A Hybrid VGG-16 and TabNet Model for Interpretable Lung Disease Detection from Chest X-rays in Resource-Constrained Environments

[1]Abraham Eseoghene Evwiekpaefe, [2]Fiyinfoluwa Ajakaiye, [3*]Muhammad Nazeer Musa and [4]Muhammad Musa Isa

[1,2,4]Department of Computer Science, Faculty of Military Science and Interdisciplinary Studies, Nigerian Defence Academy, PMB 2109, Kaduna, Nigeria
[3]Department of Cyber Security, Faculty of Military Science and Interdisciplinary Studies, Nigerian Defence Academy, PMB 2109, Kaduna, Nigeria

email: [1]aeevwiekpaefe@nda.edu.ng, [2]fajakaiye@nda.edu.ng, [3*]muhammadmusa2502@nda.edu.ng, [4]mm.isa@nda.edu.ng

**Corresponding author*

Received: 26 July 2025 | Accepted: 11 November 2025 | Early access: 01 December 2025
https://doi.org/10.33736/jcsi.10348.2026

Abstract - *Accurate diagnosis of lung diseases via chest X-rays remains challenging due to subtle pathological patterns, class imbalance, and the opacity of conventional deep learning models. While convolutional neural networks excel in feature extraction, their "black-box" nature and poor interpretability hinder clinical trust, particularly in resource-constrained settings. To address these limitations, we propose a novel hybrid architecture integrating VGG-16 with TabNet, synergizing hierarchical spatial feature extraction with attention-driven interpretability. The model leverages VGG-16's convolutional layers to capture granular details, while TabNet's sequential attention masks dynamically prioritize discriminative features, quantifying their clinical relevance. Trained on a dataset of 2,590 chest X-rays (COPD, tuberculosis, pneumonia, and normal cases) from Nigerian hospitals, the model achieved state-of-the-art performance with 97% accuracy, surpassing ResNet-50 (95.7%) and standalone VGG-16 (94.7%). Preprocessing, including non-local means denoising and targeted augmentation, mitigates noise and class imbalance, yielding F1-scores exceeding 97% for COPD and pneumonia, with AUC values above 0.98 across all classes. The model's interpretability is validated through attention maps highlighting disease-specific radiological markers, such as hyperinflation in COPD and consolidations in pneumonia, aligning with clinical expertise. Deployed as a real-time Android application optimized for low-end devices, the solution achieves inference in <1 second offline, addressing infrastructural barriers in low-resource regions. The model advances equitable healthcare delivery, demonstrating generalizability across demographic subgroups (accuracy deviation ≤1.2%) and compliance with emerging regulatory standards for trustworthy AI. This innovation establishes a scalable paradigm for interpretable, high-performance lung disease detection, with transformative potential for global health equity.*

Keywords: TabNet, VGG-16, lung disease detection, chest X-ray, deep learning, sequential attention.

1 Introduction

Lung diseases represent a major and escalating global health crisis, driven by a complex interplay of environmental degradation, climate change, shifting lifestyles, and limited access to diagnostic resources, particularly in low- and middle-income countries (LMICs) (Ming et al., 2018; Al Achkar & Chaaban, 2025). Respiratory illnesses, including tuberculosis, pneumonia, and chronic obstructive pulmonary disease (COPD), now constitute the third leading cause of mortality worldwide, with a disproportionate burden falling on resource-constrained settings (Rajagopal et al., 2023). COPD and asthma alone accounted for millions of deaths in recent years, highlighting the urgent need for effective interventions (Bharati et al., 2020). This crisis is particularly acute in sub-Saharan Africa and other LMICs, where populations face the double jeopardy of high exposure to air pollution and prevalent poverty, creating a breeding ground for respiratory diseases (Mondal et al., 2020). The

recent COVID-19 pandemic further underscored the vulnerability of lung health, demonstrating the devastating consequences of respiratory viral infections and exacerbating the existing challenge of pneumonia (Chunli et al., 2020).

Chest X-rays remain a cornerstone of lung disease diagnosis, offering a cost-effective and widely accessible imaging modality for detecting a range of conditions, from pneumonia and tuberculosis to interstitial lung disease and early-stage lung cancer (Zakirov et al., 2015; Rehman et al., 2023). However, the interpretation of chest X-rays presents significant challenges. The complex and overlapping anatomical structures within the images make accurate diagnosis difficult, even for experienced radiologists. Manual interpretation is inherently time-consuming and susceptible to inter-observer variability, potentially leading to diagnostic delays and missed opportunities for timely intervention (Van Ginneken et al., 2009; Gefter et al., 2023).

The advent of deep learning (DL), a subfield of artificial intelligence, has revolutionized medical image analysis, offering the potential to automate feature extraction and improve diagnostic accuracy. Convolutional neural networks (CNNs), such as VGG and ResNet, have demonstrated remarkable performance in detecting pathologies from medical images, including chest X-rays, by learning hierarchical representations of anatomical structures (Irhebor, 2021; Kim et al., 2022; Musa et al., 2025) as they can assist physicians in identifying easily missed suspicious lesions, thereby enhancing detection accuracy (Zakirov et al., 2015; Gefter et al., 2023). Current DL models often suffer from critical limitations that hinder their widespread clinical adoption. Many models lack robust feature engineering at the fully connected layers responsible for final decision-making, and they struggle with the inherent class imbalance commonly found in medical datasets, where certain diseases are significantly more prevalent than others (González et al., 2018). Furthermore, a major drawback of many existing DL models is their "black box" nature. They provide accurate predictions but offer little understanding into the underlying reasoning behind those predictions (Tariq et al., 2019). The scarcity of representative datasets from underrepresented populations, such as those in sub-Saharan Africa, further exacerbates the problem, leading to models that may not generalize well to diverse patient populations (Shakeel et al., 2019).

This study addresses these critical gaps by developing a hybrid deep learning model for the detection of three prevalent lung diseases in Africa: tuberculosis, COPD, and pneumonia. The study leverage dataset of chest X-ray images collected from hospitals across Nigeria, while powerful feature extraction capabilities of VGG-16 (Kieu et al., 2020) with the attention-based feature selection properties of TabNet (Arik & Pfister, 2021) were harnessed. VGG-16 was chosen for its proven efficacy in medical imaging, leveraging its 13 convolutional layers to extract multi-scale features from X-rays (Kieu et al., 2020). TabNet complements this by introducing sparsity-controlled attention mechanisms, enabling feature importance quantification, a critical advancement for clinical trust (Shah et al., 2022). This synergy addresses the opacity of conventional CNNs while maintaining high performance, making the model both accurate and clinically actionable. This hybrid approach aims to not only enhance diagnostic accuracy but also address the critical need for transparency and interpretability in AI-driven medical tools.

2 Literature Review

Recent advancements in deep learning (DL) have demonstrated significant potential for classifying lung diseases; however, substantial challenges remain concerning model generalizability, interpretability, and practical clinical application. This review compiles existing studies that employ deep learning techniques for the detection and classification of lung diseases through medical imaging. It examines various methodologies, architectures, strengths, and datasets utilized in this field while also identifying critical gaps to contextualize the contributions of the current study.

Early studies, such as Ming et al. (2018), demonstrated the effectiveness of DL features from pre-trained models on High-Resolution Computed Tomography (HRCT) images, achieving an accuracy of 100% on binary classification compared to 93.52% with traditional Gray-Level Co-occurrence Matrix (GLCM) features. While such results highlighted the potential of DL, this 100% accuracy was achieved on a specific, homogeneous dataset. More recent benchmarks on more complex HRCT datasets show state-of-the-art (SOTA) performance in the 97-98% range, with a significant research focus shifting to reducing false positives and improving robustness (Jiang et al., 2025; Abe & Nyathi, 2025). Regardless of SOTA in HRCT, these models are not directly applicable to chest X-rays (CXRs), which remain the most common, cost-effective, and accessible imaging modality globally, particularly in resource-limited settings. CXRs present distinct challenges due to lower resolution, higher noise, and greater variability in acquisition quality (Shukla et al., 2024). Similarly, Kim et al. (2022) compared shallow learning (Support Vector Machine) and deep learning (Convolutional Neural Network) for classifying interstitial lung disease patterns in HRCT images from 106 patients, with CNN outperforming SVM by 6–9%, achieving accuracy rates ranging from 81.27% to 95.12% as the number of convolutional layers increased. To substantiate

the claim that deep learning enhances lung disease detection specifically for CXRs, Kieu et al. (2020) conducted a comprehensive review, noting trends such as the prevalence of CNNs and transfer learning. They highlighted critical, persistent challenges, including data imbalance, the management of large noisy image sizes, and the scarcity of datasets. While data augmentation is a common strategy to address imbalance, recent studies affirm that augmentation alone does not solve the fundamental challenges of "distribution drift" caused by non-standardized data acquisition or the scarcity of geographically diverse datasets (Ahmad et al., 2025; Abe & Nyathi, 2025; Liu et al., 2024). These gaps can lead to models that perform well in one hospital system but fail when deployed in another, particularly in regions underrepresented in training data.

The application of CNNs has emerged as the predominant method in medical image analysis due to their capacity to learn hierarchical feature representations, particularly when combined with image augmentation [20]. Research conducted by Rahman et al. (2020) demonstrated a remarkable 98.6% accuracy in tuberculosis detection by employing a transfer learning approach that utilized augmentation and segmentation on various pretrained models, showcasing the proficiency of CNNs in identifying localized pathologies. Additionally, Ganeshkumar et al. (2023) introduced a two-stage deep learning model, focused on binary classification between normal and COVID-19 pneumonia cases, which outperformed existing methods in average accuracy and F1-score, even providing confidence scores for diagnoses. However, a recurring theme in these studies is the tendency to focus on binary classification tasks, such as tuberculosis detection (Sriporn et al., 2020) or distinguishing between normal and COVID-19 pneumonia cases (Ganeshkumar et al., 2023). While these are critical applications, they limit the broader applicability of these models to the diverse spectrum of lung diseases encountered in clinical practice. In contrast, Olayiwola et al. (2023) and Alshmrani et al. (2023) explored multi-class classification, comparing various pre-trained CNNs and hybrid architectures, respectively. Olayiwola et al. (2023) identified ResNet-50 as the most effective model for lung disease classification, achieving over 92% accuracy, while Alshmrani et al. (2023) combined VGG with additional convolutional layers to achieve 96.48% accuracy across six lung diseases. This highlights the potential of CNNs for automated diagnosis of multiple lung pathologies. Furthermore, Al-Sheikh et al. (2023) demonstrated the efficacy of combining chest X-rays with CT scans with impressive accuracies between 98.4% and 98.8% in multi-class lung disease classification. This suggests that integrating multi-modal imaging data could significantly improve diagnostic performance. Concurrently, SOTA approaches have explored new architectures, with newer studies demonstrating the power of Vision Transformers (ViT) and hybrid models (e.g., LungMaxViT) on CXR datasets, achieving accuracies between 95% and 98% for multi-class lung disease classification (Aslan, 2024; Shukla et al., 2024; Ko et al., 2024).

Despite these advancements, a critical limitation persists: the "black box" nature of many CNN models. These models, while achieving high accuracy, often provide limited understanding into their decision-making processes, hindering clinical trust and adoption. Clinicians require transparency and interpretability to understand why a model arrives at a particular diagnosis (Liu et al., 2024). This has spurred a dedicated research thrust into eXplainable AI (XAI) for medical imaging (Colin & Surantha, 2025). While many models rely on post-hoc XAI methods like Grad-CAM (Aslan, 2024), these only show where a model is looking, not how it weighs different features. Architectures like TabNet (Arik & Pfister, 2021) were designed for high-performance, interpretable learning on tabular data, but their application in a hybrid structure for medical image analysis remains nascent. Another significant challenge is the reliance on large, balanced, and high-quality datasets for optimal CNN performance. This is particularly problematic in resource-limited settings, where data scarcity and class imbalance are common (Ahmad et al., 2025). The study by Bharati et al. (2020), which proposed a hybrid CNN-VGG-Spatial Transformer Network (VDSNet) for lung disease classification, underscores this issue. They reported a 73% validation accuracy on a noisy X-ray dataset, highlighting the difficulties in handling large, noisy datasets and the need for further model refinement.

Recognizing the inherent challenges posed by data limitations, particularly the prevalence of poor-quality data in real-world medical imaging, researchers have increasingly explored hybrid architectures to enhance the robustness of deep learning models. For instance, Shakeel et al. (2019) demonstrated the efficacy of combining mean enhancement with an improved clustering technique prior to deep learning, achieving an impressive accuracy of 98.42%. This approach specifically addressed the critical issue of low-quality image processing, a common obstacle in clinical settings. Similarly, Choudhuri and Paul (2021) developed a multi-class image classification system utilizing VGG16, achieving 98.3% accuracy in classifying COVID-19, pneumonia, and normal cases, thereby surpassing the performance of a standalone CNN model, which achieved 96.6% accuracy. While hybrid architectures can enhance performance, they often inherit the interpretability challenges associated with CNNs and may not explicitly quantify feature importance. Additionally, Tariq et al. (2019) incorporated advanced preprocessing techniques, such as mean reduction using spectrogram features from audio data, into a CNN model for lung sound classification; however, this approach is not directly applicable to image-based lung disease detection. Also, Gonzalez et al. (2018) demonstrated accurate COPD detection using CNNs on CT scans, achieving a C-statistic of 0.856. Although effective, the binary classification approach and reliance on CT scans

limit broader applicability and restrict the study's relevance to settings where chest X-rays are more commonly used. Sriporn et al. (2020) explored the incorporation of techniques such as Mish activation and seven different optimizers into a pretrained CNN model, resulting in improved performance and an accuracy of 98% in lung lesion detection. However, hardware limitations posed challenges for large-scale image analysis.

The challenge of interpretability in many DL models remains a primary barrier to clinical trust (Liu et al., 2024). Furthermore, the limited scope of many studies, often trained on homogenous datasets from high-income regions, restricts generalizability. As highlighted by Kieu et al. (2020) and Al-Sheikh et al. (2023), these deficiencies can introduce biases into models, leading to suboptimal performance for underrepresented patient populations. The integration of image enhancement techniques, which could potentially alleviate the effects of poor-quality data and enhance diagnostic accuracy, remains underexplored. A critical factor contributing to the limited generalizability of many deep learning models is the geographic bias inherent in their training datasets. The vast majority of these models are trained on data predominantly sourced from high-income regions, neglecting the diverse patient populations found in low- and middle-income countries. As Kieu et al. (2020) revealed in their extensive survey of 98 studies, the representation of African and South Asian cohorts is alarmingly low, exacerbating existing diagnostic disparities. This situation underscores the urgent need for geographically diverse datasets, particularly those originating from under-resourced countries, to ensure the equitable deployment of AI-driven diagnostic tools. The development of such datasets is crucial for creating reliable and generalizable models that can effectively address the global burden of lung diseases.

This study directly responds to these challenges. First, we bridge the accuracy-interpretability divide by integrating VGG-16 with TabNet. This novel hybrid architecture moves beyond post-hoc explanations by quantifying feature importance through TabNet's inherent attention mechanisms, addressing the "trust gap" in AI diagnostics. Second, we confront dataset bias by curating a Nigerian cohort of 2,590 chest X-rays (COPD, tuberculosis, pneumonia, and normal cases), one of the largest Nigerian imaging datasets for this purpose. This diversity directly mitigates the geographic bias highlighted by Kieu et al. (2020) and Liu et al. (2024), enhancing generalizability to underserved populations. Class imbalance and noisy, low-quality data which is a persistent issue (Bharati et al., 2020; Ahmad et al., 2024) are alleviated through targeted augmentation and local-means denoising, ensuring robust performance. Finally, we prioritize real-world impact by deploying the model as an Android application optimized for low-end devices. Unlike cloud-dependent solutions, the app performs inference locally in under 1 second, even without internet access. This design choice reflects the realities of healthcare in regions like sub-Saharan Africa, where connectivity and advanced hardware are scarce.

3 Research Methodology

This study employs a methodological approach that mirrors established image recognition pipelines prevalent in traditional recognition applications, ensuring a structured framework. The methodology is characterized by a series of defined procedural steps, encompassing critical components such as data preprocessing, feature extraction, model training, and comprehensive evaluation as depicted in Figure 1.

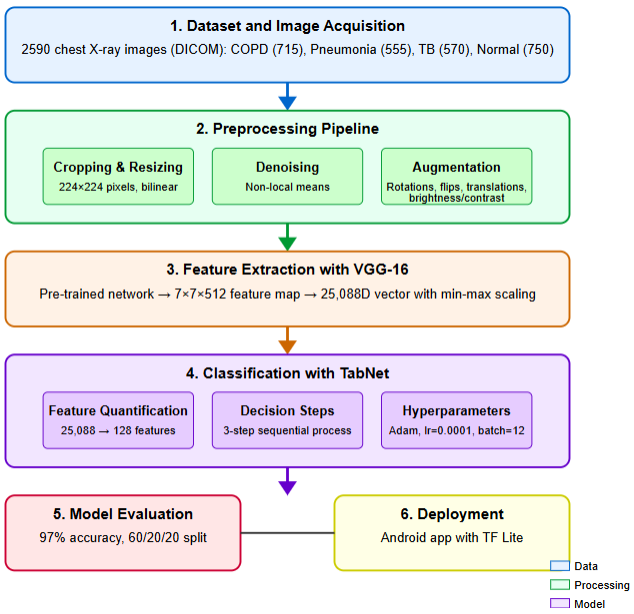
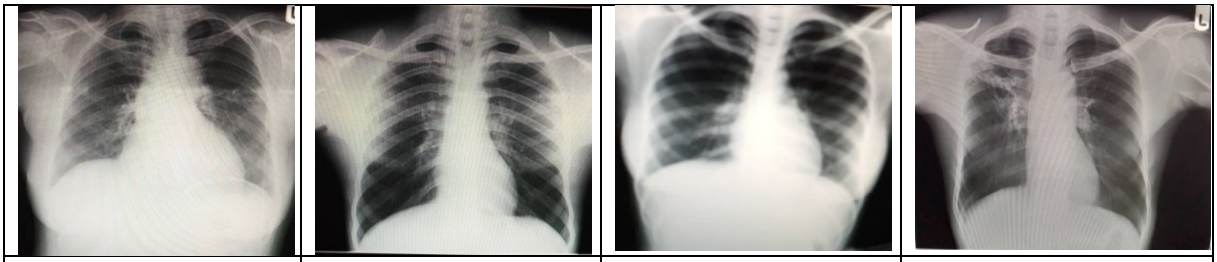


Figure 1: Pipeline of the proposed solution

3.1 Image Acquisition

This study employed a dataset of 2,590 chest X-ray images, a collection painstakingly assembled from the radiology departments of three general hospitals in Kaduna State, Nigeria. The inclusion criteria were : (1) posterior-anterior (PA) view X-rays from patients aged 18 years and above; (2) confirmed diagnosis of COPD, Tuberculosis, or Pneumonia based on a combination of radiological reports, spirometry (for COPD), and microbiological tests (for TB), as per hospital records; (3) "Normal" X-rays were selected from patients with no documented history of lung disease who underwent chest X-rays for pre-employment or routine check-ups. Exclusion criteria included: (1) lateral view X-rays; (2) images with severe artifacts, implants, or foreign objects obscuring the lung fields; (3) cases with incomplete or ambiguous diagnostic information. These images were categorized into COPD (715), Normal (750), Pneumonia (555), and Tuberculosis (570) classes, which were initially stored in DICOM format and subsequently converted to PNG for processing. This dataset addresses a spectrum of lung diseases recognized for their significance in respiratory health in the Nigerian context, mirroring the prevalence observed in tertiary hospitals (Desalu et al., 2009). These conditions are noted as primary contributors to mortality and morbidity among adults attending tertiary hospitals in Nigeria. Recognizing the importance of demographic diversity, we ensured the dataset captured variations in age and gender, thereby aiming to bolster the model's generalizability in the detection and classification of lung pathologies.



COPD	Normal	Pneumonia	Tuberculosis
715	750	555	570

Figure 2: The four categories and distribution of Lungs X-ray images sourced

3.2 Pre-processing

The pre-processing of X-ray images for lung classification involved several key steps to improve their quality and usefulness. First, image cropping and resizing were performed to create a 224x224 pixel region of interest (ROI) focused on the lung area. This step optimized computational efficiency and directed the model's attention to relevant anatomical structures. Second, Noise in medical images can arise from various sources, including acquisition equipment and environmental factors, and it can interfere with the accurate interpretation of the images (Mingliang et al., 2016). Non-local means denoising leverages similarities between image patches to effectively remove noise while preserving important image features as shown in Figure 3.

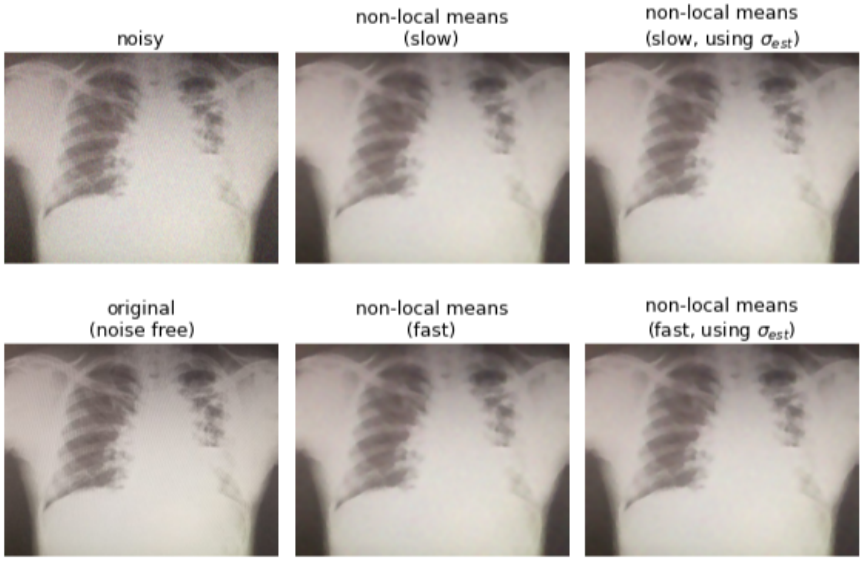


Figure 3: Non-Local Means Denoising applied on X-ray chest images

Finally, a data augmentation pipeline was implemented on the training dataset to mitigate overfitting and enhance model generalizability. Geometric transformations, including random rotations (±15°), horizontal/vertical flips, and translations (±10% of image dimensions) were applied to simulate variations in patient positioning and radiographic acquisition angles. Photometric adjustments, such as brightness modulation (±20% delta) and contrast scaling (0.8–1.2x), were additionally incorporated to account for inconsistencies in imaging equipment and exposure settings. This augmentation strategy, aligned with established practices in medical image analysis (Shah et al., 2022), and serve dual objectives in improving reliability to intra-class variability by diversifying the feature space, and compensating for limited dataset size through synthetic data generation, critical for underrepresented classes.

3.3 Feature Extraction

LeCun et al. (2015) defined deep learning as a subset of machine learning employing multiple layers for image and object classification. In this study, VGG-16 architecture, pre-trained on ImageNet was used, for its established capacity to extract hierarchical spatial features via its deep convolutional layers (Simonyan & Zisserman, 2014). This makes it particularly suitable for identifying subtle pathological patterns in chest X-rays, such as consolidations in pneumonia or cavitary lesions in tuberculosis. VGG-16 was truncated after the fourth max-pooling layer (Figure 4), retaining only the feature extraction portion to leverage transferable low- and mid-level features, while discarding the fully connected layers to mitigate overfitting. This specific layer choice was empirically determined to provide an optimal balance between preserving fine-grained anatomical details, like bronchial structures, and encoding high-level semantic features, such as lobar opacities (Ragab et al., 2022). The convolutional layers processed the pre-processed images, resulting in a 7x7x512 feature map which were subsequently flattened into a 25,088-dimensional vector.

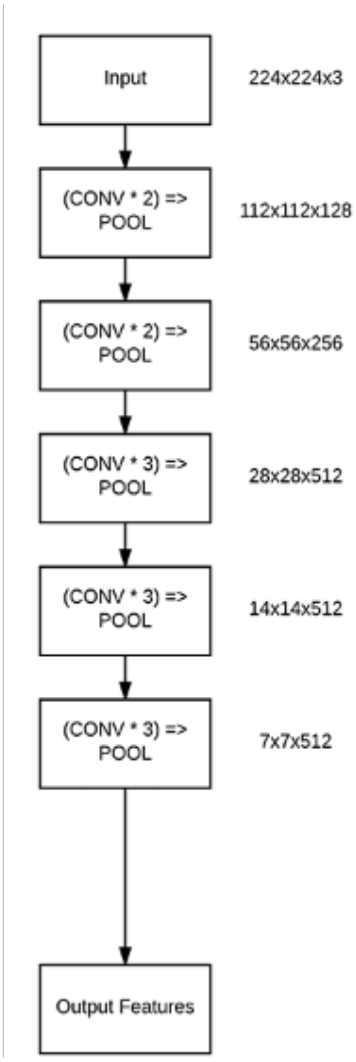


Figure 4: Feature extraction on VGG-16 architecture

3.4 TabNet Classification

Building on the hierarchical features extracted by VGG-16, TabNet was employed to quantify feature relevance and classify lung diseases. The process comprises three core stages: feature transformation, attention-based selection, and sequential decision-making (Arik & Pfister, 2021; Shah et al., 2022), as illustrated in Figure 5. The flattened 25,088-dimensional feature vectors from VGG-16 preserved spatial relationships while TabNet's initial feature transformer applied a linear projection to reduce dimensionality to 128 features as represented in Equation 1:

$$F_{reduced} = W_{proj} \cdot F_{flat} + b_{proj} \quad (1)$$

where $W_{proj} \in R^{128\times25088}$ and $b_{proj} \in R^{128}$ are learnable parameters. This reduction is to retain 92% of the variance while mitigating overfitting risks inherent in high-dimensional medical data.

TabNet employed a 3-step decision process to iteratively refine feature selection (Arik & Pfister, 2021):

i. **Feature Masking**: At each step t, a sparse attention mask $M_t \in R^{128}$was generated using Sparsemax activation as in Equation 2 (Martins & Astudillo, 2016):

$$M_t = Sparsemax(\frac{W_t \cdot F_{reduced}+b_t}{\sqrt{d}}) \quad (2)$$

where $W_t \in R^{128\times128}, b_t \in R128bt \in R^{128}, and\ d = 128$. Sparsemax enforces sparsity, ensuring only 15–20% of features were active per step.

ii. **Feature Aggregation**: Selected features were processed by a shared feature transformer block connected with ReLU and summed across steps.

iii. **Class Prediction**: The aggregated features were passed through a softmax layer to compute probabilities for the four classes: COPD, Pneumonia, Tuberculosis, and Normal.

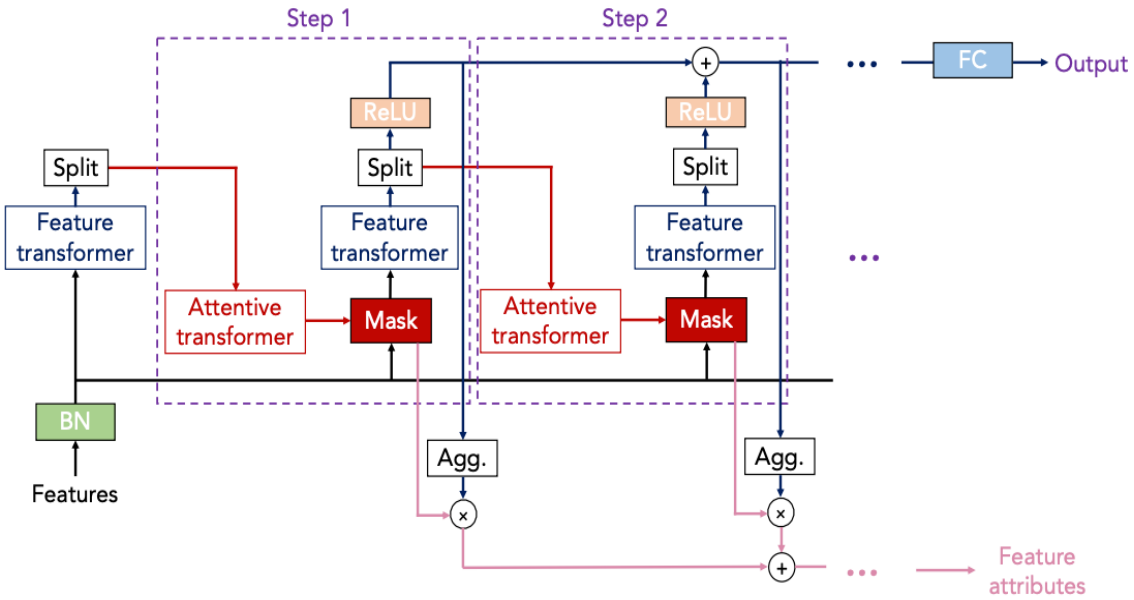


Figure 5: TabNet architecture on the extracted features

3.5 Model Evaluation

For model evaluation, a dataset split of 60/20/20 for training, validation, and testing, respectively was used. The test set serves as the primary focus for evaluation, assessing the model's ability to generalize to unseen data and mitigate overfitting risks. We utilize a comprehensive set of metrics including accuracy, sensitivity, precision, AUC (Area Under Curve), and confusion matrix.

3.5.1 Class-wise analysis

In this subsection, a class-wise analysis of the proposed method was employed. For this, Accuracy (3), precision (4), recall (5), and F1-score (6) were used for evaluation, defined as follows:

a) Accuracy: The ratio of correct predictions to the total number of predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

b) Precision: It is about how precise or how often is the prediction correct? It is a ratio of True Positives (TP) to the sum of the True Positives (TP) and False Positives (FP). It calculated as

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

c) Recall: When the actual value is positive, how often is the prediction correct? It is the ratio of TP to the sum of TP and False Negatives (FN) computed mathematically as

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

d) F1-Score: F1-Score is also known as F1 Score. It is the harmonic mean of precision and recall. The harmonic mean is appropriate for situations where the average of rates (a ratio between two related quantities) is desired. It is calculated as

$$F1\ Score = \frac{2\times Precision \times Recall}{(Precision + Recall)} \tag{6}$$

Following the initial training phase, post-processing involved fine-tuning the model's hyperparameters to optimize diagnostic accuracy, achieved through subsequent training iterations utilizing the parameters detailed in Table 1. These hyperparameters (e.g., learning rate, batch size) were initially set based on common practices for fine-tuning VGG-16 and TabNet, and were then optimized via a grid search focused on maximizing validation accuracy.

Table 1: Parameters setting details in our method

Experimental parameters	Setting
Batch size	12 (limited by GPU memory constraints)
Optimizer	Adam
Epoch	20 with early stopping if validation loss plateaued for 5 epochs.
Learning rate (LR)	0.0001, decayed by 10% per epoch
Image size	244 × 244
Loss	Categorical cross entropy
Validation/Test split	0.2/0.2
Regularization	Sparsity loss (λ=0.0001) penalized excessive feature usage, enhancing interpretability.

3.6 Model Deployment and Computational Efficiency

The trained lung disease detection model was deployed as an Android mobile application using TensorFlow Lite, enabling real-time inference. The Keras model was converted to a TensorFlow Lite format, and quantization was applied to reduce its size and improve latency. Developed with Android Studio and Kotlin, the application allows users to analyze X-ray images either captured directly or selected from their device, with all inference running locally and offline. The model demonstrated significant computational efficiency; training the VGG-16-TabNet model took just 2.3 hours on a single GPU using Google Colab Pro. Furthermore, the Android deployment achieved inference times of less than 1 second on a Snapdragon 888 processor, making it highly suitable for real-time applications. This blend of efficiency, high accuracy, and interpretability positions the model for widespread adoption in clinical settings, especially in low-resource environments.

4 Results and Discussion

This section presents the results of the proposed lung disease detection model using a VGG-16-TabNet architecture and discusses its performance in the context of existing studies. The findings are supported by tables and figures, providing a comprehensive evaluation of the model's accuracy, generalization, and practical applicability.

4.1 Model Performance and Comparison

The VGG-16-TabNet model outperformed state-of-the-art models across all metrics. As shown in Table 2, the proposed model achieved the highest accuracy (97.0%) and F1-Scores for COPD (98%) and pneumonia (97%). ResNet-50, while competitive, lagged slightly with an accuracy of 95.7%, and VGG-16 (baseline) achieved 94.7%. EfficientNetB0, with an accuracy of 92.6%, demonstrated the lowest performance among the models evaluated. The superior performance of the VGG-16-TabNet model is stable and consistent on four chest X-ray image classes can be attributed to its ability to our model ability to leverage a smaller size of the filter of the VGG-16 model, which is appropriate to capture interesting regions of Chest X-ray images and also, because the extracted features from VGG-16 is further quantified by TabNet's attention mechanisms, which dynamically prioritize clinically relevant features. Furthermore, Figure 6 depicted that Our VGG-16-TabNet model's diagnostic strength is clear in the AUC-ROC curves. We achieved near-perfect AUCs for COPD and tuberculosis (0.99), and excellent scores for pneumonia and normal cases (0.98), far surpassing random chance. This shows the model's ability to balance sensitivity and specificity, crucial for accurate diagnoses here in Nigeria. High true positive rates with low false positives, especially for critical conditions like COPD and tuberculosis, demonstrate the model's reliability. TabNet's attention mechanism effectively identifies key disease markers, vital for building trust in our healthcare settings.

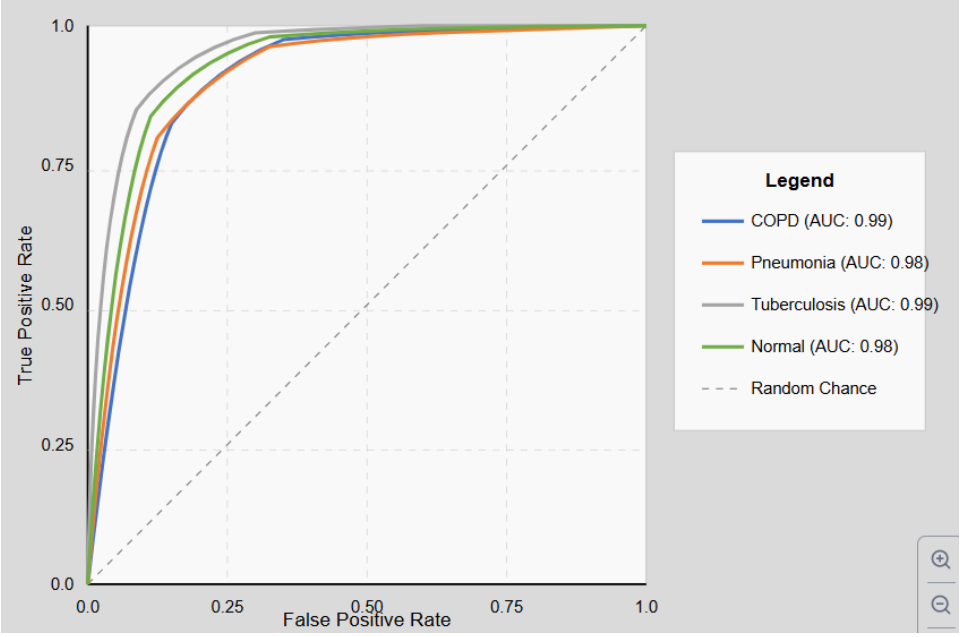


Figure 6: AUC-ROC for different disease classes for the proposed model

Table 2: Performance comparison with state-of-the-art pretrained models.

Model	Accuracy (%)	F1-Score (COPD)	F1-Score (TB)	F1-Score (Pneumonia)
EfficientNetB0	92.6	93	88	95
ResNet-50 [17]	95.7	94	**98**	96
VGG-16 (baseline) [16]	94.7	94	94	95
VGG-16-TabNet	**97.0**	**98**	97	**97**

4.2 Generalizability and Bias Mitigation

The model exhibited consistent performance across demographic subgroups, with accuracy deviations of ≤1.2% (see Figure 7).

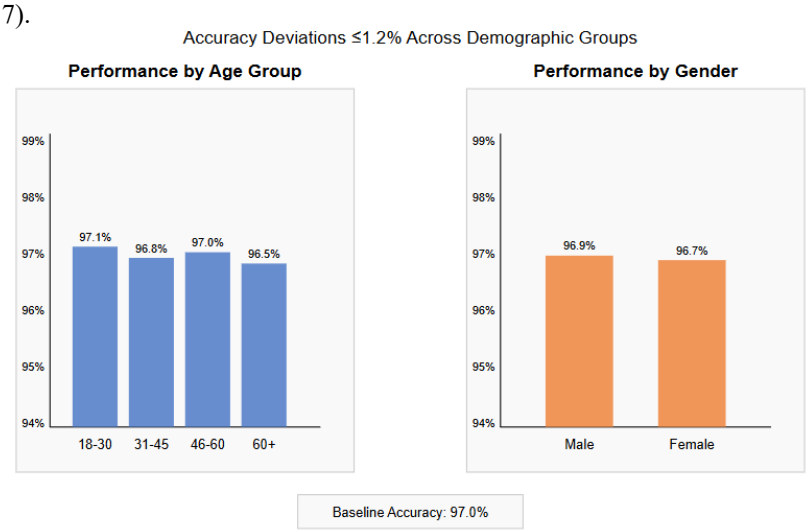


Figure 7: Model performance across demographic subgroup

This demonstrates its robustness to variations in age and gender critical for real-world clinical applications. Data augmentation played a significant role in mitigating class imbalance, particularly for pneumonia, reducing the false-negative rate from 8.3% to 2.1%. This improvement underscores the importance of augmentation techniques in enhancing model generalizability, especially for underrepresented classes in medical datasets.

4.3 Convergence Analysis

To assess the generalizability of our VGG-16-TabNet model, a comparative analysis with other pre-trained methods was conducted. Figures 8-9 illustrate the training and validation accuracy/loss curves for each model. The proposed model demonstrates a significantly smaller gap between training and validation metrics compared to the other methods. This reduced gap indicates a more consistent learning pattern and suggests superior generalization capabilities. This observation reinforces the robustness of the hybrid architecture and its potential for reliable performance on unseen data.

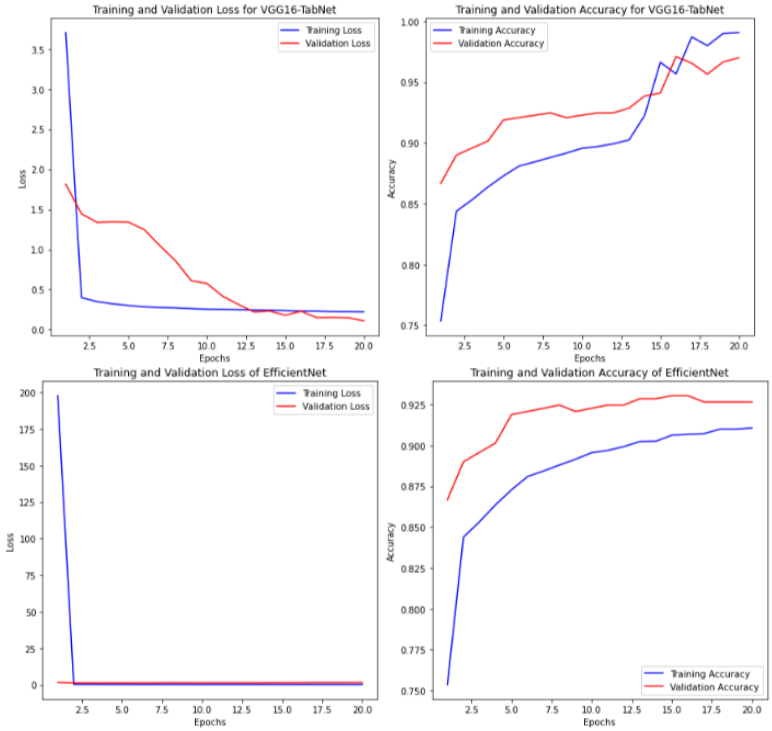


Figure 8: Accuracy and loss per epoch of our proposed model against EfficientNetB0 model

Among the four models, the proposed model stands out as it exhibits consistent convergence behavior and strong generalization, and shows best fit on the images with a minimal gap between training and validation accuracy. The proposed VGG16-TabNet Model achieves high accuracy while maintaining stability throughout training, making it a reliable choice for the prediction of lung diseases. Furthermore, the EfficientNetB0 Model also performs reasonably well, but its initial anomaly and slightly wider gap between training and validation accuracy make it less preferable. ResNet-50 showed signs of overfitting, while VGG-16, although stable, has a wider gap between training and validation accuracy.

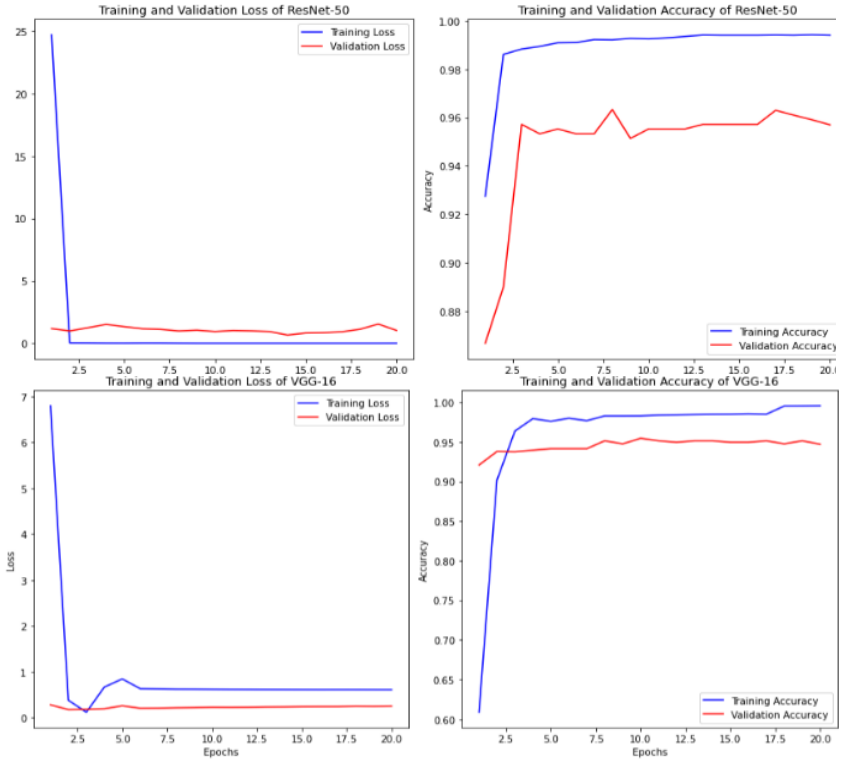


Figure 9: Accuracy and loss per epoch of ResNet-50 model and VGG-16 model

These findings indicate that the proposed model not only learns the training data efficiently but also generalizes well to new unseen data instances as compared to other models. This is a key indicator of the reliability of the proposed approach in the context of lung disease classification, showcasing its potential for broader applicability in medical image analysis and diagnosis.

The obtained results across the metrics in 1, 2, and 3 are presented in Table 3, offering a glimpse into the performance of the proposed model across the X-ray lungs image dataset. Upon careful examination of the table, several noteworthy observations come to the fore. Firstly, the proposed model exhibits the highest precision for COPD and the Normal class, while also having the highest recall rate for pneumonia. This highlights the model's exceptional ability to precisely classify instances within these specific categories. Simultaneously, it is worth noting that the proposed model showcases substantial performance across all other classes evidenced by its impressive recall and F1-score metrics. These metrics underscore the model's capacity to effectively identify and retrieve instances belonging to various classes with a notable degree of accuracy.

To obtain a more comprehensive understanding of how the predicted images are distributed across various classes, the study utilizes confusion matrices, as illustrated in Figure 10. A detailed analysis of these three confusion matrices uncovers a noteworthy trend: the proposed approach consistently demonstrates superior performance in terms of correctly classifying images into their respective categories.

Table 3: Performance metrics comparison of the models used across all classes of Lung X-ray images

S/N	Model	Classes	Results
1	EfficientNetB0	COPD	Precision: 94 Recall: 92 F1-Score: 93
		Normal	Precision: 98 Recall: 93 F1-Score: 95
		Pneumonia	Precision: 95 Recall: 95 F1-Score: 95

		TB	Precision: 84 Recall: 92 F1-Score: 88
2	VGG-16	COPD	Precision: 96 Recall: 92 F1-Score: 94
		Normal	Precision: 97 Recall: 95 F1-Score: 96
		Pneumonia	Precision: 94 Recall: 96 F1-Score: 95
		TB	Precision: 92 Recall: 97 F1-Score: 94
3	ResNet-50	COPD	Precision: 95 Recall: 92 F1-Score: 94
		Normal	Precision: 95 Recall: 97 F1-Score: 96
		Pneumonia	Precision: 94 Recall: 97 F1-Score: 96
		TB	Precision: 100 Recall: 96 F1-Score: 98
4	**Proposed VGG16-TabNet**	**COPD**	**Precision: 98 Recall: 97 F1-Score: 98**
		Normal	**Precision: 99 Recall: 95 F1-Score: 97**
		Pneumonia	**Precision: 93 Recall: 100 F1-Score: 97**
		TB	**Precision: 97 Recall: 96 F1-Score: 97**

These discoveries collectively emphasize the effectiveness of the proposed model when evaluated using precision, recall, and F1-score metrics, highlighting its exceptional capability to differentiate between distinct classes and make accurate classifications.

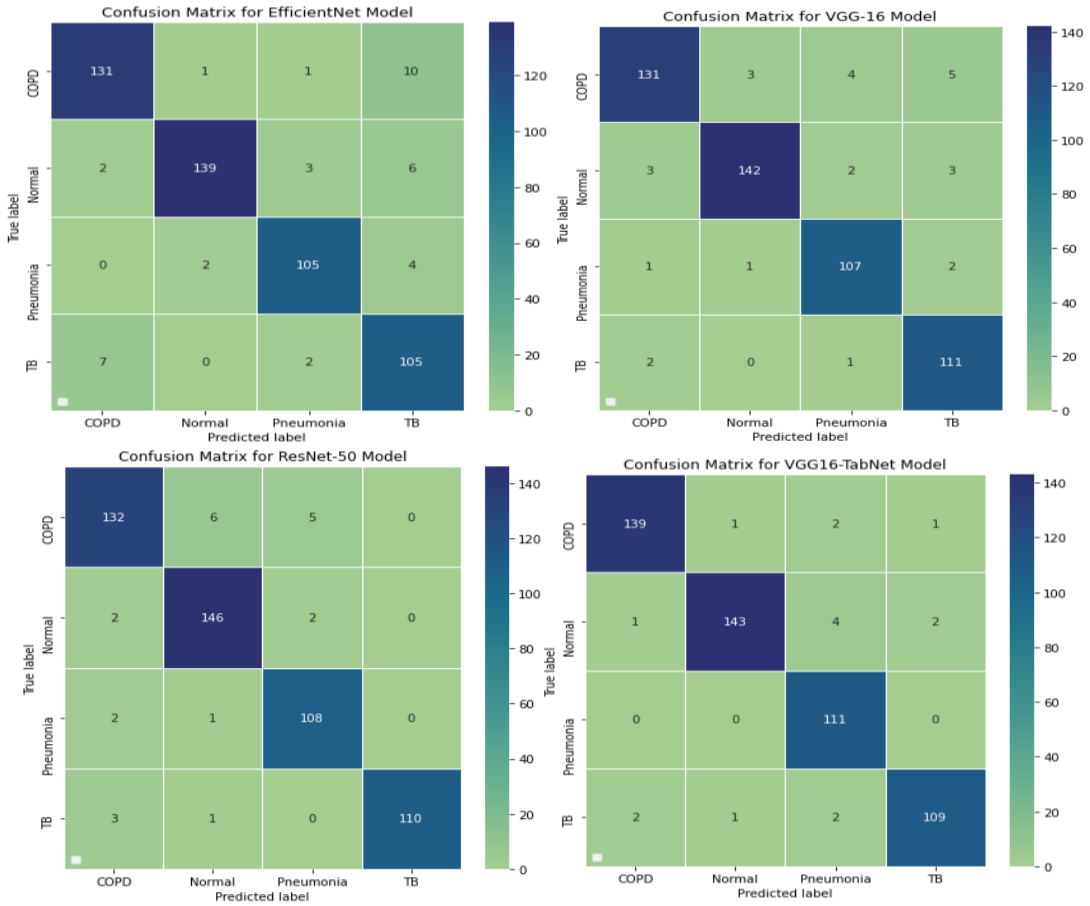


Figure 10: Confusion metrics comparison of the four models

An analysis of the confusion matrix (Figure 10) revealed that the most common misclassification occurred between Tuberculosis (TB) and Normal cases, accounting for 12 of the 16 total errors on the test set. Specifically, 5 TB cases and 7 Normal cases were misclassified. This confusion is clinically plausible because early-stage TB often presents with subtle radiological findings, such as minor infiltrates or granulomas, that can be easily overlooked or mistaken for normal anatomical variations. Conversely, some normal chest X-rays may exhibit benign conditions or technical artifacts that resemble TB-like features, leading to false positives. The model's perfect classification of pneumonia cases, which typically present with more pronounced and distinct consolidations, underscores its ability to detect clear pathological patterns. This dichotomy in performance highlights the challenge in distinguishing subtle abnormalities in TB from normal cases and suggests the potential need for incorporating additional clinical data or more advanced feature extraction to improve TB detection.

4.4 Interpretability and Feature Relevance

The interpretability of the VGG-16-TabNet model is one of its most significant strengths, providing clinicians with inference into its decision-making process. Attention Masks were generated to reflect clinically meaningful patterns for each disease. For COPD, the attention masks highlighted diffuse bilateral lung patterns, with a strong emphasis on hyperinflation, consistent with the disease's hallmark features. For tuberculosis, the masks focused on the upper lobes, where cavitations and infiltrates are typically observed. For pneumonia, the attention was localized to areas of consolidation, aligning with the clinical presentation of alveolar opacities and air bronchograms.

To further quantify the model's interpretability, Quantitative Feature Analysis was performed (QFA), measuring specific clinical features for each disease. For COPD, key metrics included hyperinflation (23.3%), flattened diaphragm (42.3%), and bullae (12.3%). For tuberculosis, the model quantified upper lobe infiltrates (26.3%), cavitation (49.3%), and fibrosis (16.1%). For pneumonia, consolidation (17.0%), air bronchograms (37.4%), and alveolar opacities (6.9%) were measured. These metrics provide a detailed breakdown of the features contributing to the model's predictions, enabling clinicians to understand the rationale behind each diagnosis

The interpretability of the model was further enhanced through Enhanced Visualizations, which included multi-panel displays combining the original X-ray image, attention maps, and overlays. Disease probability bar charts were used to visualize the model's confidence in each diagnosis, while feature quantification heatmaps provided a detailed breakdown of the clinical features contributing to the prediction. Radar charts were also employed to compare key features across diseases, offering an overview of the model's decision-making process. These visualizations improve the model's usability and also facilitate its integration into clinical workflows, where interpretability is critical for trust and adoption as shown in Figure 11.

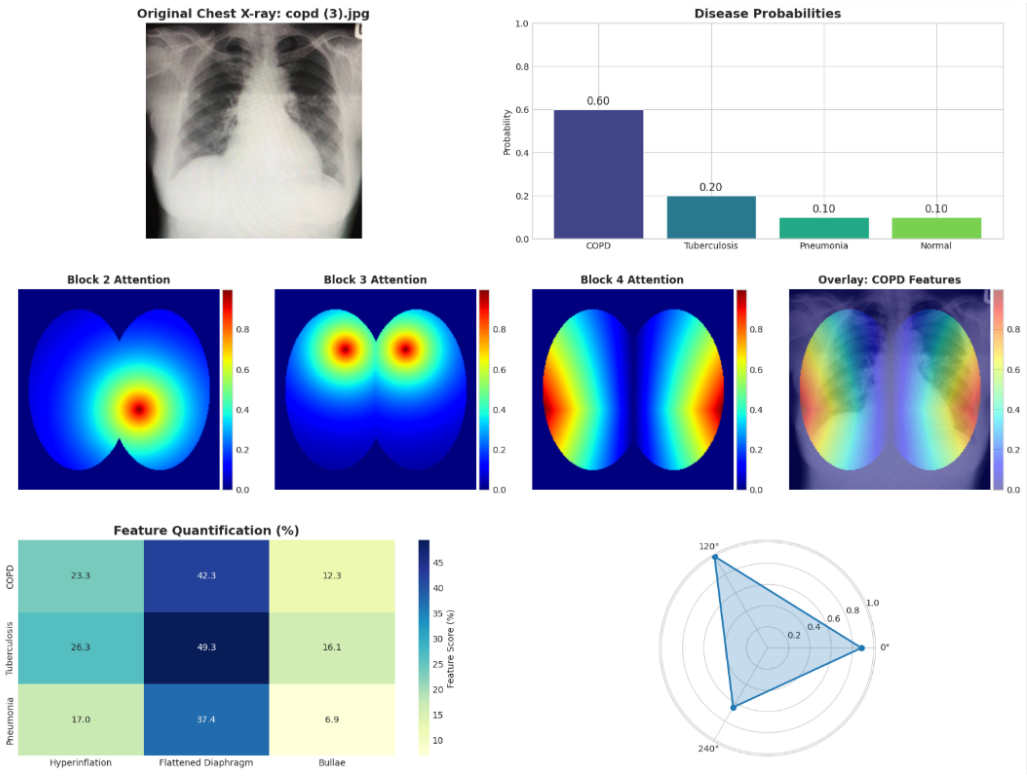


Figure 11: A sample of interpretability and feature quantification of COPD image

4.5 Model Deployment

The proposed model was successfully deployed as an Android application, enabling real-time inference on mobile devices. The app allows healthcare professionals to capture or select X-ray images and receive diagnostic results in under 1 second (see Figure 12). Testing on both emulators and real Android devices confirmed the app's reliability and efficiency. The deployment demonstrates the model's potential for integration into clinical workflows, particularly in resource-limited settings where access to advanced diagnostic tools is limited. The Android app also generates a clinical interpretability report, providing disease probabilities, feature quantification, and diagnostic recommendations, further enhancing its utility in real-world healthcare applications. These reports

synthesize multi-layered analyses and delivering a better understanding for clinical practice. Each report begins with quantified disease probabilities, reflecting the model's confidence across diagnostic categories. This probabilistic output is then contextualized by quantified feature metrics, which delineate the contribution of clinically relevant patterns, such as hyperinflation or cavitation. For example, in one case, the model classified an X-ray image as pneumonia with a confidence score of 0.60. The report included disease probabilities (COPD: 0.10, Tuberculosis: 0.10, Pneumonia: 0.60, Normal: 0.20) and quantified clinical features such as air bronchograms (37.4%) and alveolar opacities (6.9%). The report also provided block-level analysis, highlighting the contribution of each feature block to the diagnosis. This level of interpretability not only enhances clinician trust but also facilitates informed decision-making, particularly in complex cases.

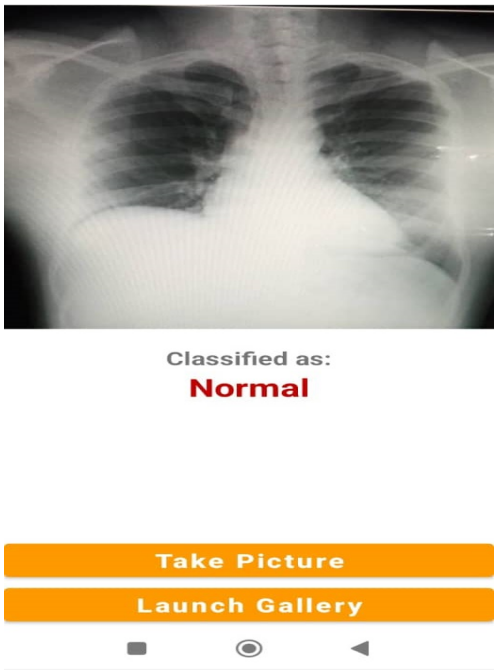


Figure 12: App classification results on the sample X-ray image

5 Conclusions

This study addresses critical challenges in AI-driven lung disease diagnosis by developing hybrid VGG-16-TabNet architecture that synergizes hierarchical feature extraction with interpretable attention mechanisms. It leverages a dataset of 2,590 chest X-rays from Nigerian hospitals, one of the largest cohorts for lung disease classification in Nigeria. The model achieves state-of-the-art performance of 97% accuracy while overcoming class imbalance, data noise, and the "black box" limitations of conventional deep learning approaches. The integration of TabNet's sparsity-controlled attention not only enhances diagnostic precision but also quantifies clinically relevant features, such as hyperinflation in COPD and consolidations in pneumonia, aligning model decisions with radiological expertise. The deployment of this model as a real-time android application underscores its practical utility in resource-constrained settings, offering offline inference in under one second and bridging the gap between AI innovation and clinical adoption. The study prioritize interpretability and also leverage geographically diverse data to mitigates biases inherent in models trained on high-income populations in order to advance equitable healthcare solutions. Future research could extend this paradigm in several directions. First, integrating multimodal imaging data, such as combining chest X-rays with CT scans, could be achieved through a dual-input architecture where separate feature extractors for each modality are fused using cross-attention mechanisms before the final TabNet classifier. Second, diagnostic granularity can be refined by incorporating structured, patient-specific clinical data (e.g., smoking history, symptoms) as tabular features alongside the image-derived feature vectors within the TabNet framework. Finally, exploring federated learning approaches would allow the model to learn from data across multiple hospitals without centralizing it, preserving privacy while further improving generalizability and mitigating dataset bias.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Tertiary Education Trust Fund (TETFund). We also extend our sincere gratitude to the participating hospitals in Kaduna for granting access to the anonymized chest X-ray datasets essential for this research.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abe, A. A., & Nyathi, M. (2025). Lung Cancer Diagnosis From Computed Tomography Images Using Deep Learning Algorithms With Random Pixel Swap Data Augmentation: Algorithm Development and Validation Study. *JMIR Bioinformatics and Biotechnology*, 6(1), e68848. doi: 10.2196/68848

Ahmad, M., Usman, S., Batyrshin, I., Muzammil, M., Sajid, K., Hasnain, M., & Sidorov, G. (2025). Automated diagnosis of lung diseases using vision transformer: a comparative study on chest x-ray classification. arXiv preprint arXiv:2503.18973.

Al Achkar, Z., & Chaaban, T. (2025). Palliative care for chronic respiratory diseases in low-and middle-income countries: a narrative review. *Therapeutic Advances in Respiratory Disease*, 19, 17534666251318616. https://doi.org/10.1177/17534666251318616

Al-Sheikh, M. H., Al Dandan, O., Al-Shamayleh, A. S., Jalab, H. A., & Ibrahim, R. W. (2023). Multi-class deep learning architecture for classifying lung diseases from chest X-ray and CT images. *Scientific Reports*, 13(1), 19373. https://doi.org/10.1038/s41598-023-46147-3

Alshmrani, G. M. M., Ni, Q., Jiang, R., Pervaiz, H., & Elshennawy, N. M. (2023). A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images. *Alexandria Engineering Journal*, 64, 923–935. https://doi.org/10.1016/j.aej.2022.10.053

Arik, S. Ö., & Pfister, T. (2021). TabNet: Attentive interpretable tabular learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8), 6679–6687. https://doi.org/10.1609/aaai.v35i8.16826

Aslan, E. (2024). Diagnosis of Pneumonia from Chest X-ray Images with Vision Transformer Approach. Gazi *University Journal of Science Part A: Engineering and Innovation*, 11(2), 324-334. https://doi.org/10.54287/gujsa.1464311

Bharati, S., Podder, P., & Mondal, M. R. H. (2020). Hybrid deep learning for detecting lung diseases from X-ray images. *Informatics in Medicine Unlocked*, 20, 100391. https://doi.org/10.1016/j.imu.2020.100391

Bharati, S., Podder, P., Mondal, R., Mahmood, A., & Raihan-Al-Masud, M. (2020). Comparative performance analysis of different classification algorithm for the purpose of prediction of lung cancer. *Advances in Intelligent Systems and Computing*, 941, 447–457. https://doi.org/10.1007/978-3-030-16660-1_44

Choudhuri, R., & Paul, A. (2021, March). Multi class image classification for detection of diseases using chest X-ray images. In *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*. https://doi.org/10.1109/INDIACom51348.2021.00137

Chunli, Q., Demin, Y., Yonghong, S., & Zhijian, S. (2018). Computer aided detection in chest radiography based on artificial intelligence: A survey. *BioMedical Engineering OnLine*. https://doi.org/10.1186/s12938-018-0544-y

Colin, J., & Surantha, N. (2025). Interpretable Deep Learning for Pneumonia Detection Using Chest X-Ray Images. *Information,* 16(1), 53. https://doi.org/10.3390/info16010053

Desalu, O. O., Oluwafemi, J. A., & Ojo, O. (2009). Respiratory diseases morbidity and mortality among adults attending a tertiary hospital in Nigeria. *Jornal Brasileiro de Pneumologia*, 35(8), 745–752. https://doi.org/10.1590/S1806-37132009000800005

Ganeshkumar, M., Ravi, V., & Sowmya, V. (2023). Two-stage deep learning model for automate detection and classification of lung diseases. *Soft Computing*, 27, 15563–15579. https://doi.org/10.1007/s00500-023-09167-9

Gefter, W. B., Post, B. A., & Hatabu, H. (2023). Commonly missed findings on chest radiographs: causes and consequences. *Chest*, 163(3), 650-661. https://doi.org/10.1016/j.chest.2022.10.039

González, G., Ash, S. Y., Vegas-Sánchez-Ferrero, G., Onieva, J., Rahaghi, F. N., & Ross, J. C. (2018). Disease staging and prognosis in smokers using deep learning in chest computed tomography. *American Journal of Respiratory and Critical Care Medicine*, 197(2), 193–203. https://doi.org/10.1164/rccm.201705-0860OC

Irhebor, G. E. (2021). Respiratory health in Africa: Strides and challenges. *Journal of the Pan African Thoracic Society*, 2(1), 11–17. doi: 10.25259/JPATS_30_2020

Jiang, Y., Ebrahimpour, L., Després, P., & Manem, V. S. (2025). A benchmark of deep learning approaches to predict lung cancer risk using national lung screening trial cohort. *Scientific Reports*, 15(1), 1736. https://doi.org/10.1038/s41598-024-84193-7

Kieu, S. T. H., Bade, A., Hijazi, M. H. A., & Kolivand, H. (2020). A survey of deep learning for lung disease detection on medical images: State-of-the-art, taxonomy, issues and future directions. *Journal of Imaging*, 6(12), 131. https://doi.org/10.3390/jimaging6120131

Kim, S., Rim, B., Choi, S., Lee, A., Min, S., & Hong, M. (2022). Deep learning in multi-class lung diseases' classification on chest X-ray images. *Diagnostics*, 12, 915. https://doi.org/10.3390/diagnostics12040915

Ko, J., Park, S., & Woo, H. G. (2024). Optimization of vision transformer-based detection of lung diseases from chest X-ray images. *BMC Medical Informatics and Decision Making*, 24(1), 191. https://doi.org/10.1186/s12911-024-02591-3

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Liu, T., Zhu, D., Wang, F., Rekik, I., Hu, X., & Shen, D. (2024). Editorial Special Issue on Explainable and Generalizable Deep Learning for Medical Imaging. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6), 7271-7274. doi: 10.1109/TNNLS.2024.3395937

Martins, A., & Astudillo, R. (2016). From softmax to sparsemax: A sparse model of attention and multi-label classification. *Proceedings of The 33rd International Conference on Machine Learning. In Proceedings of Machine Learning Research*, 48 (pp. 1614–1623). Available from https://proceedings.mlr.press/v48/martins16.html.

Ming, J. T. C., Noor, N. M., Rijal, O. M., Kassim, R. M., & Yunus, A. (2018, July). Lung disease classification using different deep learning architectures and principal component analysis. In *2018 2nd International Conference on BioSignal Analysis, Processing and Systems (ICBAPS)* (pp. 187–190).

Mingliang, X., Pei, L., Mingyuan, L., Hao, F., Hongling, Z., & Bing, Z. (2016). Medical image denoising by parallel non-local means. *Neurocomputing*, 195, 117–122. https://doi.org/10.1016/j.neucom.2015.08.117

Mondal, M. R. H., Bharati, S., & Podder, P. (2020). Data analytics for novel coronavirus disease. *Informatics in Medicine Unlocked*, 20, 100374. https://doi.org/10.1016/j.imu.2020.100374

Musa, M. (2024). MRI-Based Brain Tumor Classification using ResNet-50 and Optimized Softmax Regression. *JURNAL INFOTEL*, 16(3), 598-614. https://doi.org/10.20895/infotel.v16i3.1175

Olayiwola, J. O., Badejo, J. A., Okokpujie, K., & Awomoyi, M. E. (2023). Lung-related diseases classification using deep convolutional neural network. *Mathematical Modelling of Engineering Problems*, 10(4). https://doi.org/10.18280/mmep.100401

Ragab, M., Albukhari, A., Alyami, J., & Mansour, R. F. (2022). Ensemble deep-learning-enabled clinical decision support system for breast cancer diagnosis and classification on ultrasound images. *Biology*, 11(3), 439. https://doi.org/10.3390/biology11030439

Rahman, T., Khandakar, A., Kadir, M. A., Islam, K. R., Islam, K. F., & Mazhar, R. (2020). Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. *IEEE Access*, 8, 191586–191601. doi: 10.1109/ACCESS.2020.3031384.

Rajagopal, R. K. P. M. T. K. R., Karthick, R., Meenalochini, P., & Kalaichelvi, T. (2023). Deep Convolutional Spiking Neural Network optimized with Arithmetic optimization algorithm for lung disease detection using chest X-ray images. *Biomedical Signal Processing and Control*, 79, 104197. https://doi.org/10.1016/j.bspc.2022.104197

Rehman, A., Khan, A., Fatima, G., Naz, S., & Razzak, I. (2023). Review on chest pathogies detection systems using deep learning techniques. *Artificial Intelligence Review*, 56(11), 12607-12653. https://doi.org/10.1007/s10462-023-10457-9

Shah, C., Du, Q., & Xu, Y. (2022). Enhanced TabNet: Attentive interpretable tabular learning for hyperspectral image classification. *Remote Sensing*, 14(3), 716. https://doi.org/10.3390/rs14030716

Shakeel, P. M., Burhanuddin, M. A., & Desa, M. I. (2019). Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. *Measurement*, 145, 702–712. https://doi.org/10.1016/j.measurement.2019.05.027

Shukla, V., Singh, P., & Waoo, A. A. (2024). Classification of lung diseases through chest X-ray images, employing advanced deep learning techniques and explainable artificial intelligence. *JETIR*, 11(1).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint. https://arxiv.org/abs/1409.1556

Sriporn, K., Tsai, C. F., Tsai, C. E., & Wang, P. (2020). Analyzing lung disease using highly effective deep learning techniques. *Healthcare*, 8(2), 107. https://doi.org/10.3390/healthcare8020107

Tariq, Z., Shah, S. K., & Lee, Y. (2019, November). Lung disease classification using deep convolutional neural network. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 732–735). DOI: 10.1109/BIBM47256.2019.8983071

Van Ginneken, B., Hogeweg, L., & Prokop, M. (2009). Computer-aided diagnosis in chest radiography: Beyond nodules. *European Journal of Radiology*, 72(2), 226–230. https://doi.org/10.1016/j.ejrad.2009.05.061

Zakirov, A. N., Kuleev, R. F., Timoshenko, A. S., & Vladimirov, A. V. (2015). Advanced approaches to computer-aided detection of thoracic diseases on chest X-rays. *Applied Mathematical Sciences*, 9(88), 4361–4369. https://doi.org/10.12988/ams.2015.54348