EP-JOURNALS

GLOBAL JOURNAL OF ENGINEERING AND TECHNOLOGY REVIEW

Available on: http://gjetr.ep-journals.org/ E-ISSN: 3051-3782, P-ISSN: 3051-3774

Global Journal of Engineering and Technology Review

Volume: 01 Issue: 03 November 2025

Deep Learning in Medical Imaging: Using Densenet121 for Automated Tuberculosis Detection from Chest X-Rays

Emmy Danny Ajik
Department Information Technology, Federal University Dutsin-Ma, Katsina, Nigeria

Aminu Bashir Suleiman
Department of Cyber Security, Federal University Dutsin-Ma, Katsina, Nigeria

Stephen Luka
Department of Software Engineering, Federal University Dutsin-Ma, Katsina, Nigeria

Mukhtar Umar Shitu
Department of Computer Science, Federal University Dutsin-Ma, Katsina, Nigeria

Joseph Nda Ndabula
Department of Software Engineering, Federal University Dutsin-Ma, Katsina, Nigeria

Published Date: November 03, 2025 **Article DOI: 10.65150/EP-gjetr/V1E3/2025-01**

ABSTRACT: Millions of reported cases and associated deaths highlight the annual global threat posed by Tuberculosis (TB). Added to this, limited diagnostic services, particularly in rural Nigeria, worsen the prevalence of TB in the country. To address these challenges, this research explores the deployment of the deep learning model DenseNet121 to automate TB diagnosis from chest X-rays in low-resource settings such as Nigeria. The model aims to facilitate earlier TB detection in communities with inadequate access to diagnostic services. The absence of qualified TB radiologists in these communities further enhances the model's potential. Based on analysis of a database comprising 4,200 chest X-ray images, the model achieved the following diagnostic metrics: 97.14% accuracy, 0.94 precision, 0.93 recall, and 0.90 F1 score. Such results provide sufficient evidence that the model will significantly improve the timely diagnosis and detection of TB cases. This illustrates the power of Artificial Intelligence tools in constraining and limited environments.

KEYWORDS: Tuberculosis (TB), Chest X-rays, DenseNet121, Medical Imaging, Deep Learning.

1.0 INTRODUCTION

The 2024 Global Tuberculosis Report indicates Tuberculosis (TB) is still a challenge worldwide, as it states there were 10.8 million new TB cases and 1.25 million TB related deaths in the year 2023 (Estaji et al., 2025). 55% of new TB cases were in men, 33% in women, and 12% in children; thus, these 3 demographics covered over 66% of TB cases worldwide (Mohammed et al., 2025). New cases in Nigeria account for around 4.5% of the total cases. Nigeria is also the country in Africa with the highest number of cases of TB/HIV co-infection, multi-drug-resistant TB, as well as TB (Ugwu et al., 2021).

TB is still persistent in the world due to socio-economic challenges, weakened health systems, and a lack of diagnostic health services. In the 2023 WHO report, it estimated a 4.6% increase in TB incidence cases from 2020 (Suvvari et al., 2025). Nigeria illustrates this type of burden by reporting 371,019 TB cases. In 2023, it is estimated that there were 495,125 cases, which means 25% of cases were undiagnosed. The NTBLCP has a plan to address its follow-up issue of 25% of TB cases within Nigeria, which is due to a lack of diagnostic health services and poor coverage of TB control health services. The 2023 NTBLCP invigorates the diagnostics and TB programme with the provision that 4% of health facilities in the country still lack basic TB control services (Alege et al., 2024).

Effective control of TB is largely dependent on timely and accurate TB diagnosis; however, there still continue to be undiagnosed cases in low and middle-income countries. As with all forms of manual data interpretation, reviewing radiographic images of the chest is labour-intensive, prone to inconsistency, and variable depending on the calibrations of the interpreting radiologist. Moreover, TB (Tuberculosis) lesions are frequently classified inaccurately, as they can be considerably confused with other pulmonary pathologies (Maheswari et al., 2024). Particularly in the rural localities of countries like Nigeria, there are only a handful of registered radiologists, leading to delays and variable quality of reports on chest images. In contrast with the TB diagnosed clinically, 82% of the TB diagnosed in Nigeria is ascertained bacteriologically, advocating for a rare practice of incorporating clinically and radiologically (Eneogu et al., 2024).

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Currently, the most efficient methods being investigated for TB screening incorporate AI (Artificial Intelligence). Particularly, for image classification and analysis for chest radiographs, Convolutional Neural Networks (CNNs) are a strong performer (Mahmud et al., 2025). In TB classification, a good-performing shallow CNN using explainability methods achieves classification performance combined with strong ROC and AUC measures (Maheswari et al., 2024). Separate from the academic debates, hybrid models have also shown remarkable results (Siddharth et al., 2025).

Notably, despite the absence of radiologists and the required imaging devices, the delays in diagnosing TB internationally, even with AI technologies, are largely political. The standard interpretation of chest X-rays is also a tedious process (Hwang et al., 2024). The demand for the technologies described is the most considerable, and the least developed of the other measures described. These tools are intended to improve the identification of cases in people without symptoms and assist health workers who have less experience.

This study proposes the application of DenseNet121, a neural network architecture, to the automation of the detection of tuberculosis from chest X-rays. The principal aim here is to design, develop, and assess a CNN-based model suitable for the Nigerian setting, which is a resource-limited environment. The goals of the present study include the following:

(1) to automate the detection of TB from chest X-rays,

(2) to target use in facilities that lack comprehensive radiology services, and

(3) to evaluate the model accuracy performance.

Deploying portable, AI-powered X-ray devices and the use of convolutional neural networks (CNNs) for feature extraction and providing actionable insights is anticipated to enhance TB case detection, significantly decreasing diagnostic delays and improving case detection.

2.0 LITERATURE REVIEW

In research pertaining to the use of AI and machine learning for TB detection, core performance metrics remain indispensable for the assessment of the TB detection accuracy of the given technologies. Hansun et al. (2025) synthesized 152 studies and determined that, in order of frequency, the evaluation metrics most commonly utilized are accuracy, recall/sensitivity, and AUC-ROC, followed by specificity and F1-score. The use of deep learning models, and especially CNN architectures VGG-16, ResNet-50, and DenseNet-121, surpasses the performance levels of traditional machine learning. The report authors cautioned that the use of accuracy as a measure in unbalanced datasets could be misleading, and thus recommended the use of a combination of other metrics, AUC, sensitivity, specificity, and F1-score, in conjunction with accuracy to assess the overall effectiveness of the model in diagnosis.

Showkatian et al. (2022) conducted a comparison of CNNs where some were trained from scratch while others used pre-trained architectures (Xception, Inception_v3, ResNet50, VGG16, and VGG19) in the analysis of the Shenzhen and Montgomery chest X-ray datasets. Their assessment utilized accuracy, precision, recall, F1-score, and AUC as metrics. Every model surpassed the accuracy threshold of 87%, with Xception and ResNet50 achieving precision and recall values of approximately 0.91. The authors determined that the F1-score is the most informative measure in the assessment of precision-recall balance, which is crucial in medical screening to minimize unnecessary follow-up procedures; thus, the authors noted the importance of the AUC in determining model discrimination. Their analysis affirmed that transfer-learning architectures strengthen the model's ability to discriminate and dependability, particularly with smaller datasets.

Zulfiqar et al. (2025) created a comprehensive guide for the construction and assessment of CNN-based models, chiefly Ethereum-focused on performance monitoring aimed at reducing overfitting throughout the training, validation, and testing of the model. During this process, the model recorded 92% training accuracy, 89% validation accuracy, and 89% testing accuracy, which corresponds to other studies that record AUC score results of between 87% and 96%. They concluded that test accuracy should not be the sole measure of a model's performance, and that validation loss and AUC score provide a better measure for evaluating generalization. The study emphasized the need for the integration of the metrics (numerical score of accuracy, precision, recall, F1-score and AUC) with the visual tool of learning curves to test the consistency of a model. They argued that this is critical to assess any model's robustness for potential integration in clinical practice.

Nijiati et al. (2022) implemented deep learning algorithms on 3D CT scans and used 3D VGG-16, 3D EfficientNet, and 3D ResNet-50 to classify TB as either active or non-active. For performance evaluation, they used accuracy, recall, F1-score, and AUC-ROC. 3D ResNet-50 was found to have the best performance with an AUC of 0.961 on the test set and 0.946 on external validation. The researchers pointed out the importance of recall as the primary clinical determinant for identifying active TB and noted that AUC-ROC analysis aims to provide threshold-independent diagnostic certainty. The best model out of the 11 provided diagnostic accuracy that was comparable to and in some cases exceeded that of the radiologists, while being considerably faster. The study demonstrated that the most informative assessment of performance in the context of medical diagnosis, where the most relevant features are AUC, recall and F1-score, is likely to provide the most informative assessment of performance.

In their study, Sharma et al. (2024) analyzed TB detection from chest X-rays using UNet and Xception architectures. They reported segmentation and classification accuracies of 96.35% and 99.29% respectively, with AUCs of 0.99 and 0.999, demonstrating exceptional performance. In contrast, other models such as VGG16, ResNet, and MobileNet did not perform as well, with AUCs ranging between 0.96 and 0.998. This suggests greater discrimination power with deeper architectures. The authors explained why precision and recall provide greater class-level balance, while the F1-score offers a composite measure of the two. For assessing the generalization of a model, AUC is the best measure. Furthermore, best practice in the field states that machine learning models perform well in medical imaging. Literature ought to include AUC, recall or sensitivity and F1-score to provide sufficient information on the clinical relevance and utility of the model.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

3.0 METHODOLOGY

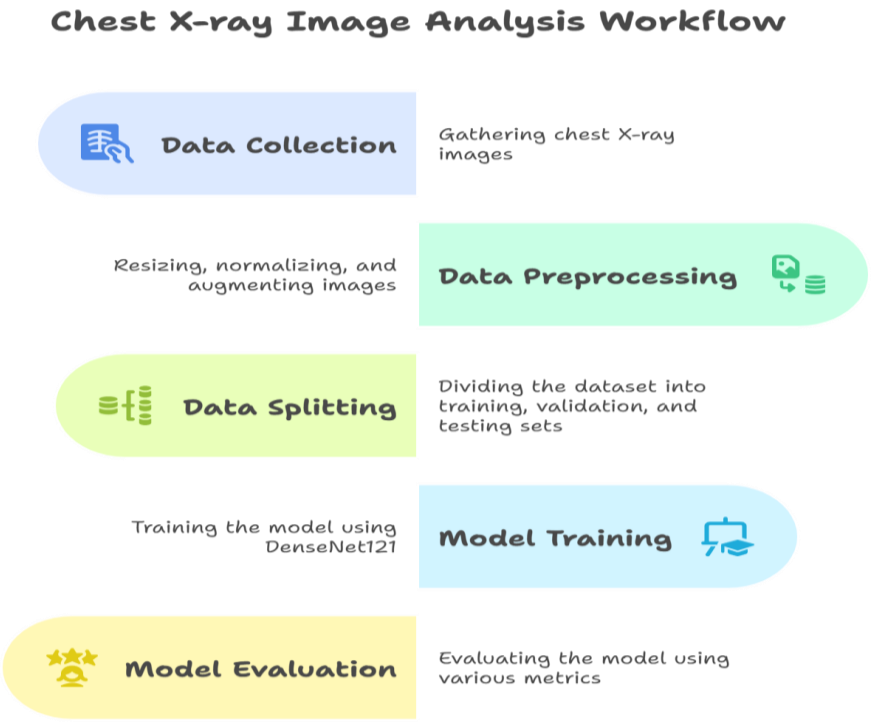


Figure 1: Architectural Framework for Tuberculosis Detection.

3.1 Data Collection and Annotation

The study's dataset was obtained from medical imaging repositories that are openly accessible, specifically the Tuberculosis Chest Radiography Database. 3,500 normal cases and 700 cases with tuberculosis were among the 4,200 chest X-ray photos in this collection. The photos came with metadata in Microsoft Excel format, including source URLs, file names, formats, and image sizes. While the cases that tested positive for tuberculosis came from Tuberculosis by and other associated sources, the normal cases were mainly taken from the National Library of Medicine's (NLM) open database. One annotated dataset was constructed by concatenating the metadata from both sources. A new column called Tuberculosis was then created, with a value of 0 indicating a normal radiograph and a value of 1 indicating tuberculosis. The combined dataset was then randomly shuffled to remove any ordering bias that might affect model learning.

3.2 Data Preprocessing

Preprocessing was a crucial step in determining the suitability of the data to input into the neural network. Each chest X-ray was resized to 320 × 320 pixels to meet the input requirement of the DenseNet121 model. To ensure uniform contrast and brightness over the images, the pixel values were normalised, scaled to the standard deviation and centred around the mean. The single-channel (grayscale) X-ray images were converted to three-channel (RGB) images by assigning the single intensity value to each colour channel. This was necessary since the models trained on ImageNet expect three-channel input.

To enhance the training dataset's variability and the model's generalization capacity, data augmentation was used in addition to normalization. Each training image was randomly rotated by up to 5 degrees, magnified by up to 10%, and shifted by up to 10% horizontally or vertically using Keras' ImageDataGenerator. These enhancements mimic real-world conditions typical in clinical imaging settings, including slight patient movement, changes in radiographic posture, and exposure variations. Thus, instead of focusing on visual artifacts, the network learns to focus on the real lung structure and disease patterns.

3.3 Data Splitting

Subsequent to preprocessing, the dataset was divided into three subsets, namely, training, validation, and testing sets, which were allocated in an 80:10:10 proportion. Such stratification and splitting was performed to ensure that there were equal amounts of normal images and tuberculosis positive images in the subsets. Details of the partitioning are illustrated in Table 1.

Table 1: Data Distribution after Splitting the Dataset

Subset	Normal	Tuberculosis	Total
Training	2800	560	3360
Validation	350	70	420
Testing	350	70	420

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

The images were organized into corresponding directories: training, validation, and testing, with separate subfolders for the *Normal* and *Tuberculosis* categories. To reduce memory load and increase computational efficiency, this structure made it easier to use Keras' data generators, which load photos in batches directly from folders during training. The test set was kept exclusively for the final assessment to evaluate the model's performance in the real world. In contrast, the validation set was used to adjust the model's hyperparameters and monitor overfitting.

3.4 Handling Class-Imbalance

Corrective action was taken to avoid the model from biasing toward the majority class because the dataset was extremely imbalanced, with normal photos substantially outnumbering tuberculosis images by a ratio of roughly 5:1. Each class's contribution to the total loss was modified according to its frequency in the training data using a specially designed weighted binary cross-entropy loss function. Because the positive class (tuberculosis) was given a higher weight than other classes, misclassifying tuberculosis cases resulted in a greater penalty than misclassifying normal instances. Despite their smaller number, the model was able to learn various aspects of tuberculosis-positive cases since the loss function mathematically balanced the positive and negative contributions. The model's sensitivity for detecting tuberculosis was greatly increased by this method.

3.5 Model Development and Training

Model development was built upon the DenseNet121 architecture. DenseNets were cited to be some of the best convolutional neural networks for their effective feature propagation and low parameter count. To use the pre-trained DenseNet121 model as a feature extractor, the top (fully connected) layers were removed, which were trained on the ImageNet dataset. The last layer, Global Average Pooling layer, was used to condense spatial information per feature map, and binary classification was performed on a Dense output layer with a sigmoid activation for the purpose of classification.

The model development utilized the previously formulated weighted binary cross-entropy loss function and the Adam optimizer for adaptive learning rate modulation. Accuracy and the area under the receiver operating characteristic curve (AUC) were used as evaluation metrics. For faster computation, a GPU (Tesla P100 with 16 GB of memory) was utilized under Google Colab's environment. To combat overfitting and retain the ideal model weights, the model was trained for three epochs with early stopping and checkpointing callbacks at a batch size of eight. Each training epoch contained 100 steps, and the results were validated using the validation set after each epoch.

3.6 Model Evaluation

To evaluate the predictive performance of the DenseNet121-based classifier after the completion of model training, the independent test set, which was not provided to the model during training or validation, was utilised. This step assesses the model's generalisation ability i.e., the ability to correctly classify unseen chest X-ray images into one of the two categories, TB-positive or normal. To thoroughly evaluate the classifier's performance, several metrics designed to measure different aspects of the classifier including accuracy, precision, recall (sensitivity), specificity, the F1 score, and the AUC (area under the receiver operating characteristic curve) were calculated.

Accuracy denotes the total proportion of instances, whether positive or negative, that have been correctly recognized in relation to the total number of samples evaluated over the given period. It is defined mathematically as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

TP refers to true positives (that is, tuberculosis images accurately recognized as tuberculosis). TN indicates true negatives (normal images accurately recognised as normal). FP indicates false positives (normal images that were inaccurately recognised as tuberculosis). FN indicates false negatives (tuberculosis images that were inaccurately recognised as normal). Although accuracy provides a general measure of performance, in cases of imbalanced datasets such as this case where normal cases outnumber tuberculosis cases, accuracy may not be fully truthful.

In response to this limitation, the model's diagnostic sensitivity and precision were assessed in isolation. Precision, or positive predictive value, measures the trustworthiness of positive predictions and is defined as:

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

A model demonstrating high precision implies that whenever it predicts tuberculosis, it is usually correct. Recall, often referred to as sensitivity, gauges the model's capacity to accurately recognise all positive cases of tuberculosis and is calculated as:

$$Recall\ (Sensitivity) = \frac{TP}{TP+FN} \qquad (3)$$

In medical applications, having a high recall value is important since, in the case of tuberculosis, not identifying positive cases (false negatives) can have serious public health ramifications. On the other hand, specificity, complementary to sensitivity, measures the model's correct identification of negatives and is computed as follows:

$$Specificity = \frac{TN}{TN + FP} \qquad (4)$$

Specificity reflects how well the model avoids false alarms, ensuring that normal chest X-rays are not misclassified as tuberculosis.

Because precision and recall often trade off against one another, the **F1-score** was computed as the harmonic mean of both metrics, providing a balanced measure of accuracy for imbalanced datasets:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (5)$$

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

This score is particularly informative when the dataset exhibits uneven class distribution, as in this study. The F1-score combines the model's ability to detect tuberculosis (recall) with its ability to minimize false alarms (precision).

A **confusion matrix** was also constructed to visualize classification outcomes and assess the distribution of predictions. In a two-by-two grid, the confusion matrix displays actual vs predicted classifications. Correct predictions (TP and TN) are represented by diagonal entries, whereas misclassifications (FP and FN) are represented by off-diagonal entries. The visualization made it easier to spot particular flaws in the model's predictions, including its propensity to overlook some TB cases.

4.0 RESULT AND DISCUSSIONS

In this section, the outcomes derived from the proposed method are described. The study was conducted using four thousand two hundred (4,200) samples, which were further subdivided into two groups: Normal X-ray Images and Tuberculosis X-ray Images using binary classification, where the two classes are identified as 0 and 1.

In this study, the dataset was divided in an 80-10-10 split (i.e. 80% for training, 10% for validation, and 10% for testing the model). The proposed model's performance was assessed using the metrics of precision, recall, total accuracy, as well as the confusion matrix and ROC curve which demonstrate model performance as shown in Figures 2 and 3.

	precision	recall	f1-score	support
0.0	0.90	1.00	0.95	350
1.0	0.97	0.87	0.85	70
accuracy			0.97	420
macro avg	0.94	0.93	0.90	420
weighted avg	0.91	0.90	0.89	420

Accuracy of the Model: 97.1428571 %

Figure 2: Model Classification Performance by Class

4.1 Class-Specific Metrics

This section breaks down the performance for each of the two classes. The model was evaluated on a test dataset with a total of 420 samples (350 for class 0 and 70 for
class 1).

Class 0

The predicted outcomes for class 0 were correct 90% of the time, indicating a precision of 0.90. Thus, the model exhibits low False Positive (FP) for this class (i.e., Class 1 instances misclassified as 0). The recall of 1.0 indicates there are no missed instances of class 0, meaning there are no false negatives. Therefore, the model perfectly classifies all instances, achieving a 100% classification.

Moreover, the model recorded 0.95 for the F1-Score which indicates a strong balance between precision and recall for class 0. The F1-score reflects this balance as a single summative measure of strength, demonstrating a strong classification performance for Class 0. Support (350) indicates there are 350 actual instances of class 0 present in the test set.

Class 1

All predictions for Class 1 were 97% correct (Precision 0.97) all the time. This indicates that the model encounters zero False Positives (FP) for this class. (i.e., no instances of Class 0 were incorrectly classified as 1.0). For the recall (0.87) this means that 87% of all the actual Class 1 instances were correctly identified by the model. This indicates that for this class there is a low rate of False Negatives (FN). Specifically, 2% of the true Class 1 instances were incorrectly classified as 0 (FN).

The model also attained a high F1-Score of 0.85. The support (70) indicates that there are 70 actual instances of class 1 in the test set.

4.2 Overall Averages

The model attained an overall accuracy rate of about 97%. While this sounds impressive, it comes with the caveat of being potentially misleading, particularly in the case of an imbalanced classification problem, where a "dummy" classifier which always predicts the majority class (0) would still achieve 350/420 predictions correct, 83.3% accurate. The 97.14% accuracy is an improvement, but should always be considered alongside the other metrics.

Regarding the Precision (0.94), Recall (0.93), and F1-Score (0.90) at the Macro Average, these metrics denote the unweighted mean of the respective metric across all classes and, as such, each class is treated the same regardless of class support. The Macro Recall (0.93) is lower than Macro Precision (0.94) because class (1) has poor Recall. This is a decent, unbiased reflection of the model's performance overall across both classes.

The Weighted Average, also described as the Average of class metrics, refers to the results of the metrics when class metrics are weighted by class support. These results are closer to the metrics of the majority class (0) because it comprises a larger portion of the test set (350/420).

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

4.3 Confusion Matrix Analysis

A confusion matrix is a tool that presents the results of a classification model on a test dataset with known target values. This table helps one visualise how well a given algorithm performs. Each row of the matrix corresponds to the cases of a given class that was encountered in the test set, while each column corresponds to the cases of a class assigned to that instance by the predictive model.

In this example, the model performs binary classification, with the classes determined by the matrix labeled 0 (Negative) and 1 (Positive). As illustrated in Figure 3, the confusion matrix captures performance by listing the values of true and false predictions.

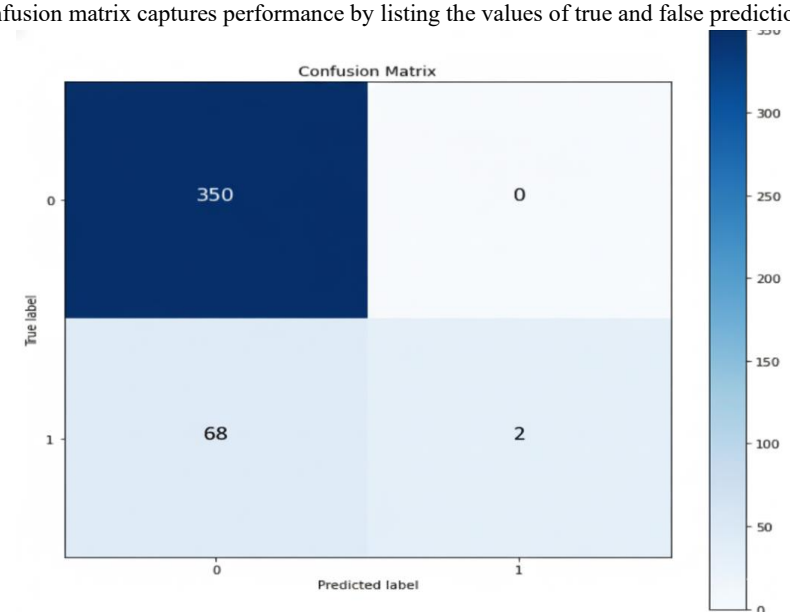


Figure 3: Confusion Matrix

The proportion of total correct predictions (TP + TN) out of all predictions (Total Samples): The overall accuracy is high, suggesting a good model. The breakdown of the Confusion Matrix is provided in Table 2.

Table 2. Confusion Matrix Breakdown

	Predicted 0	Predicted 1	Total Actual
Actual 0	350 (True Negative, TN)	0 (False Positive, FP)	350
Actual 1	68 (False Negative, FN)	2 (True Positive, TP)	70
Total Predicted	418	2	420 (Total Samples)

True Negatives (TN) = 350: The model correctly classified 350 instances that were actually class 0.

False Positives (FP) = 0: The model incorrectly classified 0 instances of class 0 as class 1.

False Negatives (FN) = 2: The model incorrectly classified 2 instances of class 1 as class 0. This is the main area of model failure.

True Positives (TP) = 68: The model correctly classified 68 instances that were actually class 1.

4.4 Comparison of Classification Results with Previous Studies

The proposed model shows strong overall performance, especially in recall for class 0 and precision for class 1. Compared to Showkatian et al., the current model has higher accuracy and slightly better macro precision and recall. The works of Zulfiqar and Sharma report higher or comparable accuracy, but lack detailed metrics, making it hard to judge their models' robustness. As for Nijiati's study, it provides no performance data, so it can't be evaluated. The summary of the comparison is given in table 3.

Table 3: Comparison of classification results with previous studies

No	Studies	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	Proposed work	97	0.97	0.87	0.85
2	Showkatian et al. (2022)	87	0.91	0.91	0.91
2	Zulfiqar et al. (2025)	89	-	-	-
3	Nijiati et al. (2022)	-	-	-	-
4	Sharma et al. (2024)	99	-	-	-

CONCLUSION

The DenseNet121 deep learning model has provided 'promising results' with respect to its implementation with a 97.14% accuracy rate. This model has great significance for low-resource contexts such as rural Nigeria, where access to professional radiologists is limited. The model uses Artificial Intelligence and can help save valuable time, enabling early detection of TB. Rapid early detection is crucial to ensure TB can be treated and help reduce its transmission to others. The model also demonstrates impressive TB case identification, as evidenced by high precision and

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

recall. The value of deep learning technology in transforming healthcare is evident in regions with scarce healthcare resources. Future work should focus on obtaining more diverse and balanced datasets and on the global clinical integration of TB diagnosis.

REFERENCES

1) Alege, A., Hashmi, S., Eneogu, R., Meurrens, V., Budts, A. L., Pedro, M., ... & Agbaje, A. (2024). Effectiveness of using AI-driven hotspot mapping for active case finding of tuberculosis in Southwestern Nigeria. *Tropical Medicine and Infectious Disease*, *9*(5), 99.
2) Eneogu, R. A., Mitchell, E. M., Ogbudebe, C., Aboki, D., Anyebe, V., Dimkpa, C. B., ... & Gidado, M. (2024). Iterative evaluation of mobile computer-assisted digital chest x-ray screening for TB improves efficiency, yield, and outcomes in Nigeria. *PLOS Global Public Health*, *4*(1), e0002018.
3) Estaji, F., Kamali, A., & Keikha, M. (2025). Strengthening the global Response to Tuberculosis: Insights from the 2024 WHO global TB report. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, *39*, 100522.
4) Hansun, S., Argha, A., Bakhshayeshi, I., Wicaksana, A., Alinejad-Rokny, H., Fox, G. J., Liaw, S.-T., Celler, B. G., & Marks, G. B. (2025). Diagnostic performance of artificial intelligence–based methods for tuberculosis detection: Systematic review. *Journal of Medical Internet Research, 27*, e69068.
5) Hwang, E. J., Jeong, W. G., David, P. M., Arentz, M., Ruhwald, M., & Yoon, S. H. (2024). AI for detection of tuberculosis: Implications for global health. *Radiology: Artificial Intelligence*, *6*(2), e230327.
6) Maheswari, B. U., Sam, D., Mittal, N., Sharma, A., Kaur, S., Askar, S. S., & Abouhawwash, M. (2024). Explainable deep-neural-network supported scheme for tuberculosis detection from chest radiographs. *BMC Medical Imaging*, *24*(1), 32.
7) Mahmud, A. I., Suleiman, A. B., & Yahaya, U. (2025). Modelling functional brain aging in Alzheimer's disease using neural ODEs on rs-fMRI graphs. *International Journal of Research Publication and Reviews, 6*(7), 2482–2488.
8) Mohammed, A., Aboagye, R. G., Duodu, P. A., Adnani, Q. E. S., Wongnaah, F. G., Seidu, A. A., & Ahinkorah, B. O. (2025). Sex-related absolute inequalities in tuberculosis incidence in 47 countries in Africa. *BMC medicine*, *23*(1), 324.
9) Nijiati, M., Zhou, R., Damaola, M., Hu, C., Li, L., Qian, B., Abulizi, A., Kaisaier, A., Cai, C., Li, H., & Zou, X. (2022). *Deep learning based CT images automatic analysis model for active/non-active pulmonary tuberculosis differential diagnosis. Frontiers in Molecular Biosciences, 9*, 1086047. https://doi.org/10.3389/fmolb.2022.1086047
10) Sharma, V., Nillmani, Gupta, S. K., & Shukla, K. K. (2024). *Deep learning models for tuberculosis detection and infected region visualization in chest X-ray images. Intelligent Medicine, 4*, 104–113. https://doi.org/10.1016/j.imed.2023.06.001
11) Showkatian, E., Salehi, M., Ghaffari, H., Reiazi, R., & Sadighi, N. (2022). Deep learning-based automatic detection of tuberculosis disease in chest X-ray images. *Polish Journal of Radiology, 87*, e118–e124. https://doi.org/10.5114/pjr.2022.113435.
12) Siddharth, G., Ambekar, A., & Jayakumar, N. (2025). Enhanced CoAtNet based hybrid deep learning architecture for automated tuberculosis detection in human chest X-rays. *BMC Medical Imaging*, *25*(1), 379.
13) Suvvari, Tarun Kumar. "The persistent threat of tuberculosis− Why ending TB remains elusive?." *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases* 38 (2025): 100510.
14) Ugwu, K. O., Agbo, M. C., & Ezeonu, I. M. (2021). PREVALENCE OF TUBERCULOSIS, DRUG-RESISTANT TUBERCULOSIS AND HIV/TB CO-INFECTION IN ENUGU, NIGERIA. *African Journal of Infectious Diseases (AJID)*, *15*(2), 24–30. https://doi.org/10.21010/Ajid.v15i2.5
15) Zulfiqar, R., Akhtar, R., Khan, T. A., Jawad, M. M., & Khan, F. u. (2025). Tuberculosis diagnosis using deep learning. *Bahria University Journal of Information & Communication Technologies, 13*(2), 42–46.

License: This is an open access article under the CC BY 4.0 license: https://creativecommons.org/licenses/by/4.0/

Corresponding Author: Aminu Bashir Suleiman