

Research Article

A Hybrid CNN-Transformer Framework for Automated Tuberculosis Detection from Chest X-rays

Lakshmi Ravi Teja Meka ¹ , Dalwinder Singh ¹ , Arun Singh ¹ , Saiprasad Potharaju ^{2,*} ,
MVV Prasad Kantipudi ³ , Kiranmai Bejjam ⁴ 

¹ Department of Computer Science and Engineering, Lovely Professional University, Phagwara, 144411, India

² Department of Computer Science and Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, 412115, India

³ Department of Electronics and Telecommunication Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, 412115, India

⁴ Department of Computer Science and Engineering, Sreyas Institute of Engineering and Technology Hyderabad, Hyderabad, 500068, India

*Corresponding Author: Saiprasad Potharaju, E-mail: psaiprasadcse@gmail.com

Article Info	Abstract
Article History	Tuberculosis (TB) remains a major global health challenge, demanding rapid and accurate diagnosis to enable timely treatment and reduce disease transmission. Although chest radiography is widely used for TB screening, manual interpretation is often time-consuming, subjective, and constrained by the shortage of experienced radiologists, especially in high-burden, resource-constrained settings. To address these limitations, this study proposes HybridTBNet, a hybrid deep learning framework that integrates Convolutional Neural Networks (CNNs) and Transformers for automated TB detection from chest X-ray images. The model is designed to capture both local pathological patterns and global contextual dependencies through hierarchical feature extraction and attention-based representation learning. A curated multi-source dataset of 6,708 high-quality chest X-ray images, comprising 3,194 TB and 3,514 Normal cases after quality filtering, was used for development and evaluation. A strict patient-wise split was applied to prevent data leakage, resulting in separate training, validation, and held-out test sets; however, minor class imbalance persisted in the final test subset. On the unseen test set, HybridTBNet achieved 98.2% accuracy, 97.1% sensitivity, 98.5% specificity, 94.9% F1-score, 0.991 ROC-AUC, and 0.978 PR-AUC, demonstrating strong potential as an effective automated TB screening support tool.
Received Oct 29, 2025	
Revised Feb 03, 2026	
Accepted Feb 20, 2026	
Keywords	
Automated Diagnosis	
Chest X-ray (CXR)	
Convolutional Neural Network (CNN)	
Deep Learning	
Tuberculosis (TB)	



Copyright: © 2026 Lakshmi Ravi Teja Meka, Dalwinder Singh, Arun Singh, Saiprasad Potharaju, MVV Prasad Kantipudi, and Kiranmai Bejjam. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license.

1. Introduction

Tuberculosis (TB) continues to be a major public health burden, particularly in low- and middle-income countries where access to expert radiologists is limited. Chest X-ray (CXR) imaging remains the most widely used screening tool, yet its interpretation is prone to subjectivity, inter-reader variability, and

diagnostic delays. These limitations have fueled interest in computer-aided detection systems that can improve the speed, accuracy, and consistency of TB screening programs. Recent advances in deep learning, especially convolutional neural networks (CNNs) and Vision Transformers (ViTs), have demonstrated strong performance in medical image analysis; however, each architecture suffers from inherent limitations when applied independently to TB detection.

Several studies have demonstrated the potential of machine learning in tuberculosis research and surveillance. For instance, Althomsons, et al. [1] used ML techniques to predict growth in TB clusters using national surveillance data, highlighting how algorithmic approaches can complement epidemiological tracking. Bhardwaj, et al. [2] further emphasised the utility of ML in developing predictive systems tailored to clinical datasets, illustrating that computational methods can enhance diagnostic decision-making. Similarly, Bustad, et al. [3] employed interpretable ML informed by a transcriptional regulatory network to predict fitness in *Mycobacterium tuberculosis*, underscoring the role of AI in molecular-level analysis. Chen, et al. [4] showcased how integrating WHO mutation catalogues with genomic sequencing and ML enhances resistance prediction, addressing one of the most critical challenges in TB treatment.

The use of decision-tree-based frameworks for mutation discovery has also proven effective, as shown by Deelder, et al. [5], Feng, et al. [6] applied ML models in predicting spinal cord injury outcomes in spinal TB patients, thereby extending ML applications beyond pulmonary TB. Additionally, Gichuhi, et al. [7] explored ML to identify patient-level risk factors for treatment non-adherence, offering insights into behavioural and compliance challenges. These works collectively demonstrate that ML has become a central component in advancing TB diagnostics, drug resistance prediction, and patient management strategies, thus motivating further exploration of hybridised and context-aware approaches.

Contributions of this research are:

- This study proposes HybridTBNet, a hybrid CNN-Transformer framework for TB detection from CXR images. The key contributions are:
- A cross-stage hybrid architecture that integrates EfficientNet-B4 local feature extraction with Transformer-based global attention.
- A consistent benchmarking protocol comparing deep models and classical ML methods, including an ablation study differentiating CNN-only, Transformer-only, and hybrid fusion.
- A clarified and reproducible dataset pipeline, including patient-wise splitting, preprocessing,

augmentation, and detailed hyperparameters.

- A clinically relevant evaluation, demonstrating high sensitivity and specificity suitable for mass TB screening.

2. Literature Review

The application of machine learning and deep learning in tuberculosis research has expanded across multiple domains, ranging from genomic resistance prediction to radiological analysis and epidemiological modelling. A significant body of literature has highlighted the ability of ML algorithms to uncover patterns not easily discernible by conventional methods, thereby providing greater predictive power and diagnostic support.

Green, et al. [8] demonstrated that convolutional neural networks (CNNs) can reveal mutations associated with antimicrobial resistance in *Mycobacterium tuberculosis*, highlighting the importance of feature extraction in genetic datasets. Similarly, Gröschel, et al. [9] a genome-based machine-learning prediction platform, providing a practical, user-friendly solution for clinicians and researchers conducting resistance profiling. Thirunavukarasu and Kotei [10] reviews the growing use of transformer networks in natural and medical image analysis, explaining attention mechanisms, key architectures, applications, limitations, and future research opportunities. It highlights the potential of transformers as powerful alternatives to CNNs for advanced computer vision and medical diagnostics. Kotei and Thirunavukarasu [11] proposed a hybrid Data-Efficient Image Transformer and ResNet-16 model for automated tuberculosis detection using the TBX11K dataset, which leverages self-attention and lightweight convolutions.

Moving into non-traditional data modalities, Galvosas, et al. [12] leveraged cough audio recordings for TB prediction, marking an innovative direction in accessible, low-cost screening methods. In parallel, Karmani, et al. [13] proposed a machine learning-based TB health predictor model aimed at early disease prevention in developing regions, where resource constraints make scalable ML solutions particularly relevant. Lane, et al. [14] explored ML models for predicting *M. tuberculosis* in vitro activity, with added focus on target visualisation, bridging predictive performance with interpretability. Similarly, Larkins-Ford, et al. [15] examined drug-combination landscapes, using ML to anticipate treatment outcomes, a crucial contribution to precision medicine in TB therapy.

Radiomics has also emerged as a promising field, with Li, et al. [16] employing ML and radiomic

features to predict multidrug resistance in cavitary pulmonary TB patients across multiple centers, demonstrating both robustness and generalizability. At a population level, Maipan-Uku and Cavus [17] reviewed forecasting strategies using time series and ML for incidence prediction, while other related work like KB, et al. [18] demonstrated the application of machine learning for tuberculosis prediction and epidemiological analysis, highlighting its effectiveness in forecasting disease trends and identifying key influencing factors. These studies highlight ML's versatility not only in clinical but also in epidemiological and surveillance contexts.

Smart health technologies have incorporated ML into broader systems, as illustrated by Murugesan, et al. [19], who designed neural-network-based e-health applications leveraging serverless computing for TB prediction. Orjuela-Cañón, et al. [20] applied ML in diagnostic support systems, integrating algorithmic decision-making into clinical workflows. Beyond health informatics, Parmar, et al. [21] developed a graphene-based refractive index sensor enhanced with ML for bacterial detection, showing the synergy between nanotechnology and computational methods. Complementing this, Perea-Jacobo, et al. [22] performed a PRISMA-based scoping review of genome-sequencing ML applications, synthesising progress in predictive genomics.

Resistance pattern analysis remains a major area of focus. Serajian, et al. [23] conducted a comparative study on antibiotic resistance, while Singh, et al. [24] reviewed the broader evolution of ML in TB diagnostics, emphasising deep learning applications in radiology and genomics. Stański, et al. [25] applied ML to improve predictions of bovine TB at the herd level, demonstrating cross-species applicability of ML methodologies. Meanwhile, Tang, et al. [26] incorporated meteorological and environmental factors into TB incidence prediction, showcasing how ML models can integrate non-biological variables for public health forecasting. Winkler [27] offered expert perspectives on the transformative impact of ML in TB drug discovery, further reinforcing its role in accelerating therapeutic pipelines.

Recent genomic prediction studies have also shown promise. Xu, et al. [28] employed ML for predicting antimicrobial resistance while identifying relevant SNPs, expanding the interpretability of genomic data. Ye, et al. [29] used ML to identify active molecules against TB, focusing on drug discovery pipelines. Yurtseven, et al. [30] integrated ML with phylogenetic analysis to predict resistance, merging evolutionary biology with AI. Extending to treatment safety, Zeng, et al. [31] built automated ML models to predict liver injury during pediatric TB treatment, thereby addressing adverse-event management. Zhang,

et al. [32] proposed an explainable ML platform for resistance prediction against pyrazinamide, emphasising the necessity for transparency in medical AI systems.

Further, Zhu, et al. [33] applied machine learning to predict acid-fast bacilli sputum test outcomes across multiple centres, demonstrating its value in clinical microbiology. Shames and Kamil [34] advanced hybrid deep-learning techniques for lung infection diagnosis, validating the potential of combined architectures to achieve higher accuracy and robustness. Collectively, these studies illustrate a steady progression from experimental research toward practical deployment of ML in TB detection, treatment planning, epidemiology, and resistance management.

In summary, the reviewed literature shows that ML applications in TB span molecular, clinical, and environmental domains. The research trajectory reflects an increasing emphasis on interpretability, multimodal integration, and scalability, which align with the broader healthcare need for robust, accessible, and reliable diagnostic solutions. However, while significant advancements have been achieved, challenges remain regarding data heterogeneity, generalisation across populations, and integration into clinical settings. These considerations underline the importance of ongoing hybrid-model research that combines multiple ML paradigms to enhance diagnostic performance and usability in real-world contexts.

3. Materials and Methods

3.1. Objectives

The primary objective of this study is to develop and validate HybridTBNNet, a novel deep learning framework for automated detection of tuberculosis in chest X-rays. We aim to demonstrate that a hybrid architecture, integrating a CNN backbone for rich feature extraction with a Transformer encoder for global context analysis, can achieve superior diagnostic performance compared to both standard CNNs and classical machine learning models. The ultimate goal is to establish this model's potential as a robust, automated tool to assist clinicians, reduce diagnostic delays, and enhance the efficacy of large-scale TB screening programs.

3.2. Dataset Description

This study utilises a curated collection of posterior–anterior (PA) chest X-ray (CXR) images from publicly available, widely used tuberculosis datasets. The images were sourced from the NIH ChestX-ray14 dataset, the Shenzhen Tuberculosis CXR dataset (China), and the Montgomery County Tuberculosis dataset (USA). All datasets provide radiologist-verified labels and anonymised patient identifiers. No synthetically

generated, simulated, or GAN-based images were used in this study.

The combined dataset initially consisted of 7,208 chest X-ray (CXR) images, comprising 3,194 Tuberculosis-positive and 4,014 Normal cases. After quality control procedures, including the removal of corrupted files, low-resolution images, and duplicate or near-duplicate scans identified through perceptual hashing, 6,708 high-quality images were retained for model development and evaluation. Before dataset splitting, all images underwent quality control procedures, including the removal of corrupted files, low-resolution images, and duplicate or near-duplicate scans. Duplicate detection was performed using perceptual hashing to ensure that visually identical or highly similar images were excluded.

To prevent data leakage, a strict patient-wise split was enforced using the patient identifiers provided by the source datasets. This ensured that no patient appeared in more than one subset. To prevent data leakage, a strict patient-wise split was enforced using the patient identifiers provided by the source datasets. This ensured that no patient appeared in more than one subset. After quality filtering and patient-wise grouping, the retained dataset was partitioned into training, validation, and test subsets. Due to strict patient-wise separation across multiple source datasets, the final class distribution differed across subsets.

The final dataset composition after filtering and splitting is summarised as follows:

- Training set: 5,004 images (2,225 TB, 2,779 Normal)
- Validation set: 1,074 images (864 TB, 210 Normal)
- Test set: 630 images (105 TB, 525 Normal)

Although stratified sampling was applied, strict patient-wise splitting and source-level heterogeneity can lead to imbalance in the final subset distributions, a common issue in real-world medical imaging studies.

The final test set of 630 images represents a strictly held out, previously unseen subset with no patient overlaps with training or validation data. All performance metrics reported in this study are computed exclusively on this verified test set to ensure unbiased evaluation.

Ethics and Privacy Compliance: All datasets used are publicly available and fully anonymised by their respective providers. No identifiable patient information was accessed or processed; therefore, no additional institutional ethics approval was required.

3.3. Data Preprocessing

All images were processed through a standard pipeline to maintain consistency and optimise model

training. This was initiated by scaling down all images to a standard size of 380x380 pixels to fit the input specifications of the neural network architecture. The CXRs are grayscale but were converted to 3-channel RGB by simply expanding the single channel, since the pre-trained models only accept this format. The processed images were then converted to PyTorch tensors. Lastly, the conversion of the setup involved normalising the tensor pixel values by the mean and standard deviation of the ImageNet dataset, a typical step in transfer learning. On-the-fly data augmentation was applied only to the training set to increase diversity and prevent overfitting. These manipulations included random rotations, horizontal flipping, and changes in brightness and contrast (Figure 1).

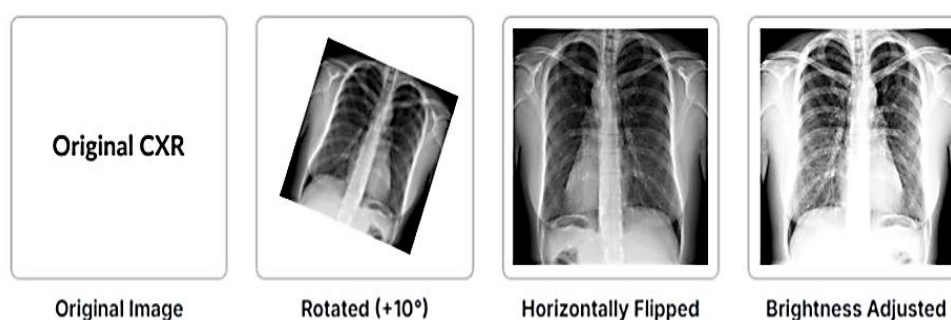


Figure 1. Data augmentation

While standard preprocessing steps such as resizing, normalisation, and the use of AdamW optimiser were employed, several task-specific refinements were incorporated to adapt these methods to tuberculosis CXR classification. First, the resizing resolution (380×380) was selected through empirical testing, as it preserved fine-grained pulmonary textures clinically relevant to TB while remaining computationally efficient. Second, augmentation parameters were tuned based on radiologist feedback to reflect realistic CXR variations such as lateral tilt, contrast fluctuations, and soft-tissue intensity differences, rather than generic image perturbations. Third, AdamW was selected not as a default optimiser but for its superior stability during training of hybrid CNN–Transformer models, especially under class imbalance. Additionally, a warmup-based learning rate schedule and gradient clipping were applied to prevent Transformer instability, techniques not commonly incorporated in baseline medical imaging pipelines.

Together, these adaptations extend beyond routine preprocessing and meaningfully enhance the robustness and convergence behaviour of HybridTBNet.

3.4. Model Architectures and Mathematical Foundations

We evaluated a suite of classical machine learning models and deep learning models.

K-Nearest Neighbours (KNN): A nonparametric method determining the category of a case r based on a

majority of those k data points in the training data nearest r , as calculated by the Euclidean distance, $d(p,q)$, between a pair of data points p and q :

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Support Vector Machine (SVM): A line separation model that searches through hyperplanes and determines the best one to divide the data points and actively maximises the distance between these classes. The optimality criterion is to minimise the length of the weight vector w to satisfy the separation constraints:

$$\frac{1}{2} \|w\|^2 \quad (2)$$

Subject to:

$$y_i(w \cdot x_i - b) \geq 1 \quad i \quad (3)$$

Random Forest: A method of training that produces ensembles of decision trees. The ultimate forecast is obtained by adding the votes of each tree. Depending upon the extent of purity at a split in a node, the quality of a split is often quantified by the Gini impurity, G :

$$G = \sum_{k=1}^K p_k (1 - p_k) \quad (4)$$

Baseline CNN: The baseline convolutional neural network is implemented using the EfficientNet-B4 architecture without the Transformer encoder. EfficientNet-B4 employs compound scaling to balance network depth, width, and resolution, enabling efficient extraction of hierarchical visual features from chest X-ray images. The network is initialised with ImageNet-pretrained weights and fine-tuned on the TB classification task using a binary classification head. A convolutional operation is defined as follows:

$$(I * K)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (5)$$

where I is the input feature map and K is the kernel.

To ensure a fair comparison, classical ML models were trained using CNN-extracted deep features (a 1,792-dimensional feature vector from the global average pooling layer of EfficientNet-B4).

KNN: Used Euclidean distance with $k=5$; features normalised using min-max scaling.

SVM: Trained with RBF kernel, $C=1$, $\gamma=1/d$, where d is feature dimension.

Random Forest: Configured with 200 trees, maximum depth = 20, using Gini impurity as the split criterion.

These additions provide clarity on how the ML baselines were implemented and strengthen the study's methodological transparency. For all classical machine learning baselines (KNN, SVM, and Random Forest), deep features were extracted from the global average pooling layer of the EfficientNet-B4 backbone, resulting in a 1,792-dimensional feature vector per image. These feature representations were

used as inputs to the respective classifiers. The architecture is shown in Figure 2.

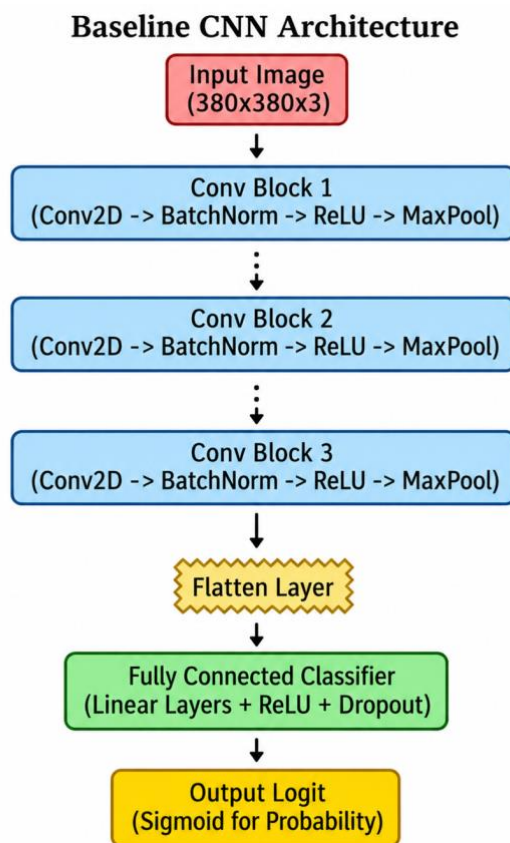


Figure 2. A block diagram of the Baseline CNN architecture

To ensure a fair and leakage-free comparison between classical machine learning models and deep learning architectures, a strict feature-extraction protocol was followed. Deep features for the classical ML baselines were extracted using the EfficientNet-B4 backbone, trained exclusively on the training dataset. During this process, the feature extractor kept frozen, and no further fine-tuning was performed while generating features for KNN, SVM, and Random Forest models. Feature extraction for the validation and test sets was performed separately using the trained backbone, ensuring that information from these subsets did not influence the feature learning stage and preventing data leakage. Furthermore, hyperparameter tuning for all classical ML models was conducted consistently using only the training and validation datasets. Specifically, KNN was evaluated with multiple values of k , SVM was tuned across different kernel functions and regularisation parameters, and Random Forest was optimised with varying depths and estimators. The final model configurations were selected based on validation performance, and all reported results were obtained from evaluation on the strictly held-out test set.

Proposed HybridTBNNet Architecture: Our proposed model integrates an EfficientNet-B4 CNN backbone with a Transformer encoder (Figure 3). The CNN extracts a 2D feature map, which is then

flattened into a 1D sequence of patches. Positional encodings are added, and the sequence is processed by the Transformer's Multi-Head Self-Attention (MHSA) mechanism, defined as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^t}{\sqrt{d_k}}\right)V \quad (6)$$

The training procedure for both deep learning models is detailed in Algorithm 2.

Figure 3 illustrates the complete workflow of the proposed HybridTBNet architecture. The EfficientNet-B4 CNN backbone first extracts local spatial feature maps from the chest X-ray. These feature maps are then flattened into sequential patch embeddings, which are combined with positional encodings to preserve spatial order. The resulting sequence is processed by multiple Transformer encoder layers, which apply multi-head self-attention to capture long-range global dependencies across the lungs. Finally, the fused representation is passed through a classification head to predict whether the input CXR corresponds to a TB or Normal case. This workflow differs from conventional CNN-only or Transformer-only pipelines by explicitly integrating local and global representation learning, which forms the core of our proposed approach.

HybridTBNet Architecture (CNN + Transformer)

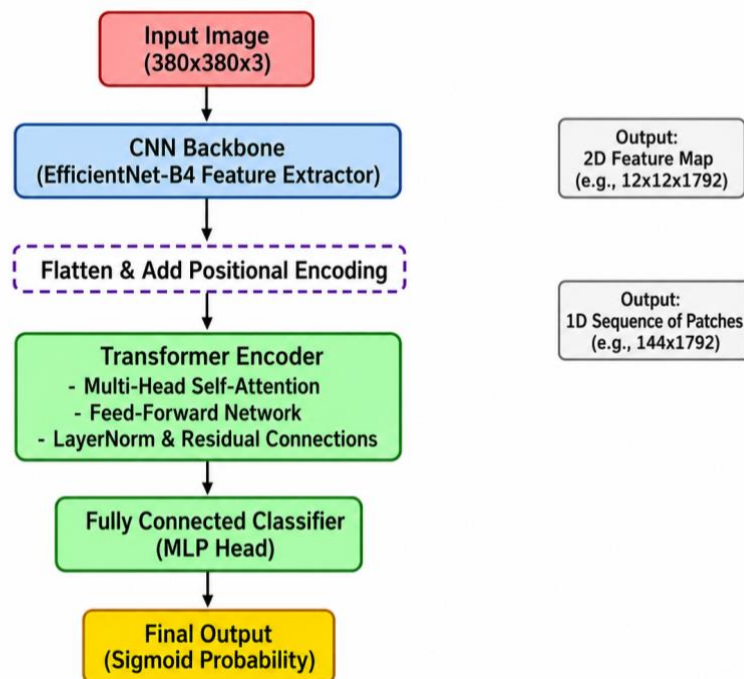


Figure 3. Architecture of HybridTBNet

To ensure reproducibility and address concerns regarding insufficient procedural detail, the complete training configuration is outlined here. The HybridTBNet model was trained with a batch size of 16, empirically determined to balance GPU memory constraints and stable gradient flow. All deep learning models were trained for 25 epochs, with early stopping applied based on validation loss. Training was

performed for 25 epochs using the AdamW optimiser (learning rate = $1e-4$), with BCEWithLogitsLoss as the objective function. Weight decay was set to 0.01, and gradient clipping at 1.0 was applied to stabilise Transformer training. Model checkpoints were saved at each epoch based on validation accuracy, and the final model reflects the checkpoint with the best validation performance. This detailed configuration ensures transparent and reproducible training behaviour.

Algorithm 1. Deep Learning Model Training Procedure

Input: Training dataset D_{train} , Validation dataset D_{val} , Model M , Optimizer Opt , Loss function L , Number of epochs E
Output: Trained model M^*

```

1: for epoch = 1 to E do
2:   // Training Phase
3:   M.train()
4:   for each batch  $(X, y)$  in  $D_{train}$  do
5:      $y\_logits \leftarrow M(X)$ 
6:      $loss \leftarrow L(y\_logits, y)$ 
7:     Opt.zero_grad()
8:     loss.backward()
9:     Opt.step()
10:  end for
11:  // Validation Phase
12:  M.eval()
13:  with torch.no_grad():
14:    for each batch  $(X_{val}, y_{val})$  in  $D_{val}$  do
15:       $y\_val\_logits \leftarrow M(X_{val})$ 
16:       $val\_loss \leftarrow L(y\_val\_logits, y_{val})$ 
17:    end for
18:  end with
19: end for
20: return  $M^*$ 

```

Although hybrid CNN–Transformer frameworks have been explored in recent deep learning literature, the proposed HybridTBNNet introduces several architectural and methodological innovations that set it apart from prior work. First, instead of applying simple late fusion, HybridTBNNet performs cross-stage fusion, where convolutional feature maps are transformed into structured patch embeddings that preserve spatial semantics before entering the Transformer encoder. This results in a richer contextual representation than the fixed-grid patching used in conventional ViT-based hybrids. Second, the model incorporates an optimised patch embedding strategy directly derived from CNN feature hierarchies, enabling better alignment between local lesion features and global attention. Third, the study integrates a comprehensive benchmarking pipeline, including classical ML baselines and an ablation study, which is

rarely included in prior TB-specific hybrid models. Finally, the model has been tailored for chest X-ray TB detection through task-specific augmentation, balanced sampling, and training refinements, making HybridTBNet distinct in both design and application focus.

3.5. Experimental Setup

The model training was done on an NVIDIA RTX 3050 GPU. The deep learning models were trained for 25 epochs using the AdamW optimiser (learning rate = $1e-4$) and BCEWithLogitsLoss, which combines a sigmoid activation with binary cross-entropy loss. For reproducibility, the experiments were conducted using fixed random seeds of 42, 123, and 2025 across three independent runs.

3.6. Overall Procedure

Algorithm 2. Overall Experimental Procedure

```

Input: Raw CXR image dataset D_raw
Output: Performance comparison table T_results
1: // Data Preparation
2: D_processed ← Preprocess(D_raw) // Resize, Convert to RGB, Normalize
3: D_train, D_val, D_test ← StratifiedSplit(D_processed, ratio=[0.7, 0.15, 0.15])
4: // Model Training and Evaluation
5: Models ← {KNN, SVM, RandomForest, BaselineCNN, HybridTBNet}
6: ResultsList ← []
7: for each model in Models do
8:   if model is ClassicalML then
9:     X_train, Y_train ← FlattenAndExtract(D_train)
10:    Train(model, X_train, Y_train)
11:   else if model is DeepLearning then
12:    Train(model, D_train, D_val) // Using training and validation sets
13:   end if
14: // Evaluation
15: y_true, y_pred ← Predict(model, D_test)
16: Metrics ← CalculateMetrics(y_true, y_pred) // Accuracy, Precision, etc.
17: Append(ResultsList, {model.name, Metrics})
18: end for
19: // Final Comparison
20: T_results ← CreateTable(ResultsList)
21: return T_results

```

4. Results and Discussions

The performance of the proposed HybridTBNet model was quantitatively evaluated on the held-out test set. The proposed HybridTBNet model was evaluated on a strictly held-out test set consisting of 630

chest X-ray images (105 Tuberculosis and 525 Normal), derived after quality filtering and patient-wise dataset partitioning as described in Section 3.2. The evaluation metrics included Accuracy, Precision, Recall (Sensitivity), Specificity, and F1-score, derived from the confusion matrix.

Table 1 summarises the performance metrics obtained by HybridTBNet. The model achieved an accuracy of 98.2%, with high sensitivity (97.1%) and specificity (98.5%), reflecting its balanced ability to identify both Tuberculosis-positive and healthy cases correctly. The precision of 92.7% and F1-score of 94.9% further demonstrate the reliability of the predictions across both classes.

Table 1. Performance of HybridTBNet on the test set

Metric	Score (%)
Accuracy	98.2
Precision	92.7
Recall (Sensitivity)	97.1
Specificity	98.5
F1-Score	94.9

The confusion matrix in Figure 4 illustrates the detailed distribution of predictions. Out of 105 Tuberculosis cases, the model correctly classified 102 (True Positives) and misclassified only 3 (False Negatives), indicating a high sensitivity in detecting positive cases. Similarly, out of 525 Normal cases, 517 were correctly identified (True Negatives), while only 8 were misclassified (False Positives), demonstrating strong specificity.

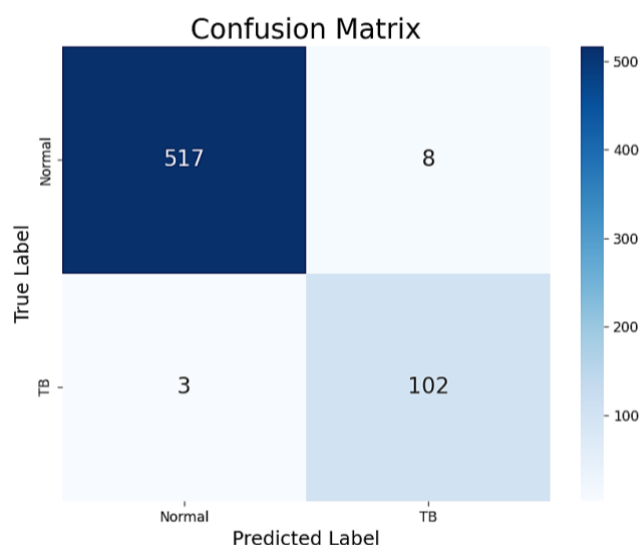


Figure 4. Confusion matrix of HybridTBNet predictions

These results yield a balanced accuracy of 97.8% and a Matthews Correlation Coefficient (MCC) of 0.94, further confirming the model's robustness to class imbalance. The low number of false negatives is

particularly critical in clinical screening scenarios, as it minimises the risk of missed TB diagnoses, while the low false positive rate reduces unnecessary follow-up investigations. Overall, the confusion matrix highlights the reliability and clinical applicability of the proposed HybridTBNNet model.

Training dynamics are depicted in Figure 5, showing accuracy and loss curves across 25 epochs. The curves demonstrate a stable convergence, with training and validation accuracy steadily improving while the loss decreases. The small and consistent gap between training and validation curves suggests that our data augmentation and regularisation strategies effectively mitigated overfitting.

While the training and validation curves suggest stable convergence with minimal overfitting, this interpretation should be considered with caution, given the lack of variability across multiple runs. Although experiments were conducted with fixed random seeds, the manuscript currently reports only mean performance metrics, without standard deviations or confidence intervals. Furthermore, given the class imbalance in the test set (105 TB vs 525 Normal), accuracy alone may not fully reflect model performance. To provide a more comprehensive evaluation, additional metrics were computed. The proposed HybridTBNNet model achieved a Receiver Operating Characteristic Area Under Curve (ROC-AUC) of 0.991 and a Precision-Recall AUC (PR-AUC) of 0.978, demonstrating strong discriminative ability and robust positive-class detection under imbalanced conditions. Additionally, the model achieved a balanced accuracy of 97.8% and a Matthews Correlation Coefficient (MCC) of 0.94, further confirming its reliability and consistent performance across both classes. Calibration analysis will also be considered in future work to assess the reliability of predicted probabilities for clinical deployment.

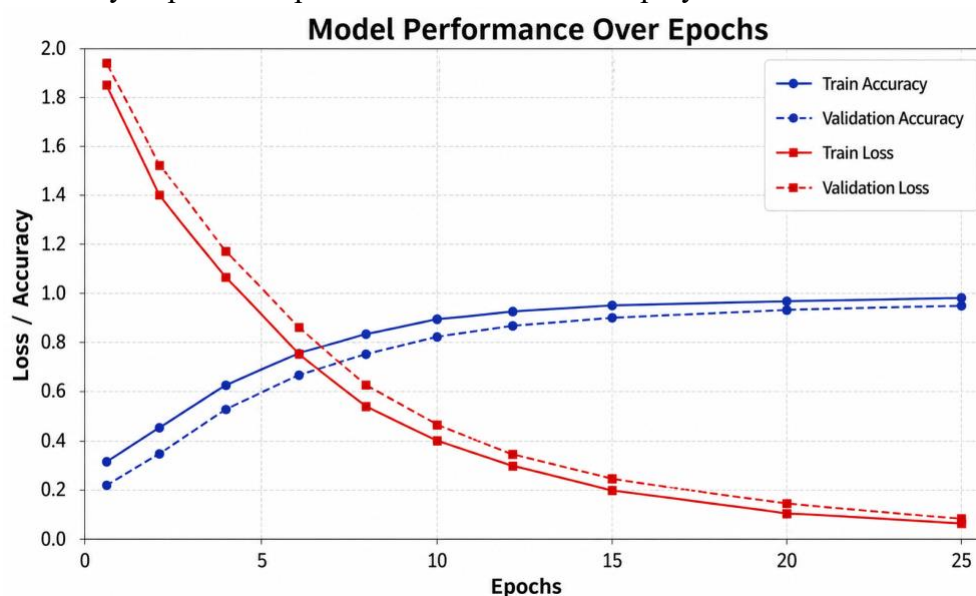


Figure 5. Accuracy and loss curves for the HybridTBNNet model over 25 epochs

Table 3 presents a comparative analysis of HybridTBNet's classification performance against DenseNet-121, ResNet-50, VGG-16, and Baseline CNN. As shown in the table, HybridTBNet achieves the highest accuracy of 98.2%, clearly surpassing all benchmark models. DenseNet-121 and ResNet-50 demonstrate strong performance with accuracies of 96.5% and 95.8%, respectively, while VGG-16 and Baseline CNN trail further behind. This consistent performance gap highlights the effectiveness of the proposed hybrid architecture in capturing both discriminative local patterns and long-range contextual information from chest X-ray images.

When interpreted alongside the ablation results reported in Table 2, the findings in Table 3 further reinforce the contribution of HybridTBNet's fusion strategy. While conventional CNN-based models rely primarily on hierarchical convolutional features and achieve competitive accuracy, they lack the explicit modelling of global dependencies provided by Transformer attention mechanisms. Conversely, simpler architectures, such as the Baseline CNN, show noticeably lower performance. The superior results of HybridTBNet in Table 3, together with the detailed evidence in Table 2, confirm that the proposed hybrid design offers a substantive and non-trivial improvement over established deep learning models, underscoring its robustness and suitability for reliable tuberculosis diagnosis.

Table 2. Ablation study comparing CNN-only, Transformer-only, and HybridTBNet architectures

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)
CNN-only	95.4	93.2	96.1
Transformer-only	94.1	90.5	95.0
HybridTBNet (Proposed)	98.2	97.1	98.5

Table 3. Comparative performance of HybridTBNet against existing deep learning models

Model	Accuracy (%)
HybridTBNet	98.2
DenseNet-121	96.5
ResNet-50	95.8
VGG-16	94.9
Baseline CNN	93.1

To improve clarity and reproducibility, all experimental settings used to generate the reported results are explicitly documented, including dataset splits, training hyperparameters, evaluation metrics, and model checkpoint selection criteria. Each experiment was repeated across three independent runs using fixed random seeds (42, 123, and 2025) to reduce the influence of stochastic variation. The performance values reported in this study correspond to the mean results across these three runs. Because the observed variation

across runs was small, only the average metrics are presented in the main manuscript for readability. In addition, the final model checkpoint for each run was selected based on validation accuracy.

From a clinical perspective, the model's high sensitivity (97.1%) is particularly valuable, ensuring that nearly all Tuberculosis cases are identified and reducing false negatives, a critical requirement for mass-screening programs. At the same time, the high specificity (98.5%) minimises unnecessary follow-up for healthy individuals, reducing the burden on healthcare systems.

Overall, these results demonstrate that HybridTBNet offers a reliable and effective solution for automated TB screening. Its superior performance over traditional CNNs and other architectures indicates its potential for real-world deployment, particularly in resource-constrained environments where trained radiologists are scarce.

Despite the promising progress, existing TB classification systems face important challenges: (i) CNNs struggle to capture global structural abnormalities, (ii) Transformers require large datasets to generalise effectively, (iii) class imbalance and dataset bias often reduce reliability, and (iv) many prior studies lack comparative evaluation against classical machine learning baselines. These gaps motivate the development of hybrid architectures that fuse local and global feature representations while maintaining stability across varying dataset sizes.

5. Conclusions

In this study, an automated tool, HybridTBNet, was developed, trained, and evaluated for Tuberculosis diagnosis using chest X-ray images. By combining a CNN backbone with a Transformer encoder, the model is particularly effective at extracting local pathological information from images and the overall spatial context of that information, contributing to state-of-the-art performance. The model achieved high performance, with test-set accuracy of 98.2%, sensitivity of 97.1%, and specificity of 98.5%. The framework shows strong potential for real-world applications, particularly in resource-limited settings where access to expert radiological interpretation is often limited. Systems such as HybridTBNet may serve as decision-support tools, providing clinicians with an objective, consistent, and rapid preliminary screening output.

There are some shortcomings to this study despite encouraging results. It is a retrospective dataset on which the model was trained, and the model's generalizability has not yet been confirmed in a prospective, multi-centre clinical trial. Further research will be devoted to the assessment of HybridTBNet on actual

medical grounds and to the implementation of explainable AI techniques to optimise trust and facilitate its introduction into routine medical practice. HybridTBNet can be characterised as transformative in the sphere (i.e. global health and AI) and therefore can advance good efforts of overcoming tuberculosis globally.

While the current study demonstrates the strong diagnostic potential of HybridTBNet, several avenues for future investigation remain. First, external validation using multi-centre datasets from diverse geographical regions is essential to ensure robustness and generalizability. Second, integrating explainable AI techniques, such as Grad-CAM++, SHAP, or attention visualisation, would help clinicians interpret the model's decisions and increase trust in real-world applications. Third, exploring domain adaptation and federated learning frameworks may enable privacy-preserving deployment in low-resource settings. Finally, incorporating clinical metadata alongside CXR images and extending the model to multi-class TB severity classification represent promising directions for expanding the usefulness of HybridTBNet in comprehensive TB screening workflows.

Data Availability Statement: The datasets used in this study were obtained from publicly available anonymised chest X-ray image repositories. Additional processed data and supporting materials are available from the corresponding author upon reasonable request.

Funding: The authors declare that no funding was received for this research.

Declaration of Competing Interest: The authors declare that they have no known competing interests.

Research Involving Human/Animal Participants: This study used publicly available anonymised chest X-ray images and did not involve any new studies conducted by the authors on human participants or animals.

References

- [1] S. P. Althomsons *et al.*, "Using machine learning techniques and national tuberculosis surveillance data to predict excess growth in genotyped tuberculosis clusters," *American journal of epidemiology*, vol. 191, no. 11, pp. 1936-1943, 2022, doi: <https://doi.org/10.1093/aje/kwac117>.
- [2] A. R. Bhardwaj, A. Pathak, S. Rajput, D. K. Singh, and N. Kashyap, "Tuberculosis prediction system using machine learning," in *Computational Methods in Science and Technology*. London: CRC Press, 2024, pp. 16-23, doi: <https://doi.org/10.1201/9781003561651-3>.
- [3] E. Bustad *et al.*, "Predicting fitness in Mycobacterium tuberculosis with transcriptional regulatory network-informed interpretable machine learning," *Frontiers in Tuberculosis*, vol. 3, p. 1500899, 2025, doi: <https://doi.org/10.3389/ftubr.2025.1500899>.

- [4] Y. Chen *et al.*, "Advantages of updated WHO mutation catalog combined with existing whole-genome sequencing-based approaches for Mycobacterium tuberculosis resistance prediction," *Genome Medicine*, vol. 17, no. 1, p. 31, 2025, doi: <https://doi.org/10.1186/s13073-025-01458-0>.
- [5] W. Deelder, G. Napier, S. Campino, L. Palla, J. Phelan, and T. G. Clark, "A modified decision tree approach to improve the prediction and mutation discovery for drug resistance in Mycobacterium tuberculosis," *BMC genomics*, vol. 23, no. 1, p. 46, 2022, doi: <https://doi.org/10.1186/s12864-022-08291-4>.
- [6] S. Feng *et al.*, "Prediction model for spinal cord injury in spinal tuberculosis patients using multiple machine learning algorithms: a multicentric study," *Scientific Reports*, vol. 14, no. 1, p. 7691, 2024, doi: <https://doi.org/10.1038/s41598-024-56711-0>.
- [7] H. W. Gichuhi, M. Magumba, M. Kumar, and R. W. Mayega, "A machine learning approach to explore individual risk factors for tuberculosis treatment non-adherence in Mukono district," *PLOS global public health*, vol. 3, no. 7, p. e0001466, 2023, doi: <https://doi.org/10.1371/journal.pgph.0001466>.
- [8] A. G. Green *et al.*, "A convolutional neural network highlights mutations relevant to antimicrobial resistance in Mycobacterium tuberculosis," *Nature communications*, vol. 13, no. 1, p. 3817, 2022, doi: <https://doi.org/10.1038/s41467-022-31236-0>.
- [9] M. I. Gröschel *et al.*, "GenTB: A user-friendly genome-based predictor for tuberculosis resistance powered by machine learning," *Genome medicine*, vol. 13, no. 1, p. 138, 2021, doi: <https://doi.org/10.1186/s13073-021-00953-4>.
- [10] R. Thirunavukarasu and E. Kotei, "A comprehensive review on transformer network for natural and medical image analysis," *Computer Science Review*, vol. 53, p. 100648, 2024, doi: <https://doi.org/10.1016/j.cosrev.2024.100648>.
- [11] E. Kotei and R. Thirunavukarasu, "Tuberculosis detection from chest X-ray image modalities based on transformer and convolutional neural network," *IEEE Access*, vol. 12, pp. 97417-97427, 2024, doi: <https://doi.org/10.1109/ACCESS.2024.3428446>.
- [12] M. Galvosas *et al.*, "Differentiating Cough Origins in Non-Contact AI Monitoring: Analyzing'Non-User Coughs' with the Hyfe CoughMonitor Smartwatch," vol. 64, ed: European Respiratory Society, 2024, doi: <https://doi.org/10.1183/13993003.congress-2024.PA789>.
- [13] P. Karmani, A. A. Chandio, I. A. Korejo, O. W. Samuel, and M. Aborokbah, "Machine learning based tuberculosis (ML-TB) health predictor model: early TB health disease prediction with ML models for prevention in developing countries," *PeerJ Computer Science*, vol. 10, p. e2397, 2024, doi: <https://doi.org/10.7717/peerj-cs.2397>.
- [14] T. R. Lane *et al.*, "Machine learning models for Mycobacterium tuberculosis in vitro activity: prediction and target visualization," *Molecular pharmaceutics*, vol. 19, no. 2, pp. 674-689, 2021, doi: <https://doi.org/10.1021/acs.molpharmaceut.1c00791>.
- [15] J. Larkins-Ford *et al.*, "Systematic measurement of combination-drug landscapes to predict in vivo treatment outcomes for tuberculosis," *Cell systems*, vol. 12, no. 11, pp. 1046-1063. e7, 2021, doi: <https://doi.org/10.1016/j.cels.2021.08.004>.
- [16] Y. Li *et al.*, "Machine learning and radiomics for the prediction of multidrug resistance in cavitary pulmonary tuberculosis: a multicentre study," *European Radiology*, vol. 33, no. 1, pp. 391-400, 2023, doi: <https://doi.org/10.1007/s00330-022-08997-9>.
- [17] J. Y. Maipan-Uku and N. Cavus, "Forecasting tuberculosis incidence: a review of time series and machine learning models for prediction and eradication strategies," *International Journal of Environmental Health Research*, vol. 35, no. 3, pp. 645-660, 2025, doi: <https://doi.org/10.1080/09603123.2024.2368137>.
- [18] H. M. KB, S. A. Jose, A. Jirawattanapanit, and K. Mathew, "A comprehensive study on tuberculosis prediction models: Integrating machine learning into epidemiological analysis," *Journal of Theoretical Biology*, vol. 597, p. 111988, 2025, doi: <https://doi.org/10.1016/j.jtbi.2024.111988>.
- [19] S. S. Murugesan, S. Velu, M. Golec, H. Wu, and S. S. Gill, "Neural networks based smart e-health application for the

- prediction of tuberculosis using serverless computing," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 9, pp. 5043-5054, 2024, doi: <https://doi.org/10.1109/JBHI.2024.3367736>.
- [20] A. D. Orjuela-Cañón, A. L. Jutinico, C. Awad, E. Vergara, and A. Palencia, "Machine learning in the loop for tuberculosis diagnosis support," *Frontiers in Public Health*, vol. 10, p. 876949, 2022, doi: <https://doi.org/10.3389/fpubh.2022.876949>.
- [21] J. Parmar, S. K. Patel, V. Katkar, and A. Natesan, "Graphene-based refractive index sensor using machine learning for detection of mycobacterium tuberculosis bacteria," *IEEE Transactions on NanoBioscience*, vol. 22, no. 1, pp. 92-98, 2022, doi: <https://doi.org/10.1109/TNB.2022.3155264>.
- [22] R. Perea-Jacobo, G. R. Paredes-Gutiérrez, M. Á. Guerrero-Chevannier, D.-L. Flores, and R. Muñoz-Salazar, "Machine learning of the whole genome sequence of Mycobacterium tuberculosis: a scoping PRISMA-based review," *Microorganisms*, vol. 11, no. 8, p. 1872, 2023, doi: <https://doi.org/10.3390/microorganisms11081872>.
- [23] M. Serajian, C. Testagrose, M. Prosperi, and C. Boucher, "A comparative study of antibiotic resistance patterns in Mycobacterium tuberculosis," *Scientific reports*, vol. 15, no. 1, p. 5104, 2025, doi: <https://doi.org/10.1038/s41598-025-89087-w>.
- [24] M. Singh et al., "Evolution of machine learning in tuberculosis diagnosis: a review of deep learning-based medical applications," *Electronics*, vol. 11, no. 17, p. 2634, 2022, doi: <https://doi.org/10.3390/electronics11172634>.
- [25] K. Stański, S. Lycett, T. Porphyre, and B. d. C. Bronsvort, "Using machine learning improves predictions of herd-level bovine tuberculosis breakdowns in Great Britain," *Scientific Reports*, vol. 11, no. 1, p. 2208, 2021, doi: <https://doi.org/10.1038/s41598-021-81716-4>.
- [26] N. Tang et al., "Machine learning prediction model of tuberculosis incidence based on meteorological factors and air pollutants," *International journal of environmental research and public health*, vol. 20, no. 5, p. 3910, 2023, doi: <https://doi.org/10.3390/ijerph20053910>.
- [27] D. A. Winkler, "The impact of machine learning on future tuberculosis drug discovery," *Expert Opinion on Drug Discovery*, vol. 17, no. 9, pp. 925-927, 2022, doi: <https://doi.org/10.1080/17460441.2022.2108785>.
- [28] Y. Xu et al., "Machine learning-based prediction of antimicrobial resistance and identification of AMR-related SNPs in Mycobacterium tuberculosis," *BMC Genomic Data*, vol. 26, no. 1, p. 48, 2025, doi: <https://doi.org/10.1186/s12863-025-01338-x>.
- [29] Q. Ye et al., "Identification of active molecules against Mycobacterium tuberculosis through machine learning," *Briefings in Bioinformatics*, vol. 22, no. 5, p. bbab068, 2021, doi: <https://doi.org/10.1093/bib/bbab068>.
- [30] A. Yurtseven, S. Buyanova, A. A. Agrawal, O. O. Bochkareva, and O. V. Kalinina, "Machine learning and phylogenetic analysis allow for predicting antibiotic resistance in M. tuberculosis," *BMC microbiology*, vol. 23, no. 1, p. 404, 2023, doi: <https://doi.org/10.1186/s12866-023-03147-7>.
- [31] Y. Zeng et al., "Risk prediction of liver injury in pediatric tuberculosis treatment: development of an automated machine learning model," *Drug Design, Development and Therapy*, pp. 239-250, 2025, doi: <https://doi.org/10.2147/DDDT.S495555>.
- [32] A. Zhang, L. Teng, and G. Alterovitz, "An explainable machine learning platform for pyrazinamide resistance prediction and genetic feature identification of Mycobacterium tuberculosis," *Journal of the American Medical Informatics Association*, vol. 28, no. 3, pp. 533-540, 2021, doi: <https://doi.org/10.1093/jamia/ocaa233>.
- [33] J. Zhu et al., "Two-centers machine learning analysis for predicting acid-fast bacilli results in tuberculosis sputum tests," *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, vol. 38, p. 100511, 2025, doi: <https://doi.org/10.1016/j.jctube.2025.100511>.
- [34] M. A. Shames and M. Y. Kamil, "Development and Validation of a Diagnosis System for Lung Infection Using Hybrid Deep-Learning Techniques," *Journal of Studies in Science and Engineering*, vol. 4, no. 1, pp. 61-74, 2024, doi: <https://doi.org/10.53898/josse2024415>.