Sensitivity-Optimized Tuberculosis Triage Using Nigerian Chest X-Rays

Wisdom Jonathan[1], Aondona Moses Iorumbur[2,3,4], Confidence Raymond[2,5,6]

[1] Department of Medicine, Niger Delta University, Bayelsa State, Nigeria

[2] Medical Artificial Intelligence Laboratory, Lagos, Nigeria

[3] Department of Medical Physics, College of Medicine, University of Lagos, Nigeria

[4] Department of Physics, Federal University of Technology Minna, Minna, Nigeria

[5] Montreal Neurological Institute, McGill University, Montreal, Canada

[6] Department of Biomedical Engineering, McGill University, Montreal, Canada

wisdomJonathan@ndu.edu.ng

Abstract. Tuberculosis (TB) remains a major public health problem in resource-constrained settings, where delayed chest X-ray interpretation may slow triage and referral for confirmatory testing. This study evaluated whether external TB-specific pretraining could improve sensitivity and reduce missed TB cases when applied to Nigerian chest X-rays. A Nigerian four-class chest X-ray dataset containing COVID-19, normal, pneumonia, and TB images was reframed as a binary triage task: TB-positive versus non-TB. DenseNet-121 was trained using two pathways: Nigerian-only training and external TB/normal pretraining followed by Nigerian fine-tuning. Both models were evaluated on the same Nigerian test set of 600 images. External pretraining reduced false-negative TB predictions from 4 to 1, lowering the false-negative rate from 2.67% to 0.67%, while maintaining high specificity (0.99). A threshold selected on the validation set and fixed before independent test evaluation achieved 98.67% sensitivity and 99.78% specificity on the Nigerian test set. These findings suggest that TB-specific pretraining combined with local fine-tuning can improve the safety of AI-assisted TB triage in Nigerian and similar resource-constrained settings.

Keywords: Tuberculosis, chest X-ray, deep learning, sensitivity, Nigerian dataset; resource-constrained settings

1 Introduction

Tuberculosis (TB) remains one of the leading infectious causes of morbidity and mortality worldwide. Although the disease is preventable and curable, delayed detection can prolong transmission and postpone treatment. Early identification of presumptive TB cases is therefore essential for confirmatory testing, treatment initiation, and infection-control measures [1]-[3].

Chest X-ray (CXR) is widely used as a screening and triage tool for pulmonary TB, particularly in high-burden settings where it is relatively accessible, non-invasive, and low-cost [1], [4], [5]. However, limited radiology capacity remains a barrier in many Nigerian health facilities. Nigeria has been reported to have only about 250-300

registered radiologists for a population above 200 million people, creating major delays in specialist interpretation [6]. AI-assisted CXR triage may help prioritize suspected TB cases for further review and confirmatory testing.

Many AI studies for TB detection report high accuracy or AUROC (Area Under the Receiver Operating Characteristic), but these metrics alone may hide clinically important errors. In TB triage, false-negative predictions are especially concerning because they may incorrectly exclude true TB cases from further evaluation. In addition, many models are developed using simplified TB-versus-normal datasets, whereas real triage must distinguish TB from other abnormal non-TB CXRs, including pneumonia or COVID-19. Models trained in high-resource settings may also underperform in African settings because of domain shift in image quality, equipment, patient population, and disease distribution [7].

This study addresses these gaps by reframing a Nigerian four-class CXR dataset from Aminu Kano Teaching Hospital, Kano, Nigeria as a binary TB triage task. The study compares DenseNet-121 trained directly on the Nigerian binary dataset with the same architecture pretrained on an external TB/normal dataset and then fine-tuned on Nigerian data. The central aim is to determine whether external TB-specific pretraining can improve sensitivity and reduce missed TB cases in a Nigerian screening context.

2 Related Work

2.1 Chest X-ray AI for Pulmonary Disease Detection

Deep learning has been widely applied to CXR classification, detection, segmentation, and triage. Convolutional neural networks can learn radiographic patterns relevant to pulmonary disease, supporting automated analysis where specialist interpretation is limited [3]. Local work using Nigerian CXR data reported test accuracies above 90% for DenseNet-121, EfficientNet-B0, and ResNet-50, with DenseNet showing strong class-wise performance for TB, pneumonia, and COVID-19 classification [8].

Other studies have explored architectures such as CenterNet, EfficientNet, Vision Transformers, and hybrid CNN-transformer models for pulmonary disease detection [9], [10]. Ahsan et al. also showed that transfer learning with VGG16 could support TB classification from CXRs with minimal preprocessing [11]. These studies demonstrate the promise of AI for CXR interpretation, but reported performance varies with dataset size, disease mix, preprocessing, and evaluation design.

2.2 AI-Based Tuberculosis Detection and Triage

Several studies have focused directly on automated TB detection from CXRs. Recent approaches using filtering, normalization-free networks, and deep learning classifiers have reported high accuracy or AUC for binary TB classification [12]-[14]. Computer-aided detection studies have also shown that AI-assisted CXR interpretation may expand access to TB screening where trained human resources are limited [15], [16].

Despite these advances, many studies still evaluate models in simplified TB-versus-normal settings. This is less realistic for clinical triage because non-TB patients may have abnormal images caused by pneumonia, COVID-19, or other lung diseases. A model intended for triage should therefore be assessed against both normal and abnormal non-TB images, with emphasis on sensitivity and false-negative reduction rather than accuracy alone.

2.3 Transfer Learning, External Pretraining, and Domain Shift

Transfer learning is practical in medical imaging because local datasets are often small and expensive to label. Models pretrained on larger datasets can be fine-tuned for local tasks, improving convergence and reducing the need for training from scratch [8], [11]. However, transfer learning does not automatically solve domain shift. CXR models trained on external datasets may underperform in African settings because of differences in acquisition conditions, image quality, equipment, and population characteristics [7].

External TB-specific pretraining may provide a useful middle ground: a model first learns broad TB-related radiographic features from an external TB/normal dataset, then adapts to local Nigerian CXR data through fine-tuning. However, few studies have examined whether this pathway specifically reduces missed TB cases in Nigerian or African CXR triage. This study therefore evaluates external pretraining using false-negative reduction, sensitivity, and threshold-based screening performance as the central outcomes.

3 Methodology

3.1 Study Design and Experimental Framework

This study was designed as a supervised deep learning experiment for CXR-based TB triage. DenseNet-121 was selected because of its strong performance in CXR classification tasks [8]. The experiment tested whether adding an external TB pretraining step could help reduce false-negative predictions when the model was applied to Nigerian CXRs.

Two training pathways were compared. In the first pathway, DenseNet-121 was trained directly on the Nigerian binary dataset, serving as the Nigerian-only baseline. In the second pathway, the same DenseNet-121 architecture was first trained on an external TB-versus-normal dataset, then transferred and fine-tuned on the Nigerian binary task. Both pathways used the same architecture and were evaluated on the same Nigerian test set. The Nigerian-only pathway involved 15 epochs of training on the Nigerian training set, whereas the external-pretraining pathway involved an additional 10 epochs of external TB/normal pretraining followed by 15 epochs of Nigerian fine-tuning.

3.2 Dataset

The main dataset was a publicly available Nigerian CXR dataset containing 2,600 images from patients at Amino Kano Teaching Hospital, Kano, Nigeria [7]. The dataset was equally distributed across COVID-19, normal, pneumonia, and TB classes, with 650 images per class. For the binary triage task, all TB images were labelled TB-positive, while COVID-19, normal, and pneumonia images were grouped as non-TB. This produced 650 TB-positive and 1,950 non-TB images. The external pretraining dataset was the Tuberculosis (TB) Chest X-ray Database, which contained 3,500 normal CXRs and 700 TB CXRs from high-income countries [17].

Table 1. Dataset splits used for model development and evaluation.

Database	Type	Total CXR images	Train	Validation	Test
Nigerian CXR dataset	Non-TB	1,950	1,200	300	450
Nigerian CXR dataset	TB	650	400	100	150
External CXR dataset	Non-TB	3,500	2,450	525	525
External CXR dataset	TB	700	489	106	105

3.3 Preprocessing and Model Training

All images were resized to 224 x 224 pixels, converted to tensors, and normalized using ImageNet mean and standard deviation values. Training images were augmented using random horizontal flipping with probability 0.5 and random rotation up to 10 degrees. Validation and test images were resized and normalized without random augmentation.

The Nigerian-only baseline was trained for 15 epochs. The externally pretrained model was first trained on the external TB/normal dataset for 10 epochs, after which the learned weights were fine-tuned on the Nigerian binary dataset for 15 epochs. Class-weighted cross-entropy loss was used to reduce class imbalance. AdamW optimization was used with batch size 32 and weight decay of 1×10^{-4}. The learning rate was 1×10^{-4} for Nigerian-only training and external pretraining, and 5×10^{-5} for Nigerian fine-tuning.

3.4 Evaluation and Error Analysis

Final evaluation was performed on the Nigerian test set containing 450 non-TB and 150 TB-positive images. The primary endpoint was the number of false-negative TB predictions because missed TB cases are the most clinically important error in triage. Secondary metrics included accuracy, sensitivity, specificity, false-negative rate, false-positive rate, precision, F1-score, AUROC, and AUPRC. Threshold-based analysis assessed whether sensitivity could be improved while maintaining specificity above clinically useful levels.

For threshold analysis, the validation set was used to select an operating threshold, which was then fixed before evaluation on the independent Nigerian test set. The selected threshold was 0.634114. In addition, predictions from the Nigerian-only and externally pretrained models were compared pairwise on the same test images using an exact McNemar test to assess whether their classification errors differed significantly.

4 Results

4.1 Model Performance

Both training strategies were evaluated on the same Nigerian test set. The Nigerian-only DenseNet-121 model missed 4 TB-positive cases, while the externally pretrained and Nigerian fine-tuned DenseNet-121 model missed 1 TB-positive case. The confusion matrix comparison is shown in Fig. 1.

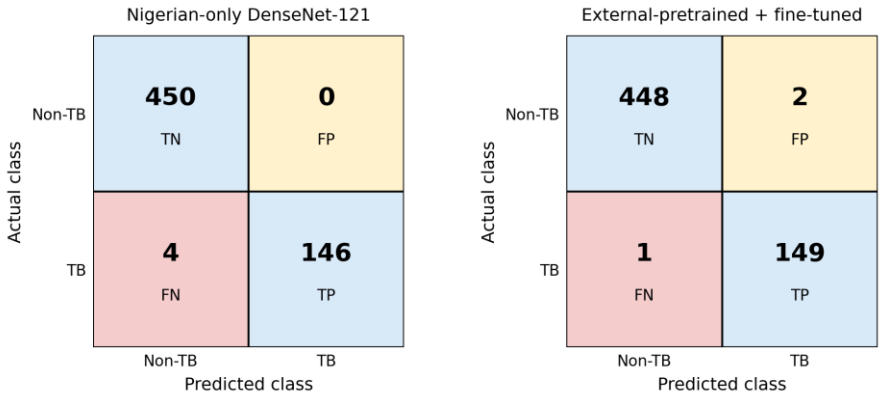

Fig. 1. Confusion matrix comparison on the Nigerian test set. External TB pretraining reduced false-negative TB predictions from 4 to 1.

The main screening and discrimination metrics are shown in Table 2. After external TB pretraining and Nigerian fine-tuning, sensitivity increased from 0.97 to 0.99, while specificity remained high at 0.99. Precision was 1.0000 for the Nigerian-only model and 0.98 for the externally pretrained model.

Table 2. Nigerian test-set performance of both training strategies.

Model	Sens.	Spec.	FNR	Acc.	F1	AUROC	AUPRC
N-only	0.97	1.00	0.026	0.99	0.98	0.99	0.99
Ext + FT	0.99	0.99	0.01	0.99	0.99	0.99	0.99

N-only = Nigerian-only DenseNet-121; Ext + FT = external-pretrained and Nigerian fine-tuned DenseNet-121; Sens. = sensitivity; Spec. = specificity; FNR = false-negative rate; Acc. = accuracy; AUROC = area under the receiver operating characteristic curve; AUPRC = area under the precision-recall curve.

4.2 Effect of External TB Pretraining and Threshold Adjustment

External TB pretraining reduced missed TB-positive cases from 4 to 1, lowering the false-negative rate from 2.67% to 0.67%. Although the externally pretrained model produced 2 false-positive predictions, this trade-off is acceptable in a triage setting because false-positive cases can undergo confirmatory microbiological testing, whereas false-negative cases may delay further assessment and treatment.

In the paired comparison, the externally pretrained model correctly classified three cases that were misclassified by the Nigerian-only model, while the Nigerian-only model correctly classified two cases that were misclassified by the externally pretrained model. Consequently, external pretraining reduced the number of missed TB-positive cases from 4 to 1, corresponding to a 75% reduction in the false-negative rate (2.67% to 0.67%). The exact McNemar test yielded $p = 1.000$, indicating that the difference in

paired classification errors did not reach statistical significance in this test set. Nevertheless, the reduction in false-negative cases is directly relevant to the sensitivity-first objective of TB triage and supports further evaluation of external TB-specific pretraining on larger and more diverse datasets. A threshold selected using the Nigerian validation set was fixed before evaluation on the independent test set. The selected threshold was 0.63, at which the externally pretrained model achieved 98.67% sensitivity and 99.78% specificity on the test set, with a false-negative rate of 1.33%.

4.3 Computational Efficiency

The final externally pretrained and Nigerian fine-tuned model had a saved size of 27.14 MB. Inference on the 600 Nigerian test images took 1.6561 s on a Tesla T4 GPU, corresponding to 0.002760 s per image and a throughput of 362.31 images/s. This suggests computational feasibility for rapid CXR triage, although real-world speed will depend on local hardware and workflow integration.

5 Discussion

This study evaluated whether external TB-specific pretraining could reduce missed TB cases in a Nigerian CXR triage task. The main finding was that external pretraining followed by Nigerian fine-tuning reduced false-negative TB predictions from 4 to 1, lowering the false-negative rate from 2.67% to 0.67%. Although both models achieved high overall performance, this reduction in missed TB cases is the most clinically important result because the model is intended for triage rather than definitive diagnosis.

The result shows why accuracy alone is not sufficient for evaluating TB screening models. The Nigerian-only model achieved high accuracy, but it still missed more TB-positive cases than the externally pretrained model. In TB triage, false-negative predictions are clinically dangerous because they may delay confirmatory testing, treatment, and infection-control measures. Therefore, sensitivity, false-negative rate, and threshold-based performance should be emphasised when assessing AI tools for TB screening.

This study analysis was retrospective and based on a relatively small dataset, which may limit generalizability. The study also evaluated a single model architecture, DenseNet-121. Future work should validate the model on larger multi-centre African datasets and assess prospective performance in real clinical workflows.

6 Impact in Resource-Constrained Settings

This work is relevant to resource-constrained settings where radiology expertise is limited and delays in CXR interpretation may affect TB detection. An AI-assisted triage model could help identify patients who should be prioritized for confirmatory testing, clinical review, or infection-control measures. The model is not proposed as a replacement for radiologists or microbiological diagnosis, but as a decision-support tool for earlier identification of suspected TB cases.

The study also highlights the importance of local adaptation in AI-assisted medical imaging. Models developed only on external datasets may not perform reliably in African clinical environments because of domain shift. By combining external TB-specific pretraining with Nigerian fine-tuning, this approach uses the advantage of

broader external disease exposure while still adapting the model to local imaging conditions.

Acknowledgement

The author acknowledges SPARK Academy and the following funding sources and contributors: Lacuna Fund for Health and Equity (PI: Udunna Anazodo, 0508-S-001); National Science and Engineering Research Council of Canada (NSERC) Discovery Launch Supplement (PI: Udunna Anazodo, DGECR-2022-00136); Digital Research Alliance of Canada (the Alliance); McGill University Doctoral Internship Program; University of Washington Azure GenAI for Science Hub; and SPARK Academy 2025 Instructors.

Conflict of Interest

The author declares no competing interests.

References

[1] A. Vasiliu et al., "Tuberculosis prevention: current strategies and future directions," Clin. Microbiol. Infect., vol. 30, no. 9, pp. 1123-1130, Sep. 2024, doi: 10.1016/j.cmi.2023.10.023.

[2] World Health Organization, "Global tuberculosis report 2025," Geneva, 2025. [Online]. Available: https://www.who.int/publications/i/item/9789240116924

[3] J. Yayan, K.-J. Franke, M. Berger, W. Windisch, and K. Rasche, "Early detection of tuberculosis: a systematic review," Pneumonia, vol. 16, no. 1, p. 11, Jul. 2024, doi: 10.1186/s41479-024-00133-z.

[4] A. S. Gaudencio, M. Carvalho, P. G. Vaz, J. M. Cardoso, and A. Humeau-Heurtier, "Tuberculosis detection on chest X-rays using two-dimensional multiscale symbolic dynamic entropy," Biomed. Signal Process. Control, vol. 111, p. 108346, Jan. 2026, doi: 10.1016/j.bspc.2025.108346.

[5] B. Tripathy et al., "TL-PneuNet: a transfer learning-based pneumonia classification framework," Sci. Rep., vol. 15, no. 1, p. 40307, Nov. 2025, doi: 10.1038/s41598-025-24180-8.

[6] B. M. Idowu and T. A. Okedere, "Diagnostic Radiology in Nigeria: A Country Report," J. Glob. Radiol., vol. 6, no. 1, p. 1072, Jun. 2020, doi: 10.7191/jgr.2020.1072.

[7] A. Musa, M. Ibrahim Adamu, H. A. Kakudi, M. Hernandez, and Y. Lawal, "Analyzing Cross-Population Domain Shift in Chest X-Ray Image Classification and Mitigating the Gap with Deep Supervised Domain Adaptation," in Medical Image Computing and Computer Assisted Intervention - MICCAI 2024, vol. 15003, Cham: Springer Nature Switzerland, 2024, pp. 585-595, doi: 10.1007/978-3-031-72384-1_55.

[8] W. Jonathan and D. Fidelis, "Explainable Transfer Learning for Multi-Class Pulmonary Disease Classification Using Nigerian Chest X-ray Dataset," Blue Apple Limited. Accessed: Jun. 04, 2026. [Online]. Available: https://engineersforum.com.ng/wp-content/uploads/2026/03/Explainable-Transfer-Learning-for-Multi-Class-Pulmonary-Disease-Classification-Using-Nigerian-Chest-X-ray-Dataset.pdf

[9] S. Albahli and T. Nazir, "AI-CenterNet CXR: An artificial intelligence (AI) enabled system for localization and classification of chest X-ray disease," Front. Med., vol. 9, p. 955765, Aug. 2022, doi: 10.3389/fmed.2022.955765.

[10] L. T. Duong, N. H. Le, T. B. Tran, V. M. Ngo, and P. T. Nguyen, "Detection of tuberculosis from chest X-ray images: Boosting the performance with vision transformer and transfer learning," Expert Syst. Appl., vol. 184, p. 115519, Dec. 2021, doi: 10.1016/j.eswa.2021.115519.

[11] 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: IEEE, May 2019, pp. 427-433, doi: 10.1109/EIT.2019.8833768.

[12] L. Eisentraut, C. Mai, J. Hosch, A. Benecke, P. Penava, and R. Buettner, "Deep Learning-Based Detection of Tuberculosis Using a Gaussian Chest X-Ray Image Filter as a Software Lens," IEEE Access, vol. 13, pp. 36065-36081, 2025, doi: 10.1109/ACCESS.2025.3544923.

[13] V. Acharya et al., "AI-Assisted Tuberculosis Detection and Classification from Chest X-Rays Using a Deep Learning Normalization-Free Network Model," Comput. Intell. Neurosci., vol. 2022, pp. 1-19, Oct. 2022, doi: 10.1155/2022/2399428.

[14] C. Dasanayaka and M. B. Dissanayake, "Deep Learning Methods for Screening Pulmonary Tuberculosis Using Chest X-rays," Comput. Methods Biomech. Biomed. Eng. Imaging Vis., vol. 9, no. 1, pp. 39-49, Jan. 2021, doi: 10.1080/21681163.2020.1808532.

[15] M. Muyoyeta et al., "The Sensitivity and Specificity of Using a Computer Aided Diagnosis Program for Automatically Scoring Chest X-Rays of Presumptive TB Patients Compared with Xpert MTB/RIF in Lusaka Zambia," PLoS ONE, vol. 9, no. 4, p. e93757, Apr. 2014, doi: 10.1371/journal.pone.0093757.

[16] D. S. Sule et al., "Diagnostic accuracy of AI-assisted chest radiographs in tuberculosis screening: A Ghanaian clinical study," PLOS One, vol. 21, no. 3, p. e0342988, Mar. 2026, doi: 10.1371/journal.pone.0342988.

[17] 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, doi: 10.1109/ACCESS.2020.3031384.