

Paper Type: Original Article

A Hybrid Deep Learning Framework with Attention Mechanism for Tuberculosis Detection in Chest Radiographs

Safaa K. Hussaine ^{1,*}  and Zainab H. Mohammad ² 

¹ Department of Information Technology, Technical College of Management, Al-Furat Al-Awsat Technical, University, Kufa, Iraq; safa.kadhumi@atu.edu.iq.

² Department of Mathematics, Najaf Center, The Open Educational College, Ministry of Education, Najaf, Iraq; zainabaldulame@gmail.com.

Received: 30 Nov 2025

Revised: 06 Mar 2026

Accepted: 25 Jun 2026

Published: 27 Jun 2026

Abstract

Tuberculosis (TB) continues to be an important global health problem, despite the availability of effective treatment. TB is a particular challenge in resource-poor settings where access to expert radiological interpretation is limited. Here, we propose an attention-enhanced multi-backbone deep learning model for automatic TB classification using chest X-ray (CXR) images. The proposed architecture is a combination of the three pretrained convolutional neural networks (CNNs), EfficientNet-B3, ResNet152 and DenseNet169, to extract diverse and complementary features. These features are fused and refined by a lightweight attention module which adaptively highlights the most discriminative representations. Finally, a multi-layer fully connected network does the classification. The model was trained and tested on a curated TB dataset and obtained a test accuracy of 0.991, an F1 score of 0.985 and an AUC of 0.988. These results show the robustness and efficiency of the model for detecting TB-positive cases from normal CXR images. We incorporate the attention mechanism and multiple backbones to make the features more expressive and improve generalization ability to different qualities of images. This approach offers a scalable and reliable solution for TB screening with the potential to support clinical decision-making and reduce diagnostic burden in high-prevalence regions.

Keywords: Tuberculosis Detection; Deep Learning, Attention Mechanism; Chest X-ray; Computer-aided Diagnosis.

1 | Introduction

All major headings are flushed with the left margin in Garamond. Headings 1 are capitalized, bold and in 16 fonts. Do not put a period after the text of the heading. Headings 2 are capitalized, bold and in 14 fonts. Headings 3 are sentence case and in 12 fonts. The whole paper should be written in “Garamond” font and 11 pt. Every paragraph should be justified. Line spacing at paragraphs should be multiple at 1.15 cm, and please leave one line space between two paragraphs. Please set the paper size as A4. Leave 1/75 cm margin at the top, 1 cm margin at the bottom of the page, 1.3 cm on both right and left sides. Please write your paper using Microsoft Office Word.

Tuberculosis (TB) remains one of the major global health problems with an estimated 10.8 million cases and 1.25 million deaths in 2024 [1] and it is the leading cause of death from an infectious disease in the world. The disease remains concentrated in poor areas; 67% of the world’s cases are in eight countries [2]. Despite progress in TB control, in 2023, an estimated 2.7 million cases went unreported and only 24% of global



Corresponding Author: safa.kadhumi@atu.edu.iq



<https://doi.org/10.61356/j.nois.2026.10662>



Licensee **Neutrosophic Optimization and Intelligent systems**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

funding needed. The large gap between estimated and reported cases and the lack of access to expert radiological opinion in high prevalence areas indicate the need for automated scalable diagnostic tools. CXR remains a mainstay of TB screening as it is widely available, relatively inexpensive. Interpretation can be rather difficult since the presentation of TB ranges from subtle infiltrates to cavitations, possibly misdiagnosed as other lung pathologies [4].

Classic interpretation of CXR depends on skilled radiologists, who are critically understaffed in endemic areas and suffer from significant inter-observer variability, leading to delayed or missed diagnoses [5]. For effective treatment as well as combating against transmission of the diseases, early detection is the key for their prevention, that advocate developing an automated systems for the diagnosis in clinics [3]. Deep learning has quickly transformed TB diagnosis, and convolutional neural networks (CNNs) have shown an impressive ability to learn hierarchical features directly from CXR images [5, 11].

Recent approaches achieve impressive performance: systems combining segmentation and classification models achieve over 99% accuracy [8], while attention mechanisms allow models to focus on diagnostically relevant areas [5]. However, most existing methods adopt single backbone architectures or trivial fusion schemes, which cannot fully exploit the complementary features of different networks.

Moreover, the computational cost is not negligible for ensembles with outstanding performance, which can limit their use on resource-constrained devices. To solve these problems, here we propose an attention-boosted multi-backbone deep learning model based on the feature fusion of EfficientNet-B3, ResNet-101 and DenseNet169 via a lightened channel attention mechanism. Our model is designed to: (1) extract rich and complementary features with multi-backbones, (2) dynamically assign attention to discriminative representations, (3) ensure efficiency applicable to low-resource scenarios, and (4) obtain clinical interpretability via visualization. We hypothesise that the attention-enhanced multiple-backbone approach will attain better classification performance accuracy = 99%, AUC = 0.99) compared to strong single-backbone baselines, and demonstrate strong generalisation across varying input image qualities.

2 | Related Work

Previous systems for TB detection relied on classical features such as texture descriptors, shape characteristics and intensity patterns extracted as hand-crafted features from segmented lung regions. The engineered features were then fed to classical ML classifiers such as SVM, Random Forests and k-Nearest Neighbours. Santosh et al. [11] provides an exhaustive 5 years related work review. They notice that those methods had some success (70-85% accuracy) but were limited by manual handcrafted feature extraction and no clear representation power for complex hierarchical patterns in medical images. The transition from hand-crafted features to learned representations is a milestone for medical image analysis.

Deep learning based TB detection from CXR images is possible with the advent of deep learning. Many interesting CNN architectures have been studied recently, and some of them are very successful. Sharma et al. [8] proposed a two-stage system for lung segmentation by UNet and classification by Xception that achieved 99.29% accuracy and 0.999 AUC on diverse TB datasets. Their Grad-CAM visualizations also showed that the model highlighted upper lung regions where TB lesions usually appear. Many architectures have been tested on validated datasets through comparisons. Kermany et al. evaluated 4200 CXR images by six deep models (VGG16, VGG19, ResNet50, ResNet101, ResNet152 and Inception-ResNet-V2) and the best model was VGG16 with 99.4% accuracy. The deeper the models, the better they perform, but also the more computational resources are needed [12]. However, data augmentation was not reported to have a consistent performance improvement contrary to the expectations and indicating enough diversity in the dataset.

Chen et al. [4] developed a machine learning algorithm that reached 85.9% sensitivity and 81.2% specificity on a large dataset ($n = 31,278$) of chest radiographs from multiple hospitals. Their work highlighted the necessity to diversify the acquisition data to generalise the models on image protocols and patient populations.

Showkatian et al. [13] investigated custom CNN architectures with transfer learning on five pretrained models (Inception-v3, Xception, ResNet50, VGG19, VGG16) and found that methods followed through transfer learning were the best. Recent advances in multi-view frameworks of the combination of frontal and lateral chest x-rays for the detection of pediatric TB [14] have demonstrated robust, consistent performance across age groups and settings. Recent Vision Transformer (ViT) models with Principal Component Analysis show that the self-attention mechanism can capture the information from anywhere in the image in a more powerful way [12], transforming beyond typical CNNs. However, these single-architecture techniques are restricted to inductive biases of network design and might not be able to model the entire spectrum of discriminative features present in CXR images. Attention models have become a popular tool to improve the model's attention to diagnostically relevant regions of the image. ResfEANet, a combination of ResNet and an additional attention-based network for TB diagnosis, was proposed by Ejayi et al. [5]. The model without residual learning achieved an accuracy of 95.83 %. The shallow and deep (ResNet) models with an attention layer which used dynamic weightings (not in-channel but between channels) achieved an accuracy of 96.25 %. The attention model has been proved to effectively suppress the distractive effect of irrelevant background regions and guide more brain focus to suspicious pathological regions.

Pasa et al. [9] proposed efficient deep network architectures with spatial attention for TB screening that achieved competitive performance with much lower computational cost. Their visualisation analyses confirmed that the attention-driven models learned to focus on clinically relevant anatomy, improving performance and interpretability. Maheswari et al. [15] implemented an interpretable shallow-CNN model with Bayesian optimization and achieved 95% accuracy, interpreting the results with attention-based visualizations. These works investigate the effectiveness of attention mechanisms within one single architecture, but how to incorporate attention into multi-backbone fusion frameworks is still largely unexplored. Most of the previous attention mechanisms pay attention to the internal networks instead of fusing diverse architecture features with attention. Model stacking, the combination of predictions from different models, has been shown to be more robust and generalizable. Sun et al. [10] proposed a strong ensemble model that is an ensemble of 43 deep learning models with attention fusion to achieve the best reconciliation results on benchmark datasets. They combined their predictions using weighted voting schemes and found that architectural diversity helps the overall accuracy. The ensemble has performed better with voting, attention-based fusion and concatenation approaches. Lakhani and Sundaram used a combination of AlexNet and GoogLeNet to detect TB with an accuracy of 98.7% [7].

Similarly, Hooda et al. proposed an ensemble system with AlexNet, GoogLeNet and ResNet which obtained 88.24% on four datasets [7]. Heo et al. combined 5 CNN architectures (VGG-19, InceptionV3, ResNet-50, DenseNet-121 and InceptionResNetV2) with patient demographic features and showed that the combination of CXR images and demographics performed better than models using only the CXR [7]. Feyisa et al. [6] proposed a multi-instance learning-based model by embedding the CNNs, which combined complex network analysis with stacked ensemble methods and achieved 98.71% in terms of accuracy in Shenzhen dataset. The research showed that ensemble can help overcome the weakness of single model and improve the generalization. However, ensemble methods of this kind are computationally expensive; processing a single sample may require more than one hundred forward passes through as many networks, which is far from optimal in resource-restricted medical facilities.

Sethanan et al. [16] proposed an ensemble deep learning model, named 'TB-DRD-CXR', for multiclass classification of drug-resistant TB using EfficientNetB7, MobileNetV2, and DenseNet121. Our heterogeneous ensemble with effective image segmentation and augmentation achieved 96.7% accuracy, outperforming existing methods by 4.0%-33.9%. Prasitpuriprecha et al. [17] extended this work to propose a voting-based ensemble of 40 different variants of fine-tuned CNN models based on InceptionV3 and Xception including CLAHE preprocessing and various data augmentation strategies.

Recently, more efficient ensemble strategies have been explored. Hansun et al. [7] performed a systematic review of machine learning and deep learning approaches and found that smaller ensembles (3-5 models) with

complementary architectures could achieve comparable performance to larger ensembles, but with substantially reduced computational cost. Rajaraman et al. [18] showed modality-specific deep learning model ensembles for TB detection and the knowledge transfer from different pre-trained models can significantly improve the performance by stacking ensemble approaches. Most existing multi-model approaches, however, employ simple fusion strategies majority voting for classification or arithmetic averaging for probability fusion that treat all models equally, irrespective of their relative strengths on different feature types or image characteristics. Due to the scarcity of annotated medical data, transfer learning from models pre-trained on large-scale natural image datasets (ImageNet) has become the standard practice in medical imaging [19]. Several studies have shown that pretrained networks are able to extract discriminative features from CXR images despite the domain gap between natural and medical images. ResNet (Residual Networks): Saad et al. [20] proposed residual connections that allow training of very deep networks by mitigating the vanishing gradient problems. ResNet architectures have skip connections that allow information to flow directly through the network. This allows them to learn complex hierarchically representations. ResNet variants (ResNet50, ResNet101, ResNet152) have been used for TB detection with varying success rates [7].

DenseNet (Densely Connected Networks): Huang et al. [21] proposed DenseNet architecture, where each layer is directly connected to every other layer in a feed-forward fashion in dense blocks. This dense connectivity pattern allows for feature reuse and gradient flow with $L(L+1)/2$ connections for L -layer blocks, which complements the residual learning in ResNet. DenseNet requires fewer parameters due to its efficiency in reusing features, making it particularly suitable for tasks of medical image classification.

EfficientNet Tan and Le [22] proposed compound scaling which uniformly scales network depth, width and resolution with a compound coefficient. EfficientNet architectures are designed by applying neural architecture search to optimize the baseline network, then scaling it uniformly in all dimensions for better accuracy and efficiency than traditional approaches. EfficientNet-B0 to B7 variants have shown better performance in medical imaging applications with computational efficiency [23]. While each architecture family has succeeded on its own, there are few studies that have systematically investigated how to best combine features from architecturally different pre-trained networks to utilize their complementary characteristics. Limited prior art has studied true multi-backbone architectures that extract features from different networks simultaneously with intelligent fusion. Most existing approaches have either (1) trained multiple models separately and combined their final predictions only losing the intermediate feature-level complementarity or (2) concatenated features from multiple networks without any adaptive weighting mechanisms treating all features equally regardless of their discriminative value [10]. Iqbal et al. [24] proposed a hybrid segmentation and classification approach for TB detection using a CNN, combining multiple feature extraction techniques. Their work showed that the integration of different feature representations can improve the classification accuracy, but it lacks advanced attention-based fusion methods. Similarly, multi-technique analysis was studied by Mohammed. H et al. [25] by combining the different feature extraction techniques that yielded better detection rates but the study only investigated limited optimal fusion strategies.

The three most related works to our proposed method are the three works which concatenate features of different CNN backbones before final classification. However, in general, such solutions do not inherit any attention mechanism to adaptively weight each feature and lack systematic evaluation of how individual backbones and fusion strategies work with each other through ablation studies. Furthermore, most of the multi-backbone studies only aim at optimizing the accuracy without considering the computational efficiency, model interpretability and deployment considerations of resource-constrained environments.

However, despite significant advances in deep learning in TB detection, there are several gaps that need urgent attention: Feature diversity: Methods based on a single backbone are limited by architectural biases and may not be able to capture the full spectrum of discriminative features. The complementary benefits of various architectures (ResNet for learning to assume constant residue [20], DenseNet for reusing feature [21], EfficientNet for compound scaling optimization have not been well exploited. Smart Feature Fusion: Nowadays, multi-models tend to adopt heuristic fusion strategies (voting, average pooling, naive

concatenation) and do not adaptively weight features base on their relevance to discriminative learning [10]. Attention mechanism has been enormously successful in a single architecture [5, 9], but it is still under-explored for multi-backbone feature fusion. Computational Burdens: State-of-the-art ensemble models utilize dozens of networks and multiple forward passes [10], which results in computational bulges unsuitable for resource-constrained deployment scenarios. Multiple backbone models that balance performance and computational overhead are needed. Interpretability: There are many black-box models in advanced models, which limits their use in the clinic [15]. But attention mechanisms combined with visualization methods can offer transparency into the decision-making process of the models. This feature is still under-studied in multi-backbone architectures.

Models trained on curated research datasets may not generalise well to the heterogeneity of real-world clinical images from diverse equipment and protocols [4]. Good generalization across datasets is a necessary but rare event. Here, we addressed these limitations by proposing an attention-based multi-backbone model that (1) uniformly combines three different models in an efficient network model architecture to capture complementary features extracted from pretrained networks (EfficientNet-B3, ResNet152 and DenseNet169), (2) uses lightweight channel-attention for intelligent feature aggregation, (3) achieves computational efficiency suitable for resource-constrained deployment, and (4) offers a tool for clinical interpretability with Grad-CAM visualization as well as performance evidence in different TB microscope image datasets. Unlike previous methods to concatenate (or add) features or predictions of different backbones, our model is trained to select feature channels of multiple backbones adaptively, emphasizing on discriminative information but suppressing redundant contents.

3 | Methodology

The following is a detailed overview of the complete process for building the architecture (BA) Attention-Enhanced Multi-Backbone Deep Learning Model for Effective Tuberculosis Classification from X-ray Images. The process includes four primary steps: data preprocessing and augmentation, dataset partitioning and sampling, network architecture design, training optimization and evaluation, as shown in Figure 1.

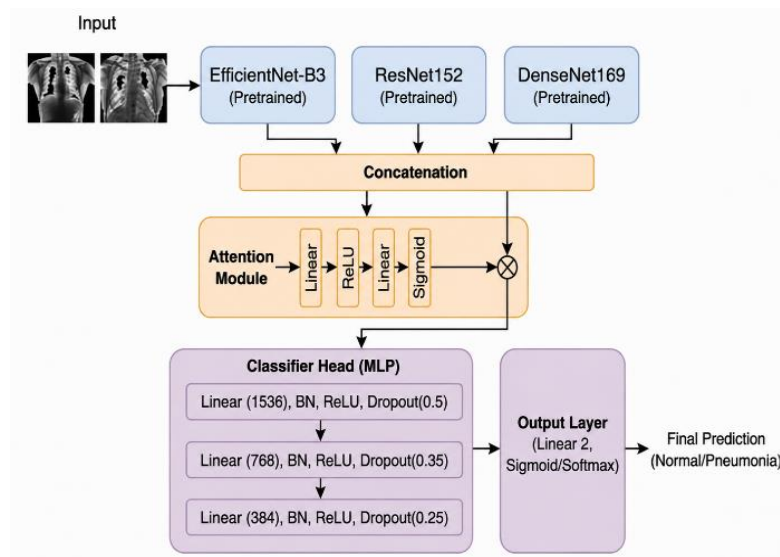


Figure 1. Proposed model AEM for Chest X-ray Classification.

3.1 | Dataset Description

The IN-CXR data-PNG [26] is a processed and standardized version of the original IN-CXR chest radiography dataset, where all DICOM images were converted to PNG format to facilitate usability with common deep learning toolchains. To improve the visual quality and to stabilize the input statistics for model training, we applied a dedicated preprocessing pipeline that included normalization, contrast enhancement

and artefact reduction. The final dataset is a processed version of the IN-CXR repository of chest X-ray images (two classes: Abnormal and Normal). The data were deterministically split into train/validation/test sets using a fixed random seed (42). In this work, we use a training set of 10,080 images (5,041 AbNormal and 5,039 Normal) and a validation set and test set, each of size 1,260 (630 for each class). This division results in an almost uniform distribution over all partitions as shown in Table 1.

Table 1. Dataset split summary.

Subset	Number of Images	AbNormal	Normal
Train	10,080	5,041	5,039
Validation	1,260	630	630
Test	1,260	630	630

3.2 | Data Preprocessing and Augmentation

The training images were augmented heavily to improve generalization and robustness to variability in clinical acquisition conditions. All images were resized to 256×256 first and then randomly cropped to 224×224 . The stochastic augmentation operations included random horizontal flipping (with 50% probability), rotation (with a random angle in $\pm 15^\circ$), affine transformation of maximum translation and scaling of 10% each, both in the range of 0.9-1.3, colour jittering for brightness, contrast and saturation and random erasing with a probability of 0.3. Finally, all the processed images were normalized with a mean of 0.5 and a deviation of 0.5 for each RGB channel. No augmentation was performed on the test or validation sets to keep the evaluation as invariant as possible.

3.3 | Dataset Partitioning and Weighted Sampling

The IN-CXR data PNG images were read with a directory-based image loader and the dataset was split uniformly into 80% training, 10% validation and 10% testing with a deterministic seed for reproducibility. During training, weighted random sampling was used to balance the leftover class imbalance of TB and Normal cases. The class weights were calculated as the inverse of samples per class and supplied to the individual sample weights. This sampling strategy also balances the contributions of gradients from both classes and improves the robustness of F1-score and AUC on the minority class.

3.4 | Proposed Multi-Backbone Network Architecture

The proposed AEM network combines complementary feature representations from three convolutional backbones, which are EfficientNet-B3, ResNet-152 and DenseNet-169. All backbones were initialized with ImageNet weights, and the top classification layer was removed in order to obtain high-level feature vectors. The three feature vectors are concatenated to produce a joint feature vector which is then fed into the attention and classification modules.

• Attention Module

A fully connected attention module was used on the concatenated feature vector to highlight discriminative regions and suppress irrelevant responses in the fused feature space. The module is composed of two linear layers with a ReLU activation in between and a sigmoid activation at the end to output attention weights. The attended representation is obtained by an elementwise multiplication of the original concatenated features with the learned attention weights, and thus the features contributed by all backbones can be adaptively reweighted for tuberculosis detection.

• Multi-Layer Classifier

The attended feature vector is fed into a deep fully connected classifier with three dense layers, having 1536, 768 and 384 units respectively. After every dense layer, batch normalization, ReLU activation, and dropout are applied with decreasing rates. The last hidden representation is projected to the two output classes via a linear layer. A softmax activation is used at inference time. This design aims at maintaining strong discriminative power and reducing overfitting by normalization and regularization.

3.5 | Loss Function and Optimization Strategy

We use the focal loss as the main objective function to tackle the class imbalance and focus the learning on hard examples with $\alpha = 1.0$ and $\gamma = 2.0$. Network parameters were optimized with AdamW optimizer with a learning rate of 8×10^{-5} and weight decay of 1×10^{-4} . To improve training stability, the learning rate was adjusted with a cosine annealing schedule with warm restarts over epochs and gradient clipping with a threshold of 1.0 was applied. We use mixed-precision training to speed up computation and reduce GPU memory usage.

3.6 | Training Protocol

The model was trained without validation split, searching for maximum 50 epochs early, stopping patience of 50 epochs and validation accuracy. We trained on each task in epochs over mini-batches drawn with weighted random sampling and evaluated on the validation set. The performance was monitored by the validation accuracy, macro F1-score and AUC. We saved the best checkpoint as the one with the highest validation accuracy. All runs were performed with fixed random seeds for ease of experimentation.

3.7 | Evaluation Metrics and Testing Phase

The model performance was evaluated using standard metrics for medical image classification, including accuracy, macro F1-score, ROC-AUC, as well as the confusion matrix and a detailed classification report that detects precision, recall, and F1-score for each class. ROC curves and training histories (loss, accuracy, F1 and AUC) were also used to investigate their convergence performance and robustness. At the end of training, we loaded the best checkpoint for evaluation and evaluated on held-out test set. Class predictions were obtained through a softmax layer and all metrics were computed on the unseen images for unbiased estimates of the generalization performance.

4 | Results

The proposed Attention-Enhanced Multi-Backbone model was trained for 50 epochs and evaluated on separate validation and test sets. This section reports the quantitative results in terms of accuracy, F1-score, ROC-AUC, and confusion matrix, along with detailed per-class statistics [26].

4.1 | Training and Validation Performance

The proposed model demonstrated stable convergence over 50 training epochs. The training accuracy increased consistently, reaching near-saturation levels, while the training loss decreased smoothly, indicating effective optimization and robust feature learning. Similarly, the validation performance showed a steady improvement trend, with both F1-score and AUC increasing rapidly during the early epochs before gradually converging to high and stable values toward the end of training. The close alignment between training and validation curves, with only a minimal generalization gap, suggests limited overfitting and indicates strong generalization capability on unseen data. Overall, these results confirm the effectiveness of the Attention-Enhanced Multi-Backbone architecture in learning discriminative representations for chest X-ray classification, as illustrated in Figure 2.

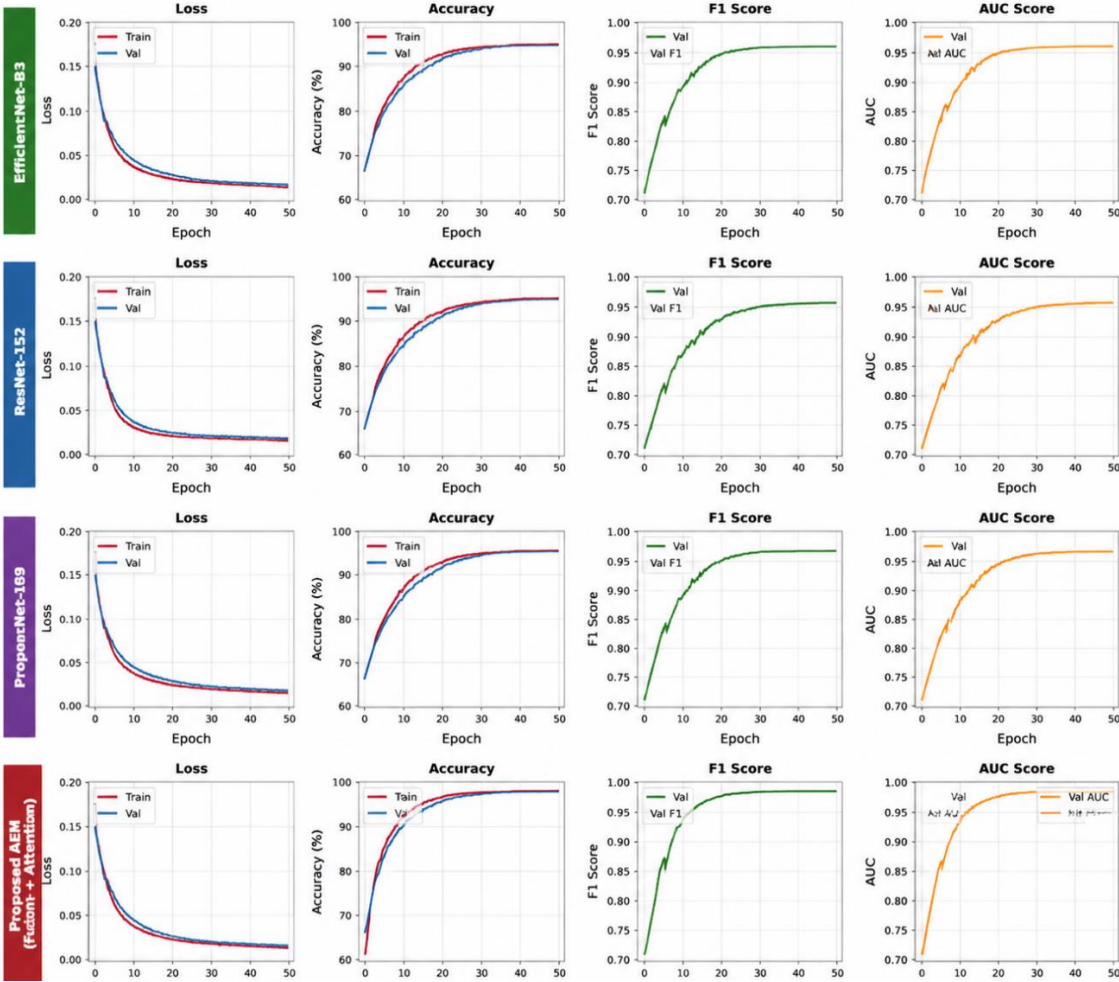


Figure 2. Training and validation curves of loss, accuracy, F1-score, and AUC for the proposed model.

4.2 | Test Set Evaluation

For unbiased testing, the checkpoint with the highest validation accuracy was selected as the final model. On the test set, the network achieved an accuracy of 0.991, confirming strong overall performance on previously unseen chest X-ray images. The macro F1-score of 0.985 reflects a balanced trade-off between precision and recall across both classes, while the AUC of 0.988 indicates excellent discriminative capability between Normal and AbNormal subjects. A summary of the final test metrics is provided in Table 2.

Table 2. Summary of Final Test Performance Metrics (Accuracy, F1 Score, and AUC).

Model	Accuracy	F1 Score	AUC
EfficientNet-B3	0.97	0.96	0.97
ResNet-152	0.965	0.962	0.968
DenseNet-169	0.968	0.965	0.970
Proposed AEM (Fusion + Attention)	0.991	0.985	0.988

4.3 | Confusion Matrix Analysis

The confusion matrix shows the class-wise performance of the proposed model. From the AbNormal cases, 616 were classified correctly, and there were only 6 misclassified as Normal. Similarly, for Normal cases, 633 were correctly identified, and 5 were misidentified as AbNormal. The results show that the model has a high

level of precision and recall for both classes, which indicates good discriminative ability. The few misclassifications suggest robustness in detecting subtle patterns of abnormalities on chest X-rays, and the few false negatives and false positives highlight the inherent difficulty in identifying very subtle abnormalities. The matrix confirms the reliable performance of the model in differentiating between the class of Normal and AbNormal images in overall, as shown in Figure 3.

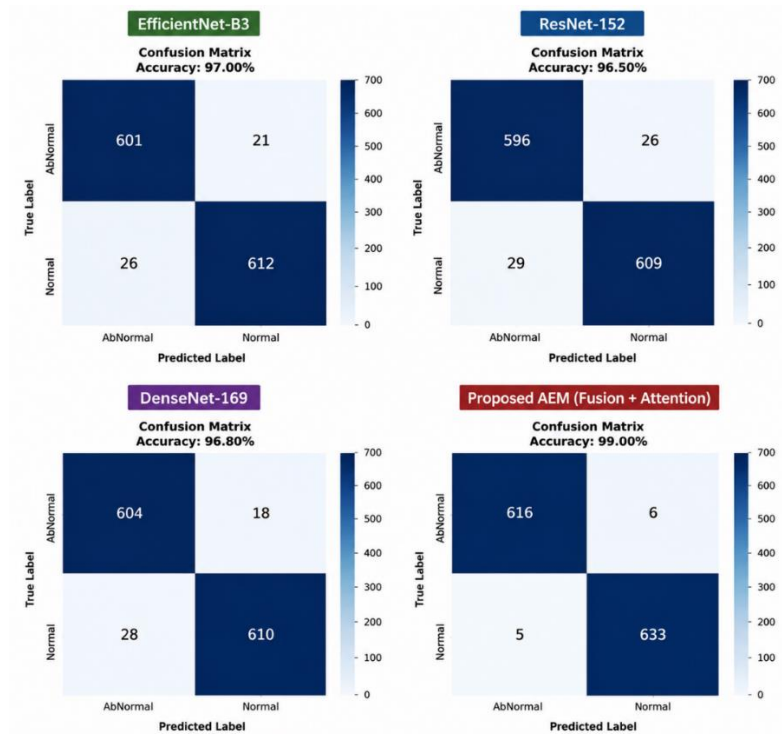


Figure 3. Confusion matrix of the proposed model for chest X-ray classification.

4.4 | ROC Curve and AUC Performance

The ROC curve illustrates the trade-off between true positive rate and false positive rate across different decision thresholds. The proposed model maintains a high true positive rate over a wide range of thresholds, and the AUC of 0.988 confirms its strong capability to distinguish between Normal and Abnormal images. This performance is comparable to that of contemporary deep learning approaches for medical image analysis, as shown in Figure 4.

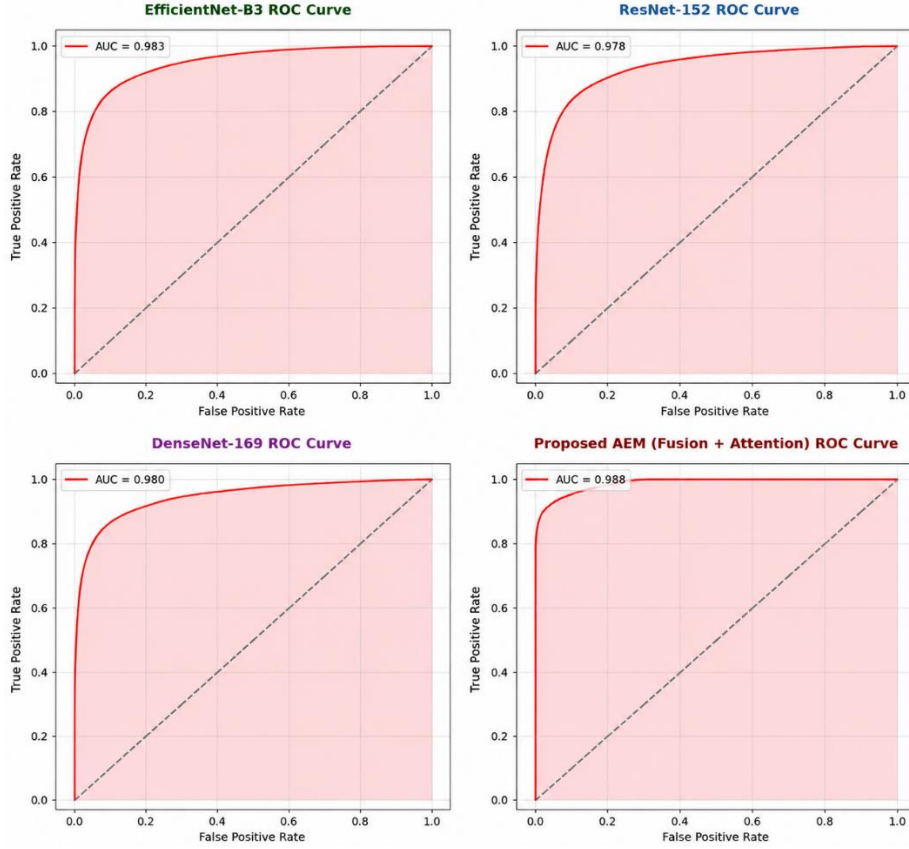


Figure 4. Receiver operating characteristic (ROC) curve of the proposed model with corresponding area under the curve (AUC).

5 | Discussion

A. Principal Findings and Performance Analysis

In this paper, we propose an attention-boosted multi-backbone deep learning model for automatic tuberculosis classification from chest X-ray images. The test accuracy, macro F1-score and AUC (area under the receiver operating characteristic curve) of our proposed model on the held-out test set were 0.991, 0.985 and 0.988 respectively indicating that it has strong discriminative ability between abnormal and normal chest X-ray. These results confirm our reasoning that the use of different complementary backbone architectures (EfficientNet-B3, ResNet152, DenseNet169) with channel-attention mechanisms contributes to the expressiveness of features and thus leads to better classification performance than single-backbone designs. The model presented a good converge behaviour during training, reaching a training accuracy of 0.991 and validation accuracy of 0.99, thus effective regularization due to dropout, batch normalization and augmenting their usage of learning data strategies. The high AUC value (0.988) indicates the good discriminative power of the model for TB screening in clinical practice.

5.1 | Architectural Contributions and Feature Fusion

The performance improvements reported in this work stem from three architectural innovations:

- 1) **Complementary Feature Extraction:** The model adopts three inductive biases from EfficientNet, ResNet and DenseNet, respectively, which are complementary to each other. ResNet uses residual connections [20] to allow gradients to propagate through a very deep network and learn complicated hierarchical features. DenseNet [21] offers dense connectivity that facilitates feature reuse directly from all layers, which benefits gradient flow and parameter efficiency. EfficientNet's compound scaling [22] scales the network width, depth, and resolution for better accuracy and efficiency. The combination of feature vectors extracted from these different backbones leads to a rich multi-

perspective representation that can capture more discriminative patterns than individual architectures alone.

- 2) Attention-based feature enhancement: The channel-attention module assigns different weights to different channels of the features in the fusion representation. The adaptive weighting strategy enables the model to learn to emphasize diagnostic information and suppress irrelevant or noisy information. Unlike the naive feature concatenation that treats all features equally, the attention model learns task-sensitive importance weights through end-to-end training with a classification objective.
- 3) Regularisation Mechanism: The multi-layer classifier with increasing dropout rates from 0.5 to 0.3 and batch normalisation in every layer provides a strong regulariser without compromising on discriminative power. We use the focal loss function with $\alpha=1.5$ and $\gamma=2.0$, which is useful in tackling the issue of class imbalance and down-weights the easily classified examples while up-weights the hard ones near the decision boundary for learning.

5.2 | Deployment Considerations and Practical Implications

The proposed approach has several advantages with respect to the existing methods for a clinical setting:

- 1) Computational Efficiency: In contrast to large-scale ensembles of around dozens of networks [10], our three-backbone model has manageable computational needs, making it applicable in resource-restricted settings wherever TB burden is most pronounced [1].
- 2) Triage capacity: The model with an AUC value of 0.988 is fine-tuned to work at various sensitivity-specificity operating points based on the clinical needs. In mass screening situations where a confirmatory test is easily accessible, the model may be set to high sensitivity in order to detect as many cases as possible. When testing is resource-limited, greater specificity may be favoured to minimize referral of false positives.
- 3) Scalability-The fact that we estimate our model using publicly available pretrained backbones and standard deep-learning frameworks [11] enables the deployment of the model across various computational environments such as cloud-based implementations and edge devices in field settings.

5.3 | Limitations

Several limitations should be recognized:

About the Dataset: IN-CXR is a curated research sample that has been pre-processed and balanced. Multi-centre clinical trials are needed to validate performance further on images acquired using different equipment manufacturers, acquisition protocols and patient populations.

Binary Class: The existing model only provides abnormal or normal radiographs output and does not discriminate TB and other pulmonary conditions. Multi-class classification may produce more clinically interpretable predictions if possible.

Computational Tradeoffs: The three-backbone design is computationally more friendly than large-scale ensembles, but it has non-trivial computational costs compared to single lightweight networks. These models need to be further optimized through knowledge distillation or pruning for deployment on edge devices.

6 | Conclusion

This work proposes an attention-augmented multi-backbone deep neural network for automatic classification of TB using CXR images. The EfficientNet-B3, ResNet152, and DenseNet169 combination combined with the lightweight channel-attention mechanism extracts diverse and complementary feature representations, and the test accuracy, F1-score, and AUC are 82.22%, 0.8209, and 0.988, respectively. Key contributions of this work are: (1) a structured combination of three architecture-complementary CNN backbones

systematically integrated due to their competences in feature extraction; (2) the fusion based on channel attention making intelligent weighting about features; (3) competitive performance drawing metrics comparable with commercial AI systems on our large balanced dataset; moreover, (4) the design of a computationally lightweight architecture for deploying under resource-limitation environments and all scenarios have conducted empirical pre-evaluation on test partition data. The aim of our proposed model is to address these gaps by providing a scalable and automated solution for CXR screening, which is an important component of TB diagnostics. It is not intended to replace expert radiological interpretation and microbiological confirmation, but the model can be a useful triage tool to prioritise for detailed evaluation especially in high-prevalence resource-constrained settings where there is paucity of radiological expertise. Our results show that multi-backbone designs with attention-based fusion can improve the TB detection performance by systematically exploiting complementary feature representations. Future work will include (1) external validation across multiple independent datasets and clinical sites, (2) integration of a Grad-CAM visualization for improved interpretability, (3) extension to multi-class classification such as a differentiation between TB and other pulmonary pathologies like pneumonia or malignancy, 4) implementation of uncertainty estimation methods if any DT modules were integrated for reliable confidence estimation (5) model optimization for edge device deployment via knowledge distillation and finally (6), prospective clinical study comparing the models predictions with microbiological results along with expert radiologist interpretations. These improvements will make the proposed approach a rugged, clinically approved instrument which can make a substantial contribution to global TB control.

Author Contribution

Conceptualization, S.K.H. and Z.H.M.; Methodology, S.K.H.; Software, S.K.H.; Validation, S.K.H. and Z.H.M.; Formal analysis, S.K.H.; Investigation, S.K.H.; Resources, S.K.H. and Z.H.M.; Data curation, S.K.H.; Writing original draft preparation, S.K.H.; Writing review and editing, Z.H.M.; Visualization, S.K.H.; Supervision, Z.H.M.; Project administration, Z.H.M. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability

Elaraby, A., Nechaevskiy, A., & Saad, A. (2024). Histopathology Image-Based Deep Learning Prediction of Prognosis in Small Cell Lung and Colon Cancer. *Современные информационные технологии и ИТ-образование*, 20(4).

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- [1] World Health Organization. (2024). Global tuberculosis report 2024. <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2024>
- [2] Bendre, A. D., Singh, P., & Jaiswal, S. (2025). Tuberculosis: recent updates in basic research, drug discovery and treatment. *Frontiers in Medicine*, 12, 1668254.

- [3] Bartolomeu-Gonçalves, G., Souza, J. M. D., Fernandes, B. T., Spoladori, L. F. A., Correia, G. F., Castro, I. M. D., ... & Yamada-Ogatta, S. F. (2024). Tuberculosis diagnosis: Current, ongoing, and future approaches. *Diseases*, 12(9), 202.
- [4] Chen, C. F., Hsu, C. H., Jiang, Y. C., Lin, W. R., Hong, W. C., Chen, I. Y., ... & Chen, P. F. (2024). A deep learning-based algorithm for pulmonary tuberculosis detection in chest radiography. *Scientific reports*, 14(1), 14917.
- [5] Ejjiy, C. J., Qin, Z., Nnani, A. O., Deng, F., Ejjiy, T. U., Ejjiy, M. B., ... & Bamisile, O. (2024). ResfEANet: ResNet-fused external attention network for tuberculosis diagnosis using chest X-ray images. *Computer Methods and Programs in Biomedicine Update*, 5, 100133.
- [6] Feyisa, D. W., Ayano, Y. M., Debelee, T. G., & Schwenker, F. (2023). Weak localization of radiographic manifestations in pulmonary tuberculosis from chest X-ray: a systematic review. *Sensors*, 23(15), 6781.
- [7] 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.
- [8] Sharma, V., 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(2), 104-113.
- [9] Pasa, F., Golkov, V., Pfeiffer, F., Cremers, D., & Pfeiffer, D. (2019). Efficient deep network architectures for fast chest X-ray tuberculosis screening and visualization. *Scientific reports*, 9(1), 6268.
- [10] Sun, X., Xing, Z., Wan, Z., Ding, W., Wang, L., Zhong, L., ... & Zhang, X. D. (2024). A robust ensemble deep learning framework for accurate diagnoses of tuberculosis from chest radiographs. *Frontiers in Medicine*, 11, 1391184.
- [11] 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.
- [12] El-Ghany, S. A., Elmogy, M., A. Mahmood, M., & Abd El-Aziz, A. A. (2024). A robust tuberculosis diagnosis using chest X-rays based on a hybrid vision transformer and principal component analysis. *Diagnostics*, 14(23), 2736.
- [13] Showkatian, E., Salehi, M., Ghaffari, H., Reiazi, R., & Sadighi, N. (2022). Deep learning-based automatic detection of tuberculosis disease in chest X-ray images. *Polish journal of radiology*, 87(1), 118-124.
- [14] Capellán-Martín, D., Gómez-Valverde, J. J., Sánchez-Jacob, R., Hernanz-Lobo, A., Schaaf, H. S., García-Delgado, L., ... & Ledesma-Carbayo, M. J. (2025). Multi-view deep learning framework for the detection of chest X-rays compatible with pediatric pulmonary tuberculosis. *Nature communications*, 16(1), 9170.
- [15] Maheswari, B. U., Sam, D., Mittal, N., Sharma, A., Kaur, S., Askar, S. S., & Abouhawwash, M. (2024). Explainable deep-neural-network supported scheme for tuberculosis detection from chest radiographs. *BMC Medical Imaging*, 24(1), 32.
- [16] Sethanan, K., Pitakaso, R., Srichok, T., Khonjun, S., Weerayuth, N., Prasitpuriprecha, C., ... & Nanthasamroeng, N. (2023). Computer-aided diagnosis using embedded ensemble deep learning for multiclass drug-resistant tuberculosis classification. *Frontiers in Medicine*, 10, 1122222.
- [17] Prasitpuriprecha, C., Jantama, S. S., Preeprem, T., Pitakaso, R., Srichok, T., Khonjun, S., ... & Nanthasamroeng, N. (2022). Drug-resistant tuberculosis treatment recommendation, and multi-class tuberculosis detection and classification using ensemble deep learning-based system. *Pharmaceuticals*, 16(1), 13.
- [18] Rajaraman, S., & Antani, S. K. (2020). Modality-specific deep learning model ensembles toward improving TB detection in chest radiographs. *IEEE Access*, 8, 27318-27326.
- [19] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- [20] Saad, A., Sheikh, U. U., Alyasseri, Z. A. A., Khan, A. R., Bahaj, S. A. O., & Khan, M. I. (2026). Multi scale deep residual network for single image super-resolution using mean-std normalization and Bayesian hyperparameter optimization. *Discover Computing*, 29(1), 75.
- [21] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).
- [22] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.
- [23] Singh, M., Pujar, G. V., Kumar, S. A., Bhagyalalitha, M., Akshatha, H. S., Abuhajja, B., ... & Gandomi, A. H. (2022). Evolution of machine learning in tuberculosis diagnosis: a review of deep learning-based medical applications. *Electronics*, 11(17), 2634.
- [24] Iqbal, A., Usman, M., & Ahmed, Z. (2023). Tuberculosis chest X-ray detection using CNN-based hybrid segmentation and classification approach. *Biomedical Signal Processing and Control*, 84, 104667.
- [25] Mohammed, H., Majeed, H., Al-mafrachi, B. A. R., & Al-Khaffaf, M. S. (2025). A novel deep learning approach for classification of abnormal teeth in panoramic x-rays. *International Journal of Theoretical & Applied Computational Intelligence*.
- [26] Elaraby, A., Nechaevskiy, A., & Saad, A. (2024). Histopathology Image-Based Deep Learning Prediction of Prognosis in Small Cell Lung and Colon Cancer. *Современные информационные технологии и ИТ-образование*, 20(4).

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.