology and Medicine, vol. 168, Art. no. 107667, 2024, doi: 10.1016/j.compbiomed.2023.107667.

M. H. Rahman, M. S. Islam, M. R. Islam, and M. A. Hossain, "An automated deep learning approach for tuberculosis detection using chest X-ray images," BMC Pulmonary Medicine, vol. 24, no. 1, Art. no. 202, 2024, doi: 10.1186/s12880-024-01202-x.

A. M. Ayalew, N. W. Asnake, G. Demil, and M. Oussalah, "An explainable hybrid deep learning system for tuberculosis detection with Grad-CAM," Discover Computing, vol. 28, art. no. 262, 2025, doi: 10.1007/s10791-025-09791-z.

M. A. Khan, A. Rehman, S. Kadry, and Y. Nam, "Explainable Deep Learning Framework for Tuberculosis Detection Using Chest X-ray Images," arXiv preprint arXiv:2510.18819, Oct. 2025

A. Dosovitskiy *et al.*, "An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale," *Proc. Int. Conf. Learning Representations (ICLR)*, 2021. [Online]. Available: https://arxiv.org/abs/2010.11929

T. Rahman *et al.*, "Reliable Tuberculosis Detection Using Chest X-ray With Deep Learning, Segmentation and Transfer Learning," *IEEE Access*, vol. 8, pp. 191586–191601, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9208795

K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7780459

J. Johnson *et al.*, "Image Pre-processing Techniques for Chest X-Ray Images in Deep Learning- Based Medical Diagnosis," *Computers in Biology and Medicine*, vol. 145, Art. no. 105470, 2022.

S. He, X. Wang, and Y. Luo, "Transferring Vision Transformers for Chest X-Ray Classification," *IEEE Access*, vol. 10, pp. 58945–58956, 2022.

R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional Neural Networks: An Overview and Application in Radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, 2018.