LIGHTWEIGHT CNN ARCHITECTURES FOR THREE-CLASS TUBERCULOSIS X-RAY CLASSIFICATION

Esbergenov Hamidulla Seidabullaevich[1], Nurimov Paraxat Baymuratovich[2]

[1]Tashkent University of Information Technologies named after Muhammad ibn Musa al-Khwarizmi
[2]National Research University "Tashkent Institute of Irrigation and Agricultural Mechanization Engineers"
E-mail: hamidesbergenov@gmail.com

KEYWORDS

Tuberculosis detection, chest X-ray classification, deep learning, transfer learning, TBX11K, convolutional neural networks, Grad-CAM, class imbalance, MobileNet-V3, EfficientNet-B3, DenseNet-121.

ABSTRACT

Tuberculosis (TB) remains one of the leading infectious diseases worldwide, causing approximately 1.3 million deaths annually. Automated computer-aided diagnosis (CAD) systems based on chest X-ray (CXR) imaging offer a promising solution for early TB screening, particularly in resource-limited settings. We present a systematic comparison of three pretrained CNN backbones - DenseNet-121, EfficientNet-B3, and MobileNet-V3 - for three-class CXR classification (Healthy, Sick-but-non-TB, and TB) using the TBX11K benchmark dataset under identical training conditions. MobileNet-V3 achieves the highest macro AUC of 99.98% with a TB sensitivity of 99.37%, while EfficientNet-B3 achieves perfect TB sensitivity of 100.00%. Grad-CAM visualizations confirm clinically meaningful activation patterns, and external validation on a confirmed TB case demonstrates real-world applicability.

Introduction

Tuberculosis (TB) is an airborne infectious disease caused by Mycobacterium tuberculosis, and despite being preventable and curable, it remains one of the top 10 causes of death worldwide. According to the WHO Global Tuberculosis Report 2023, approximately 10.6 million people fell ill with TB in 2022, and 1.3 million deaths were recorded [3]. The burden of TB is disproportionately concentrated in low- and middle-income countries, where access to specialist radiologists is limited, including the Karakalpakstan Autonomous Republic of Uzbekistan - a region severely affected by the Aral Sea ecological disaster and its respiratory health consequences.

Chest X-ray (CXR) imaging is the primary tool for TB screening due to its low cost and wide availability. However, manual interpretation is time-consuming, subjective, and prone to inter-reader variability. Computer-aided diagnosis (CAD) systems powered by deep learning have shown great promise in automating CXR analysis [4]. The TBX11K dataset [1], introduced by Liu et al. at CVPR 2020, contains 11,200 annotated images and enables the clinically important three-class classification task distinguishing TB from other pulmonary diseases - a critical distinction for treatment planning.

While previous works have explored deep learning for TB detection, a controlled comparison of modern lightweight backbones on TBX11K's three-class problem - with consistent training protocols, class imbalance handling, and interpretability analysis - remains lacking. This paper addresses that gap with architectures practical for deployment in resource-limited clinical settings.

1.1 Contributions

1. A fair, controlled comparison of DenseNet-121, EfficientNet-B3, and MobileNet-V3 on the three-class TBX11K task under identical training conditions.
2. A principled class imbalance strategy combining Focal Loss with WeightedRandomSampler.
3. Grad-CAM heatmap visualizations demonstrating interpretability and qualitative alignment with clinically relevant lung regions.

© H.S.Esbergenov, P.B.Nurimov

4. External validation on a confirmed miliary tuberculosis case demonstrating real-world applicability.
5. Inference time analysis relevant to deployment in low-resource settings.

2. Related Work

Early approaches to automated TB detection relied on hand-crafted features with classical classifiers. With deep CNNs, performance improved substantially: Hooda et al. achieved 82.09% accuracy on binary TB classification [11], while transfer learning from ImageNet-pretrained models further advanced results. DenseNet-121, as the backbone of CheXNet [5], became a widely adopted baseline for CXR analysis, with Nguyen et al. achieving AUC values of 0.94–0.82 on Shenzhen and Montgomery datasets [11]. EfficientNet [6] achieves state-of-the-art accuracy through compound scaling, and has been applied effectively to lung disease detection [9]. For TBX11K specifically, Liu et al. established baselines using object detectors (SSD: 84.7%, Faster R-CNN: 89.7%) [1], and a subsequent study using MobileNetV2 with an Orthogonal Softmax Layer achieved 98.17% accuracy [2]. Despite these advances, few studies perform a controlled multi-backbone comparison on TBX11K's three-class problem with consistent protocols and interpretability analysis. Our work fills this gap.

3. Dataset

The TBX11K dataset [1] contains 11,200 chest X-ray images at 512×512 pixels, annotated by clinical radiologists across three categories: Healthy, Sick-but-non-TB, and TB (active or latent). A subset of 8,400 images is used in this study. The dataset exhibits a notable class imbalance: approximately 45.2% Healthy, 45.2% Sick-but-non-TB, and only 9.5% TB, as shown in Figure 1. This reflects real-world clinical distribution but poses challenges for standard training, as models may fail to reliably detect the clinically critical TB minority class.

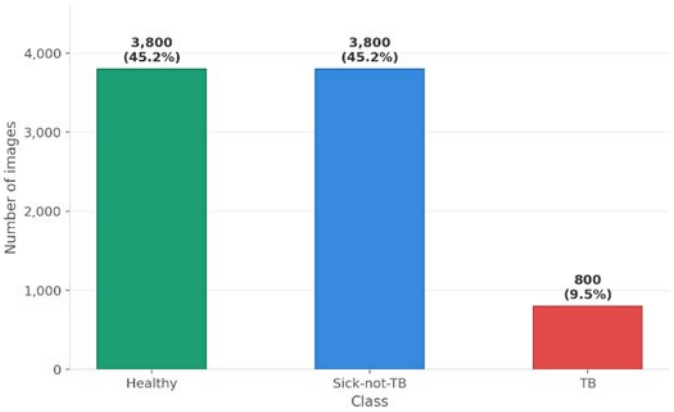


Figure 1. *Class distribution of the TBX11K dataset. The TB class represents a significant minority at approximately 9.5% of total images.*

The dataset is split into training (80%, 6,720 images) and validation (20%, 1,680 images) sets using stratified sampling to preserve class distribution across both splits.

4. Methodology

4.1 Preprocessing and Data Augmentation

All images are resized to 512×512 pixels with ImageNet normalization (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]). Training augmentation includes random horizontal flipping (p=0.5), random rotation (±15°), and color jitter (brightness/contrast ±0.2). No augmentation is applied during validation.

4.2 Class Imbalance Handling

The 9.5% TB class representation poses a fundamental challenge: naïvely trained models can achieve high accuracy by predicting majority classes while failing on TB cases. We address this with a two-component strategy: at the data level, WeightedRandomSampler assigns sampling weights inversely proportional to class frequency; at the loss level, Focal Loss [10] (gamma=2.0) down-weights easy examples and focuses gradient updates on hard cases. Together, WeightedRandomSampler controls what the model sees, while Focal Loss controls what the model learns from.

(cc) BY © H.S.Esbergenov, P.B.Nurimov

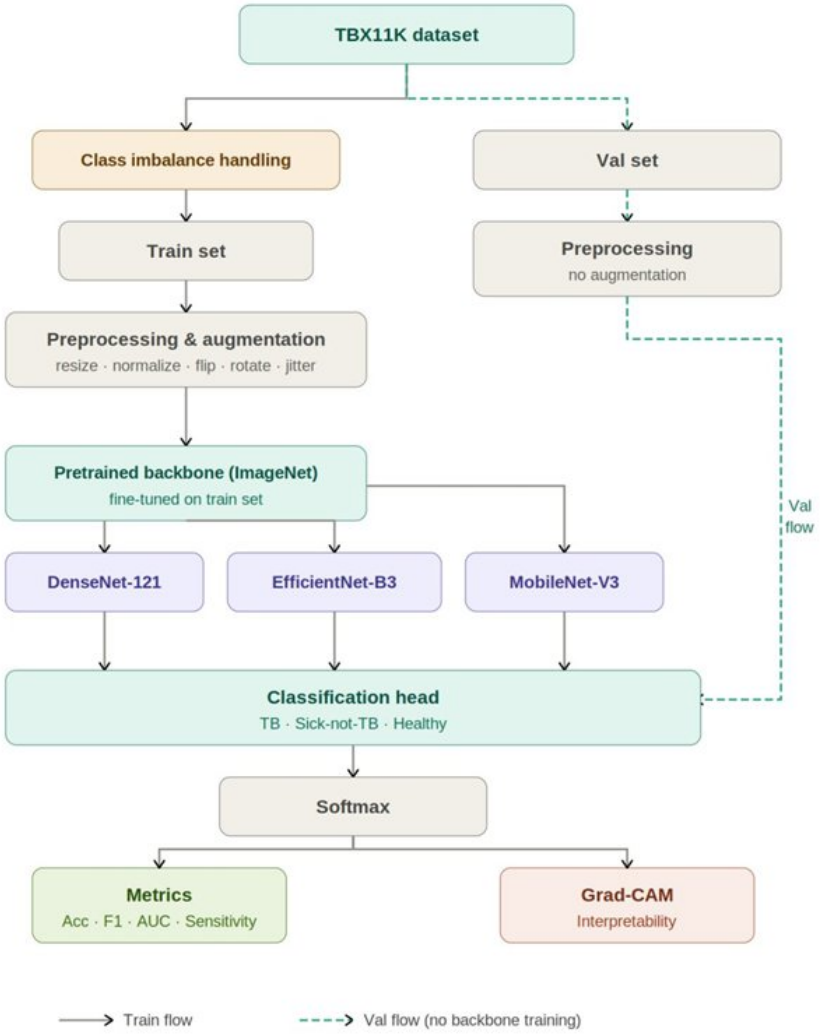


Figure 2. *Full classification pipeline.*

4.3 Backbone Architectures

We evaluate three ImageNet-pretrained backbones, all fine-tuned end-to-end on TBX11K. For each, the final layer is replaced with Dropout followed by a fully connected layer with 3 outputs and Softmax activation. The pipeline is illustrated in Figure 2.

DenseNet-121 [7] connects each layer to all subsequent layers, promoting feature reuse with 7.0M parameters. As the backbone of CheXNet [5], it is a well-established baseline for CXR classification.

EfficientNet-B3 [6] uses compound scaling to balance depth, width, and resolution, achieving high accuracy with 10.7M parameters and strong results in lung disease classification [9].

MobileNet-V3 [8] is designed for efficient inference with 4.2M parameters. It incorporates hard-swish activations and Squeeze-and-Excitation attention blocks, making it substantially more capable than its parameter count suggests and well-suited for mobile clinical deployment.

4.4 Training Configuration

All models are trained for up to 5 epochs using the Adam optimizer (lr=1e-4, weight decay=1e-4) with CosineAnnealingLR scheduling (eta_min=1e-6) and early stopping (patience=3). The checkpoint with the lowest validation loss is used for evaluation. All experiments are conducted on an NVIDIA T4 GPU; total training time for all three backbones is approximately 80 minutes.

(cc) BY © H.S.Esbergenov, P.B.Nurimov

Table 1.

Backbone comparison on TBX11K validation set.

Backbone	Params (M)	Accuracy	Macro F1	Macro AUC	TB Sens.	TB Spec.	Infer (ms)
DenseNet-121	7.0	98.93%	98.18%	99.95%	99.37%	99.21%	19.7
EfficientNet-B3	10.7	98.99%	98.24%	99.96%	100.00%	99.14%	13.5
MobileNet-V3	4.2	98.99%	98.30%	99.98%	99.37%	99.28%	6.5

5. Experiments and Results

5.1 Backbone Comparison

Table 1 presents performance of all three backbones on the TBX11K validation set.

MobileNet-V3 achieved the highest macro AUC (99.98%) with the fastest inference (6.5ms) and fewest parameters (4.2M), making it the best overall choice for resource-limited deployment. EfficientNet-B3 achieved perfect TB sensitivity (100.00%), making it the most clinically conservative option - a false negative in TB screening carries greater public health consequences than a false positive. DenseNet-121 performed competitively (AUC 99.95%) but at 19.7ms inference is nearly three times slower than MobileNet-V3. Training convergence is shown in Figure 3.

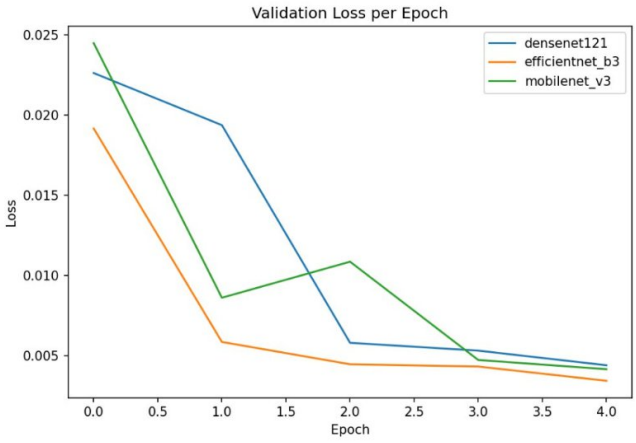


Figure 3. *Validation loss per epoch for all three backbone architectures.*

5.2 Confusion Matrix Analysis

Figure 4 presents the confusion matrix for MobileNet-V3 on the TBX11K validation set. The healthy class achieves precision 1.00 and recall 0.99; sick-not-TB achieves precision and recall of 0.99; the TB class achieves precision 0.94 and recall 0.99. Occasional misclassification between sick-not-TB and TB is expected given visual similarity of pulmonary pathologies.

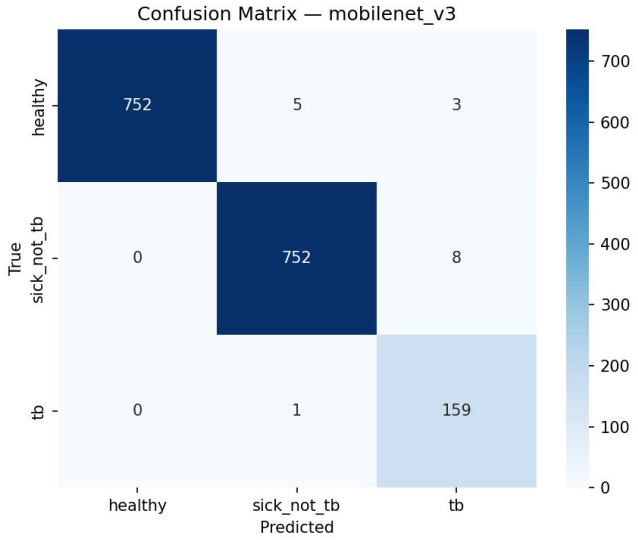


Figure 4. *Confusion matrix of MobileNet-V3 on the TBX11K validation set.*

5.3 Grad-CAM Visualization

Grad-CAM [12] generates heatmaps highlighting image regions most influential for the predicted class. Figure 5 shows a representative Healthy-class example, where activations are distributed uniformly across the lung fields,

(cc) BY © H.S.Esbergenov, P.B.Nurimov

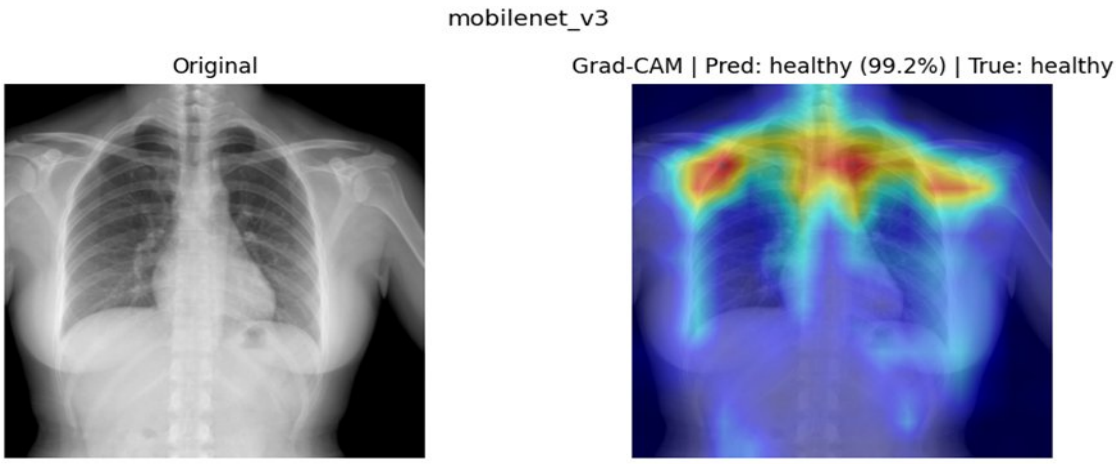


Figure 5. *Grad-CAM visualization: original X-ray is Healthy. The model predicts the correct class with 99.2% confidence.*

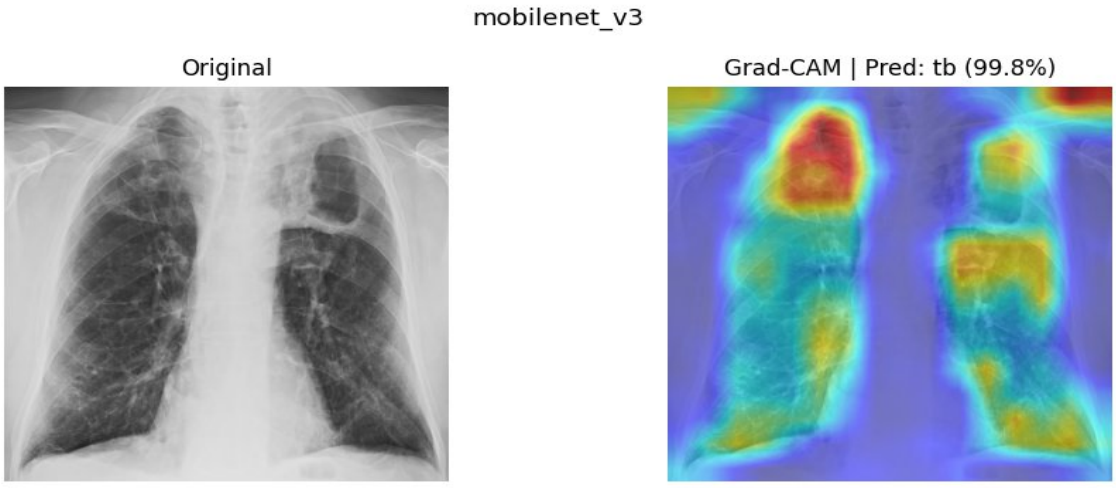


Figure 6. *Grad-CAM heatmap for the external TB case. The model correctly classified this image as TB with bilateral lung field activations consistent with miliary TB.*

consistent with absence of focal pathology - suggesting the model has learned clinically meaningful spatial features from image-level labels alone.

5.4 External Validation

MobileNet-V3 was tested on an external chest X-ray depicting confirmed miliary tuberculosis, sourced from Wikimedia Commons [13]. The model correctly classified it as TB with high confidence. The Grad-CAM heatmap highlights bilateral lung field activations consistent with the diffuse miliary pattern, as shown in Figure 6, demonstrating generalization beyond the TBX11K training distribution.

6. Discussion

The results challenge the common assumption that larger models are necessary for medical imaging: on structured, high-contrast CXR images with moderate dataset sizes, a well-designed lightweight architecture can match or exceed heavier models. MobileNet-V3's combination of accuracy (AUC 99.98%), speed (6.5ms), and compactness (4.2M parameters) makes it particularly compelling for deployment as a mobile application - enabling community health workers in areas such as Karakalpakstan to perform rapid preliminary TB triage on standard smartphones without specialist infrastructure.

EfficientNet-B3's perfect TB sensitivity (100.00%) is clinically significant: in high-prevalence screening settings, a clinician may prefer zero missed cases over marginally lower false-positive rates, accepting the trade-off. DenseNet-121's 19.7ms inference may limit its practicality in high-throughput screening.

(cc) BY © H.S.Esbergenov, P.B.Nurimov

The two-component imbalance handling strategy proved effective across all three backbones. Focal Loss ensured hard TB examples dominated gradient updates, while WeightedRandomSampler ensured sufficient TB exposure per batch - addressing the imbalance at two complementary points in the training pipeline.

Grad-CAM activations concentrated in the upper lung lobes and hilar regions for TB-positive images are consistent with the radiological presentation of pulmonary TB, lending clinical credibility to model decisions - a prerequisite for clinical adoption of AI-assisted diagnostic tools.

6.1 Limitations

First, models are trained for only 5 epochs due to computational constraints; performance may improve with extended training. Second, evaluation is limited to the TBX11K validation set; testing on independent external datasets (e.g., Montgomery County, Shenzhen) would strengthen generalizability claims. Third, only image-level classification is addressed; incorporating TB area localization using TBX11K bounding box annotations would produce more clinically useful outputs.

7. Conclusion

We presented a systematic comparison of DenseNet-121, EfficientNet-B3, and MobileNet-V3 for three-class TB classification from chest X-rays using TBX11K. All three architectures achieved strong performance (macro AUC ≥99.95%). MobileNet-V3 is the recommended choice for resource-limited deployment - best overall AUC (99.98%) and fastest inference (6.5ms). EfficientNet-B3 is the most clinically conservative option with perfect TB sensitivity (100.00%). Our two-component imbalance strategy and Grad-CAM interpretability analysis provide practical guidance for future TB CAD system design.

Future work will extend to TB area detection using TBX11K bounding box annotations, semi-supervised learning on unlabeled CXR data, evaluation across multiple external datasets, and investigation of extended training convergence.

References

1. Liu, Y., Wu, Y.H., Ban, Y., Wang, H., Cheng, M.M. (2020). Rethinking Computer-Aided Tuberculosis Diagnosis. CVPR, pp. 2735–2744.
2. Das, P.K., et al. (2025). OSLTBDNet: Orthogonal softmax layer-based tuberculosis detection network with small dataset. Biomedical Signal Processing and Control. https://doi.org/10.1016/j.bspc.2025.000953
3. World Health Organization. (2023). Global Tuberculosis Report 2023. WHO, Geneva.
4. Kotei, E., Thirunavukarasu, R. (2024). A comprehensive review on deep learning techniques for automatic detection of tuberculosis from chest X-ray images. Archives of Computational Methods in Engineering, 31, 455–474.
5. Rajpurkar, P., et al. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv:1711.05225.
6. Tan, M., Le, Q.V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. ICML, pp. 6105–6114.
7. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q. (2017). Densely connected convolutional networks. CVPR, pp. 4700–4708.
8. Howard, A., et al. (2019). Searching for MobileNetV3. ICCV, pp. 1314–1324.
9. Ravi, V., Acharya, V., Alazab, M. (2023). A multichannel EfficientNet deep learning-based stacking ensemble for lung disease detection. Cluster Computing, 26, 1181–1203.
10. Lin, T.Y., et al. (2017). Focal loss for dense object detection. ICCV, pp. 2980–2988.
11. Mirugwe, A., Tamale, L., Nyirenda, J. (2025). Improving Tuberculosis Detection Through Transfer Learning and Deep Learning. JMIRx Med, 6, e66029.
12. Selvaraju, R.R., et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. ICCV, pp. 618–626.
13. Häggström, M. (2014). Chest radiograph of miliary tuberculosis [Radiograph]. Wikimedia Commons.

(cc) BY © H.S.Esbergenov, P.B.Nurimov