Enhancing Tuberculosis Detection from Chest X-Ray Images Using Deep Learning: Evaluating Multi-Architecture Performance and Efficiency

Deden Witarsyah
Faculty of Computer Science, Universiti Tun Hussein Onn Malaysia, Malaysia | Faculty of Computer Science, Universitas Brawijaya, Indonesia
deden@uthm.edu.my

Hadi Almohab
Faculty of Engineering, Computer, and Design, Nusa Putra University, Indonesia
hadi.almohab@nusaputra.ac.id (corresponding author)

Riswan Septriayadi Sianturi
Faculty of Computer Science, Universitas Brawijaya, Indonesia
rsianturi@ub.ac.id

Dedy Syamsuar
School of Information Systems, Bina Nusantara University, Indonesia
dedy.syamsuar@binus.ac.id

Received: 17 November 2025 | Revised: 8 December 2025, 25 December 2025, 13 January 2026, 1 February 2026, and 2 February 2026 | Accepted: 4 February 2026

 DOI: https://doi.org/10.48084/etasr.16321

ABSTRACT

Tuberculosis (TB) remains a critical global health challenge, with early detection often hindered by limited diagnostic resources and variability in manual Chest X-Ray (CXR) interpretation. This study evaluates the diagnostic performance and computational efficiency of four deep learning architectures—VGG16, DenseNet121, MobileNetV2, and a custom lightweight CNN—trained for 10 to 100 epochs on a multi-source dataset comprising 4,200 CXR images. Stratified data splitting and data augmentation were employed to address class imbalance and improve generalization. Among the evaluated models, VGG16 achieved the highest classification performance, with 99.21% accuracy, an F1-score of 0.98, and a ROC-AUC of 0.99. In contrast, MobileNetV2 provided a favorable trade-off between accuracy (98.73%) and inference efficiency, achieving the fastest prediction time of 12 ms per image, supporting its potential use in resource-constrained environments. Although the results indicate that transfer learning with optimized training protocols can support accurate and efficient AI-assisted TB screening, this study is limited by the lack of patient-level data separation, statistical cross-validation, and model explainability. Future work will address these limitations by incorporating explainable AI methods, statistically robust validation strategies, and evaluation on independent and diverse clinical datasets.

Keywords-tuberculosis classification; deep learning; chest X-ray; transfer learning; Convolutional Neural Networks (CNNs); medical image analysis

I. INTRODUCTION

Tuberculosis (TB) remains a major global health concern, with 10.8 million cases and 1.25 million deaths reported in 2023 [1]. Despite diagnostic advances, early detection in low-income regions is hindered by a scarcity of radiology expertise and limited infrastructure [2]. Chest X-ray (CXR) imaging remains a cost-effective screening method; however, manual interpretation is suffering from inter-observer variability and scalability issues, underscoring the need for accurate automated solutions [3].

Deep learning, particularly Convolutional Neural Networks (CNNs), has significantly increased TB detection from CXRs. Self-trained models, such as [4], achieved 96.57% accuracy, while pre-trained architectures such as MobileNetV2, VGG16, and VGG19 reported accuracies between 98.90% and 99.14% [5]. Fine-tuned VGG16 models have reached 98% accuracy [6], demonstrating the scalability and efficiency of deep learning for TB screening and reducing reliance on radiologists [7]. Training protocols also play a critical role in model performance. Optimized strategies have yielded improved results in diverse domains, including scene text recognition [8] and time-series regression [9], whereas improper transfer learning can lead to overfitting in small datasets [10]. Standardized processes enhance generalizability in human activity recognition [11], yet such systematic protocols remain underexplored in TB detection. Although dynamic learning rate scheduling and augmentation methods—such as rotation and brightness adjustment—have shown promise [12, 13], their performance varies across architectures [14]. Several recent studies have investigated CNN architectures and transfer learning for TB detection. In [15], VGG-19, ResNet50, and DenseNet121 were compared on multi-source CXR datasets, with VGG-19 achieving the highest AUC scores (0.89–0.95). In [16], 91% accuracy and F1-scores were reported for Xception and ResNet50, while in [17], 99.20% accuracy was achieved for TB using a tailored CNN model. Transfer learning has proven particularly valuable for limited datasets [18]. In [19], DenResCov-19 combined DenseNet121 and ResNet50 for multi-class CXR classification, achieving 94.7% accuracy. Mobile applications have demonstrated how AI diagnostics can extend into care coordination, particularly in resource-limited regions [20]. Further studies showed that ML-driven TB prediction models maintain strong performance even when datasets vary in quality [21], and reviews have emphasized the role of explainable AI in enhancing clinician trust and adoption [22]. Robust dataset construction remains pivotal for building generalizable models. In [23], adaptive preprocessing was introduced to mitigate domain shifts in TB detection. In [24], CNNs were integrated with traditional classifiers using the NIH dataset, emphasizing class balancing. In [25], the potential of semi-supervised learning for TB screening was demonstrated by achieving 95% accuracy with radiologist-guided labels. This study aggregates multi-institutional datasets (Belarus, NIAID, NLM, RSNA) and applies stratified splitting along with targeted augmentation techniques (rotation, flipping, brightness adjustment) to address class imbalance.

Deep learning model performance is significantly influenced by training strategies. In [26], the benefits of L2 regularization were highlighted for small medical datasets, while in [27], a comprehensive survey was provided on regularization strategies critical to generalization. In [28, 29], population-based training was used with regularization, demonstrating the value of adaptive optimization methods. This study employed a structured training protocol, with Adam optimization, learning rate scheduling, L2 regularization (λ = 0.01–0.001), and dropout (0.3–0.5) across 10–100 epochs to balance training efficiency and performance. Recent evaluations suggest that VGG-based models remain highly competitive for TB detection across multiple datasets. In [16], 90.0% accuracy and 91.0% AUC-ROC were reported for VGG16, outperforming deeper models such as ResNet152, and in [30], this was validated with a hyperparameter-tuned VGG19 achieving 98% accuracy. However, many studies overlook computational trade-offs, particularly in low-resource settings. MobileNetV2, known for its lightweight design, remains understudied for TB detection despite its success in other medical imaging applications.

This study evaluates VGG16, DenseNet121, MobileNetV2, and a custom lightweight CNN across 10–100 training epochs using a multi-source dataset of 4,200 CXRs. Stratified splitting and adaptive augmentation are applied to enhance generalization and benchmark accuracy, ROC-AUC, F1-score, and inference speed. The contributions of this study are as follows:

- Detailed quantitative analysis of performance-efficiency trade-offs across training durations (10–100 epochs)
- Benchmarking of lightweight architectures (MobileNetV2, custom CNN) against established models (VGG16, DenseNet121)
- Integration of multi-source CXR datasets with stratified splitting and adaptive augmentation
- Systematic evaluation of regularization strategies (λ = 0.01–0.001) across different training phases.

These contributions bridge the gap between high-performing deep learning models and their real-world deployment in resource-constrained clinical environments. A quantitative comparison with representative and recent TB detection models from the literature contextualizes the performance of the proposed approaches.

II. METHODOLOGY

This study assesses the performance of CNNs in detecting TB in CXR images. It compares a custom CNN architecture with three transfer learning models (VGG16, DenseNet121, and MobileNetV2) across training durations of 10, 20, 50, and 100 epochs to identify optimal trade-offs between efficiency and diagnostic accuracy. This study differs from previous VGG16- or MobileNet-based TB studies by evaluating adaptive augmentation strategies and systematically analyzing the effect of different training durations on model performance.

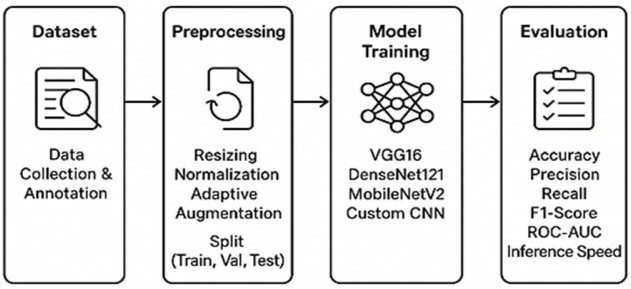

Fig. 1. Workflow for TB Detection Using CNNs and Transfer Learning.

A. Dataset and Preprocessing

1) Dataset

This study used the TB Chest X-ray Database, a publicly available composite dataset for tuberculosis detection. The dataset contains 4,200 chest X-ray images, including 700 TB-positive and 3,500 normal cases [31]. The dataset can be accessed in [32]. The dataset internally aggregates CXR images from multiple established public sources, including the NLM Montgomery and Shenzhen datasets, the Belarus Tuberculosis dataset, the NIAID TB Portal, and TB-related images from the RSNA Pneumonia Detection Challenge.

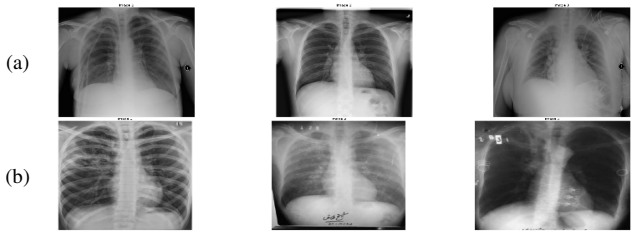

Fig. 2. Sample chest X-ray images: (a) normal, (b) tuberculosis.

2) Data Preparation

Images were resized to 128×128 pixels for the custom CNN and 224×224 pixels for the transfer learning models. Pixel values were normalized to the [0, 1] range. Adaptive augmentation was applied only to the training set to address class imbalance, including rotation (±30°), horizontal flip, and brightness adjustment (±20%). This also helped simulate domain variation across datasets. The dataset was stratified into training (2,940; 70%), validation (630; 15%), and test (630; 15%) sets, preserving the ratios of TB-positive and normal samples. The validation set included a higher share of images from the Montgomery and Belarus datasets to simulate deployment scenarios with unseen equipment and patient demographics.

TABLE I. SUMMARY OF PREPROCESSING STEPS

Step	Description
Resizing	Custom CNN: 128×128 px; Transfer Learning: 224×224 px
Normalization	Pixel values scaled to [0, 1] range
Augmentation	Applied only to training set: Rotation (±30°), Horizontal Flip, Shift (±20%), Brightness (±20%)
Stratified splitting	Total: 4,200 images; Train: 2,940 (70%), Validation: 630 (15%), Test: 630 (15%)

B. Model Architecture

1) Custom CNN

A custom CNN was developed to balance computational efficiency and diagnostic accuracy in TB detection. It consists of two convolutional blocks: the first with 64 filters and the second with 128 filters, both using 3×3 kernels. Batch normalization was applied after each convolutional layer to stabilize training and accelerate convergence, while max pooling with a 2×2 filter was used to reduce the spatial dimensions and computational requirements. The fully connected layers include 128 units with L2 regularization (λ = 0.01) and a dropout rate of 0.5 to mitigate overfitting. Sigmoid activation was applied on the final output layer for binary classification, as shown in Figure 3. This architecture was intentionally kept shallow to enable fast inference, suitable for low-resource settings or portable diagnostic devices.

2) Transfer Learning Models

Three pre-trained CNNs—VGG16, DenseNet121, and MobileNetV2—were selected to provide a representative comparison between high-capacity and lightweight architectures. VGG16 was chosen for its robust hierarchical feature extraction, as it has demonstrated strong performance in prior radiological image classification tasks. DenseNet121 was included for its dense connectivity pattern, which promotes feature reuse, improves gradient flow, and captures fine-grained image features with fewer parameters than other deep architectures. MobileNetV2 was selected for its inverted residual design and computational efficiency, which enable near real-time inference on embedded or mobile devices without substantial performance degradation. Fine-tuning strategies were tailored to each model. For VGG16, the final four convolutional blocks were unfrozen and followed by a global average pooling layer and two fully connected layers (512 and 256 units) with dropout rates of 0.5 and 0.3. DenseNet121 was fine-tuned by unfreezing the last 50 layers, with classification performed via global average pooling and L2-regularized dense layers (λ = 0.001). MobileNetV2 training proceeded in two phases: initially freezing the base layers for 30 epochs, followed by unfreezing the final 30 layers to enhance domain adaptation for tuberculosis detection.

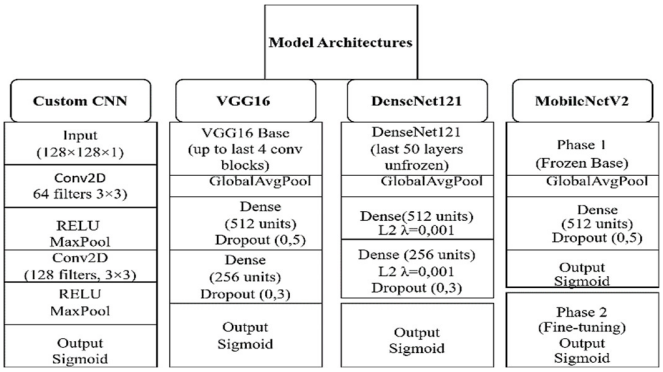

Fig. 3. Model architecture diagrams.

C. Training Protocol

1) Hyperparameters

All models were trained using the Adam optimizer. Learning rate reduction on plateau was employed (factor 0.5, patience 2 epochs). The batch size was 32. Training durations of 10, 20, 50, and 100 epochs were systematically evaluated to determine optimal performance and generalization. Dropout (0.3–0.5) and L2 regularization (λ = 0.01 for custom CNN, λ = 0.001 for DenseNet121) were applied to prevent overfitting.

2) Hardware and Software

All experiments were run on an NVIDIA Tesla V100 GPU (16 GB VRAM). TensorFlow 2.8 and Keras 2.6 were used for

implementation, while scikit-learn 1.0.2 was used for performance evaluation. Random seeds were fixed using NumPy and TensorFlow to ensure reproducibility.

D. Evaluation Metrics

Model performance was assessed using multiple metrics, including training accuracy and loss, to track learning behavior and identify overfitting or underfitting. Overall accuracy, precision, recall, and F1-score were used to evaluate class-wise prediction quality, supported by a confusion matrix showing correct and incorrect classifications. The ROC curve and ROC-AUC quantified the models' discrimination ability. Additionally, inference time (ms/image) was measured to determine each model's suitability for fast diagnostic use in high-incidence, resource-limited settings.

III. RESULTS

This section presents the evaluation of CNN models for classifying CXR images as Normal or Tuberculosis (TB). Inference speed was measured to evaluate real-time clinical applicability.

A. Model Performance Across Epochs

Table II summarizes the classification performance of the proposed models at different training epochs (10, 20, and 100). The VGG16 model achieved the highest performance across all key evaluation metrics.

TABLE II. MODEL PERFORMANCE METRICS ACROSS TRAINING EPOCHS

Model	Accuracy (%)	Precision (N)	Precision (TB)	Recall (N)	Recall (TB)	F1-score (TB)
VGG16	99.21	1.00	0.98	1.00	0.98	0.98
DenseNet121	97.94	0.97	0.99	1.00	0.86	0.92
MobileNetV2	98.73	0.99	1.00	1.00	0.95	0.98
Custom CNN (10 Epochs)	98.57	0.99	0.96	0.99	0.96	0.96
Custom CNN (20 Epochs)	96.67	0.97	0.95	0.99	0.85	0.89
Custom CNN (100 Epochs)	97.46	0.98	0.94	0.99	0.90	0.92

B. ROC Curve Analysis

The ROC curve for the VGG16 model illustrates its ability to distinguish between normal and TB cases (Figure 4). VGG16 achieved an ROC-AUC score of 0.99, exhibiting excellent discriminatory capability with both high sensitivity and specificity. This strong result reinforces its suitability for medical diagnostic applications where accurate separation between classes is critical.

C. Confusion Matrices

Figure 5 presents the confusion matrices for each model to analyze the classification outcomes, showing True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN):

- VGG16: TP = 103, TN = 523, FP = 2, FN = 2
- DenseNet121: TP = 90, TN = 524, FP = 1, FN = 15
- MobileNetV2: TP = 100, TN = 525, FP = 0, FN = 5
- Custom CNN (10 Epochs): TP = 100, TN = 521, FP = 4, FN = 5
- Custom CNN (20 Epochs): TP = 89, TN = 520, FP = 5, FN = 16
- Custom CNN (100 Epochs): TP = 95, TN = 519, FP = 6, FN = 10

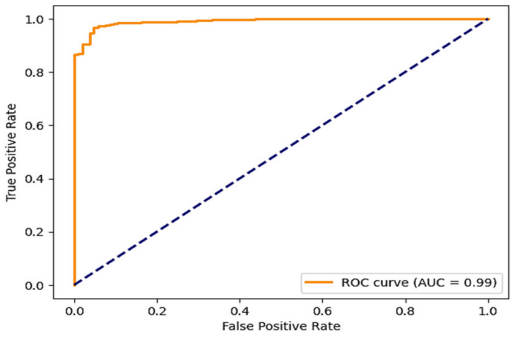

Fig. 4. ROC curve for VGG19.

D. Accuracy and Loss Trends

The training and validation accuracy/loss curves are summarized in Figures 6–9.

- VGG16: VGG16 achieved 99% training accuracy with a training loss of 0.025 within 8 epochs. Validation accuracy stabilized at 98% with a validation loss of 0.05, indicating strong generalization and minimal overfitting (Figure 6).
- MobileNetV2: This model reached 98% validation accuracy and 0.05 validation loss in just 8 epochs, highlighting its efficiency as a lightweight architecture (Figure 7).
- DenseNet1: Training accuracy improved steadily from 65% to 95%, while training loss decreased from 0.7 to 0.10, demonstrating gradual but effective convergence (Figure 8).
- Custom CNN: In its standard setup, the model achieved 98% training accuracy and a loss of 0.05, but its validation accuracy plateaued at 88%, suggesting underfitting. The 100-epoch configuration achieved 100% training accuracy and a loss of 0.00, indicating clear overfitting.
- Inference speed was measured as the average time per test image on an NVIDIA RTX 3060 GPU (16 GB RAM) to assess real-time deployment potential in low-resource settings. Table III shows the inference speed of the models

TABLE III. INFERENCE SPEED IN MS PER IMAGE.

Model	Inference speed (ms/image)
VGG16	~18
DenseNet121	~23
MobileNetV2	~12
Custom CNN (10 Epochs)	~9
Custom CNN (20 Epochs)	~9.2
Custom CNN (100 Epochs)	~9.5

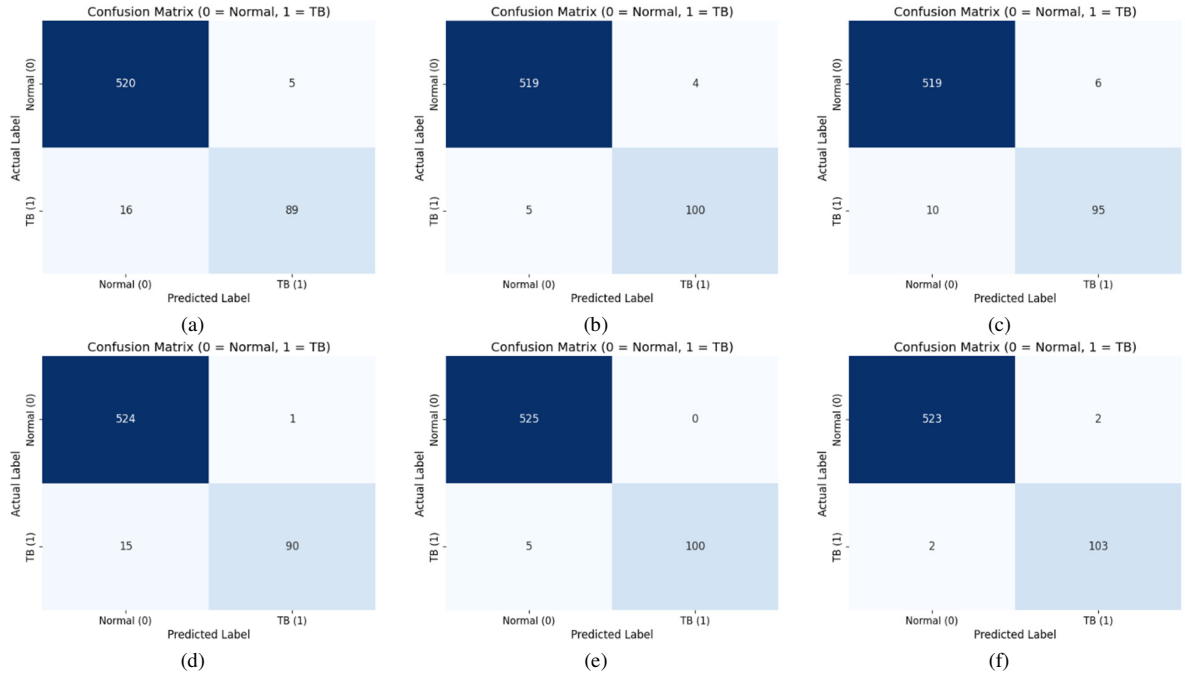

Fig. 5. Confusion matrices for each model: (a) 20 Epochs CNN, (b)10 Epochs CNN, (c)100 Epochs CNN, (d) DenseNet121, (e) MobileNetV2, (f) VGG16.

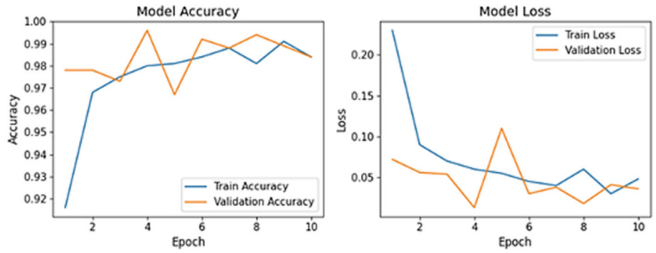

Fig. 6. VGG16 training/validation trends.

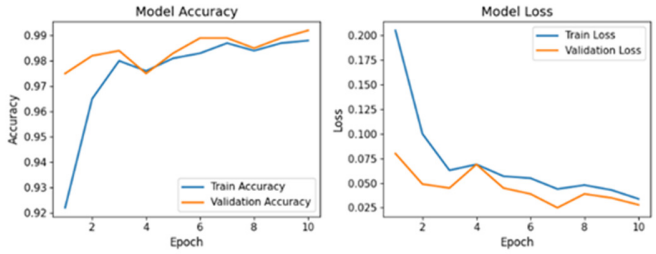

Fig. 7. MobileNetV2 training/validation trends.

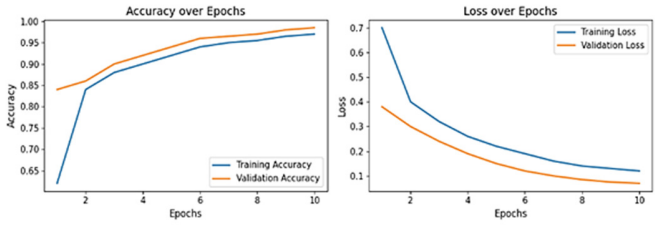

Fig. 8. DenseNet121 training/validation trends.

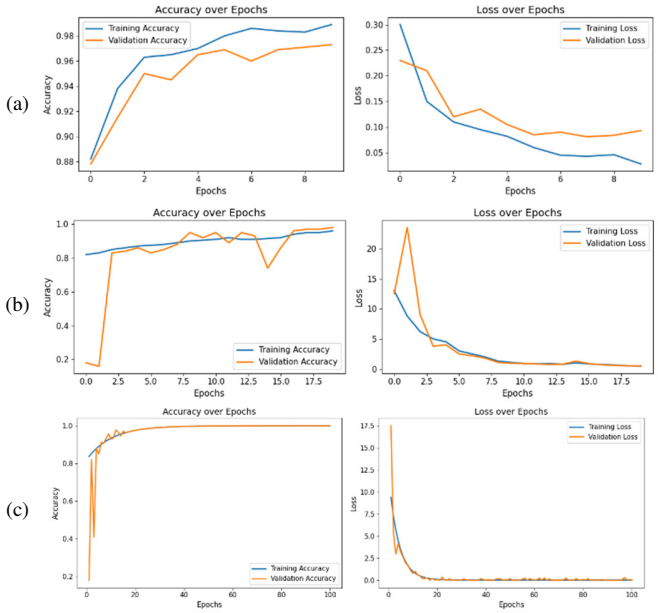

Fig. 9. Training and validation performance of Custom-CNN for: (a) 10 epochs, (b) 20 epochs, and (c) 100-epochs.

MobileNetV2 and the custom CNN offer the fastest inference, making them attractive for real-time screening scenarios in high-patient-volume or low-resource environments. Although VGG16 and DenseNet121 are slower, their superior accuracy justifies their use in settings where maximum diagnostic precision is prioritized over speed.

IV. DISCUSSION

This study investigates the effectiveness of deep learning and transfer learning approaches for automated TB detection from CXR images. The experimental results demonstrate that pretrained CNNs, particularly VGG16 and MobileNetV2, achieve highly competitive performance in terms of accuracy, generalization, and robustness when combined with appropriate preprocessing, data augmentation, and regularization strategies. These findings provide insights into architectural selection, training behavior, and practical deployment considerations for TB screening applications.

A. Performance of Deep Learning Architectures

The VGG16 model attained state-of-the-art performance, achieving a test accuracy of 99.21%, outperforming existing models such as RF-HOGADM [33] (97.0%) and ResNet50 [16] (91.0%), as summarized in Table IV. This performance is attributed to VGG16's deep hierarchical architecture, which facilitates the extraction of complex and abstract features from chest X-ray images. Furthermore, rigorous regularization techniques—including dropout and data augmentation—effectively mitigated overfitting and enhanced model generalization. The model's classification metrics underscore its clinical reliability, achieving a precision of 1.00 for normal cases and 0.98 for TB-positive cases. These results indicate an extremely low false positive rate, a crucial requirement in clinical contexts where overdiagnosis can lead to unnecessary interventions.

MobileNetV2 also demonstrated strong performance, achieving a test accuracy of 98.73% despite its lightweight design. This suggests its suitability for deployment in computationally constrained environments, where efficiency and diagnostic accuracy must be balanced.

B. Comparison with Prior Works

The comparative analysis focuses primarily on studies that employed the same or comparable TB CXR datasets. When different datasets were used, the reported results are discussed to provide contextual rather than strictly numerical comparisons. Compared to previous approaches, the proposed models demonstrated superior or comparable performance in key evaluation metrics. For example, the hybrid CNN-SVM model [34] achieved an accuracy of 96.5%, while our VGG16 implementation outperformed it by 2.71%. Likewise, transfer learning-based models such as InceptionV3, DenseNet201, and VGG16 reported accuracies of 95.72%, 95.07%, and 90.0%, respectively [16, 31], all of which were surpassed by our VGG16 model's 99.21% (see Table IV). The RF-HOGADM method [33], while achieving a competitive accuracy of 97.0%, lacks the end-to-end learning capability of deep CNNs. In contrast, the proposed transfer learning-based models—particularly VGG16 and MobileNetV2—utilize automated feature extraction and hierarchical representation learning, offering enhanced robustness and scalability across heterogeneous datasets. These comparative results confirm the effectiveness of deep learning and transfer learning, especially when paired with domain-specific preprocessing and robust regularization strategies.

C. Training Efficiency and Generalization

VGG16 and MobileNetV2 achieved similar high validation performance (≈98% accuracy) despite their architectural differences, indicating that strong generalization is not limited to lightweight models. VGG16 benefited from effective regularization, while MobileNetV2 achieved comparable performance with significantly lower computational cost. DenseNet121 exhibited slower convergence due to its depth, and the custom CNN showed overfitting at extended training duration. Overall, VGG16 and MobileNetV2 provide the best balance between accuracy and efficiency, making them suitable for deployment in both high-resource and low-resource clinical settings.

D. Model Limitations and Clinical Implications

The custom CNN exhibited performance variability, achieving 98.57% accuracy at 10 epochs but declining to 97.46% at 100 epochs due to overfitting. Despite perfect training accuracy (100%), its lower test performance highlights the limited generalization of shallow, non-pretrained models for complex medical imaging tasks. The absence of residual connections and limited regularization further constrained robustness, which is consistent with prior findings reporting only 89% accuracy for shallow CNNs [36]. In contrast, all evaluated models demonstrated high recall values (0.98–1.00) for both normal and TB-positive classes, with VGG16 achieving an ROC-AUC of 0.99 and a low false negative rate of 0.38%, which is critical for minimizing missed TB cases. MobileNetV2, with only 3.4 million parameters and 98.73% accuracy, offers an efficient solution for deployment in low-resource settings. These results support the integration of deep learning-based diagnostic tools into national TB control programs, particularly in regions with limited radiology expertise.

E. Inference Speed, Limitations, and Future Work

Although accuracy remains the primary metric for medical image classification, inference speed is a crucial factor for real-world deployment, particularly in high-volume patient environments or with resource constraints. The results in Table III show that MobileNetV2 and the custom CNN achieved the fastest inference times, which makes them suitable for rapid screening applications. In contrast, VGG16 and DenseNet121, while slower, delivered the highest diagnostic accuracy, making them more appropriate for settings where precision outweighs processing speed.

Despite the strong performance of the proposed models, several limitations remain. Patient-level data separation could not be strictly enforced due to the absence of reliable patient identifiers introduce a potential risk of data leakage. Additionally, no external validation dataset or k-fold cross-validation was employed, which may limit the generalizability and statistical robustness of the results. Finally, model interpretability was not addressed in this study; future work will incorporate explainability techniques such as Grad-CAM to enhance clinical trust and adoption.

TABLE IV. COMPARATIVE PERFORMANCE OF VARIOUS MODELS

Model	Dataset	Approach	Accuracy (%)	Precision	Recall	F1-Score
ConvNet [16]	Multi-source	Supervised CNN	87.0	88.0	87.0	87.0
VGG16 [16]	Multi-source	Transfer learning	90.0	91.0	91.0	91.0
CNN + SVM [34]	Dataset XYZ	Hybrid	96.5	97.4	95.9	--
CNN [35]	Dataset ABC	Supervised	81.3	80.0	81.0	80.0
RF-HOGADM [33]	Dataset DEF	Feature-based ML	97.0	97.0	97.0	97.0
CNN [36]	Multi-source	Transfer learning	89.0	85.0	91.0	89.0
InceptionV3 [31]	Multi-source	Transfer learning	95.72	95.92	--	95.73
VGG19 [31]	Multi-source	Transfer learning	95.80	95.92	--	95.80
DenseNet201 [31]	Multi-source	Transfer learning	95.07	95.27	--	95.07
MobileNet [31]	Multi-source	Transfer learning	94.33	94.65	--	94.32
CNN [36]	multi-source	TSSG-CNN	98.75	97.44	--	98.70
CNN [37]	multi-source	Supervised CNN	97.55	--	--	--
CNN + ViT (CoAtNet) [38]	Multi-source	Hybrid	86.39	89.24	82.70	85.85
VGG16 (Proposed)	Multi-source	Transfer learning	**99.2**	**98.0**	**98.0**	**98.0**
MobileNetV2 (Proposed)	Multi-source	Transfer learning	**98.7**	**100.0**	**95.0**	**98.0**
CNN (Proposed)	Multi-source	Lightweight CNN	**98.6**	**96.0**	**96.0**	**96.0**

V. CONCLUSION

This study demonstrates the effectiveness of deep learning models, particularly VGG16 and MobileNetV2, for high-accuracy TB detection from chest X-ray images. VGG16 achieved state-of-the-art performance with 99.21% accuracy, an F1-score of 0.98, and a ROC-AUC of 0.99, surpassing existing methods by 2.21–17.91%. MobileNetV2, with 98.73% accuracy and an inference speed of ~12 ms/image, offers an optimal balance between diagnostic precision and computational efficiency, making it highly suitable for deployment in resource-constrained or high-throughput clinical settings. The custom CNN achieved the fastest processing (~9–9.5 ms/image) but exhibited overfitting during extended training. The integration of multi-source datasets, stratified data splitting, and adaptive augmentation strategies enhanced model generalizability across diverse imaging sources. These results highlight the importance of combining architectural depth, efficient inference, and robust training protocols to bridge the gap between research and practical deployment. Future work should explore hybrid architectures that unify accuracy and speed, integrate explainable AI techniques to strengthen clinical trust, and conduct large-scale, multi-center validation to ensure equitable and reliable AI-driven TB diagnostics.

REFERENCES

[1] "Global Tuberculosis Report 2023". Geneva: World Health Organization, 2023. [Online]. Available: https://iris.who.int/items/60aa1a1d-04f9-46e2-9897-9836594b8d21.

[2] T. K. A. Teibo *et al.*, "Barriers That Interfere with Access to Tuberculosis Diagnosis and Treatment across Countries Globally: A Systematic Review," *ACS Infectious Diseases*, vol. 10, no. 8, pp. 2600–2614, Aug. 2024, https://doi.org/10.1021/acsinfecdis.4c00466.

[3] R. Kancherla, A. Sharma, and P. Garg, "Diagnosing Respiratory Variability: Convolutional Neural Networks for Chest X-ray Classification Across Diverse Pulmonary Conditions," *Journal of Imaging Informatics in Medicine*, vol. 38, no. 5, pp. 2677–2687, Oct. 2025, https://doi.org/10.1007/s10278-024-01355-9.

[4] K. Sarawagi *et al.*, "Self-Trained Convolutional Neural Network (CNN) for Tuberculosis Diagnosis in Medical Imaging," *Cureus*, vol. 16, no. 6, June 2024, https://doi.org/10.7759/cureus.63356.

[5] M. H. Rabby, O. Islam, Assaduzzaman, and M. Dutta, "Tuberculosis Disease Detection from Chest X-rays Using Deep Learning Techniques," in *2023 26th International Conference on Computer and Information Technology (ICCIT)*, Cox's Bazar, Bangladesh, Sept. 2023, pp. 1–6, https://doi.org/10.1109/ICCIT60459.2023.10441031.

[6] R. Rani and S. Gupta, "Deep Learning-Based Tuberculosis Detection Using Fine-Tuned VGG16 on Chest X-Ray Images," in *2024 3rd International Conference for Advancement in Technology (ICONAT)*, GOA, India, Sept. 2024, pp. 1–5, https://doi.org/10.1109/ICONAT61936.2024.10775087.

[7] R. Sivaramakrishnan *et al.*, "Comparing deep learning models for population screening using chest radiography," in *Medical Imaging 2018: Computer-Aided Diagnosis*, Houston, TX, USA, Feb. 2018, vol. 10575, pp. 322–332, https://doi.org/10.1117/12.2293140.

[8] X. Chu, Y. Wang, C. Shen, J. Chen, and W. Chu, "Training Protocol Matters: Towards Accurate Scene Text Recognition via Training Protocol Searching." arXiv, Mar. 17, 2022, https://doi.org/10.48550/arXiv.2203.06696.

[9] R. Ströbel, M. Mau, A. Puchta, and J. Fleischer, "Improving Time Series Regression Model Accuracy via Systematic Training Dataset Augmentation and Sampling," *Machine Learning and Knowledge Extraction*, vol. 6, no. 2, pp. 1072–1086, May 2024, https://doi.org/10.3390/make6020049.

[10] J. Plested and T. Gedeon, "An Analysis of the Interaction Between Transfer Learning Protocols in Deep Neural Networks," in *Neural Information Processing*, Sydney, Australia, 2019, pp. 312–323, https://doi.org/10.1007/978-3-030-36708-4_26.

[11] Y. Huang, H. Zhao, Y. Zhou, T. Riedel, and M. Beigl, "Standardizing Your Training Process for Human Activity Recognition Models – A Comprehensive Review in the Tunable Factors," in *Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2024, pp. 15–27, https://doi.org/10.1007/978-3-031-63992-0_2.

[12] P. Gang *et al.*, "Effect of Data Augmentation and Lung Mask Segmentation for Automated Chest Radiograph Interpretation of Some Lung Diseases," in *Neural Information Processing*, Sydney, Australia, 2019, pp. 333–340, https://doi.org/10.1007/978-3-030-36808-1_36.

[13] H. Almohab, A. D. W. M. Sidik, and A. P. Junfithrana, "Development and Evaluation of an AI-Driven Convolutional Neural Network (CNN) Model for Automated Pneumonia Detection Using Chest X-Ray Image," in *2024 10th International Conference on Computing, Engineering and Design (ICCED)*, Jeddah, Saudi Arabia, Sept. 2024, pp. 1–6, https://doi.org/10.1109/ICCED64257.2024.10983686.

[14] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, July 2019, Art. no. 60, https://doi.org/10.1186/s40537-019-0197-0.

[15] P. A. Priya and E. R. Vimina, "Tuberculosis Detection from CXR: An Approach Using Transfer Learning with Various CNN Architectures," in *International Conference on Communication, Computing and Electronics Systems*, Coimbatore, India, 2021, pp. 407–418, https://doi.org/10.1007/978-981-33-4909-4_31.

[16] E. Showkatian, M. Salehi, H. Ghaffari, R. Reiazi, and N. Sadighi, "Deep learning-based automatic detection of tuberculosis disease in chest X-ray images," *Polish journal of radiology*, vol. 87, pp. e118–e124, Jan. 2022, https://doi.org/10.5114/pjr.2022.113435.

[17] K. K. M. Rahman, S. Zulaikha, B. Dhafer, and R. Ahmed, "Advancing tuberculosis screening: A tailored CNN approach for accurate chest X-ray analysis and practical clinical integration," *Intelligence-Based Medicine*, vol. 11, Jan. 2025, Art. no. 100196, https://doi.org/10.1016/j.ibmed.2024.100196.

[18] R. Chelghoum, A. Ikhlef, A. Hameurlaine, and S. Jacquir, "Transfer Learning Using Convolutional Neural Network Architectures for Brain Tumor Classification from MRI Images," in *Artificial Intelligence Applications and Innovations*, Neos Marmaras, Greece, 2020, pp. 189–200, https://doi.org/10.1007/978-3-030-49161-1_17.

[19] M. Mamalakis *et al.*, "DenResCov-19: A deep transfer learning network for robust automatic classification of COVID-19, pneumonia, and tuberculosis from X-rays," *Computerized Medical Imaging and Graphics*, vol. 94, Dec. 2021, Art. no. 102008, https://doi.org/10.1016/j.compmedimag.2021.102008.

[20] M. Cabanillas-Carbonell, "Mobile Application to Improve the Follow-up and Control Process in Patients with Tuberculosis.," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 3, Feb. 2024, Art. no. 75, https://doi.org/10.3991/ijim.v18i03.46875.

[21] P. Dasarwar, U. Yadav, K. Morris, N. Chavhan, S. Bondre, and S. Kalamkar, "Revolutionizing Tuberculosis Prediction: A Cutting-Edge Approach," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 22929–22936, June 2025, https://doi.org/10.48084/etasr.10449.

[22] A. Azizi, M. Azizi, and M. Nasri, "Artificial Intelligence Techniques in Medical Imaging: A Systematic Review," *International Journal of Online & Biomedical Engineering*, vol. 19, no. 17, Dec. 2023, Art. no. 66, https://doi.org/10.3991/ijoe.v19i17.42431.

[23] W. Chokchaithanakul, P. Punyabukkana, and E. Chuangsuwanich, "Adaptive Image Preprocessing and Augmentation for Tuberculosis Screening on Out-of-Domain Chest X-Ray Dataset," *IEEE Access*, vol. 10, pp. 132144–132152, 2022, https://doi.org/10.1109/ACCESS.2022.3229591.

[24] B. Moryani, K. Sood, and K. Chaudhary, "A Deep Learning Approach for the Classification of Tuberculosis and Pneumonia Using NIH Dataset *," in *2023 International Symposium on Networks, Computers and Communications (ISNCC)*, Doha, Qatar, July 2023, pp. 1–8, https://doi.org/10.1109/ISNCC58260.2023.10323983.

[25] T. K. Kim, P. H. Yi, G. D. Hager, and C. T. Lin, "Refining dataset curation methods for deep learning-based automated tuberculosis screening," *Journal of Thoracic Disease*, vol. 12, no. 9, pp. 5078–5085, Sept. 2020, https://doi.org/10.21037/jtd.2019.08.34.

[26] A. Adli and P. Tyrrell, "Impact of Training Sample Size on the Effects of Regularization in a Convolutional Neural Network-based Dental X-ray Artifact Prediction Model," *Journal of Undergraduate Life Sciences*, vol. 14, no. 1, Dec. 2020, https://doi.org/10.33137/juls.v14i1.35883.

[27] Y. Tian and Y. Zhang, "A comprehensive survey on regularization strategies in machine learning," *Information Fusion*, vol. 80, pp. 146–166, Apr. 2022, https://doi.org/10.1016/j.inffus.2021.11.005.

[28] J. Liang, S. Gonzalez, H. Shahrzad, and R. Miikkulainen, "Regularized evolutionary population-based training," in *Proceedings of the Genetic and Evolutionary Computation Conference*, Lille, France, Mar. 2021, pp. 323–331, https://doi.org/10.1145/3449639.3459292.

[29] A. Mirugwe, L. Tamale, and J. Nyirenda, "Improving Tuberculosis Detection in Chest X-Ray Images Through Transfer Learning and Deep Learning: Comparative Study of Convolutional Neural Network Architectures," *JMIRx Med*, vol. 6, no. 1, July 2025, Art. no. e66029, https://doi.org/10.2196/66029.

[30] R. Wajgi *et al.*, "Optimized tuberculosis classification system for chest X-ray images: Fusing hyperparameter tuning with transfer learning approaches," *Engineering Reports*, vol. 6, no. 11, 2024, Art. no. e12906, https://doi.org/10.1002/eng2.12906.

[31] T. Rahman *et al.*, "Reliable Tuberculosis Detection Using Chest X-Ray With Deep Learning, Segmentation and Visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020, https://doi.org/10.1109/ACCESS.2020.3031384.

[32] "Tuberculosis (TB) Chest X-ray Database." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset.

[33] R. Geethamani and A. Ranichitra, "Enhancing Tuberculosis Detection: Leveraging RF-HOG Model for Automated Diagnosis from Chest X-ray Images," *Procedia Computer Science*, vol. 230, pp. 21–32, Jan. 2023, https://doi.org/10.1016/j.procs.2023.12.057.

[34] A. Soni, A. Rai, and S. K. Ahirwar, "Mycobacterium Tuberculosis Detection using Support Vector Machine Classification Approach," in *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, Bhopal, India, June 2021, pp. 408–413, https://doi.org/10.1109/CSNT51715.2021.9509635.

[35] M. Ahsan, R. Gomes, and A. Denton, "Application of a Convolutional Neural Network using transfer learning for tuberculosis detection," in *2019 IEEE International Conference on Electro Information Technology (EIT)*, Brookings, SD, USA, Feb. 2019, pp. 427–433, https://doi.org/10.1109/EIT.2019.8833768.

[36] S. Bahri, R. Wajhillah, and M. F. Adiwisastra, "Diagnosa Tuberculosis Paru Berbasis Citra X-ray Menggunakan Convolutional Neural Network," *IJCIT Indonesian Journal on Computer and Information Technology*, vol.. 6, no. Sept. pp. 181–186, 2021.

[37] T. H. Kim, M. Krichen, S. Ojo, M. A. Alamro, and G. A. Sampedro, "TSSG-CNN: A Tuberculosis Semantic Segmentation-Guided Model for Detecting and Diagnosis Using the Adaptive Convolutional Neural Network," *Diagnostics*, vol. 14, no. 11, June 2024, https://doi.org/10.3390/diagnostics14111174.

[38] M. Raičević, N. Kavarić, T. Popović, D. Babić, and I. Jovović, "Tuberculosis Detection From Chest X- Rays Using Convolutional Neural Network," in *2025 29th International Conference on Information Technology (IT)*, Zabljak, Montenegro, Oct. 2025, pp. 1–4, https://doi.org/10.1109/IT64745.2025.10930261.

[39] G. Siddharth, A. Ambekar, and N. Jayakumar, "Enhanced CoAtNet based hybrid deep learning architecture for automated tuberculosis detection in human chest X-rays," *BMC Medical Imaging*, vol. 25, no. 1, Sept. 2025, Art. no. 379, https://doi.org/10.1186/s12880-025-01901-z.