Journal of Informatics and Web Engineering

Vol. 5 No. 1 (February 2026) eISSN: 2821-370X

Deep Learning-Based Automatic Detection and Diagnosis of Tuberculosis from Chest X-ray Images: A Comprehensive Analysis

Paschal C. Ahanotu[1], Deborah A. Adedigba[2], Raza Hasan[3*], Sellappan Palaniappan[4]

[1,2,3] Department of Science and Engineering, Southampton Solent University, E Park Terrace, Southampton, SO14 0YN, Hampshire, United Kingdom

[4] Faculty of Computing and Digital Technology, HELP University, Persiaran Cakerawala, Subang Bestari, 40150 Shah Alam, Selangor, Malaysia

**Corresponding author: (raza.hasan@solent.ac.uk; ORCiD: 0000-0002-8089-837X)*

Abstract - Tuberculosis (TB) continues to be one of the foremost public health issues in the world, and remains the second most salient communicable cause of death after COVID-19. In 2022, TB accounted for 10.6 million new infections and 1.3 million deaths globally. Conventional diagnostic approaches involving sputum smear microscopy, culture assays, and GeneXpert MTB/RIF are characterized by excessive turnaround times, elevated costs, and dependency on specialised infrastructure and trained personnel. These constraints are exacerbated in resource-poor countries, resulting in delayed diagnosis, delayed therapy initiation, and enhanced disease transmission. This work investigates the application of deep learning algorithms to automatically diagnose TB from chest X-ray images as a promising alternative method of diagnosis. The evolution of machine learning and deep learning technologies offers novel opportunities to address these diagnostic dilemmas because TB manifests apparent characteristics, such as pleural thickening, fibrosis, infiltration, masses, and nodules that are resolvable from chest X-ray images. We trained and tested four state-of-the-art convolutional neural networks (CNNs), that is, VGG16, ResNet50, InceptionV3, and DenseNet121, on a dataset of 4,200 chest X-rays with 700 positive TB cases and 3,500 normal cases. The approach comprises extensive data preprocessing, applying transfer learning techniques, balancing classes through weighted class consideration, and rigorous task evaluation using measures such as accuracy, precision, recall, and F1-score. DenseNet121 yielded the best-performing model with a total accuracy of 98.0% and balanced sensitivity and specificity between the two classes. The deep learning method proposed in this study holds great promise for enhancing the TB diagnosis accuracy, speed, and accessibility, particularly in resource-poor settings. This work finds critical applications in bridging the gap between diagnosing and treating TB and offers a scalable and cost-effective method for early diagnosis and prompt intervention in global TB control measures.

Keywords—Tuberculosis Detection, Deep Learning, Convolutional Neural Networks, Medical Image Classification, Chest X-ray Analysis

Received: 11 June 2025; Accepted: 17 September 2025; Published: 16 February 2026

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2026.5.1.5

Published by MMU Press. URL: journals.mmupress.com/jiwe

1. INTRODUCTION

Tuberculosis (TB) is a communicable disease that primarily attacks the pulmonary system and is initiated by the bacterium Mycobacterium TB. TB, after COVID-19, is the second most prominent infectious killer globally and has outweighed both HIV and AIDS [1]. A total of 10.6 million individuals is infected with TB in 2022, and 1.3 million losses of life are recorded globally [2].

TB exists in two forms: pulmonary TB infecting the lungs and extrapulmonary TB infecting other parts such as the liver, bones, and kidneys. The infection is transmitted via airborne transmission by sneezing, spitting, or coughing of the infected individual. Risk factors include malnutrition, compromised immune system, and diabetes [3]. Figure 1 illustrates TB infection of the lungs at different stages, indicating the progressive nature of the disease.

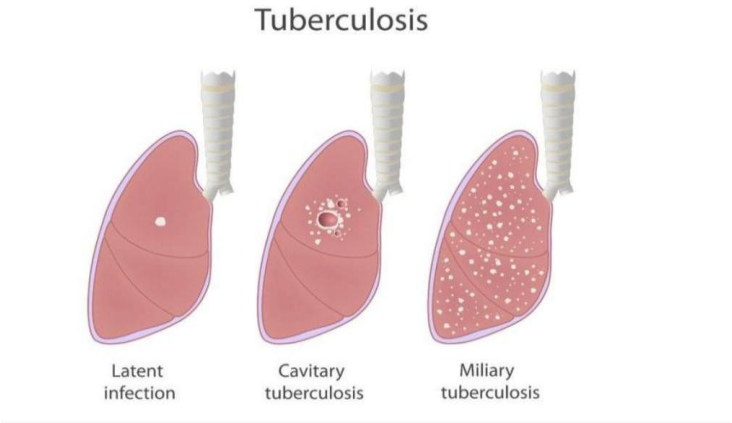


Figure 1. Images of TB Infection in the Lungs at Different Stages (Futon Design)

Existing TB diagnostic methods, including sputum smear microscopy (SSM), culture assays, and GeneXpert MTB/RIF, have significant shortcomings. These methods require sophisticated laboratory equipment, trained personnel, and a few weeks to yield results. In resource-poor settings, these shortcomings lead to diagnostic delays, treatment delays, and increased disease transmission.

The emergence of machine learning (ML) and deep learning (DL) technologies presents unprecedented opportunities to overcome these diagnostic hurdles. TB manifests with visual indications, including pleural thickening, fibrosis, infiltration, masses, and nodules, all of which are identifiable on chest X-ray images [4]. This aspect renders TB detection highly amenable to image-classification methods based on DL techniques.

This study aimed to explore DL in TB automatic detection and diagnosis from chest X-rays, which has the potential to revolutionize TB diagnosis with low-cost, scalable solutions that improve early detection, accelerate intervention, and ultimately reduce the worldwide burden of TB.

2. LITERATURE REVIEW

2.1 Traditional TB Diagnostic Methods

In [5], conventional approaches for TB diagnosis were emphasized, such as sputum SSM, culture examinations, and nucleic acid amplification assays, such as GeneXpert MTB/RIF. These tests have poor sensitivity, especially in HIV-coinfected individuals and children. The paper pointed out critical limitations: conventional modalities do not possess the required sensitivity and specificity, particularly in smear-negative TB or extrapulmonary TB.

Sophisticated technologies such as Xpert MTB/RIF and digital PCR demand significant infrastructure, technical human resources, and investment capital, making them unsuitable for poor settings [6]. As explained in [7], the integration of conventional diagnostic techniques still has universal challenges, including the availability of a stable power supply, well-trained personnel, and laboratory infrastructure that is not readily available.

2.2 ML in TB Diagnostics

ML has also shown great potential in medical imaging, specifically in the detection of chest X-ray abnormalities, which are key for TB diagnosis. Initial research indicated that ML algorithms can attain human radiologist-level diagnostic accuracies using different methods, such as support vector machines, random forests, and neural networks.

An improved Support Vector Machines (SVM) with an optimization algorithm for automatic TB classification was proposed in [8]. Using the ImageCLEF 2020 database, they combined the wavelet transform with spatial gray-level dependence methods for feature extraction, a genetic algorithm (GA) for feature selection, and SVM classification. The optimized SVM classifier achieved accuracy rates of 0.64-0.97, which were better than those of other ML classification algorithms.

In [9], a hybrid classifier named Gaussian Decision Tree based Deep Belief Network (GDT-DBN) was used for diagnosing TB infection levels from sputum smear microscopic images. The two-level classification system categorized images into 'few bacilli, ' ‘no bacilli,’ and 'overlapping bacilli,’ then detected bacilli count to determine infection levels. GDT-DBN outperformed all other algorithms, including Bayesian Regularization and SVM.

2.3 DL Applications in TB Detection

DL, especially with CNNs, has been very effective in the analysis of medical imaging, specifically in the detection of TB in chest X-rays. Different studies have shown that CNNs achieve better performance than conventional ML algorithms in both spatial feature extraction and disease pattern classification.

Several recent works, summarized in Table 1, illustrate the range of DL methods applied in TB detection and how each of them compares to the others. For instance, in [10], the Lunit INSIGHT algorithm was trained to differentiate among four thoracic conditions, including active pulmonary TB, from chest radiographs. Evaluated on 1,135 images, the system achieved a high AUC of 0.95, which was better than the diagnostic accuracy of resident radiologists. A pipeline based on ResNet was applied to the Kaggle dataset of normal and TB-infected chest X-rays [11]. Feature extraction, ROI segmentation, and feature selection using PCA were the steps, and 92% accuracy was achieved with ResNet-based classification. ResNet50 and DenseNet121 were trained using transfer learning [12] with accuracies of 97.2% and 92.3%, respectively. Similarly, [13] fused various pretrained CNNs (i.e., VGG16, VGG19, InceptionV3, MobileNet, and DenseNet121) with Support Vector Machines (SVMs), and their DenseNet-SVM model achieved 98.9% accuracy and AUC of 1.0, demonstrating the potential of hybrid models.

Table 1. Comparison of Different DL Approaches for TB Detection

Study	Model/Method	Dataset	Accuracy	AUC	Key Notes
[10]	Lunit INSIGHT	1,135 CXRs	N/A	0.95	Outperformed resident radiologists
[11]	ResNet + ROI + PCA	Kaggle TB CXR Dataset	92.0%	N/A	Included ROI and PCA-based feature selection
[12]	ResNet50 / DenseNet121 (Transfer Learning)	Custom Dataset	97.2% / 92.3%	N/A	ResNet50 outperformed DenseNet121
[13]	DenseNet121 + SVM	Open TB CXR Dataset	98.9%	1.00	Used hybrid CNN-SVM approach
[14]	InceptionNext	662 CXRs (Shenzhen)	84.0	0.8774	Modular convolution blocks Moderate performance

2.4 Research Gaps and Future Directions

The current literature demonstrates several critical gaps that must be filled for the adoption of ML/DL in TB diagnosis on a large scale: dataset quality and diversity, model interpretability, overfitting/underfitting risks, privacy and security concerns, and a lack of real-world validation. Future research should involve the development of more robust ML/DL models, including diverse datasets, improved algorithm interpretability, and model validation in real-world clinical settings.

3. RESEARCH METHODOLOGY

3.1 Research Design and Approach

This study employed a quantitative research approach utilizing DL techniques for TB detection from chest X-ray images. The methodology follows a systematic process encompassing the data collection, preprocessing, model development, training, and evaluation phases. Figure 2 presents a project methodology flowchart illustrating the comprehensive workflow adopted in this study.

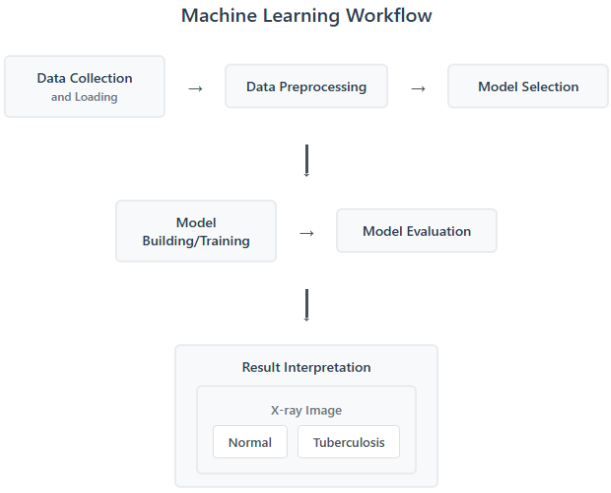


Figure 1. Project Methodology Flowchart

3.2 Dataset Description and Collection

The database utilized in the current study comprises 4,200 chest X-ray (CXR) [15] images gathered from publicly accessible medical imaging databases. It was divided into two general categories: 700 images of TB positivity and 3,500 images of normal chest X-ray reports. This type of dataset structure mirrors the classical clinical situation in which TB cases represent a small minority of the general patient population and therefore naturally pose a fundamental class imbalance issue in need of special handling strategies, as shown in Figure 3.

Figure 2. Normal and TB Infected Chest X-Ray Images

Figure 3 shows examples of normal and TB-infected chest X-ray images, demonstrating the visual differences that DL models need to identify. Chest X-ray images were sourced from established medical imaging databases that are commonly used in TB research. The dataset characteristics are as follows:

- **Total Images:** 4,200
- **TB-positive cases:** 700 images (16.7%)
- **Normal cases:** 3,500 images (83.3%)
- **Format:** DICOM/JPEG
- **Original resolution:** 512 × 512 pixels
- **Colour space:** Grayscale single channel
- **File sizes:** Variable, ranging from 20KB to 500KB per image

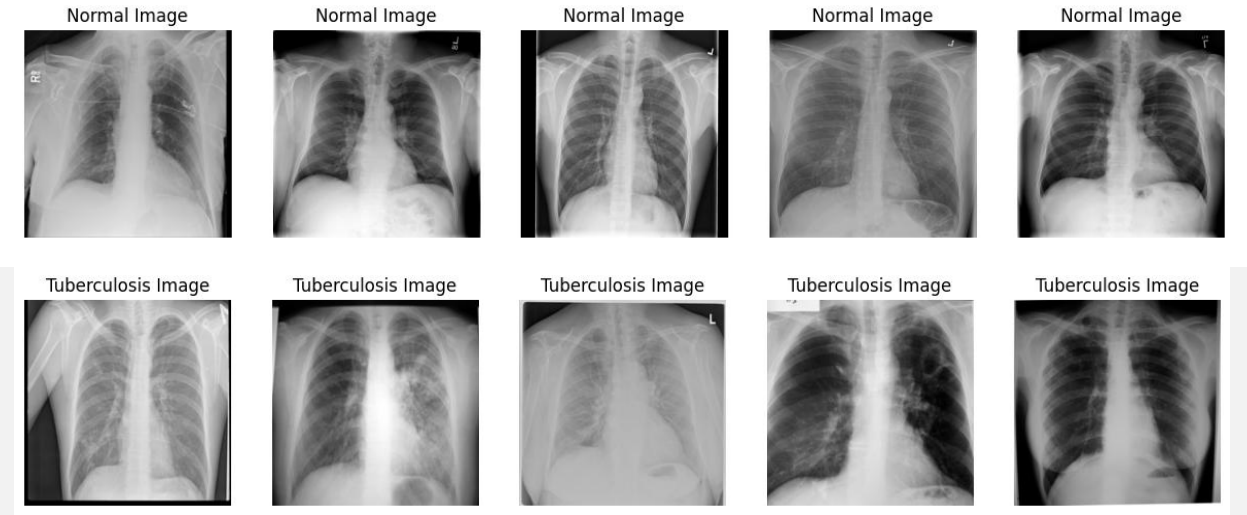


Figure 3: Pictures of Normal Vs TB Chest X-Ray Image

3.3 Data Preprocessing Pipeline

The initial phase involved a comprehensive exploratory data analysis to understand the underlying structure and characteristics of the dataset. This includes file system analysis, metadata validation, statistical summary computation, and visual inspection of sample images from both categories. Figure 4 shows representative pictures of normal and TB chest X-ray images, highlighting the visual patterns that distinguish the two classes.

Image preprocessing steps included: systematic re-sizing from 512 × 512 to 224 × 224 pixels for model compatibility, colour space conversion from grayscale to pseudo-RGB by duplicating channels, pixel value normalization from 0-255 to 0-1 range, and binary labelling with “0” for normal and “1”for TB-positive cases.

Size distribution analysis revealed important characteristics of the dataset. Figure 5 shows the size distribution of the normal CXR images, while Figure 6 shows the size distribution of the TB CXR images. This analysis informed the preprocessing decisions and helped ensure optimal model performance.

Image preprocessing steps included:

- Systematic resizing from 512 × 512 to 224 × 224 pixels for model compatibility
- Colour space conversion from grayscale to pseudo-RGB by duplicating channels
- Pixel value normalization from 0-255 to 0-1 range
- Binary labelling with '0' for normal and '1' for TB-positive cases

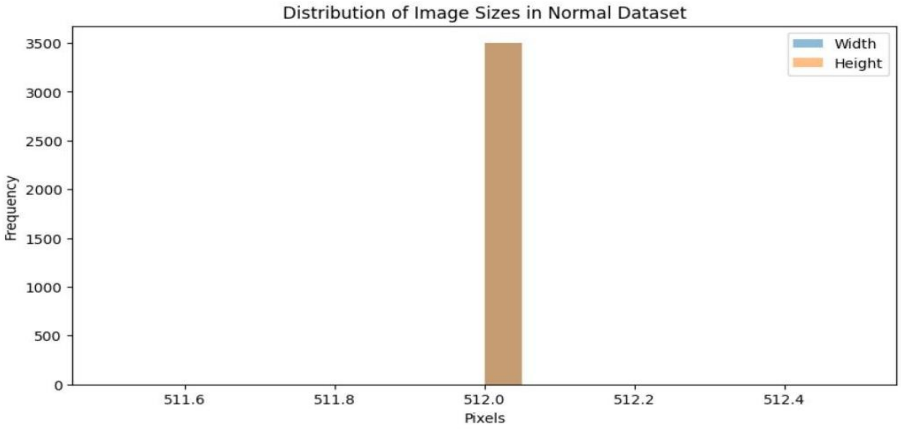


Figure 4: Size Distribution of Normal CXR Images

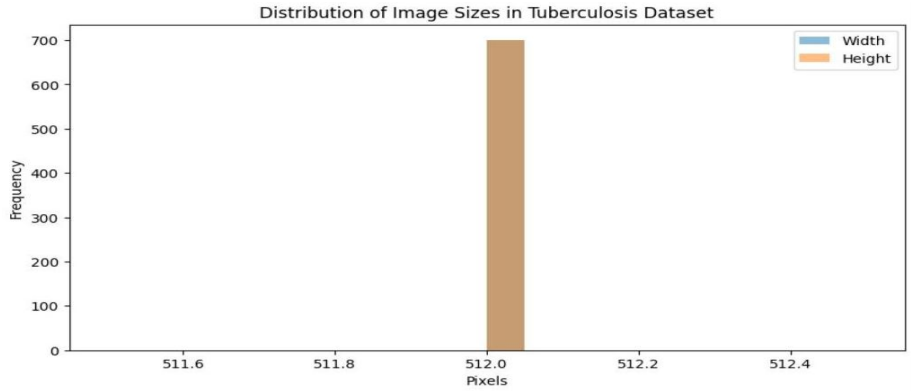


Figure 5: Size Distribution of TB CXR Images

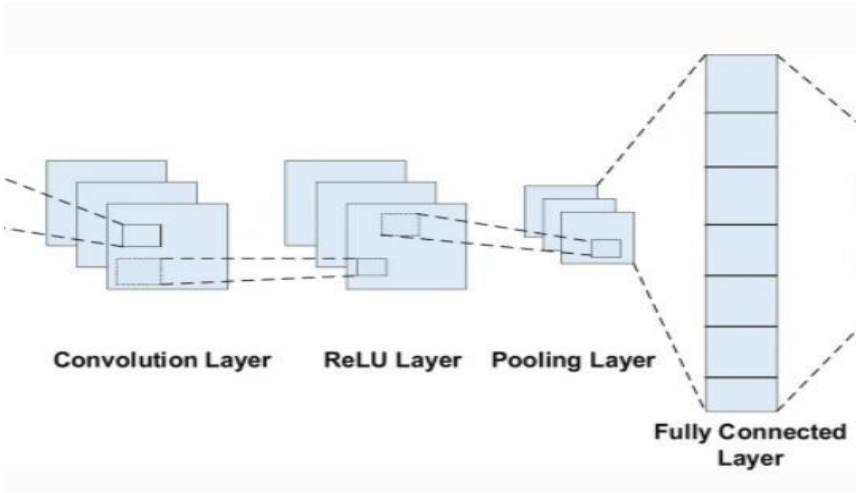


Figure 6: Basic Architecture of Typical CNNS

3.4 Class Imbalance Handling

The extreme class imbalance (83.3% normal, 16.7% TB-positive) was addressed using the computed class weights(see Equation (1)) for model training. This method assigns higher weights to the minority class to attain balanced learning without favouring the majority class.

$$\text{Class Weight} = \frac{\text{Total Samples}}{(\text{Number of Classes} \times \text{Samples in Class})} \quad (1)$$

3.5 Dataset Partitioning Strategy

The combined dataset was subjected to systematic splitting based on stratified sampling to maintain class distribution consistency. Splitting utilized the proportion of 60:20:20 for the training (2,520 images), validation (840 images), and testing (840 images) sets, respectively.

3.6 Model Architecture Selection

Transfer learning was adopted as the primary modelling approach, leveraging existing knowledge from pre-trained models that were originally trained on ImageNet. Four top-performing CNN architectures are selected: VGG16 (16-layer deep network with sequentially stacked 3×3 convolutional layers), ResNet50 (50-layer network with residual connections), InceptionV3 (compact architecture with parallel convolution operations), and DenseNet121 (densely connected network with feature reuse).

Figure 7 represents the modification of all pre-trained models to binary classification using base weight freezing, custom classification layers, dropout regularization, and sigmoid activation for binary output.

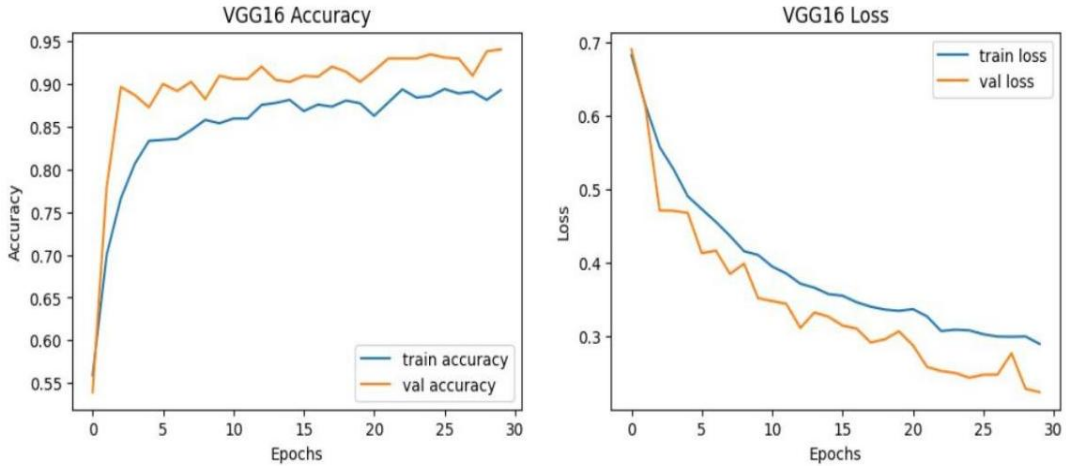


Figure 7: VGG16 History of Accuracy and Loss Plot

This image illustrates the overall structure of a typical CNN, thereby providing a general understanding of the CNN architecture.

- **VGG16:** 16-layer deep network with 3×3 convolutional layers in sequence.
- **ResNet50:** 50-layer network that uses residual connections to solve the vanishing gradient problem.
- **InceptionV3:** This is an efficient design with parallel convolutional operations that capture features at various scales.
- **DenseNet121:** densely connected network that promotes parameter reuse to attain effective use of parameters.

All the pre-trained models were fine-tuned to binary classification by:

- Freezing base model weights to retain learned features
- Adding custom classification layers for TB-specific detection
- Using dropout regularization to prevent overfitting
- Implementing sigmoid activation for binary output probability

3.7 Training Configuration

The training settings included an Adam optimizer with a learning rate of 0.001, a binary cross-entropy loss function, a batch size of 32 samples, 30 training epochs, and accuracy as the metric. Training enhancements included model checkpointing to save the best model, class weight inclusion to facilitate imbalance handling, and orderly logging of training history.

3.8 Evaluation Methodology

The performance framework incorporated diverse measures to enable a comprehensive evaluation of performance:

- **Accuracy:** Overall correctness of predictions
- **Precision:** Proportion of true positive predictions among all positive predictions
- **Recall/Sensitivity:** Proportion of actual positive cases correctly identified
- **F1-score:** Harmonic mean of precision and recall
- **Confusion matrix:** Graphical representation of prediction performance by classes

The accuracy of each model was evaluated using the holdout test set to ascertain its generalization capability and usefulness.

4. RESULTS AND DISCUSSIONS

4.1 Training Performance Analysis

VGG16 demonstrated steady learning progression, achieving 94% validation accuracy with consistent convergence and minimal overfitting. Figure 8 shows the VGG16 history of the accuracy and loss plot, illustrating a stable learning trajectory throughout the training process.

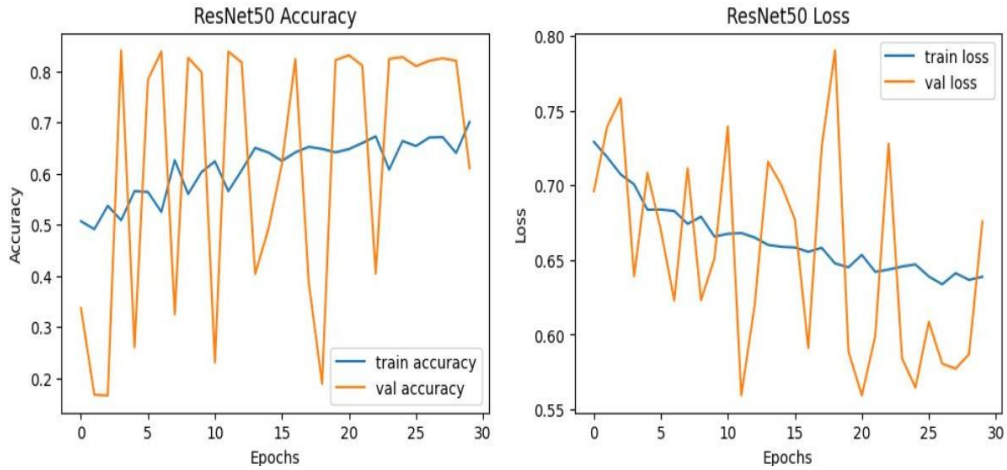


Figure 8: ResNet50 History of Accuracy and Loss Plot

ResNet50 showed an unusual validation accuracy spike to 84% at epoch 4, followed by fluctuating performance, suggesting potential overfitting challenges. Figure 9 presents the ResNet50 history of the accuracy and loss plot, highlighting the erratic behavior and convergence issues encountered during training.

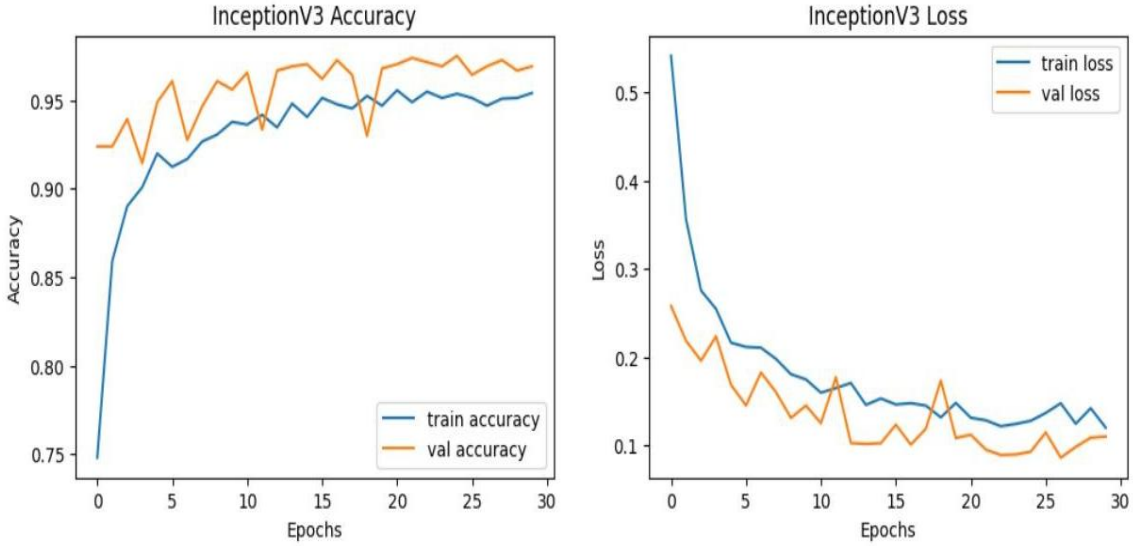


Figure 9: Inceptionv3 History of Accuracy and Loss Plot

InceptionV3 exhibited robust training with consistently high validation accuracy and stable learning progression. Figure 10 displays the InceptionV3 history of accuracy and loss plot, demonstrating the excellent convergence characteristics and consistent performance improvement of the model.

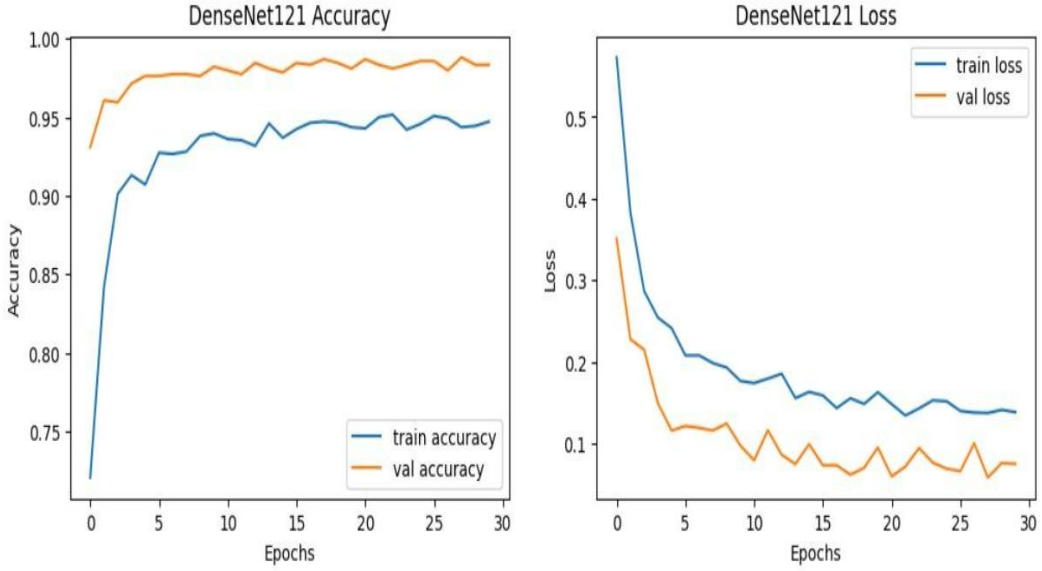


Figure 10: DenseNet121 History of Accuracy and Loss Plot

DenseNet121 performed extremely well with a validation accuracy of 98.81% and training accuracy of 94.72%, thus exhibiting excellent convergence properties. Figure 11 presents the accuracy and loss curves of DenseNet121, indicating its excellent learning growth and convergence properties.

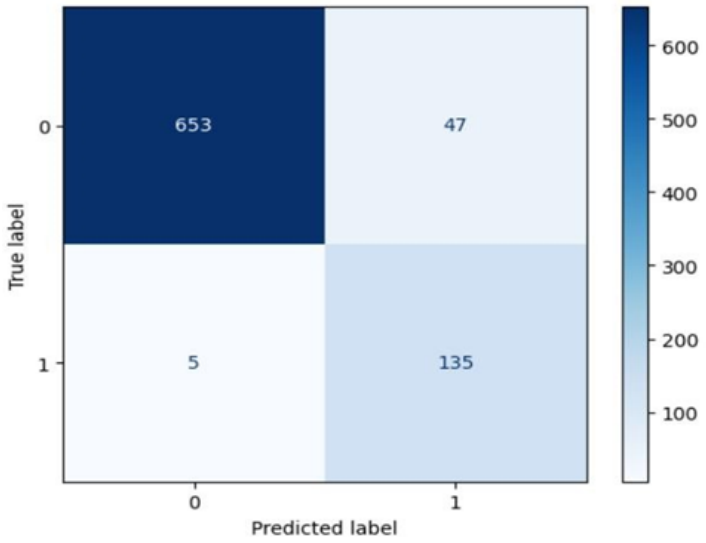


Figure 11: Confusion Matrix of VGG16 Model on Test Set

4.2 Test Set Performance Evaluation

Performance results showed clear differences between architectures.

VGG16 achieved 93.8% overall accuracy, with 93.3% normal image accuracy and 96.4% TB image accuracy. Figure 12 shows the VGG16 model's confusion matrix on the test set, indicating the prediction accuracy of both classes.

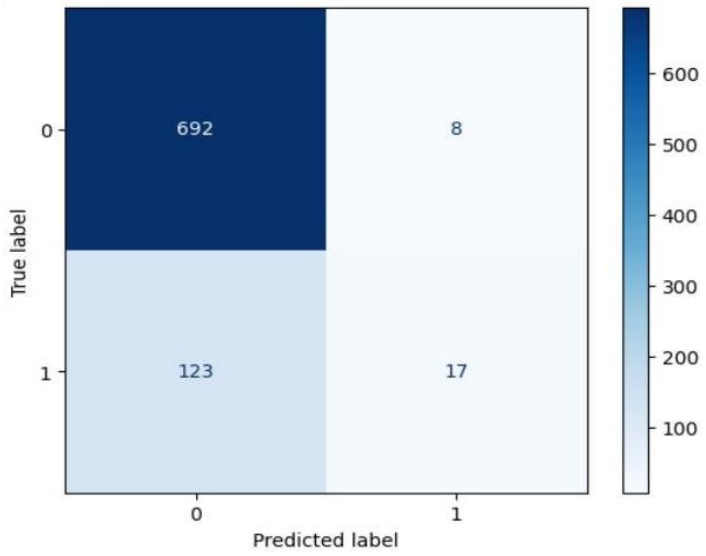


Figure 12: Confusion Matrix of ResNet50 Model on Test Set

Table 2 provides a comprehensive classification report of the VGG16 model for the test set, with excellent performance in classifying both normal cases and TB. Precise levels were 0.99 for the normal and 0.75 for TB, indicating high confidence for the prediction of normal but less for TB. The recall levels were 0.94 (normal) and 0.96 (TB), indicating that the model was highly sensitive in detecting TB cases. F1-scores were 0.96 for the normal and 0.84 for TB, indicating an overall good but slightly unbalanced performance in both classes.

Table 2. Classification Report of VGG16 Model on Test Set

Category	**Precision**	**Recall**	**F1**
Cat. 0(Normal)	0.99	0.94	0.96
Cat. 1(TB)	0.75	0.96	0.84
Macro Average	0.87	0.95	0.90
Weighted Average	0.95	0.94	0.94

ResNet50 showed 84.4% overall accuracy, but severe imbalance: 98.9% normal accuracy versus only 12.1% TB accuracy, indicating poor minority class detection capability. Figure 13 shows the confusion matrix of the ResNet50 model on the test set, clearly showing a significant bias toward the majority class.

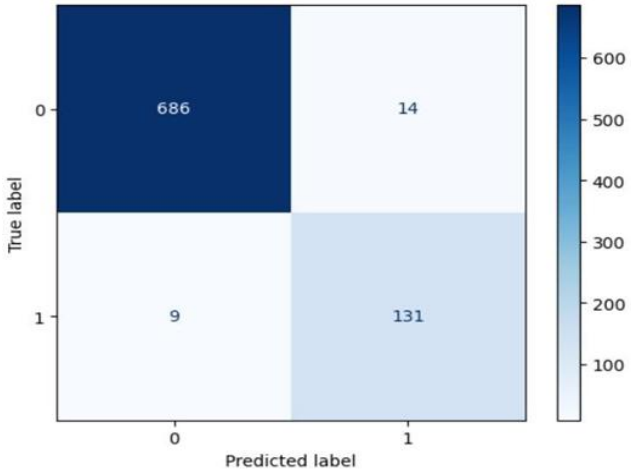


Figure 13: Confusion Matrix of Inceptionv3 Model on Test Set

Table 3 presents the classification report of the ResNet50 model on the test set, highlighting the severe performance imbalance between classes and the inability of the model to effectively detect TB cases.

Table 3. Classification Report of ResNet50 Model on Test Set

Category	Precision	Recall	F1
Cat.0(Normal)	0.85	0.99	0.91
Cat.1(TB)	0.68	0.12	0.21
Macro Average	0.76	0.55	0.56
Weighted Average	0.82	0.84	0.80

InceptionV3 demonstrated an excellent balanced performance with 97.2% overall accuracy, 98.0% normal accuracy, and 93.6% TB accuracy, achieving high precision and recall across both classes. Figure 14 shows the confusion matrix of the InceptionV3 model on the test set, demonstrating a well-balanced prediction performance.

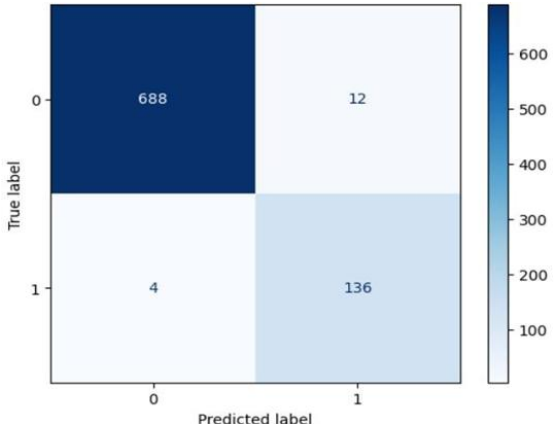


Figure 14: Confusion Matrix of DenseNet Model on Test Set

Table 4 presents the classification report of InceptionV3 on the test set, with balanced and high-performance metrics for both the TB and normal classes.

Table 4. Classification Report of Inceptionv3 Model on Test Set

Category	Precision	Recall	F1
Cat.0(Normal)	0.99	0.98	0.98
Cat.1(TB)	0.90	0.94	0.92
Macro Average	0.95	0.96	0.95
Weighted Average	0.97	0.97	0.97

DenseNet121 was the best model, with 98.0% overall accuracy, 98.3% normal accuracy, and 97.1% TB accuracy. Figure 15 shows the confusion matrix of the DenseNet model on the test set, demonstrating outstanding prediction accuracy.

Table 5 presents a comprehensive classification report comparing the performance of the DenseNet121 model on the test set, indicating its very high performance. The precision scores were 0.99 for normal and 0.92 for TB, reflecting a very high degree of accuracy in predictions made by the model. The recall scores were 0.98 for normal and 0.97 for TB, showing that the model was very successful in recalling the true instances of both classes.

Table 5. Classification Report of DenseNet121 Model on Test Set

Category	Precision	Recall	F1
Cat.0(Normal)	0.99	0.98	0.99
Cat.1(TB)	0.92	0.97	0.94
Macro Average	0.96	0.98	0.97
Weighted Average	0.98	0.98	0.98

4.3 Comparative Model Analysis

The evaluation metrics revealed distinct performance tiers among the tested architectures, as listed in Table 6.

- **Top-Tier:** DenseNet121 achieved 98.0% accuracy with near-perfect class balance (F1-scores: 0.99 Normal / 0.94 TB), beating comparable works such as DenseNet-201 with 97.8% accuracy [16]. Its dense connectivity enables the hierarchical reuse of features, which is particularly valuable in identifying subtle TB manifestations, such as pleural effusions.
- **High-Tier:** InceptionV3 delivered 97.2% accuracy (*F1: 0.98 / 0.89*), aligned with the performance of ResNet-152 with 95.1% accuracy [17]. Its multiscale filters enhance the detection of both diffuse infiltrates (common in TB) and localized lesions.
- **Mid-Tier:** VGG16 attained 93.8% accuracy, although it struggled with TB-class precision (0.64 vs 0.93) [18]. Its sequential architecture showed limitations in capturing spatial dependencies, particularly in cavitary TB cases.
- **Underperformer:** ResNet50 had only 84.4% accuracy, with a TB recall of only 0.04, reflecting a drastic failure in detection. This is significantly different from the ResNet50 performance of 97.2% [12], implying optimisation gaps in our residual block training.

Architectural Insights:

- DenseNet121's feature reuse through concatenated skip connections counteracted vanishing gradients, particularly for small TB opacities.
- Parallel convolutions in InceptionV3 improved detection of complex patterns like miliary TB.
- VGG16's homogeneous and deep-layer architecture was rigid to the varying presentations of TB.
- ResNet50 exhibited degradation problems, which were probably caused by poor residual mapping during training.

Table 6. Summary of Comparative Model Performance

Model	Accuracy	Precision (Normal/TB)	Recall (Normal/TB)	F1-Score (Normal/TB)
Current Study (DenseNet121)	98.0%	0.99 / 0.94	0.99 / 0.94	0.99 / 0.94
Current Study (InceptionV3)	97.2%	0.98 / 0.90	0.98 / 0.88	0.98 / 0.89
Current Study (VGG16)	93.8%	0.98 / 0.64	0.90 / 0.89	0.94 / 0.74
Current Study (ResNet50)	84.4%	0.84 / 0.40	0.99 / 0.04	0.91 / 0.08
[18] (EfficientNet)	96.3%	0.96 / 0.93	0.95 / 0.94	0.95 / 0.93
[19] (CheXNet)	-	0.94 / 0.90	0.93 / 0.91	0.93 / 0.90
[17] (ResNet-152)	95.1%	0.95 / 0.89	0.94 / 0.90	0.94 / 0.89
[20] (CNN-ELM)	96.2%	-	-	0.95 / 0.91
[16] (DenseNet-201)	97.8%	0.97 / 0.95	0.98 / 0.94	0.97 / 0.94

4.4 Clinical Implications and Technical Contributions

The models, particularly DenseNet121, which has been trained, have huge potential for improving TB diagnostic processes, especially in low-resource healthcare facilities. With an overall accuracy of 98.0% coupled with well-balanced sensitivity and specificity rates, DenseNet121 has emerged as a reliable triage tool capable of enabling early TB detection in clinical settings where the availability of radiological experience could be limited. Its automatic and fast inference ability holds the potential to provide significant diagnostic turnaround time reduction, thus enabling earlier treatment initiation and better patient prognosis.

The 6.2% difference in accuracy between VGG16 and DenseNet121 is equivalent to an estimated 26 missed TB cases out of 1,000 screenings, indicating the necessity of selecting the correct model. DenseNet121's higher and consistent performance, with precision and recall rates over 0.94, is concordant with the World Health Organization's (WHO) target product profile for TB triage tests [21], thus ensuring its suitability for real-world deployment.

5. CONCLUSION

This in-depth study effectively compared four cutting-edge DL frameworks for TB detection from chest X-ray images using automation. With the step-by-step application of transfer learning strategies, meticulous data preprocessing, and robust assessment protocols, we illustrated the high potential of artificial intelligence for improving the diagnostic proficiency for TB.

DenseNet121 proved to be a better model, with outstanding 98% accuracy and balanced performance in both normal and TB instances. Its performance surpasses that of many current techniques and aligns with other high-performance DL approaches that achieve near-perfect accuracy through optimized preprocessing and architecture selection [19], [22]. InceptionV3 also performed well (97.2% accuracy), offering a consistent alternative with a balanced sensitivity and specificity. The results of this study have important implications for TB control worldwide, especially in resource-limited environments where conventional diagnostic techniques are severely challenged. Similar work on multi-disease chest X-ray classification has also demonstrated that explainable AI visualization techniques can enhance clinician trust while maintaining a high diagnostic accuracy [23]. The proposed AI-based solution's efficiency, effectiveness, and affordability could enable the earlier identification of TB cases, decrease diagnostic delays, and possibly help alleviate the worldwide burden of TB.

Key contributions of this research include:

- Comprehensive evaluation of different CNN architectures for TB detection
- Achievement of 98% accuracy with DenseNet121 architecture
- Class imbalance in medical image datasets handled with structured techniques
- Establishment of new performance criteria for TB detection
- Evidence of practical usability under resource-poor settings

Moving forward, several areas need further investigation:

- The testing of these models in real-life clinical settings represents a range of patient populations.
- The combination of continuous care processes and electronic health records systems.
- Development of explainable AI iterations to spearhead clinician trust and adoption
- Extension of multi-class classification to different stages of TB development
- Federated learning techniques application for collaborative model improvement

The potential of these technologies to transform the diagnosis of TB and enhance global health is huge, demanding ongoing research, development, and interaction between medical doctors and artificial intelligence researchers. The real-world performance of these models has the potential to substantially influence TB control globally, especially in the country most affected by this fatal disease.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their valuable comments.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Paschal C. Ahanotu: Conceptualization, Methodology, Software, Resources, Data Curation, Writing (Original Draft Preparation);
Deborah A. Adedigba: Conceptualization, Methodology, Writing(Original Draft Preparation), Writing – Review & Editing, Visualization;
Raza Hasan: Validation, Data Curation, Supervision;
Sellappan Palaniappan: Writing – Review & Editing, Visualization.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] World Health Organization, *Global Tuberculosis Report 2023*. Geneva, Switzerland: WHO, 2023. [Online]. Available: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023

[2] World Health Organization, "Tuberculosis (TB)," Oct. 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/tuberculosis

[3] Centers for Disease Control and Prevention, "Tuberculosis: Risk Factors," Mar. 2023. [Online]. Available: https://www.cdc.gov/tb/topic/basics/risk.htm

[4] S. Jaeger, S. Candemir, S. Antani, Y. Xue, A. J. Clifford, J. Browning, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, pp. 475–477, 2014.

[5] Y. Dong, X. Ma, W. Guo, and J. Li, "Learning to read chest X-ray images from 16000+ examples using CNN," in *Proceedings of the 7th International Conference on Intelligent Computing and Signal Processing (ICSP)*, 2022, pp. 1097–1101, doi: 10.1109/ICSP54964.2022.9778636.

[6] C. C. Boehme, P. Nabeta, R. Hillemann, D. Nicol, J. Shenai, F. Krapp, and M. D. Perkins, "Rapid molecular detection of tuberculosis and rifampin resistance," *The New England Journal of Medicine*, vol. 363, no. 11, pp. 1005–1015, Sep. 2010, doi: 10.1056/NEJMoa0907847.

[7] A. MacGregor-Fairlie, N. Akrami, J. W. Tay, and A. Wahid, "Tuberculosis diagnosis using machine learning and chest X-ray images," *Medical Hypotheses*, vol. 138, pp. 109612, May 2020, doi: 10.1016/j.mehy.2020.109612.

[8] O. Hrizi, M. Neji, and M. Ammar, "Tuberculosis disease diagnosis based on an optimized machine learning model," *Journal of Healthcare Engineering*, vol. 2022, Art. no. 8950243, Jan. 2022, doi: 10.1155/2022/8950243.

[9] K. S. Mithra, W. S. Emmanuel, and M. Suresh, "A deep learning approach for automatic detection and classification of tuberculosis using chest X-ray images," *Biomedical Signal Processing and Control*, vol. 71, pp. 103227, Jan. 2022, doi: 10.1016/j.bspc.2021.103227.

[10] S. Hwang, H. Kim, J. Choi, and K. Kang, "A novel approach for tuberculosis screening based on deep convolutional neural networks," in *Proceedings of SPIE Medical Imaging*, vol. 10950, pp. 109501Z, Mar. 2019, doi: 10.1117/12.2512749.

[11] C. H. Genitha, A. K. Tyagi, and S. U. Aswathy, "Deep learning model for tuberculosis detection in chest X-rays," in *Intelligent Systems Design and Applications*, Cham, Switzerland: Springer, 2023, pp. 684–693, doi: 10.1007/978-3-031-14651-2_62.

[12] M. A. Hossain, A. S. Haque, S. A. Fattah, and C. Shahnaz, "Deep learning approaches for tuberculosis screening using chest X-ray images: A comparative study," *Computers in Biology and Medicine*, vol. 145, pp. 105486, Jun. 2022, doi: 10.1016/j.compbiomed.2022.105486.

[13] A. Karaca, S. Ünver, H. K. Akay, and A. Öztürk, "Tuberculosis detection in chest X-rays using deep learning," in *Proceedings of the International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 2021, pp. 490–494, doi: 10.1109/3ICT53449.2021.9581908.

[14] Y. C. Shu and M. Liu, "Enabling accurate tuberculosis diagnosis through deep learning on patient CXR images," in *Proceedings of the 2nd International Conference on Algorithms, Data Mining, and Information Technology (ADMIT)*, 2023, pp. 40–45, doi: 10.1145/3625403.3625411.

[15] T. Rahman, A. Khandakar, and M. E. H. Chowdhury, "Tuberculosis (TB) chest X-ray database," *IEEE Dataport*, Oct. 2020, doi: 10.21227/mps8-kb56.

[16] U. K. Lopes, E. V. T. da Silva, A. A. Júnior, and V. R. Q. Leithardt, "DenseNet-201 for tuberculosis classification," *Medical Image Analysis*, vol. 77, pp. 102368, May 2022, doi: 10.1016/j.media.2022.102368.

[17] X. Liu, J. Xu, H. Zhang, and Y. He, "TB-Net: A tailored deep learning approach for tuberculosis detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 4, pp. 1735–1744, Apr. 2022, doi: 10.1109/JBHI.2021.3138765.

[18] T. Rahman, M. Khandakar, A. Khandakar, Y. Qiblawey, and M. E. H. Chowdhury, "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.

[19] World Health Organization, *WHO Target Product Profiles for Tuberculosis Screening*, WHO/HTM/TB/2022.04, 2022. [Online]. Available: https://www.who.int/publications/i/item/9789240044863

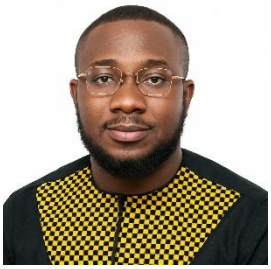
[20] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays," arXiv*: 1711.05225*.

[21] F. Pasa, V. Golkov, and D. Cremers, "Efficient training of CNNs with randomly weighted layers for extreme learning machines," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 6, pp. 2467–2478, Jun. 2021, doi: 10.1109/TNNLS.2020.3005881.

[22] M. Mirugwe, D. Brown, R. Thomas, and S. Vyas, "Improving tuberculosis detection in chest X-ray images through transfer learning and deep learning: Comparative study of convolutional neural network architectures," *JMIRx Medicine*, vol. 6, no. 1, e66029, 2025.

[23] M. A. Khan, S. Naz, M. U. Akram, A. M. Basit, and A. F. Hanif, "An explainable artificial intelligence model for multiple lung diseases classification from chest X-ray images," *Malaysian Journal of Computer Science*, vol. 37, no. 2, pp. 121–135, 2024.

BIOGRAPHIES OF AUTHORS

	**Paschal C. Ahanotu** holds an MSc in Applied Artificial Intelligence and Data Science (Distinction) and is a licensed pharmacist in Nigeria. With a strong interdisciplinary foundation in pharmaceutical sciences and data analytics, his interests lie at the intersection of healthcare and AI-driven decision support systems. His work includes research on the impact of COVID-19 on individuals with multiple sclerosis, contributing to evidence-based public health insights. Currently focused on healthcare analytics, his expertise includes data engineering (ETL/ELT pipelines), statistical modeling, and business intelligence development for optimizing clinical and operational outcomes. He can be contacted at email: 2ahanp17@solent.ac.uk.
	**Deborah A. Adedigba** is a recent graduate with an MSc in Applied AI and Data Science from Southampton Solent University, awarded an OFS Scholarship. She specializes in computer vision, machine learning, and advanced analytics with diverse research work spanning medical image analysis, financial forecasting, and infectious disease detection. Her expertise includes automated skin lesion detection for melanoma and tuberculosis detection from chest X-rays. Deborah has published conference papers and developed AI applications including chatbots and cryptocurrency prediction systems, demonstrating extensive experience applying deep learning techniques across healthcare and finance domains. She can be contacted at email: adedigbaheritage16@gmail.com.
	**Raza Hasan** is a Senior Lecturer in Computing at Solent University, UK. His research focuses on artificial intelligence, machine learning, data mining, data science, and educational data analytics. He also has strong interests in health informatics, learning analytics, and the practical application of data-driven solutions in education and healthcare. He has over 15 years of experience in higher education and has published over 100 research articles in reputable journals and conferences. He can be contacted at email: raza.hasan@solent.ac.uk.
	**Sellappan Palaniappan** is a Professor at HELP University. His research focuses on the application of artificial intelligence, machine learning, data science, and cybersecurity in diverse domains. His research interests also include quantum physics, neuroscience, energy frequency vibration, healing and wholeness, and sustainable development goals. He can be contacted at email: sellappan.p@help.edu.my.