GSJ: Volume 13, Issue 11, November 2025, Online: ISSN 2320-9186

www.globalscientificjournal.com

A TRANSFER LEARNING BASED DETECTION ALGORITHM FOR TUBERCULOSIS DIAGNOSIS USING X-RAY IMAGES

Njoku Perfect Izuchukwu[1a], Sule Haruna Sani[1b], Onodu Patrick Obinna[3], Issah Shamsudeen[4], Abdelrahman M.A. Eissa[5], Ogbu Meletus Chibuike[6], Temesgen Tesfaye Ajabo[7], Agbukata George Obinna[8]

1, 4, 6, & 7 Department of Public Health, Epidemiology and Evidence Base Medicine, I.M. Sechenov First Moscow State Medical University.

1b Department of Statistics, Federal University of Agriculture Abeokuta.

3 Medicine & Surgery, Enugu state university of science and technology.

5 Faculty Of Veterinary Medicine, University of Sadat City, Egypt

8 Chemical pathology, State University of Medical and Applied Sciences, Igbo Eno, Enugu State

ABSTRACT

Tuberculosis (TB) remains a major infectious disease worldwide and continues to pose a critical public health challenge, particularly in developing countries where access to rapid diagnostic tools is limited. Chest X-ray imaging is a widely available and cost-effective screening technique, but its interpretation requires expertise and is susceptible to human error. Recent advancements in deep learning and transfer learning provide new opportunities to automate TB detection with high accuracy. This study proposes a transfer learning-based approach for Tuberculosis Diagnosis using pretrained algorithms such as DenseNet121, VGG16, and InceptionV3. A publicly available dataset was preprocessed, augmented, and used to fine-tune these models. The models' performance was evaluated using precision, recall, F1-score, accuracy, AUC, and Cohen's Kappa. The study results show that DenseNet121 outperformed the other pre-trained architectures, achieving an accuracy of 99.52%, AUC of 99.91%. The findings confirm the effectiveness of transfer learning for TB diagnosis and highlight the potential of deep learning in enhancing medical screening, especially in low-resource regions.

KEYWORDS: Tuberculosis, Machine Learning, Deep Learning

1.0 INTRODUCTION

Tuberculosis (TB) can be described as infectious disease that is caused by an intracellular pathogen known as Mycobacterium tuberculosis (Gupta et al. 2022; Bussi et al. 2019). In 2021, approximately 10.6 million new cases of tuberculosis were reported globally, reversing prior declines in incidence. This marked a concerning 3.6% rise in TB incidence compared to 2020. Contributory factors include disruptions to healthcare due to the COVID-19 pandemic and gaps in diagnostic and treatment coverage (Sotgiu et al. 2019; Ravikoti et al. 2019). Chest X-ray (CXR) imaging is one of the primary diagnostic tools used for TB screening due to its affordability and widespread availability. However, interpreting X-ray images manually requires trained radiologists, and misdiagnosis may occur due to fatigue, limited expertise, or subtle radiographic features. Recent developments in artificial intelligence (AI), especially deep learning, have revolutionized medical image analysis. Deep convolutional neural networks (CNNs) have achieved state-of-the-art performance in various classification and detection tasks. However, training such models from scratch requires large, annotated datasets, which are often unavailable in medical domains. Transfer learning addresses this challenge by allowing the use pre-trained models that has been developed using larger datasets. This study leverages transfer learning Convolutional Neural Network Algorithms such as DenseNet121, VGG16, and InceptionV3 for detecting TB. The main aim of this study is to evaluate the performance of these algorithms and identify the most effective model for TB diagnosis using X-ray images.

2.0 LITERATURE REVIEW

Hasbib et al. (2020) proposed a unified process for evaluating CT scans using 3D CNNs for Tuberculosis detection. The study reported that the process attained an accuracy score of 67.5% and an Area Under the Curve (AUC) of 73%. Mostofa et al. (2019) utilized a VGG16 pretrained model for detecting tuberculosis using X-ray images. The study developed a comparative analytical approach reviewing the performance of the models with and without data augmentation. The results show that 80% without applying augmentation and 81.25 % with the application of augmentation. Sangheum et al. (2016) developed a Computer-Aided Diagnosis (CAD) using a CNN algorithm for Tuberculosis diagnosis using a Large X-ray Dataset. The study results reported an AUC of 0.96, 0.93, and 0.88. Syeda et al. (2019) evaluated the performance of the following pretrained CNNs, which are VGG-16, VGG-19, RestNet50, and Googlenet, in handling Tuberculosis diagnosis. The study utilized several public datasets, which include the Montgomery County and Shenzhen datasets. The study results review that VGG-16 & 19 attained the highest classification accuracy under the MC dataset of 77.14% and VGG-16 surpasses other algorithms for the SH dataset with an accuracy score of 86.74%. Chang et al. (2017) developed a CNN algorithm that was able to deal with imbalanced Tuberculosis X-ray images. The study decided to evaluate the capacity of this proposed model under a multi-classification instance. The study results indicate that this model attained a classification accuracy of 85.68%. Smit et al. (2021) fine-tuned a VGG16-CNN model for predicting tuberculosis diagnosis. The study utilized an 80:20 data split for developing the model and trained the model using two datasets, which are the Shenzhen and Montgomery datasets. The study results revealed that the model attained an accuracy of 91% (Shenzhen) and 89% (Montgomery). Xavier et al. (2021) proposed an approach that utilizes segmentation of Chest X-ray images linked to Tuberculosis diagnosis. The study trained several Machine learning algorithms, which include Logistic Regression, Naïve Bayes, Support Vector Machine, and Stacked Loopy Decision Tree (SLDT). The study results review that SLDT performed well over the three datasets considered, with an accuracy of 99.78%, 98.9%, 99.6%, and 98.96%.

3.0 METHODOLOGY

3.1 DATASET DESCRIPTION

The dataset used in this study was collated by researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh along with their collaborators from Malaysia in collaboration with medical doctors from Hamad Medical Corporation and Bangladesh have created a database of chest X-ray images for Tuberculosis (TB) positive cases along with Normal images. The dataset is made up of 3500 normal X-ray images and 700 tuberculosis images. The distribution used in training the Deep learning algorithm in this study is shown in Table 1.0 below.

Table 1.0: Dataset Distribution

Class/Image Distribution	**Training**	**Validation**	**Testing**
Normal	2450	525	525
Tuberculosis	490	105	105

3.2 DATA PREPROCESSING

For this study, the data preprocessing techniques utilized for preparation of the X-ray images for the pretrained CNN algorithms include image resizing, image normalization, and data augmentation. The programming language used in the implementation of both the data preprocessing stage, model training, and evaluation was Python. For the image resizing, we used an image input size of 224 x 224, which is suitable for the pretrained algorithms. Then, we applied normalization to the pixel values of the images to improve the model convergence. Data augmentation was used to address issues of overfitting during the training and evaluation phase.

3.3 DEEP LEARNING ALGORITHM

3.3.1 DENSENET121

DenseNet121 architecture uses a feed-forward process that connects each layer in the network by sending feature maps from preceding layers (Sankari, 2025). This variant of the Convolution neural network supports the recycling of features, supporting gradient flow into the network (Heru et al. 2024; Sankari, 2025). This mitigates issues of vanishing gradients. The key features of this model are feature maps that use concatenation and its ability to learn with fewer parameters (Sankari, 2025; Dheeraj et al., 2025).

3.3.2 INCEPTIONV3

InceptionV3 architecture is a Convolution Neural Network variant that uses batch-normalization, alongside varying filters to capture multi-scale information (Poonam Shourie et al., 2023). The transfer learning algorithm's strength lies in its ability to extract hierarchical features efficiently (Tanishq Soni et al., 2024). Also, the concatenated layers and global average pooling give it a robust generalization over handling different classification and feature extraction tasks efficiently.

3.3.3 VGG16

VGG16 comprises 16 layers with learnable weights, including 13 convolutional layers and 3 fully connected layers. This depth facilitates robust feature extraction and representation, enabling the network to recognize increasingly complex patterns in images. The architecture adheres to a simple and uniform design philosophy, relying exclusively on 3×3 convolutional filters and max-pooling layers.

3.4 MODEL EVALUATION METRICS

3.4.1 Accuracy

Accuracy measures the proportion of correctly classified instances across all instances in a dataset. The mathematical formula that describes the metrics is show below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad 1.0$$

3.4.2 Precision

Precision measures the reliability of the model in predicting positive outcomes. It quantifies the proportion of true positives among all instances predicted as positive.

$$Precision = \frac{TP}{TP + FP} \quad 2.0$$

3.4.3 Recall

Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive instances that the model successfully identifies as positive.

$$Recall = \frac{TP}{TP + FN} \quad 3.0$$

3.4.4 F1-score

The F1-score is the harmonic mean of precision and recall, providing a single metric that balances both. It emphasizes the trade-off between precision and recall and is particularly useful in datasets with class imbalance.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall}$$

4.0 RESULT FINDINGS

The sections focus on the results metrics of the pretrained CNN models trained using an image input size of 224 by 224, batch size of 32, ReLu activation, and a Sigmoid output function. These algorithms were trained using an epoch size of 30 and also initiated a checkpoint variable that enables us to save the model with the training weights based on its accuracy using the validation set. The model evaluation results are presented in the section below.

Table 2.0: DenseNet121 model performance result

	Precision	**Recall**	**F1-score**
Normal	0.9943	1.0000	0.9972
Tuberculosis	1.0000	0.9714	0.9855

Table 3.0: InceptionV3 model performance result

	Precision	**Recall**	**F1-score**
Normal	0.9849	0.9943	0.9896
Tuberculosis	0.9700	0.9238	0.9463

Table 4.0: VGG16 model performance result

	Precision	**Recall**	**F1-score**
Normal	0.9887	0.9962	0.9924

Tuberculosis	0.9802	0.9429	0.9612

Tables 2 to 4 show the transfer learning algorithms' performance over the X-ray test images across the different diagnostic classes considered in this study. The performance was evaluated using the precision, recall, and F1-score metrics defined in the previous section. Table 1 shows that the DenseNet121 model attained a recall score of 0.9714 over the tuberculosis diagnosis class, indicating that the model was able to capture 97.14% of the tuberculosis cases accurately. This capture had a precision score of 100% and an F1-score of 98.55%. DenseNet121 model also attained a recall of 100% for classifying patients with a No-tuberculosis diagnosis with a precision of 99.43% and an F1-score of 99.72%. Inception V3 model yielded a recall score of 0.9238, indicating 92.38% tuberculosis classes were detected with a precision score of 97% and an f1-score of 94.63% (See Table 2.0). The InceptionV3 model also had a higher recall score of 99.43% similar to that of DenseNet121 for detecting patients with a No-tuberculosis health status. This was attained with a precision of 98.49% and an F1-score of 98.96%. VGG16 model attained a recall score of 94.29% for the Tuberculosis class with a precision score of 98.02% and an f1-score of 96.12% (See Table 3.0). It also attained a recall score of 99.62%, precision of 98.87% and f10score of 99.24% for detecting the Normal Class. Overall, it was observed that the Transfer learning models yielded a higher recall score for detecting patients who were classified as Normal (No Tuberculosis) for all three models utilized. We also noted that the DenseNet121 attained the highest performance across all metrics across the two distinctive diagnostic classes.

Table 5.0: The model performance results

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**	**Kappa statistic**	**AUC**
DenseNet121	**99.52%**	**99.72%**	**98.57%**	**99.13%**	**98.27%**	**99.91%**
InceptionV3	98.25%	97.75%	95.90%	96.80%	93.59%	99.83%
VGG16	98.73%	98.44%	96.95%	97.68%	95.36%	99.89%

Table 5.0 shows the models summary performance with three additional metrics which are Accuracy, Kappa statistics, and AUC. The table results were consistent with that recorded earlier which indicated that DenseNet121 attained the best predictive performance with an accuracy score of 99.52%, average precision of 99.72%, average recall of 98.75% and average f1-score of 99.13%.

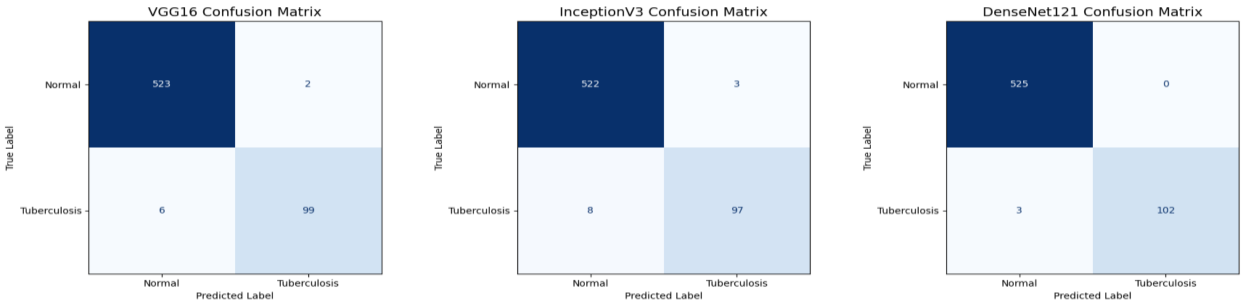

Figure 1.0: Confusion Matrix

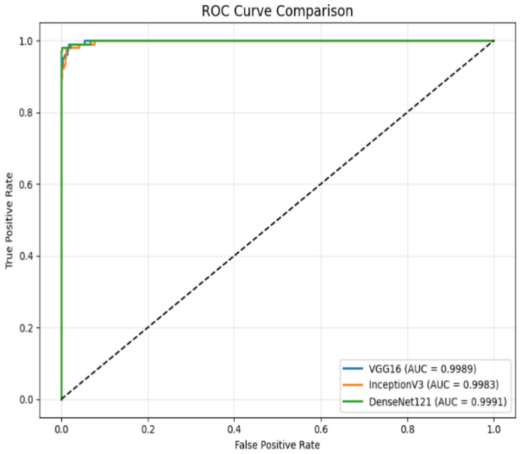

Figure 2.0: ROC Curve Comparison

Table 6.0: Comparative Analysis with recent studies

Author(s)	**Algorithms**	**Performance**
Tawsifur et al. 2020	Data preprocessing+Data Augmentation+Image Segmentation+(ChexNet, DenseNet201, InceptionV3, SqueezeNet, VGG19, MobileNet, ResNet50, ResNet101, ResNet18)	DenseNet201: (Accuracy: 98.6%, Recall:98.56%, Precision:98.57%, F1-score:98.56%)
Proposed Model	Data preprocessing + Data Augmentation + (VGG16, DenseNet121, InceptionV3)	DenseNet121: (Accuracy:99.52%, Recall:98.57%, Precision:99.72%, F1-score: 99.13%, Kappa:98.27%, AUC:99.91%)

5.0 CONCLUSION

The findings of this study demonstrate that transfer learning–based deep convolutional neural networks provide highly effective diagnostic performance for tuberculosis detection using chest X-ray images across all evaluated pretrained models considered in this study, including DenseNet121, InceptionV3, and VGG16. There was consistently strong classification capability, particularly for identifying Normal (non-tuberculosis) cases, where all models achieved recall values above 99%. This indicates their reliability in correctly identifying patients without tuberculosis, an essential requirement for reducing false alarms in clinical settings. Among the three pretrained architectures, DenseNet121 outperformed the others across all major evaluation metrics, including precision, recall, F1-score, accuracy, Kappa statistic, and AUC. The model achieved an overall accuracy of 99.52%, with a recall of 98.57%, a precision of 99.72%, and an F1-score of 99.13%, highlighting its robustness in detecting tuberculosis cases with minimal misclassification. The high Kappa statistic (98.27%) further confirms excellent alignment between predicted and true instances, while the AUC of 99.91% demonstrates outstanding discriminative ability. The comparative analysis with recent studies also shows that the proposed DenseNet121 model surpasses previously reported performance metrics, including those from Tawsifur et al. (2020), reinforcing the effectiveness of the adopted preprocessing, augmentation strategies, and transfer-learning approach. Overall, the results underscore the potential of DenseNet121 as a reliable and accurate tool for automated tuberculosis screening. Its high sensitivity and specificity suggest its suitability for integration into computer-aided diagnostic systems, particularly in resource-limited environments where rapid and accurate tuberculosis detection is critical. Future research may explore model deployment, computational optimization, and validation using larger and more diverse clinical datasets to further enhance diagnostic reliability and real-world applicability.

6.0 REFERENCES

Hasib Zunair1, Aimon Rahman2, Nabeel Mohammed3, and Joseph Paul Cohen (2020). Uniformizing Techniques to Process CT scans with 3D CNNs for Tuberculosis Prediction. Retrieved from: https://arxiv.org/pdf/2007.13224

Mostofa Ahsan, Rahul Gomes, Anne Denton (2019). Application of a Convolutional Neural Network using transfer learning for tuberculosis detection. Retrieved from: https://www.researchgate.net/profile/Mostofa-Ahsan/publication/335793363_Application_of_a_Convolutional_Neural_Network_using_transfer_learning_for_tuberculosis_detection/links/5dd0d599a6fdcc7e13877a24/Application-of-a-Convolutional-Neural-Network-using-transfer-learning-for-tuberculosis-detection.pdf

Sangheum Hwang, Hyo-Eun Kim, Jihoon Jeongb, Hee-Jin Kim (2016). A Novel Approach for Tuberculosis Screening Based on Deep Convolutional Neural Networks.

Syeda Shaizadi Meraj, Razali Yaakob, Azreen Azman, Siti Nurulain Mohd Rum, Azree Shahrel Ahmad Nazri, Nor Fadhlina Zakaria (2019). Detection of Pulmonary Tuberculosis Manifestation in Chest X-Rays Using Different Convolutional Neural Network (CNN) Models. Retrieved from: https://www.researchgate.net/profile/Syeda-Meraj/publication/337674271_Detection_of_Pulmonary_Tuberculosis_Manifestation_in_Chest_X-Rays_Using_Different_Convolutional_Neural_Network_CNN_Models/links/5de4f8b4299bf10bc337bd60/Detection-of-Pulmonary-Tuberculosis-Manifestation-in-Chest-X-Rays-Using-Different-Convolutional-Neural-Network-CNN-Models.pdf

Chang Liu, Yu Cao, Marlon Alcantara, Benyuan Liu, Maria Brunette, Jesus Peinado, Walter Curioso (2017). TX-CNN: DETECTINGTUBERCULOSISINCHESTX-RAYIMAGESUSING CONVOLUTIONALNEURALNETWORK.

Smit B. Patel, Parth H. Patel, Viral D. Jain, and Jai Prakash Verma (2021). Improvised VGG16 CNN Architecture for Predicting Tuberculosis Using the Frontal Chest X-Ray Images.

Inbaraj, X.A.; Villavicencio, C.; Macrohon, J.J.; Jeng, J.-H.; Hsieh, J.-G. A Novel Machine Learning Approach for Tuberculosis Segmentation and Prediction Using Chest-X-Ray (CXR) Images. Appl. Sci. 2021, 11, 9057. https://doi.org/ 10.3390/app11199057

Mohan Bhandaria, Tej Bahadur Shahib,∗, Birat Sikua, Arjun Neupane (2022). Explanatory classification of CXR images into COVID-19,Pneumoniaand Tuberculosis using deep learning and XAI. Retrieved from: https://pmc.ncbi.nlm.nih.gov/articles/PMC9549800/pdf/main.pdf

Arief Rachman Hakim*, Budi Warsito, Muhammad Rionando Divangga, Diah Safitri, Ardiana Alifatus Sa'adah (2025). PERFORMANCE CONVOLUTIONAL NEURAL NETWORK (CNN) AND SUPPORT VECTOR MACHINE (SVM) ON TUBERCULOSIS DISEASE CLASSIFICATION BASED ON X-RAY IMAGE

Sotgiu, G., Goletti, D., & Matteelli, A. (2019). Global tuberculosis prevention: should we start from the beginning? *European Respiratory Journal*, *54*(3), 1901394. https://doi.org/10.1183/13993003.01394-2019

Ravikoti, S., Bhatia, V., & Mohanasundari, S. K. (2025). Recent advancements in tuberculosis (TB) treatment regimens. *Journal of Family Medicine and Primary Care*, *14*(2), 521–525. https://doi.org/10.4103/jfmpc.jfmpc_1237_24

Gupta, R., Pandey, M., Pandey, A. K., Tiwari, P. K., & Amrathlal, R. S. (2022). Novel genetic polymorphisms identified in the clinical isolates of Mycobacterium tuberculosis PE_PGRS33 gene modulate cytokines expression and promotes survival in macrophages. *Journal of Infection and Public Health*, *15*(2), 245–254. https://doi.org/10.1016/j.jiph.2022.01.001

Bussi, C., & Gutierrez, M. G. (2019). *Mycobacterium tuberculosis*infection of host cells in space and time. *FEMS Microbiology Reviews*, *43*(4), 341–361. https://doi.org/10.1093/femsre/fuz006