Original Article

Check for updates

Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 3, No. 1, 2025

Explainable Deep Learning with Lightweight CNNs for Tuberculosis Classification

Teuku Rizky Noviandy [1], Ghazi Mauer Idroes [2], Teuku Zulfikar [3] and Rinaldi Idroes [4,*]

[1] Department of Information Systems, Faculty of Engineering, Universitas Abulyatama, Aceh Besar 23372, Indonesia; rizky_si@abulyatama.ac.id (T.R.N.)
[2] Department of Occupational Health and Safety, Faculty of Health Sciences, Universitas Abulyatama, Aceh Besar 23372, Indonesia; idroesghazi_k3@abulyatama.ac.id (G.M.I.)
[3] Department of Pulmonology and Respiratory Medicine, Faculty of Syiah Kuala, Universitas Syiah Kuala/Zainoel Abidin Hospital, Banda Aceh, Indonesia; teukuzulfikar@ymail.com (T.Z.)
[4] School of Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

* Correspondence: rinaldi.idroes@usk.ac.id

Article History

Received 26 February 2025
Revised 13 May 2025
Accepted 19 May 2025
Available Online 26 May 2025

Keywords:
Chest radiography
Medical image analysis
Model interpretability
Resource-limited settings
Deep feature extraction

Abstract

Tuberculosis (TB) remains a major global health threat, particularly in low-resource settings where timely diagnosis is critical yet often limited by the lack of radiological expertise. Chest X-rays (CXRs) are widely used for TB screening, but manual interpretation is prone to errors and variability. While deep learning has shown promise in automating CXR analysis, most existing models are computationally intensive and lack interpretability, limiting their deployment in real-world clinical environments. Three lightweight and explainable CNN architectures were evaluated to address this gap: ShuffleNetV2, SqueezeNet 1.1, and MobileNetV3. These models were assessed for binary TB classification using a locally sourced dataset comprising 3,008 CXR images. Transfer learning and Gradient-weighted Class Activation Mapping (Grad-CAM) were applied to provide visual explanations. MobileNetV3 and ShuffleNetV2 achieved perfect test performance, with 100% accuracy, sensitivity, specificity, precision, and F1-score. Both models also recorded AUC scores 1.00, with inference times of 94.66 and 103.63 seconds, respectively. SqueezeNet performed moderately, with a lower F1-score of 82.98% and several misclassifications. These results demonstrate that lightweight CNNs can deliver high diagnostic accuracy and transparency, supporting their use in scalable, AI-assisted TB screening systems for underserved healthcare settings.

1. Introduction

Tuberculosis (TB) remains a major global health threat and is once again the leading cause of death from a single infectious agent worldwide, after being surpassed by COVID-19 for three years. In 2023, an estimated 10.8 million people contracted TB, including 6.0 million men, 3.6 million women, and 1.3 million children, with approximately 1.25 million deaths attributed to the disease [1].

Early and accurate diagnosis of TB is critical to reducing transmission and ensuring timely treatment initiation. Chest X-rays (CXRs) remain among the most accessible and cost-effective tools for pulmonary TB screening, particularly in low-resource settings such as rural clinics, mobile health units, and correctional facilities [2]. However, interpreting CXRs accurately requires experienced radiologists, who are often scarce or unavailable in these environments [3]. Manual

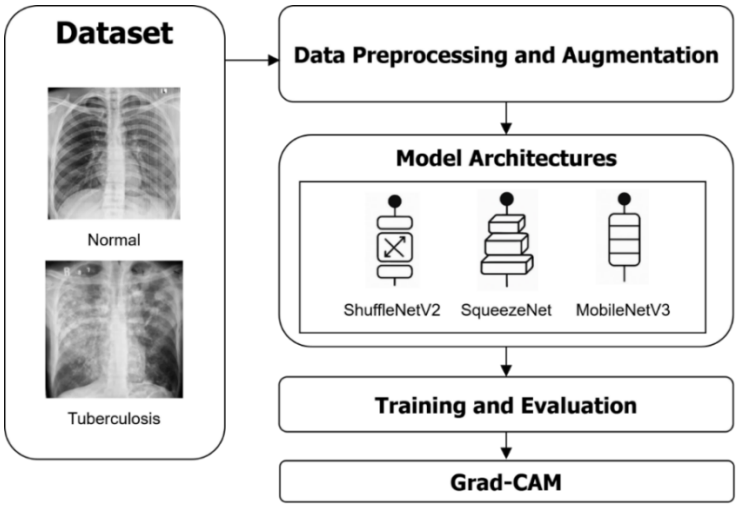


Figure 1. Workflow of the proposed TB classification system using chest X-rays.

interpretation introduces variability due to reader fatigue, differences in training, and inconsistent diagnostic standards [4]. Subtle radiographic signs of early-stage TB, such as faint infiltrates or minor cavitations, are frequently missed, increasing the risk of undiagnosed cases. In high-burden regions with limited personnel and high patient throughput, these limitations contribute to diagnostic delays and inappropriate treatment. To address this, there is a growing need for automated TB classification tools that can operate reliably on portable devices, achieve clinically acceptable sensitivity and specificity, and integrate smoothly into frontline workflows as decision-support systems, rather than replacements for radiologists.

In recent years, deep learning, particularly convolutional neural networks (CNNs), has revolutionized medical image analysis by offering automated, accurate, and scalable solutions for disease classification, lesion detection, and segmentation [5–7]. CNNs have shown strong performance in various radiological applications, including pneumonia detection [8], lung nodule classification [9], and TB screening [10]. These models learn complex hierarchical representations directly from pixel data, enabling them to outperform traditional machine learning techniques that rely on hand-crafted features [11].

Several studies have successfully applied deep CNNs such as ResNet [12], DenseNet [13], and VGG [14] to detect TB in CXRs with high accuracy. While these models achieve promising diagnostic performance, they are often computationally intensive and require substantial memory and processing power, limiting their practical use in real-time or edge-based diagnostic systems [15]. Additionally, most existing models lack transparency, providing little insight into the rationale behind their predictions, which hinders clinical adoption [16].

Although deep learning has significantly advanced TB classification from CXRs, most high-performing models rely on large, computationally intensive architectures unsuitable for mobile or resource-constrained settings [17]. At the same time, while more feasible for deployment, lightweight models often suffer from reduced diagnostic accuracy and limited interpretability [18]. Moreover, few existing approaches incorporate visual explanation techniques, such as class activation maps, to aid clinicians in understanding model predictions [16]. This combination of low efficiency, lack of transparency, and uncertain diagnostic performance in lightweight solutions underscores a critical gap: the need for TB classification models that are both interpretable and optimized for deployment in low-resource environments.

This study aims to develop and evaluate lightweight and interpretable CNN models for automated TB classification from CXR images to address existing limitations. The performance of three efficient deep learning architectures is investigated: SqueezeNet, ShuffleNetV2, and MobileNetV3. A dataset comprising TB-positive and normal CXRs is used for training and evaluation. Gradient-weighted Class Activation Mapping (Grad-CAM) is incorporated to visualize and interpret model predictions, which enhances clinical trust and supports decision-making. The objective is to provide a fast, accurate, and explainable AI solution suitable for deployment in real-world, resource-constrained healthcare settings.

2. Materials and Methods

The proposed TB classification framework is illustrated in Figure 1, outlining the overall workflow. It begins with a curated normal and TB-positive CXR image dataset, followed by systematic data preprocessing and

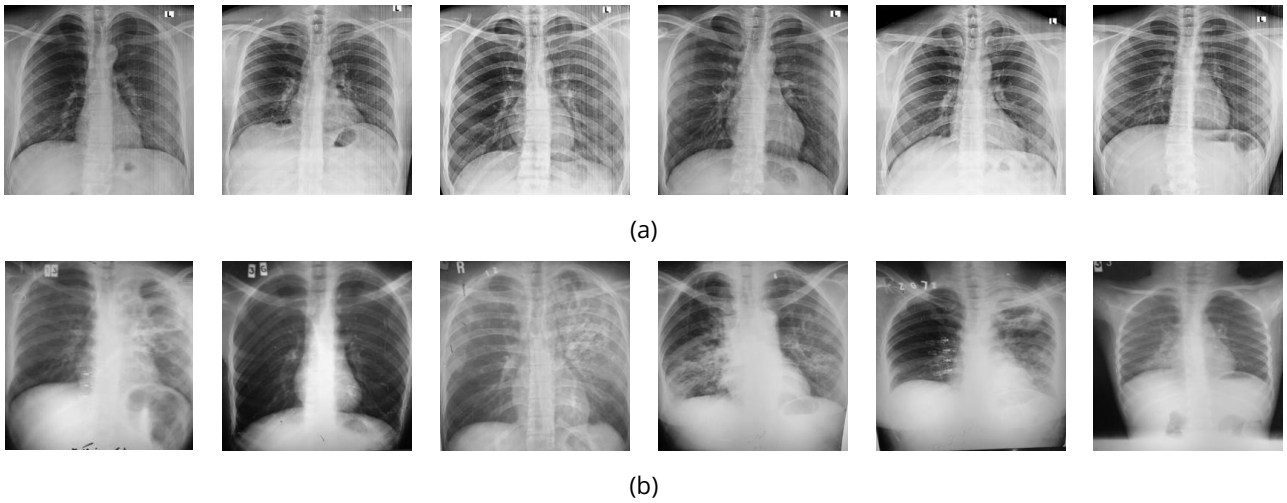


Figure 2. Sample chest X-ray images from the dataset. (a) Normal cases with clear lung fields. (b) Tuberculosis-positive cases show abnormalities such as infiltrates, opacities, or cavitary lesions.

augmentation. Three lightweight CNN architectures, ShuffleNetV2, SqueezeNet 1.1, and MobileNetV3, are utilized to perform automated TB classification. These models undergo training and evaluation using performance metrics tailored for medical image analysis. Finally, Grad-CAM visualizes model attention regions, enhancing interpretability and clinical trust.

2.1. Dataset Description

The dataset used in this study, contributed by Kiran and Jabeen via Mendeley Data [19], comprises 3,008 posterior-anterior view CXR images collected from a local hospital in Pakistan. It includes 2,494 images of confirmed TB cases and 514 normal (healthy) CXR images, annotated based on clinical diagnosis and radiological findings. Representative examples of normal and TB-positive CXR images are shown in Figure 2.

2.2. Data Preprocessing and Augmentation

All CXR images were resized to 224×224 pixels to match the input dimensions required by the selected deep learning models, which are commonly pre-trained on ImageNet at this resolution [20]. This standard size balances computational efficiency with sufficient spatial detail for feature extraction. The dataset was randomly split into 70% for training, 20% for validation, and 10% for testing, with a fixed random seed to ensure reproducibility [21].

A comprehensive data augmentation pipeline was applied exclusively to the training set to enhance model generalization and reduce overfitting. These augmentations simulate real-world variations in CXR imaging, including changes in orientation, lighting, position, and artifacts or occlusions [22]. The validation and test sets were processed using only resizing and normalization to preserve the integrity of model evaluation. Table 1 summarizes the augmentation techniques applied during training.

All images were normalized using the standard ImageNet mean and standard deviation [23]. Data were loaded using PyTorch's DataLoader, with a batch size 32. Training data were shuffled on each epoch, while validation and test sets were fixed. This preprocessing strategy ensures that the models are trained on a diverse and realistic distribution of X-ray appearances, improving their robustness and generalization performance across unseen cases.

2.3. Model Architectures

Three state-of-the-art lightweight CNN architectures were selected to balance diagnostic accuracy, computational efficiency, and model interpretability: ShuffleNetV2, SqueezeNet 1.1, and MobileNetV3. These models are specifically designed for deployment in resource-constrained environments, offering reduced model size and faster inference times without significant compromises in classification performance. Each architecture was chosen to represent a different design strategy for efficient deep learning: ShuffleNetV2 for its high throughput and low memory cost, SqueezeNet for its extreme parameter reduction, and MobileNetV3 for its advanced optimization techniques tailored to mobile platforms. This diversity allows for a comprehensive comparison of performance, efficiency, and explainability in the context of TB classification.

2.3.1. ShuffleNetV2

ShuffleNetV2 is an efficient CNN architecture tailored for high-speed and low-memory operations, particularly on mobile and embedded devices [24]. It improves upon its

Table 1. Summary of data augmentation techniques applied.

Transformation	Purpose
Random Resized Crop	Introduces spatial variability by cropping and scaling image regions
Random Horizontal Flip	Simulates anatomical symmetry and positional variation
Random Rotation (±15°)	Accounts for patient posture and slight angle differences
Color Jitter	Mimics variations in brightness, contrast, saturation, and hue
Random Affine Translation	Introduces small translations to simulate changes in image acquisition
Random Erasing	Simulates occlusions and artifacts commonly present in medical imaging
Normalization (ImageNet)	Standardizes pixel intensity values for stable training

predecessor by addressing key limitations such as memory access cost (MAC) and parallelism. The network employs channel splitting, shuffling, and pointwise group convolutions to reduce computational complexity while maintaining high feature representation capacity. In this study, the 1.0× version of ShuffleNetV2 with ImageNet pre-trained weights was implemented and subsequently fine-tuned on the TB dataset. A global average pooling layer was appended to the final convolutional output, followed by a fully connected classification head with a sigmoid activation function for binary TB classification.

2.3.2. SqueezeNet 1.1

SqueezeNet 1.1 is a compact deep learning model that significantly reduces the number of parameters (fewer than 1.25 million) while maintaining comparable accuracy to larger models like AlexNet [25]. The core building block of SqueezeNet is the Fire module, which combines a "squeeze" layer using 1×1 filters and an "expand" layer that applies a mix of 1×1 and 3×3 filters. The 1.1 version offers a smaller model size and improved inference speed, making it suitable for edge-based TB screening applications. The final architecture used in this study consists of the standard SqueezeNet 1.1 backbone with a modified classification head, including dropout regularization and a sigmoid activation to support binary classification.

2.3.3. MobileNetV3

MobileNetV3, the most advanced among the selected models, integrates innovations such as squeeze-and-excitation (SE) modules, hard-swish activations, and efficient depthwise separable convolutions [26]. These enhancements enable it to achieve superior trade-offs between speed and accuracy. The MobileNetV3-Small variant, optimized for low-latency inference on mobile devices, was adopted in this study. The model was initialized with weights pre-trained on ImageNet and fine-tuned using the TB dataset. A custom classification layer, consisting of global average pooling, a dropout layer (p=0.2), and a final dense layer with sigmoid activation, was appended to output TB probability scores.

2.4. Training Procedure

All models were trained using transfer learning, with weights initialized from ImageNet-pretrained networks. Feature extraction layers were frozen during early epochs to retain general visual representations [27]. Fine-tuning was subsequently applied to all layers to optimize performance on the TB classification task.

Model training was conducted using the Adam optimizer with a learning rate 0.001 and a weight decay of 1×10^{-6} to prevent overfitting. The objective function used was categorical cross-entropy, which is suitable for binary classification. To adaptively adjust the learning rate, a ReduceLROnPlateau scheduler was employed, which monitored the validation loss and reduced the learning rate by a factor of 0.1 if no improvement was observed for three consecutive epochs. A batch size 32 was used for all training runs, and models were trained for a maximum of 10 epochs. All experiments used PyTorch on a GPU-enabled system to accelerate training and inference.

2.5. Evaluation Metrics

A comprehensive set of evaluation metrics was employed to evaluate the performance of the proposed CNN models for TB classification. These metrics provide a holistic understanding of model effectiveness, especially in a medical context where false positives and negatives carry significant clinical implications [28].

- Accuracy measures the overall proportion of correct predictions, indicating how often the model classifies CXRs correctly.
- Precision reflects the proportion of true TB cases among all instances predicted as TB-positive. High precision is important to minimize false alarms.
- Sensitivity quantifies the model's ability to detect actual TB cases correctly. High sensitivity is critical in TB screening to avoid missed diagnoses.
- Specificity evaluates the model's ability to correctly identify normal (non-TB) cases, ensuring that healthy individuals are not misclassified.

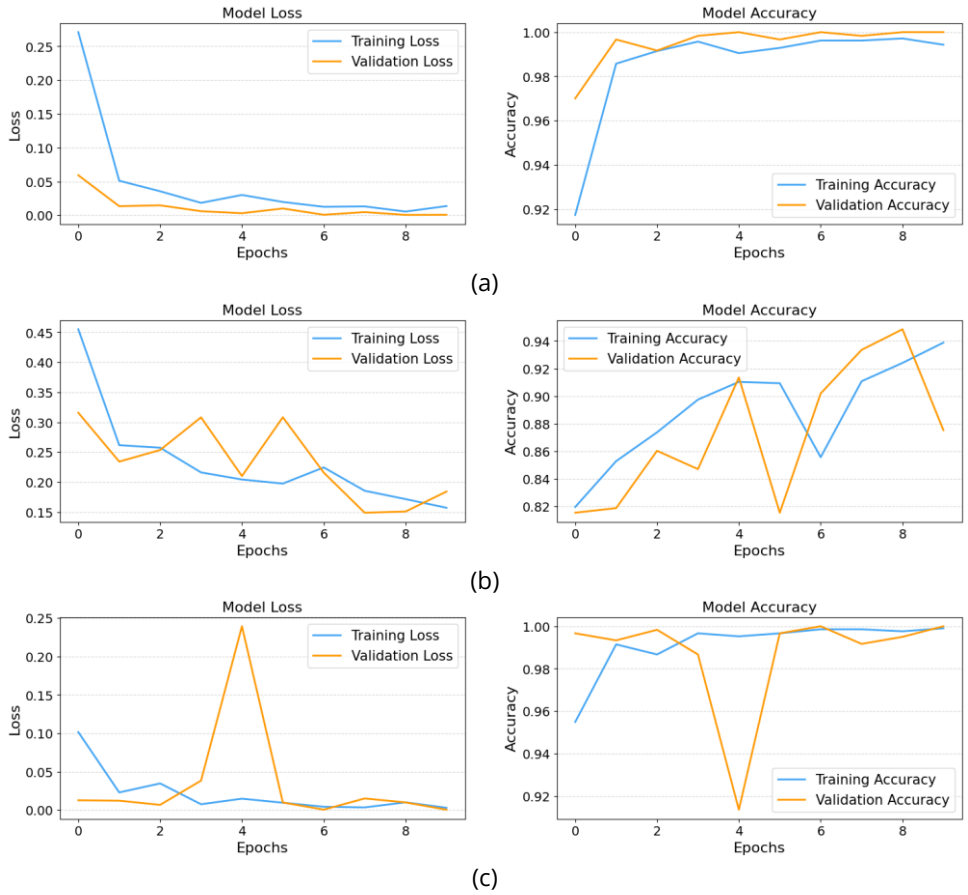


Figure 3. Model loss and accuracy for (a) ShuffleNetV2; (b) SqueezeNet; (c) MobileNetV3.

- F1-Score is the harmonic mean of precision and sensitivity. It provides a balanced measure, especially useful in datasets with class imbalance.

In addition to reporting scalar metrics, Receiver Operating Characteristic (ROC) curves were visualized for each model to provide an intuitive understanding of classification behavior across different thresholds [29, 30]. This visualization supports better clinical interpretability and trust in model decisions.

2.6. Explainability with Grad-CAM

To enhance model interpretability, Grad-CAM was used to visualize the regions of CXRs that influenced each model's predictions [31]. Grad-CAM produces heatmaps that highlight important areas in the image by tracing gradients back to the final convolutional layers [32].

For each model, including ShuffleNetV2, SqueezeNet 1.1, and MobileNetV3, Grad-CAM was applied to two representative test images: one TB-positive case and one normal case. In TB cases, the models focused on regions showing typical abnormalities such as infiltrates or cavitations. The activation was diffuse or minimal for normal cases, consistent with the absence of disease.

3. Results and Discussion

The performance of the three lightweight CNN models was evaluated over 10 training epochs using training and validation accuracy and loss curves. Figure 3 shows that ShuffleNetV2 demonstrated rapid convergence and stable learning behavior. Its training and validation losses decreased consistently, while validation accuracy quickly plateaued near 99%, indicating strong generalization and minimal overfitting. This suggests that ShuffleNetV2 is well-suited for the TB classification, balancing performance and efficiency.

SqueezeNet exhibited more fluctuations in both validation loss and accuracy. Although the model eventually achieved over 90% validation accuracy, its instability across epochs, evident in the jagged curves, suggests sensitivity to hyperparameters or limited capacity to capture complex features. This may be due to its compact architecture, which trades off depth for speed. MobileNetV3 achieved high performance with accuracy exceeding 98% on both training and validation sets. However, Figure 3 shows a noticeable spike in validation loss and dip in accuracy around epoch 5, indicating a brief overfitting or instability phase.

Table 2. Test the performance and inference time of each model.

Model	Time (s)	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)
ShuffleNetV2	103.63	100.00	100.00	100.00	100.00	100.00
SqueezeNet	101.02	94.70	79.59	86.67	96.11	82.98
MobileNetV3	94.66	100.00	100.00	100.00	100.00	100.00

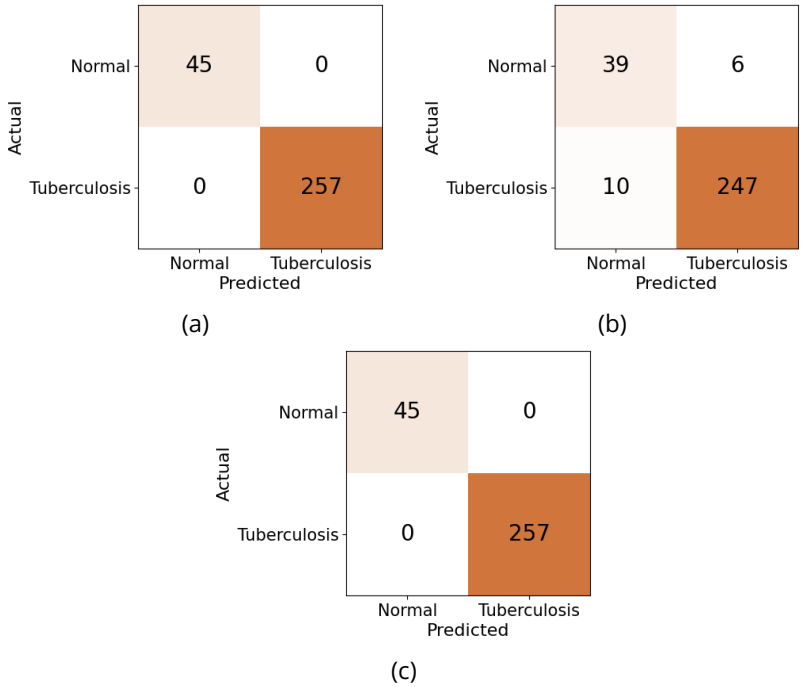


Figure 4. Confusion matrices for test set predictions using (a) ShuffleNetV2, (b) SqueezeNet, and (c) MobileNetV3.

Nevertheless, the model recovered well, showing strong final performance and generalization.

Table 2 summarizes the test set's quantitative results, including accuracy, precision, sensitivity, specificity, F1 Score, and inference time. ShuffleNetV2 and MobileNetV3 achieved perfect scores across all evaluation metrics, indicating exceptional performance distinguishing TB-positive from normal CXRs. Their sensitivity and specificity, reaching 100%, suggest no false negatives or false positives in the test samples, which is especially critical in clinical settings.

While still performing reasonably well, SqueezeNet lagged behind the other two models. It achieved an accuracy of 94.70%, with a notably lower precision of 79.59% and an F1-score of 82.98%. These values suggest the model misclassified several normal cases as TB or failed to identify all TB-positive cases, impacting its clinical reliability.

Regarding inference time, MobileNetV3 was the fastest, completing test set evaluation in 94.66 seconds, followed closely by SqueezeNet (101.02 s) and ShuffleNetV2 (103.63 s). The small difference in execution times further supports the viability of all three models for real-time or edge-based deployment, although MobileNetV3 offers a slight computational advantage.

To understand model behavior further, confusion matrices were generated for each CNN using the test set, as shown in Figure 4. ShuffleNetV2 and MobileNetV3 achieved perfect classification, correctly identifying all normal and TB-positive cases without misclassifying. This outcome reinforces the earlier quantitative metrics, confirming these models' high sensitivity and specificity.

In contrast, SqueezeNet misclassified 16 cases; six normal cases were incorrectly predicted as TB (false positives), and 10 TB cases were incorrectly labeled as normal (false negatives). This pattern suggests that SqueezeNet may struggle to distinguish subtle radiographic features that separate early-stage TB from normal findings, or vice versa. False negatives, in particular, may result from the model failing to capture less prominent signs of TB, such as faint infiltrates or minimal cavitations. Meanwhile, the false positives could indicate that the model is overly sensitive to nonspecific lung opacities or artifacts that resemble TB-like patterns. These misclassifications contribute to the lower precision

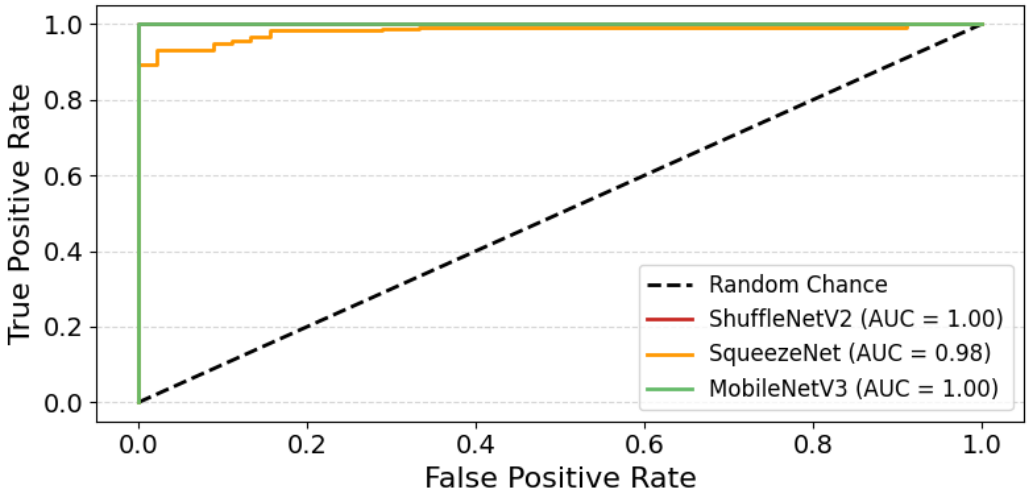


Figure 5. ROC curves for ShuffleNetV2, SqueezeNet, and MobileNetV3, with corresponding AUC values.

and F1-score observed in Table 2. Although SqueezeNet remains lightweight and reasonably accurate, its higher rate of false positives and negatives limits its suitability for clinical deployment, where diagnostic errors can delay treatment or lead to unnecessary interventions.

ROC analysis assessed each model's discriminative capability across various threshold settings. As shown in Figure 5, both ShuffleNetV2 and MobileNetV3 achieved an AUC of 1.00, indicating perfect separation between TB-positive and normal cases. This confirms their earlier performance metrics and reflects excellent model confidence and reliability.

While still performing well, SqueezeNet achieved a slightly lower AUC of 0.98. This aligns with its reduced precision and sensitivity, suggesting occasional misclassification near the decision boundary. Despite this, the AUC remains high, indicating a strong capability to distinguish between the two classes.

Grad-CAM visualizations were generated for a representative normal case and a TB-positive case to assess model interpretability, as illustrated in Figure 6. All three models, ShuffleNetV2, SqueezeNet, and MobileNetV3, correctly predicted both cases, but the confidence and attention maps varied.

For the normal case, MobileNetV3 and ShuffleNetV2 showed high confidence (100% and 99.00%, respectively). They focused activation on lung regions that appeared radiologically normal, indicating that the models did not rely on irrelevant features. SqueezeNet also predicted the case as normal but with significantly lower confidence (69.13%), and its activation map showed less concentrated attention, possibly explaining its less stable performance overall.

In the TB-positive case, all three models again provided correct predictions with high confidence, and their Grad-CAM maps highlighted pathological regions typically associated with TB, such as localized opacities. ShuffleNetV2 and MobileNetV3 produced sharper, more focused activation in the affected lung zones, while SqueezeNet's map was more diffuse, though still aligned with abnormal areas.

These visualizations confirm that the best-performing models achieved high accuracy and made predictions based on medically relevant image regions. This strengthens their case for use in clinical decision support, particularly in low-resource settings where explainability is crucial.

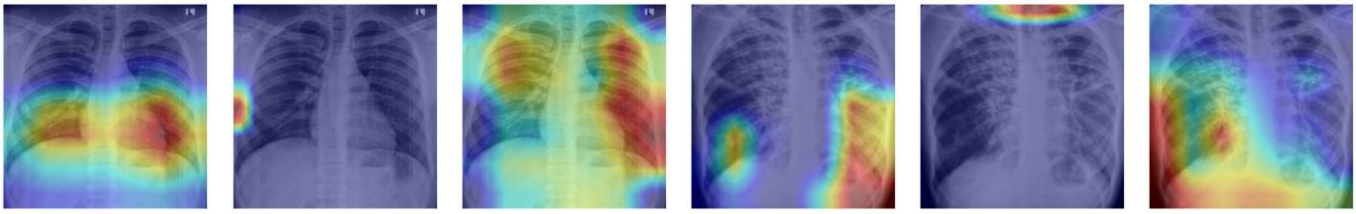
Based on the results, ShuffleNetV2 and MobileNetV3 outperformed SqueezeNet across all key evaluation metrics. Both models achieved 100% accuracy, precision, sensitivity, specificity, and F1-score on the test set. Their ROC curves showed perfect AUC values (1.00), and their confusion matrices indicated no misclassifications, highlighting their strong performance in distinguishing TB-positive from normal cases.

However, these perfect metrics should be interpreted with caution. Although the results are promising, they may not fully capture the variability and complexity of real-world clinical settings. Image quality, patient diversity, and atypical presentations can affect model reliability. Further evaluation on larger and more heterogeneous external datasets is necessary to validate these findings and assess generalizability.

In comparison, SqueezeNet performed reasonably well but with lower effectiveness. It achieved an accuracy of 94.70 percent, a precision of 79.59 percent, and an F1-score of 82.98 percent. Its confusion matrix revealed 16

Actual Class: Normal

Actual Class: Tuberculosis



ShuffleNetV2	**SqueezeNet**	**MobileNetV3**	**ShuffleNetV2**	**SqueezeNet**	**MobileNetV3**
Predicted: Normal Confidence: 99.00%	Predicted: Normal Confidence: 69.13%	Predicted: Normal Confidence: 100.00%	Predicted: Tuberculosis Confidence: 100.00%	Predicted: Tuberculosis Confidence: 99.77%	Predicted: Tuberculosis Confidence: 100.00%

Figure 6. Grad-CAM visualizations showing model predictions, confidence scores, and highlighted regions for a normal and a TB-positive chest X-ray.

misclassifications, and its AUC was slightly lower at 0.98, indicating reduced clinical reliability.

MobileNetV3 had the shortest inference time (94.66 seconds), making it the most computationally efficient option. ShuffleNetV2 was slightly slower but similarly accurate and stable during training. SqueezeNet, while efficient, showed instability in its training curves and attention maps and reduced confidence in predictions.

Grad-CAM results further supported the quantitative findings. ShuffleNetV2 and MobileNetV3 produced focused and clinically meaningful activation maps with high prediction confidence. In contrast, SqueezeNet's explanations were less localized and its confidence varied, indicating potential limitations in feature learning.

These findings indicate that ShuffleNetV2 and MobileNetV3 are highly effective for TB classification in CXRs, combining strong accuracy, speed, and interpretability. MobileNetV3 may be particularly well suited for real-time or mobile deployment scenarios due to its faster inference and stable performance. Meanwhile, ShuffleNetV2 remains a robust alternative, especially in settings where training consistency and resilience are key. SqueezeNet, despite its compact architecture, may benefit from further refinement to achieve diagnostic performance comparable to the other models.

While the proposed models demonstrated strong performance in TB classification, several limitations should be acknowledged. First, the dataset used in this study was sourced from a single hospital in Pakistan, which may limit the generalizability of the models to other populations, imaging equipment, or clinical settings. Diverse, multi-center datasets are essential to ensure broader applicability and robustness across real-world deployments. Second, although the models achieved near-perfect accuracy on the test set, the dataset was imbalanced, with a higher proportion of TB-positive cases. This imbalance may have influenced learning dynamics and should be addressed in future work through strategies such as data resampling or cost-sensitive training.

Another limitation is the binary classification framework. While useful for TB screening, real-world diagnosis often involves distinguishing TB from other lung diseases such

as pneumonia, lung cancer, or chronic obstructive pulmonary disease. Extending the models to multi-class classification would improve their clinical utility. Grad-CAM offers only coarse localization and is limited in interpreting deeper feature interactions or explaining misclassifications in complex cases. Future work could explore more advanced explainability techniques, such as layer-wise relevance propagation or attention-based models, to provide richer insights into model decisions.

Finally, while the current models are suitable for offline analysis or use in edge devices, real-time clinical deployment would benefit from integration into user-friendly mobile or web-based platforms. Future directions include deploying these models in field settings, conducting prospective validation, and incorporating clinician feedback to improve performance and usability iteratively.

4. Conclusions

This study demonstrated the effectiveness of lightweight and explainable CNN models, ShuffleNetV2, SqueezeNet 1.1, and MobileNetV3, for automated TB classification using CXR images. Among them, ShuffleNetV2 and MobileNetV3 achieved perfect classification performance with high efficiency and clinically meaningful visual explanations via Grad-CAM, making them suitable for deployment in resource-constrained healthcare settings. MobileNetV3, in particular, offered the best trade-off between speed and accuracy, while ShuffleNetV2 provided strong training stability. In contrast, SqueezeNet showed reduced reliability and confidence in predictions. The findings support integrating efficient deep learning models into TB screening workflows, especially where expert radiological support is limited.

Author Contributions: Conceptualization, T.R.N. and R.I.; methodology, T.R.N. and G.M.I.; software, T.R.N. and G.M.I.; validation, T.Z. and R.I.; formal analysis, T.R.N. and G.M.I.; investigation, T.R.N. and G.M.I.; resources, T.Z. and R.I.; data curation, T.Z. and R.I.; writing—original draft preparation, T.R.N. and G.M.I.; writing—review and editing, T.Z. and R.I.; visualization, T.R.N.; supervision, R.I.; project administration, R.I.; funding acquisition, R.I. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this study is publicly available and can be accessed through Mendeley Data at https://data.mendeley.com/datasets/8j2g3csprk/2.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. World Health Organization (WHO). (2025). Tuberculosis, World Health Organization, from https://www.who.int/news-room/fact-sheets/detail/tuberculosis, accessed 20-3-2025.
2. Yousif, D., Mesilhy, R., Aly, R., Hegazi, S., Yousif, Z., Cyprian, F. S., and Abdallah, A. M. (2024). Innovations in Tuberculosis Disease Screening, *Surveillance, Prevention, and Control of Infectious Diseases*, Springer Nature Switzerland, Cham, 97–113. doi:10.1007/978-3-031-59967-5_5.
3. Karera, A., Engel-Hills, P., and Davidson, F. (2024). Radiology Image Interpretation Services in a Low-Resource Setting: Medical Doctors' Experiences and the Potential Role of Radiographers, *Radiography*, Vol. 30, No. 2, 560–566. doi:10.1016/j.radi.2024.01.009.
4. Brady, A. P. (2017). Error and Discrepancy in Radiology: Inevitable or Avoidable?, *Insights into Imaging*, Vol. 8, No. 1, 171–182. doi:10.1007/s13244-016-0534-1.
5. Suganyadevi, S., Seethalakshmi, V., and Balasamy, K. (2022). A Review on Deep Learning in Medical Image Analysis, *International Journal of Multimedia Information Retrieval*, Vol. 11, No. 1, 19–38. doi:10.1007/s13735-021-00218-1.
6. Noviandy, T. R., Maulana, A., Zulfikar, T., Rusyana, A., Enitan, S. S., and Idroes, R. (2024). Explainable Artificial Intelligence in Medical Imaging: A Case Study on Enhancing Lung Cancer Detection through CT Images, *Indonesian Journal of Case Reports*, Vol. 2, No. 1, 6–14. doi:10.60084/ijcr.v2i1.150.
7. Al-qaness, M. A. A., Zhu, J., AL-Alimi, D., Dahou, A., Alsamhi, S. H., Abd Elaziz, M., and Ewees, A. A. (2024). Chest X-Ray Images for Lung Disease Detection Using Deep Learning Techniques: A Comprehensive Survey, *Archives of Computational Methods in Engineering*, Vol. 31, No. 6, 3267–3301. doi:10.1007/s11831-024-10081-y.
8. Yao, S., Chen, Y., Tian, X., and Jiang, R. (2021). Pneumonia Detection Using an Improved Algorithm Based on Faster R-CNN, *Computational and Mathematical Methods in Medicine*, Vol. 2021, 1–13. doi:10.1155/2021/8854892.
9. Mahmood, S. A., and Ahmed, H. A. (2022). An Improved CNN-Based Architecture for Automatic Lung Nodule Classification, *Medical & Biological Engineering & Computing*, Vol. 60, No. 7, 1977–1986. doi:10.1007/s11517-022-02578-0.
10. Alawi, A. E. B., Al-basser, A., Sallam, A., Al-sabaeei, A., and Al-khateeb, H. (2021). Convolutional Neural Networks Model for Screening Tuberculosis Disease, *2021 International Conference of Technology, Science and Administration (ICTSA)*, IEEE, 1–5. doi:10.1109/ICTSA52017.2021.9406520.
11. Al-Jabbar, M., Alshahrani, M., Senan, E. M., and Ahmed, I. A. (2023). Analyzing Histological Images Using Hybrid Techniques for Early Detection of Multi-Class Breast Cancer Based on Fusion Features of CNN and Handcrafted, *Diagnostics*, Vol. 13, No. 10, 1753. doi:10.3390/diagnostics13101753.
12. Ejiyi, C. J., Qin, Z., Nnani, A. O., Deng, F., Ejiyi, T. U., Ejiyi, M. B., Agbesi, V. K., and Bamisile, O. (2024). ResfEANet: ResNet-Fused External Attention Network for Tuberculosis Diagnosis Using Chest X-Ray Images, *Computer Methods and Programs in Biomedicine Update*, Vol. 5, 100133. doi:10.1016/j.cmpbup.2023.100133.
13. Huy, V. T. Q., and Lin, C.-M. (2023). An Improved Densenet Deep Neural Network Model for Tuberculosis Detection Using Chest X-Ray Images, *IEEE Access*, Vol. 11, 42839–42849. doi:10.1109/ACCESS.2023.3270774.
14. Mohan, R., Kadry, S., Rajinikanth, V., Majumdar, A., and Thinnukool, O. (2022). Automatic Detection of Tuberculosis Using VGG19 with Seagull-Algorithm, *Life*, Vol. 12, No. 11, 1848. doi:10.3390/life12111848.
15. Xu, Y., Khan, T. M., Song, Y., and Meijering, E. (2025). Edge Deep Learning in Computer Vision and Medical Diagnostics: A

Comprehensive Survey, *Artificial Intelligence Review*, Vol. 58, No. 3, 93. doi:10.1007/s10462-024-11033-5.

16. Zhang, Y., Weng, Y., and Lund, J. (2022). Applications of Explainable Artificial Intelligence in Diagnosis and Surgery, *Diagnostics*, Vol. 12, No. 2, 237. doi:10.3390/diagnostics12020237.
17. Kotei, E., and Thirunavukarasu, R. (2024). A Comprehensive Review on Advancement in Deep Learning Techniques for Automatic Detection of Tuberculosis from Chest X-Ray Images, *Archives of Computational Methods in Engineering*, Vol. 31, No. 1, 455–474. doi:10.1007/s11831-023-09987-w.
18. Muhammad, D., Salman, M., Keles, A., and Bendechache, M. (2025). ALL Diagnosis: Can Efficiency and Transparency Coexist? An Explainble Deep Learning Approach, *Scientific Reports*, Vol. 15, No. 1, 12812. doi:10.1038/s41598-025-97297-5.
19. Kiran, S., and Jabeen, D. I. (2024). Dataset of Tuberculosis Chest X-Rays Images, *Mendeley Data*. doi:10.17632/8j2g3csprk.2.
20. Nguyen, T. T., and Nguyen, T. P. (2024). Rescaling Large Datasets Based on Validation Outcomes of a Pre-Trained Network, *Pattern Recognition Letters*, Vol. 185, 73–80. doi:10.1016/j.patrec.2024.07.001.
21. Idroes, G. M., Maulana, A., Suhendra, R., Lala, A., Karma, T., Kusumo, F., Hewindati, Y. T., and Noviandy, T. R. (2023). TeutongNet: A Fine-Tuned Deep Learning Model for Improved Forest Fire Detection, *Leuser Journal of Environmental Studies*, Vol. 1, No. 1, 1–8. doi:10.60084/ljes.v1i1.42.
22. Maulana, A., Noviandy, T. R., Suhendra, R., Earlia, N., Bulqiah, M., Idroes, G. M., Niode, N. J., Sofyan, H., Subianto, M., and Idroes, R. (2023). Evaluation of Atopic Dermatitis Severity Using Artificial Intelligence, *Narra J*, Vol. 3, No. 3, e511. doi:10.52225/narra.v3i3.511.
23. Noviandy, T. R., Idroes, G. M., Purnawarman, A., Imran, I., Lestari, N. D., Hastuti, S., and Idroes, R. (2024). Enhancing Early Detection of Alzheimer's Disease through MRI Using Explainable Artificial Intelligence, *Indonesian Journal of Case Reports*, Vol. 2, No. 2, 43–51. doi:10.60084/ijcr.v2i2.255.
24. Ma, N., Zhang, X., Zheng, H.-T., and Sun, J. (2018). ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design, 122–138. doi:10.1007/978-3-030-01264-9_8.
25. Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2016). SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters And< 0.5 MB Model Size, *ArXiv Preprint ArXiv:1602.07360*.
26. Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., and Vasudevan, V. (2019). Searching for MobileNetV3, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1314–1324.
27. Idroes, G. M., Noviandy, T. R., Emran, T. Bin, and Idroes, R. (2024). Explainable Deep Learning Approach for Mpox Skin Lesion Detection with Grad-CAM, *Heca Journal of Applied Sciences*, Vol. 2, No. 2, 54–63. doi:10.60084/hjas.v2i2.216.
28. Noviandy, T. R., Nisa, K., Idroes, G. M., Hardi, I., and Sasmita, N. R. (2024). Classifying Beta-Secretase 1 Inhibitor Activity for Alzheimer's Drug Discovery with LightGBM, *Journal of Computing Theories and Applications*, Vol. 2, No. 2, 138–147. doi:10.62411/jcta.10129.
29. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2025). Integrating Explainable Artificial Intelligence and Light Gradient Boosting Machine for Glioma Grading, *Informatics and Health*, Vol. 2, No. 1, 1–8. doi:10.1016/j.infoh.2024.12.001.
30. Noviandy, T. R., Maulana, A., Emran, T. B., Idroes, G. M., and Idroes, R. (2023). QSAR Classification of Beta-Secretase 1 Inhibitor Activity in Alzheimer's Disease Using Ensemble Machine Learning Algorithms, *Heca Journal of Applied Sciences*, Vol. 1, No. 1, 1–7. doi:10.60084/hjas.v1i1.12.
31. Lundberg, S. M., and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions, *Advances in Neural Information Processing Systems*, Vol. 30.
32. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, *2017 IEEE International Conference on Computer Vision (ICCV)*, IEEE, 618–626. doi:10.1109/ICCV.2017.74.