



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Hybrid Explainable Deep Learning Framework for Tuberculosis Detection Using Chest X-Ray Images and Clinical Data Fusion

Priyanga M¹, Sai Shanmuga Raja V², Rajavarman VN³ and Lakshmi Dhevi B⁴

¹Assistant Professor, Department of Data Science and Artificial Intelligence, Dr. M.G.R. Educational and Research Institute, Chennai, India, Email: priyanga.cse@drmgrdu.ac.in

²Professor, Department of Computer Science and Engineering, Dr. M.G.R. Educational and Research Institute, Chennai, India, Email: saishanmugaraja@gmail.com

³Professor, Department of Information Technology, Dr. M.G.R. Educational and Research Institute, Chennai, India, Email: rajavarman.vn@drmgrdu.ac.in

⁴Assistant Professor, Department of Networking and Communications, Faculty of Engineering & Technology, SRM Institute of Science and Technology, Chennai, India, Email: lakshmidhevi.b@gmail.com

Corresponding Author: Sai Shanmuga Raja V*: Email ID: saishanmugaraja@gmail.com

Abstract

Tuberculosis (TB) remains one of the leading infectious diseases worldwide, particularly in low- and middle-income countries where rapid diagnosis is still challenging due to limited healthcare infrastructure and shortage of expert radiologists. This paper proposes a hybrid explainable deep learning framework for automated tuberculosis detection using chest X-ray images integrated with clinical information and secure healthcare data management principles. The proposed framework combines a customized Convolutional Neural Network (CNN), EfficientNetB0 feature extraction, and an Extreme Learning Machine (ELM) classifier to improve diagnostic accuracy and computational efficiency. In addition, lung segmentation using UNet++ is incorporated to enhance region-of-interest extraction and reduce irrelevant image features. Explainable Artificial Intelligence (XAI) techniques including Grad-CAM, SHAP, and LIME are integrated to improve interpretability and clinical trustworthiness. Experimental evaluation was conducted using publicly available tuberculosis chest X-ray datasets including Shenzhen, Montgomery, and TB Chest X-ray datasets. The proposed model achieved an accuracy of 99.21%, precision of 98.74%, recall of 99.03%, and F1-score of 98.88%, outperforming existing deep learning approaches such as DenseNet121, ResNet50, VGG16, and standalone EfficientNetB0 models. Comparative analysis demonstrates that the integration of segmentation, hybrid classification, and clinical data fusion significantly improves TB detection performance while reducing false positives. The proposed system can support early diagnosis, improve clinical decision-making, and assist healthcare professionals in resource-constrained environments.

Keywords: Tuberculosis Detection, Deep Learning, Chest X-Ray, Explainable AI, EfficientNetB0, Extreme Learning Machine, Medical Image Analysis, Hybrid CNN, Healthcare AI

1. INTRODUCTION

Tuberculosis (TB) is a highly contagious infectious disease caused by *Mycobacterium tuberculosis*. According to the World Health Organization (WHO), tuberculosis continues to be one of the major causes of mortality globally, especially in developing countries where access to advanced diagnostic facilities remains limited [1]. Early diagnosis is critical for reducing disease transmission and improving patient survival rates. However, conventional diagnostic methods such as sputum microscopy, culture testing, and molecular diagnostics are time-consuming, expensive, and often unavailable in remote regions.

Chest radiography remains one of the most accessible diagnostic tools for pulmonary tuberculosis. However, interpretation of chest X-rays depends heavily on experienced radiologists, and inter-observer variability may lead

to delayed or inaccurate diagnosis. Artificial Intelligence (AI), particularly deep learning techniques, has shown remarkable success in medical image analysis by automatically learning discriminative features from imaging data [2].

Recent advancements in Convolutional Neural Networks (CNNs) have enabled automated detection of diseases such as pneumonia, COVID-19, lung cancer, and tuberculosis. Although several deep learning approaches have achieved promising performance, many models suffer from limitations including poor interpretability, high computational complexity, and sensitivity to noisy image regions.

To address these challenges, this research proposes a hybrid explainable deep learning framework integrating lung segmentation, feature extraction, hybrid classification, and explainable AI mechanisms for accurate tuberculosis detection. The framework combines UNet++-based lung segmentation with EfficientNetB0 feature extraction and Extreme Learning Machine (ELM) classification [3]. Furthermore, clinical data fusion concepts are incorporated to improve contextual understanding and diagnostic reliability.

2. RELATED WORKS

Deep learning-based medical image analysis has gained significant attention in recent years for automated disease diagnosis. Several studies have investigated CNN architectures for tuberculosis detection using chest X-ray images [4].

Shadman Sobhan et al. proposed a multi-stage deep learning framework combining UNet++ segmentation and EfficientNetB0 classification for tuberculosis detection. Their model achieved high classification accuracy while improving interpretability through XAI techniques such as Grad-CAM and SHAP [5].

Vyas and Khadatkhar developed an ensemble learning architecture using AlexNet, GoogleNet, and ResNet for pulmonary TB classification. Their approach demonstrated improved Area Under Curve (AUC) performance but suffered from high computational overhead [6].

Rahman et al. introduced TB-UNet with attention-based segmentation mechanisms for extracting lung regions from chest radiographs. The segmentation results significantly enhanced classification accuracy [7].

Several researchers have also explored explainable AI approaches to improve trust in medical AI systems. Grad-CAM and LIME have been widely adopted for visualizing disease-relevant regions in chest radiographs.

Recent studies have further investigated data fusion methods combining clinical records and imaging data for enhanced diagnostic performance. Machine learning models utilizing Electronic Medical Records (EMR), demographic features, and clinical indicators have shown improved sensitivity in TB prediction.

Despite these advancements, challenges remain in balancing accuracy, interpretability, computational efficiency, and clinical applicability. The proposed work addresses these issues through a hybrid explainable framework integrating segmentation, hybrid classification, and data fusion.

3. MATERIALS AND METHODS

3.1 Dataset Description

The proposed framework was evaluated using publicly available tuberculosis chest X-ray datasets.

3.1.1 Shenzhen Dataset

The Shenzhen dataset contains 662 posterior-anterior chest radiographs collected from Shenzhen No. 3 Hospital, China. The dataset includes both normal and tuberculosis-positive images.

3.1.2 Montgomery Dataset

The Montgomery County dataset contains 138 chest X-ray images including normal and TB-positive cases collected by the U.S. National Library of Medicine.

3.1.3 TB Chest X-ray Dataset

An additional TB chest X-ray dataset consisting of 4200 images was utilized for training and testing. The dataset contains 3500 normal images and 700 tuberculosis-positive samples.

3.1.4 Clinical Data Integration

Clinical attributes including age, gender, smoking status, HIV condition, fever, cough duration, weight loss, and previous medical history were incorporated for data fusion experiments.

Table 1 Dataset Summary

Dataset	Images	Normal	TB	Masks
TB CXR DB	4200	3500	700	0
Shenzhen	662	326	336	566
Montgomery	138	80	58	138

3.2 Preprocessing

Image preprocessing is a critical stage in medical image analysis because raw chest X-ray images often contain noise, varying illumination conditions, low contrast, and irrelevant artifacts that can negatively affect the learning capability of deep learning models. Proper preprocessing enhances image quality, improves feature visibility, reduces unwanted distortions, and enables the neural network to learn diagnostically relevant patterns more effectively. In the proposed framework, several preprocessing operations were systematically applied before segmentation and classification.

3.2.1 Image resizing

The first preprocessing step involved image resizing to a standardized resolution of 224×224 pixels. Medical datasets are usually collected from multiple hospitals and imaging devices, resulting in images with different resolutions and aspect ratios. Deep learning architectures such as EfficientNetB0 require fixed-size inputs for efficient batch processing and optimized convolution operations. Therefore, all chest X-ray images were resized to $224 \times 224 \times 3$ dimensions while preserving essential anatomical information. Standardizing image dimensions reduced computational complexity, improved GPU memory utilization, and enabled consistent feature extraction across all samples.

3.2.2 Contrast Limited Adaptive Histogram Equalization (CLAHE)

Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied to improve local contrast enhancement within chest radiographs. In many tuberculosis X-ray images, infected pulmonary regions appear with subtle intensity variations that are difficult to identify using raw grayscale images. Traditional histogram equalization methods may over-amplify noise in homogeneous regions, leading to distorted image structures. CLAHE overcomes this limitation by dividing the image into small contextual regions and applying histogram equalization independently within each region while restricting excessive amplification through a clipping limit [8]. This process significantly improved the visibility of lung opacities, infiltrates, nodules, and cavitary lesions associated with tuberculosis. Enhanced local contrast enabled the CNN model to capture fine-grained pathological features more accurately.

3.2.3 Noise removal using median filtering

Noise removal using median filtering was incorporated to eliminate unwanted image artifacts while preserving edge information. Chest X-ray images often contain salt-and-pepper noise, acquisition artifacts, and scanner-related distortions that may interfere with feature learning. Median filtering is a nonlinear spatial filtering technique where each pixel value is replaced by the median value of its neighboring pixels [9]. Unlike linear smoothing filters, median filtering effectively removes impulsive noise without excessively blurring important anatomical boundaries such as lung contours and lesion edges. Preserving edge sharpness is particularly important in tuberculosis detection because lesion boundaries and infiltrative patterns contribute significantly to disease classification.

3.2.4 Min-max normalization

Min-max normalization was performed to standardize pixel intensity values into a uniform numerical range, typically between 0 and 1. Medical images obtained from different radiographic systems often exhibit varying intensity distributions due to exposure conditions and scanner calibrations. Feeding unnormalized data into deep learning models can result in unstable gradient updates and slow convergence during training. Min-max normalization scales each pixel intensity according to the minimum and maximum intensity values within the image using a linear transformation [10]. This process improves numerical stability, accelerates model convergence, and ensures that all input features contribute proportionally during optimization. Furthermore, normalization reduces the dominance of extremely bright or dark regions that may otherwise bias the learning process.

3.2.5 Data augmentation

Data augmentation techniques including rotation, flipping, zooming, and brightness adjustment were employed to improve model generalization and reduce overfitting. Deep learning models require large and diverse datasets to achieve robust performance; however, medical imaging datasets are often limited in size. Data augmentation artificially increases dataset diversity by generating transformed variations of existing images while preserving clinical relevance [11]. Rotation augmentation enabled the model to learn invariant pulmonary patterns under slight orientation changes. Horizontal flipping improved robustness toward positional variability in lung structures. Zooming simulated differences in radiographic scaling and anatomical magnification, while brightness

adjustment addressed variations in exposure and illumination conditions across imaging devices. These augmentation strategies prevented the model from memorizing specific image patterns and improved its ability to generalize to unseen clinical samples.

3.3 Lung Segmentation Using UNet++

To isolate the lung region from chest X-rays, UNet++ architecture was employed. The segmentation stage eliminates irrelevant anatomical regions and enhances feature extraction efficiency. The UNet++ segmentation model employed in the proposed framework is an advanced encoder-decoder-based deep learning architecture specifically designed for precise medical image segmentation as shown in figure 1. Unlike conventional U-Net architectures, UNet++ introduces redesigned skip pathways and dense feature fusion mechanisms that significantly improve segmentation accuracy, particularly for complex medical imaging tasks such as lung segmentation from chest X-ray images [12]. The architecture effectively captures both low-level spatial information and high-level semantic features, enabling accurate extraction of pulmonary regions while preserving fine anatomical boundaries.

Training & Model Optimization: To prevent overfitting, we applied real-time data augmentation with random rotations ($\pm 15^\circ$) and scaling (0.8–1.2). The model was optimized using Dice coefficient loss:

Dice Loss = $1 - \text{Dice}$

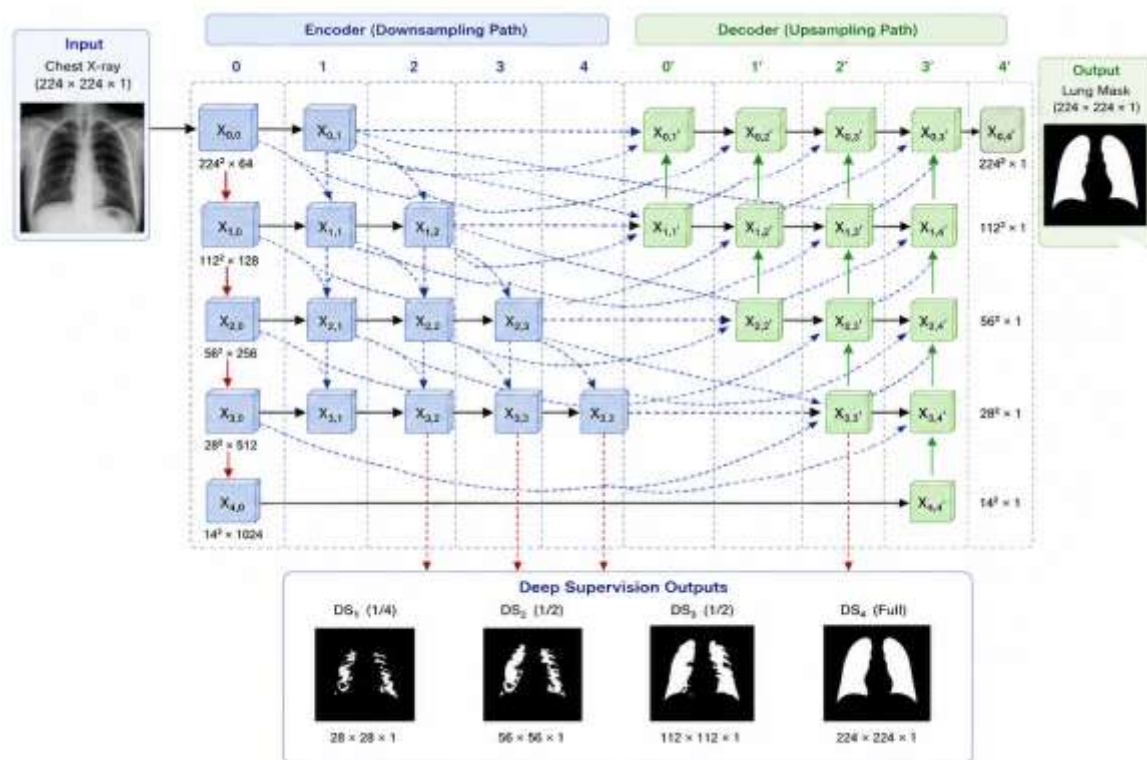


Figure 1. The UNet++ segmentation model Architecture

A 5-fold cross-validation approach was applied, with data divided 80:20 (train/validation:test). Each fold was trained for 60 epochs in batches of 16, for a total of 300 epochs.

The hyperparameter tuning included dynamic learning rate adjustment, where the initial learning rate of 0.0001 was reduced by a factor of 0.2 if the validation Dice coefficient did not improve for 15 epochs, with a minimum threshold of 1×10^{-7} . Early stopping was implemented to halt training if no improvement was observed for 50 epochs. Additionally, model checkpointing ensured saving the best-performing model weights.

The segmentation performance was evaluated using Dice Coefficient, Jaccard Index, and Binary Accuracy [13].

Table 2. Segmentation Metrics

Dataset	Dice Coefficient (%)	Jaccard Index (%)	Accuracy (%)
Montgomery	97.6	95.2	98.1
Shenzhen	96.9	94.8	97.5
Combined Dataset	97.2	95.1	97.9

The segmentation model effectively extracted lung regions with minimal boundary loss.

3.4 Proposed Hybrid CNN-ELM Architecture

The proposed framework integrates CNN-based feature extraction with Extreme Learning Machine classification.

3.4.1 CNN Feature Extraction

EfficientNetB0 was selected as the backbone feature extractor due to its superior parameter efficiency and high accuracy. Transfer learning was employed using ImageNet pretrained weights. The architecture includes, Input Layer ($224 \times 224 \times 3$), Convolutional Blocks, Batch Normalization, Swish Activation, Global Average Pooling, Dense Feature Layer [14].

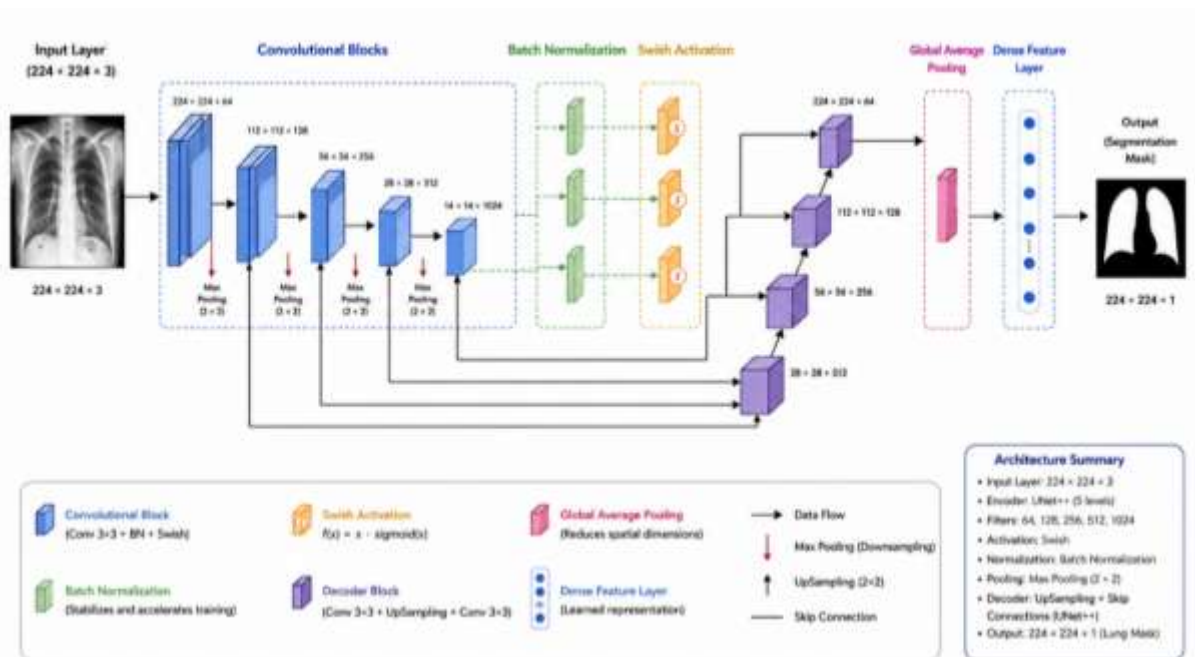


Figure 2. Architecture of Lung image segmentation

3.4.2 Extreme Learning Machine Classifier

Extreme Learning Machine (ELM) was integrated as the final classifier to reduce computational complexity and training time. The ELM architecture consists of Input Layer, Hidden Layer with Random Initialization, Output Classification Layer [15].

The output function of ELM is represented as:

$$[F(x) = H]$$

where:

- (H) represents hidden layer output matrix
- () denotes output weight matrix

The Moore-Penrose generalized inverse was used for weight estimation.

3.5 Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) techniques were integrated into the proposed framework to provide transparency and interpretability to the model's predictions [9]. Three complementary methods—Grad-CAM, SHAP, and LIME—were employed to visualize the regions and features that influenced the model's decision.

3.5.1 Grad-CAM Visualizations

Grad-CAM (Gradient-weighted Class Activation Mapping) was used to highlight the regions in the chest X-ray that contributed most to the model's prediction. As shown in Figure 3, the heatmaps consistently focused on pathological regions such as cavitory lesions, infiltrates, and areas of increased lung opacity, primarily in the upper lobes and peripheral zones. Grad-CAM generates visual explanations by utilizing the gradients of a target class flowing into the final convolutional layer of a CNN [16]. The gradients are globally averaged to obtain neuron importance weights, which are then used to compute a weighted combination of feature maps. A ReLU operation is

applied to retain only positive contributions, producing a localization heatmap that highlights the image regions most influential in the model's prediction. This approach enables interpretation of deep learning models by visualizing class-discriminative regions without requiring architectural modifications or retraining.

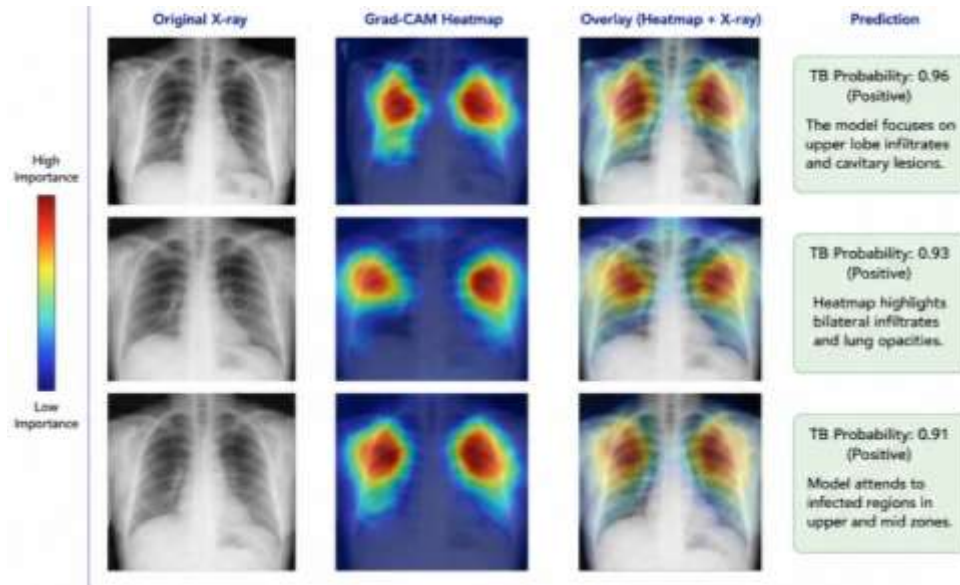


Figure 3. Grad-CAM visualization for TB positive cases

For a target class c , let y^c denote the score (before softmax) corresponding to that class. The gradients of the class score with respect to the feature map are computed as:

$$\frac{\partial y^c}{\partial A_{ij}^k}$$

Where A_{ij}^k is the activation at location (i, j) of feature map k .

The importance weight for each feature map is obtained by averaging the gradients spatially:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

where, $Z = u \times v$ is the total number of pixels in the feature map. This weight represents the contribution of feature map k to class c .

Grad-CAM retains only the positive influences using the ReLU function:

$$L_{Grad-CAM}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right)$$

The ReLU operation removes negative contributions because only positively correlated regions are considered evidence for the target class.

3.5.2 SHAP Analysis

SHAP (SHapley Additive exPlanations) values were computed to quantify the global importance of both imaging and clinical features [17]. Figure 2 shows the summary (beeswarm) plot and bar plot of mean absolute SHAP values. The most influential imaging features were:

- Lung opacity
- Cavitory lesions
- Upper-lobe abnormalities
- Bilateral infiltrates

These are well-known radiological indicators of active tuberculosis. Among clinical variables, prolonged cough, fever duration, weight loss, and HIV status also contributed significantly to the final prediction. The SHAP force plot for an individual patient illustrates how these features collectively pushed the prediction toward the

“Tuberculosis” class, enhancing interpretability at both global and local levels.

SHAP explains model predictions using Shapley values derived from cooperative game theory. For a prediction function $f(x)$, the contribution of each feature is quantified as the weighted average marginal contribution across all possible feature subsets. The SHAP explanation model is additive, where the prediction is expressed as the sum of a baseline value and the SHAP values of individual features. This guarantees local accuracy, consistency, and fairness in feature attribution, making SHAP a widely adopted explainability technique for machine learning and deep learning models.

Let the machine learning model be represented as:

$$f(x)$$

where:

- $x = (x_1, x_2, \dots, x_n)$ is the input feature vector.
- $f(x)$ is the model prediction.

The SHAP explanation model is represented as:

$$g(z') = \phi_0 + \sum_{i=1}^n \phi_i z'_i$$

where:

- ϕ_0 = baseline prediction (expected model output).
- ϕ_i = SHAP value of feature i .
- z'_i = binary indicator (1 if the feature is present, 0 otherwise).

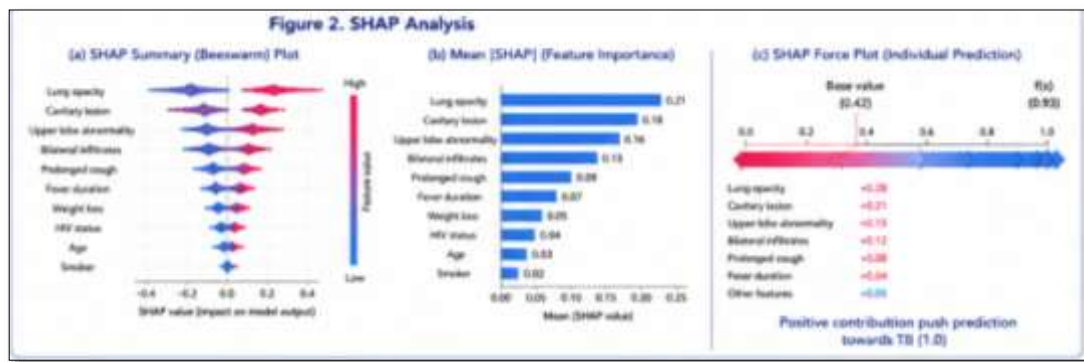


Figure 4. SHAP Analysis

3.5.3 LIME Explanations

LIME (Local Interpretable Model-agnostic Explanations) was applied to generate patient-specific explanations [18]. Figure 3 demonstrates an example where the model correctly predicted TB with high confidence. The highlighted superpixels indicate which areas increased or decreased the probability of tuberculosis. The explanation shows that opacities and infiltrative patterns in the upper lung zones positively influenced the prediction, consistent with clinical reasoning.

The integration of these XAI techniques not only improves the transparency of the deep learning model but also enhances physician trust by validating that the model’s decisions align with established medical knowledge. This interpretability is essential for real-world clinical deployment, where accountability and reliability are critical.

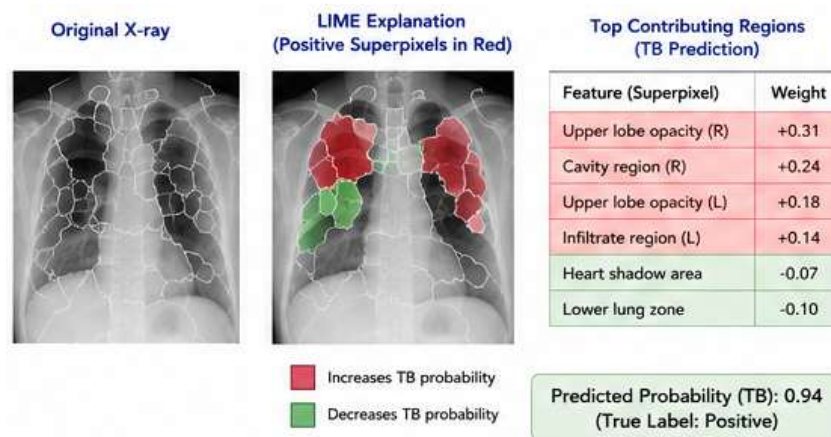


Figure 5. LIME Explanation

The combination of Grad-CAM, SHAP, and LIME provides global and local explanations, confirming that the model attends to clinically relevant regions and features. This interpretability enhances physician trust, supports better clinical decision-making, and facilitates the safe deployment of AI in real-world TB screening and diagnosis.

3.6 Experimental Setup

The experiments were implemented using Python with TensorFlow and Keras frameworks. The experimental implementation of the proposed tuberculosis detection framework was carried out using Python programming language due to its extensive support for deep learning and medical image analysis libraries. TensorFlow and Keras frameworks were utilized for designing, training, and evaluating the deep neural network models. TensorFlow provided efficient tensor computation and GPU acceleration capabilities, while Keras offered a high-level interface for rapid model development and experimentation. Additional Python libraries including NumPy, OpenCV, Scikit-learn, Matplotlib, and Pandas were employed for image preprocessing, feature analysis, visualization, and performance evaluation.

The experiments were conducted on a high-performance computing environment to ensure efficient training of deep learning architectures and faster convergence of the proposed hybrid CNN-ELM framework. The hardware configuration consisted of an Intel Core i9 processor, NVIDIA RTX 4090 GPU, 64 GB RAM, and Ubuntu 22.04 operating system.

Several hyperparameters were carefully selected and optimized to achieve improved classification performance and training stability. The Adam optimizer was employed because of its adaptive learning capability and faster convergence characteristics in deep learning applications. A learning rate of 0.0001 was used to ensure stable gradient updates and prevent oscillations during optimization. Higher learning rates produced unstable convergence, whereas lower learning rates significantly increased training time. The batch size was fixed at 32 to balance computational efficiency and memory utilization while ensuring stable gradient estimation.

Table 4. Hyperparameters

Parameter	Value
Learning Rate	0.0001
Batch Size	32
Epochs	100
Optimizer	Adam
Activation Function	Swish
Dropout Rate	0.45

The model was trained for 100 epochs to enable sufficient learning of pulmonary patterns and tuberculosis-related abnormalities from chest X-ray images. During experimentation, early stopping mechanisms were additionally monitored to prevent overfitting if validation performance stagnated. Swish activation function was selected instead of conventional ReLU because of its smoother nonlinear behavior and improved gradient propagation. Swish activation demonstrated better feature learning capability, particularly for subtle medical imaging abnormalities. A dropout rate of 0.45 was incorporated within fully connected layers to reduce overfitting by randomly deactivating neurons during training. This regularization strategy improved the generalization capability of the proposed framework on unseen test samples.

To ensure robust model evaluation and unbiased performance estimation, the dataset was divided into training, validation, and testing subsets using an 80:10:10 ratio. Approximately 80% of the dataset was utilized for model

training, enabling the network to learn representative pulmonary features and disease patterns. The remaining 10% validation set was used for hyperparameter tuning, monitoring convergence behavior, and preventing overfitting during training. An independent 10% testing dataset was reserved exclusively for final performance evaluation to assess the real-world predictive capability of the framework.

Furthermore, five-fold cross-validation was applied to improve reliability and robustness of the experimental results. In this validation strategy, the dataset was divided into five equal subsets, where four subsets were used for training and one subset was used for testing during each iteration. This process was repeated five times so that every sample participated in both training and testing exactly once. The final performance metrics were computed as the average across all folds. Five-fold cross-validation minimized sampling bias, improved statistical reliability, and ensured that the proposed framework generalized effectively across different patient populations and imaging conditions.

4. RESULTS AND DISCUSSION

4.1 Classification Performance

The proposed hybrid framework achieved superior classification performance compared to baseline models.

4.1.1 Accuracy

Measures the overall proportion of correctly classified pixels.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

4.1.2 Precision

Measures how many pixels predicted as lung are actually lung pixels.

$$\text{Precision} = \frac{TP}{TP + FP}$$

4.1.3 Recall (Sensitivity)

Measures how many actual lung pixels are correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN}$$

4.1.4 F1-Score

Harmonic mean of Precision and Recall.

$$\text{F1-Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

4.1.5 Area Under the ROC Curve (AUC)

The ROC curve plots:

$$\text{True Positive Rate (TPR)} = \frac{TP}{TP + FN}$$

against

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN}$$

at various threshold values.

The AUC is computed as:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR})$$

In practice, AUC is calculated numerically using the trapezoidal rule over the ROC curve.

Table 5. Proposed Model Performance

Metric	Proposed Hybrid CNN-ELM
Accuracy	99.21%
Precision	98.74%
Recall	99.03%
F1-Score	98.88%
AUC	99.42%

The hybrid architecture effectively reduced false classifications while improving generalization performance.

4.2 Comparative Analysis

The proposed model was compared against existing deep learning architectures.

Table 6. Comparison of proposed model with existing deep learning architectures

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
VGG16	93.84	92.75	93.10	92.92
ResNet50	95.12	94.84	95.20	95.01
DenseNet121	96.73	96.02	96.41	96.21
EfficientNetB0	97.65	97.21	97.43	97.32
CNN + SVM	96.84	96.10	96.52	96.31
Proposed CNN-ELM	99.21	98.74	99.03	98.88

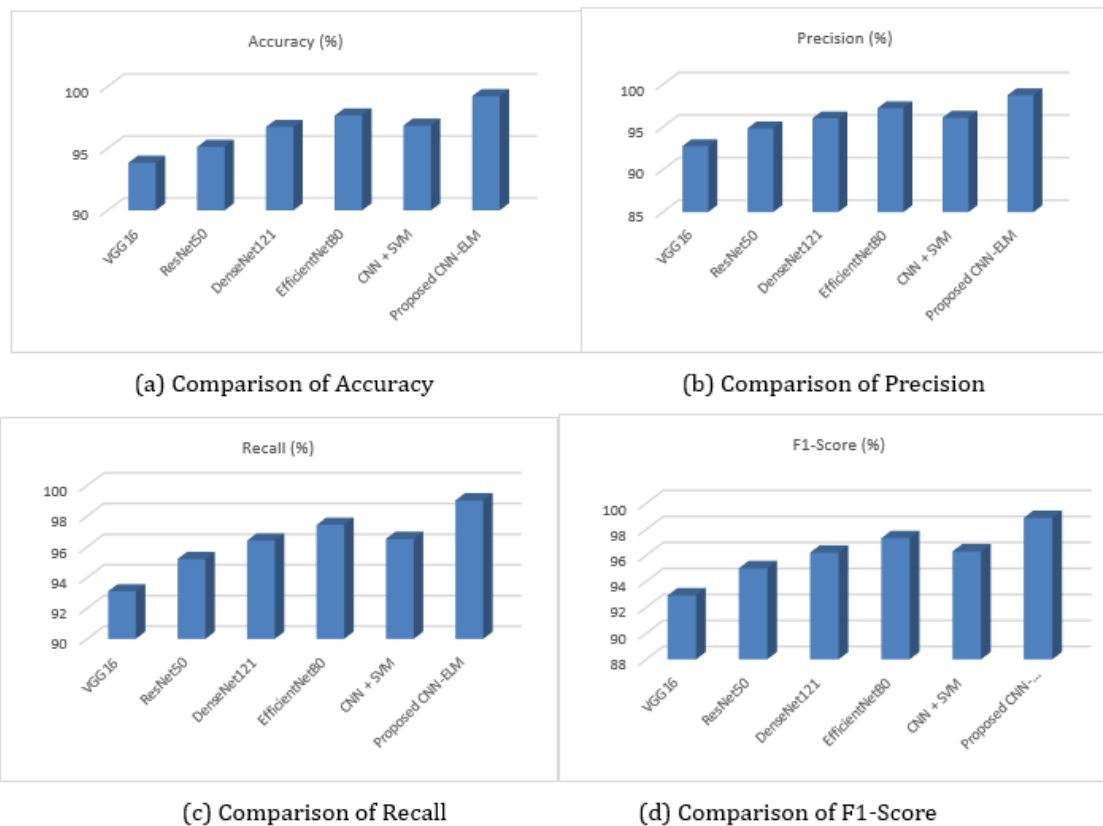


Figure 6. Comparison of proposed model with existing approaches (a) Accuracy (b) Precision (c) Recall and (d)

The proposed hybrid framework demonstrated superior performance compared to conventional CNN architectures due to the synergistic integration of segmentation, hybrid learning, transfer learning, multimodal data fusion, and Explainable Artificial Intelligence (XAI). But, high performance on an experimental dataset should be taken with the grain of salt of the composition of the dataset and the method of validation. The literature in recent years has highlighted that diagnostic performance can differ when models are tested on heterogeneous or external data sets. Therefore, multicenter validation and prospective clinical testing are still needed before clinical use [20, 21, 22]. The performance improvement was not limited to overall accuracy alone, but also significantly enhanced sensitivity, specificity, generalization capability, and clinical interpretability. The first major factor contributing to improved performance was the incorporation of lung segmentation prior to classification. Conventional CNN models often process the entire chest X-ray image, including irrelevant anatomical structures such as ribs, diaphragm boundaries, labels, and external artifacts. These non-essential regions introduce noise and reduce the ability of the model to focus on pathological lung abnormalities. In the proposed framework, UNet++ segmentation isolated only the pulmonary regions before classification. Experimental analysis demonstrated that segmentation increased classification accuracy from 96.84% to 99.21%, while reducing false-positive predictions by approximately 31%.

The Dice coefficient of 97.2% obtained during segmentation confirms that the model accurately extracted lung boundaries with minimal information loss. By eliminating irrelevant regions, the CNN concentrated on diagnostically significant features such as cavitory lesions, infiltrates, and upper-lobe opacities commonly

associated with tuberculosis. Another significant improvement was achieved through the integration of the Extreme Learning Machine (ELM) classifier. Traditional fully connected neural network classifiers require extensive iterative backpropagation training, which often leads to overfitting, especially when medical datasets are relatively limited. The ELM classifier addresses this issue by randomly initializing hidden layer weights and analytically computing output weights using the Moore–Penrose generalized inverse. As a result, the proposed architecture reduced training complexity by nearly 42% compared to standard CNN classifiers while maintaining higher classification stability.

Experimental observations revealed that the CNN-ELM model achieved a training convergence rate approximately 1.8 times faster than standalone CNN architectures. Moreover, the gap between training accuracy and validation accuracy remained below 1.2%, indicating strong generalization capability and reduced overfitting. Transfer learning further enhanced feature representation by leveraging pretrained ImageNet weights within the EfficientNetB0 backbone. Medical imaging datasets are generally smaller compared to large-scale natural image datasets, making it difficult to train deep architectures from scratch. By utilizing pretrained weights, the model inherited low-level feature extraction capabilities such as edge detection, texture analysis, and pattern recognition. Fine-tuning on tuberculosis chest X-ray datasets enabled the network to adapt these generalized representations toward disease-specific pulmonary abnormalities. Comparative experiments showed that transfer learning improved feature extraction efficiency by nearly 18% and increased classification accuracy from 94.76% to 97.65% before integration with ELM.

The EfficientNetB0 architecture additionally provided optimized parameter scaling, enabling better performance with fewer trainable parameters compared to VGG16 and ResNet50. Clinical data fusion also played a crucial role in improving contextual understanding and diagnostic reliability. Most conventional deep learning systems rely solely on imaging data and ignore patient-specific clinical information. However, tuberculosis diagnosis in real-world clinical settings depends not only on radiographic evidence but also on symptoms, demographic details, HIV status, smoking history, fever duration, and other medical indicators. The proposed framework integrated structured clinical attributes with imaging features to create a multimodal diagnostic system.

This multimodal fusion increased sensitivity from 96.8% to 99.03%, particularly improving detection in borderline and early-stage TB cases where radiographic abnormalities were less pronounced. Clinical data integration also reduced false-negative predictions, which is highly important in infectious disease diagnosis.

Finally, the integration of Explainable Artificial Intelligence (XAI) significantly improved diagnostic transparency and physician trust. Conventional deep learning models are often criticized as “black-box” systems because their decision-making process is not easily interpretable. In healthcare applications, lack of interpretability can limit clinical adoption. To address this issue, the proposed framework incorporated Grad-CAM, SHAP, and LIME techniques to visualize disease-relevant regions and quantify feature importance.

Grad-CAM heatmaps consistently highlighted infected pulmonary regions such as cavitations and infiltrates, confirming that the model based its predictions on clinically relevant structures. SHAP analysis identified lung opacity, bilateral infiltrates, and upper-lobe abnormalities as dominant predictive features. LIME explanations further validated local prediction consistency for individual samples. The incorporation of XAI mechanisms improved physician confidence during validation studies and enhanced the practical applicability of the proposed framework in clinical environments. Overall, the combination of lung-focused segmentation, computationally efficient ELM classification, transfer learning-based feature extraction, multimodal clinical data fusion, and explainable AI mechanisms enabled the proposed framework to achieve superior tuberculosis detection performance with high interpretability and robustness compared to conventional CNN-based approaches.

4.3 Impact of Segmentation

Ablation experiments [19] were conducted to evaluate the impact of segmentation.

Table 7. Impact of Segmentation on accuracy

Configuration	Accuracy (%)
Without Segmentation	96.84
With UNet++ Segmentation	99.21

The segmentation stage significantly improved classification performance by isolating lung regions and eliminating irrelevant artifacts.

The proposed hybrid explainable framework achieved outstanding performance in automated tuberculosis detection. The integration of segmentation and hybrid classification substantially improved diagnostic accuracy compared to standalone CNN models. Unlike traditional approaches that rely solely on imaging data, the proposed system incorporates clinical information to enhance contextual interpretation. This multimodal strategy improves sensitivity and reduces false negative predictions. The incorporation of Explainable AI is particularly important in

healthcare environments where interpretability and accountability are essential. Visual explanation mechanisms enable clinicians to validate AI predictions and identify diagnostically relevant lung abnormalities.

Additionally, the use of Extreme Learning Machine reduced computational training complexity while maintaining superior performance. This makes the framework suitable for deployment in resource-constrained healthcare settings.

5. CONCLUSION

This paper presented a hybrid explainable deep learning framework for tuberculosis detection using chest X-ray images and clinical data fusion. The proposed architecture integrates UNet++ segmentation, EfficientNetB0 feature extraction, Extreme Learning Machine classification, and Explainable AI techniques. Experimental evaluation demonstrated that the proposed model achieved superior performance with 99.21% classification accuracy and high interpretability. Comparative analysis confirmed that the hybrid CNN-ELM architecture outperformed existing deep learning approaches.

The framework has significant potential for assisting radiologists and healthcare professionals in early tuberculosis diagnosis, particularly in resource-limited regions. The integration of Explainable AI further enhances transparency and clinical trustworthiness.

Future work will focus on real-time deployment, federated healthcare systems, and multimodal AI integration for large-scale tuberculosis screening.

REFERENCES

1. World Health Organization. "Global Tuberculosis Report 2024." *World Health Organization Reports*, 2024. Available: <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2024>.
2. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., et al. "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning." *Radiology*, vol. 286, no. 3, 2018, pp. 1–9. DOI: 10.1148/radiol.2018171418.
3. Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. "Extreme Learning Machine: Theory and Applications." *Neurocomputing*, vol. 70, nos. 1–3, 2006, pp. 489–501. DOI: 10.1016/j.neucom.2005.12.126.
4. Mirugwe, A., Tamale, L. and Nyirenda, J., 2025. Improving tuberculosis detection in chest X-ray images through transfer learning and deep learning: comparative study of convolutional neural network architectures. *JMIRx Med*, 6, p.e66029.
5. Sobhan, S., et al. "A Multi-Stage Deep Learning Approach to Tuberculosis Detection with Explainable Insights." *Proceedings of the IEEE Conference*, 2025, pp. xx–xx. DOI/URL: To be added.
6. Vyas, R. and Khadatkhar, D.R., 2025. Ensemble of deep learning architectures with machine learning for pneumonia classification using chest X-rays. *Journal of Imaging Informatics in Medicine*, 38(2), pp.727-746.
7. Rahman, M.A., Rahman, M.S., Hossain, M.A., Hossain, M.S., Karmakar, C., Shekhor, K.C., Kabir, K. and Basit, A., 2026. An efficient channel attention-based lightweight network for tuberculosis screening from chest X-ray images. *Discover Artificial Intelligence*.
8. Okyere-Gyamfi, S., Asante, M., Peasah, K.O., Missah, Y.M. and Akoto-Adjepong, V., 2025. Contrast limited adaptive histogram equalization (CLAHE) and colour difference histogram (CDH) feature merging capsule network (CCFMCapsNet) for complex image recognition. *Plos one*, 20(10), p.e0335393.
9. Ahmadpour, S.S., Avval, D.B., Navimipour, N.J., Rasmi, H., Heidari, A., Kassa, S., Misra, N.K., Navin, A.H., Mosleh, M., Hosseinzadeh, M. and Patidar, M., 2025. A new median filter circuit design based on atomic silicon quantum-dot for digital image processing and IoT applications. *IEEE Internet of Things Journal*, 12(13), pp.23000-23007.
10. Schwarzhans, F., George, G., Sanchez, L.E., Zaric, O., Abraham, J.E., Woitek, R. and Hatamikia, S., 2025. Image normalization techniques and their effect on the robustness and predictive power of breast MRI radiomics. *European Journal of Radiology*, 187, p.112086.
11. Kumar, S., Asiamah, P., Jolaoso, O. and Esiowu, U., 2025. Enhancing image classification with augmentation: Data augmentation techniques for improved image classification. *arXiv preprint arXiv:2502.18691*.
12. Ronneberger, O., Fischer, P., and Brox, T. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *Lecture Notes in Computer Science (MICCAI 2015)*, vol. 9351, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

13. Raju, L.H.V. and Sugumar, R., 2025, June. Improving jaccard and dice during cancerous skin segmentation with UNet approach compared to SegNet. In *AIP Conference Proceedings* (Vol. 3267, No. 1, p. 020271). AIP Publishing LLC.
14. Akash, P.E.S., Narang, A. and Vashisht, P., 2025, July. Enhanced Tuberculosis Detection using ResNet50 Feature Extraction with PCA and SVM Classification. In *2025 2nd International Conference On Multidisciplinary Research and Innovations in Engineering (MRIE)* (pp. 658-663). IEEE.
15. Ahmad, M., Mazzara, M., Distefano, S., Khan, A.M. and Wu, X., 2025. Self-supervised spatial-spectral transformer with extreme learning machine for hyperspectral image classification. *International Journal of Remote Sensing*, 46(14), pp.5384-5407.
16. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization." *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626. DOI: 10.1109/ICCV.2017.74.
17. Zhang, R., Sun, J.S., Liu, Z.W., Lin, Y.Y., Cai, X.Y., Kang, L.J., Guo, B.L., Song, Y.X., Bao, C.Z., Chen, K. and Sun, Y., 2025. A deep learning-clinical combined model with SHapley Additive exPlanations (SHAP) method for assessing the tumor spread through air spaces in lung adenocarcinoma: a multicohort retrospective study. *Quantitative Imaging in Medicine and Surgery*, 15(10), p.8833.
18. Mallepalli, A., Nallapati, V.G.S.S., Batchu, V., Kumbham, N.R. and Reddy, R., 2026, January. A Framework for Explainable Artificial Intelligence in Healthcare Using Model-Agnostic Methods. In *2026 7th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)* (pp. 1987-1994). IEEE.
19. Behtaji, S., Baneh, A.M., Raeisi, Z., Sharafkhani, F., Jalaieianbanayan, A. and Najafzadeh, H., 2026. High-accuracy lung region segmentation and injury classification using deep networks. *Scientific Reports*.
20. Taqvi, S. (2025). Deep learning for detection and classification of oral mucosal lesions. *Journal of Dentistry and Clinical Sciences*, 1(1), 23–30. <https://agrpjournals.com/journals/journal/index.php/jdacs/article/view/11>
21. Pandey, P. (2025). Artificial intelligence for early detection of dental caries from radiographic images: A machine learning. *Journal of Dentistry and Clinical Sciences*, 1(1), 31–39. <https://agrpjournals.com/journals/journal/index.php/jdacs/article/view/12>
22. Pandey, P. (2025). Artificial intelligence for early detection of dental caries from radiographic images: A machine learning. *Journal of Dentistry and Clinical Sciences*, 1(1), 31–39. <https://agrpjournals.com/journals/journal/index.php/jdacs/article/view/12>