Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 8 (2026) 3466

Journal of the Nigerian Society of Physical Sciences

Privacy-preserving federated learning with MobileViT for chest X-ray classification in Nigerian hospitals

Oboti Nwamaka Peace[a], Osita Miracle Nwakeze[b,*], Sunday Stephen Okika[c], Okafor Chinedu Martin[d], Agubosim Chuka Charles[b]

[a]*Department of Computer Science, Nnamdi Azikiwe University (NAU), Awka*
[b]*Department of Computer Science, Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra State*
[c]*Department of Electrical and Electronic Engineering, State University of Medical and Applied Sciences, Igbo-Eno, Enugu State, Nigeria*
[d]*Electronics Development Institute (ELDI), Awka (a division of the National Agency for Science and Engineering Infrastructure (NASENI))*

Abstract

Nigeria faces challenges in the accurate diagnosis of respiratory diseases such as tuberculosis (TB), pneumonia, and COVID-19 because of limited radiology resources and stringent patient-privacy requirements. Deep learning models can provide strong diagnostic performance but often rely on centralised data, raising ethical and security concerns. This study presents an application-oriented privacy-preserving federated learning framework integrating Mobile Vision Transformer (MobileViT) for chest X-ray classification across simulated Nigerian healthcare environments. Using the Nigeria Chest X-ray Dataset (2,600 radiologist-labelled images across four classes), experiments were conducted in a simulated federated environment on a single compute instance with five clients as a proof of concept for multi-hospital collaboration. Differential privacy was implemented through differentially private stochastic gradient descent (DP-SGD), and performance was evaluated using accuracy, area under the curve (AUC), precision, recall, F1-score , and class-wise sensitivity. In this simulated setting, the differentially private federated learning (DP-FL) MobileViT model achieved test accuracy above 92% and a macro F1-score above 89% at moderate privacy budgets ($\varepsilon \approx 5$–6), while maintaining TB sensitivity above 85% and pneumonia sensitivity above 88%. These proof-of-concept results suggest that privacy-preserving federated learning could support reliable and ethical AI diagnostics in resource-limited Nigerian hospitals, pending real-world deployment validation.

DOI: 10.46481/jnsps.2026.3466

Keywords: Federated learning, Deep learning, MobileViT, Chest X-ray, Disease classification

Article History:
Received: 26 April 2026
Received in revised form: 31 June 2026
Accepted for publication: 17 August 2026
Available online: 07 September 2026

Communicated by: B. J. Falaye

1. Introduction

Pneumonia and tuberculosis (TB) remain leading causes of morbidity and mortality in Nigeria [1]. Early diagnosis is critical for effective treatment, yet healthcare institutions face severe shortages of trained radiologists, patient overcrowding, and inadequate diagnostic equipment [2]. Chest X-ray imaging serves as a first-line diagnostic tool, but interpretation is often subjective in resource-limited settings, leading to misdiagnosis and delayed care [3].

Deep learning has demonstrated strong performance in automated chest X-ray analysis, with both convolutional neural networks (CNNs) and Vision Transformers achieving high accuracy in pneumonia, TB, and COVID-19 detection [4–6]. However, most existing systems rely on centralised training, requiring patient data to be aggregated at a single location [7]. Such approaches conflict with Nigerian data-privacy reg-

*Corresponding Author Tel. No.: +234-803-264-3003.
Email address: ma.nwakeze@coou.edu.ng (Osita Miracle Nwakeze)

ulations, institutional data silos, and patient-confidentiality requirements [8]. To address related diagnostic limitations, ultraportable, AI-enabled X-ray devices have been implemented in Nigerian populations to enhance TB case detection , as reported by John *et al.* [9]. Cost-effectiveness analyses by Garg *et al.* [10] further support the scalability of these technologies in active case-finding campaigns. The application of artificial intelligence (AI) to detect pneumonia among paediatric patients in Nigeria also demonstrates the potential applicability of these tools to respiratory diseases such as TB [11].

Federated learning (FL) offers a solution because it allows multiple healthcare institutions to jointly train a shared diagnostic model without moving raw medical images beyond their local settings , as reviewed by Rehman *et al.* [12]. Within this model, individual hospitals can train the model using their own chest X-ray data and exchange only model updates with a central coordinating server [13, 14]. The model architecture is important for making this approach viable in the Nigerian clinical environment. Mobile Vision Transformer (MobileViT) provides a trade-off between computational cost and performance through its fusion of convolutional and transformer mechanisms , making it applicable to hospitals and clinics with limited computing capabilities while retaining global image information, as discussed by Eshraghi *et al.* [15].

More recent results have shown that lightweight transformer-based models, including RepViT-CXR and MobileViT, can detect pneumonia and tuberculosis from chest X-rays with high accuracy at low computational cost [16, 17]. In addition, privacy-preserving federated learning systems that combine secure aggregation and differential privacy have been proposed to support compliance with healthcare data regulations. Despite these advances, a critical research gap remains: no prior work has applied differentially private federated learning with a vision transformer architecture to chest X-ray classification using a Nigerian dataset. Existing studies either employ centralised training, lack formal privacy guarantees, or rely on datasets from non-African populations. This paper addresses this gap by proposing a privacy-preserving federated learning framework tailored to the Nigerian healthcare context.

This paper presents an application-oriented privacy-preserving federated learning framework for chest X-ray classification in Nigeria. Rather than proposing a novel federated optimisation algorithm, this study systematically integrates existing techniques such as MobileViT, Federated Averaging (FedAvg), and DP-SGD to evaluate feasibility, privacy–utility trade-offs, and diagnostic performance on a Nigerian dataset for the first time. Specifically, the study applies an integrated DP-FL-MobileViT framework to a Nigerian chest X-ray dataset to evaluate feasibility in a simulated multi-hospital setting. It also systematically characterises privacy–utility trade-offs across four clinically relevant classes using DP-SGD with varying privacy budgets, empirically evaluates whether a lightweight Vision Transformer maintains diagnostic utility under differential-privacy noise in a resource-constrained federated environment, and provides an ablation and baseline comparison to demonstrate the incremental value of each framework component.

2. Related work

Vision Transformer (ViT)-based models have shown good results in diagnosing respiratory diseases from chest X-rays. Several studies indicate that transformer architectures can achieve high diagnostic accuracy and robustness across institutions. For instance, multi-task ViT frameworks achieved area under the curve (AUC) values exceeding 0.95 for COVID-19 classification and sensitivities above 95%, while demonstrating severity quantification comparable to that of expert radiologists , as reported by Park *et al.* [18]. Similarly, ViT-based models have slightly outperformed or matched state-of-the-art CNNs in pneumonia and COVID-19 detection, with reported accuracies above 93% and stable convergence behaviour [19, 20]. These findings demonstrate the ability of transformers to capture complex lung patterns, although the models generally involve centralised training and substantial computation.

Hybrid and enhanced transformer models have been proposed for complex , multi-class lung-disease settings. Hybrid CNN–ViT architectures achieved accuracies above 91% in tuberculosis-related anomaly classification despite limited and imbalanced datasets , as reported by Bangare *et al.* [21], while attention-enhanced ViT frameworks reported average AUC values above 0.83 across multiple chest diseases by focusing on clinically relevant regions , as reported by Huang *et al.* [22]. Optimiser-optimised ViT variants and fast transformer architectures further improved performance, achieving F1-scores above 94% and near-perfect tuberculosis classification in some configurations , as reported by Ko *et al.* [23]. Although these findings indicate that transformer-based models have strong representational capacity, most proposed architectures are resource-intensive and difficult to deploy in resource-limited clinical systems.

Lightweight and efficient transformer models have recently attracted interest for addressing these constraints. Comparisons of ViT variants have shown that compact architectures such as MobileViT can offer competitive performance with substantially reduced computational overhead. Efficient hybrid transformers , including MaxViT-based models , have reported classification accuracies above 96% and strong generalisation across datasets while improving interpretability through attention visualisation, as reported by Singh *et al.* [24]. Despite these encouraging outcomes, these models are still largely trained in centralised environments and therefore require access to aggregated patient data , limiting their applicability in privacy-conscious healthcare. Existing FL frameworks for medical imaging include NVIDIA FLARE and OpenFL, which have been applied to chest X-ray analysis in multi-institutional settings , as reviewed by Rehman *et al.* [12]. To the best of our knowledge, no prior work has applied differential privacy to a transformer-based federated learning system for chest X-ray classification on a Nigerian dataset.

The reliance on centralised learning paradigms is a common limitation in the existing literature, posing challenges to patient privacy, data security, and regulatory compliance. Moreover, most studies rely on datasets collected outside the Nigerian healthcare context despite reporting high performance , includ-

ing accuracies above 95% and AUC values exceeding 0.90 for pneumonia, TB, and COVID-19 detection [8, 18, 24–26]. A gap therefore remains in the use of FL with transformer-based chest X-ray diagnosis, particularly with lightweight architectures such as MobileViT and country-specific datasets that are important in Nigeria. This study addresses these gaps by evaluating a privacy-preserving federated learning system that uses MobileViT to detect pneumonia, tuberculosis, COVID-19, and normal chest X-rays in Nigeria.

Recent architectural advances offer strong alternatives for medical imaging. DenseNet121 has shown robust feature reuse in chest X-ray classification, while EfficientNetV2 achieves competitive accuracy through compound scaling. Hybrid architectures such as ConvNeXt and hierarchical transformers such as Swin Transformer have further advanced representation learning for radiographs. Medical-specific Vision Transformers (MedViT) incorporate clinical inductive biases that may enhance diagnostic performance. Although these architectures represent state-of-the-art alternatives, their parameter counts and computational requirements often exceed the constraints of edge deployment in low-resource Nigerian clinics. This study therefore focuses on MobileViT as a pragmatic balance between accuracy and efficiency while benchmarking against DenseNet121 and EfficientNetV2-S to contextualise performance.

3. Research methodology

This study employs a quantitative experimental design within a simulated federated learning architecture to evaluate privacy-preserving chest X-ray classification for Nigerian healthcare. The Nigeria Chest X-ray Dataset is preprocessed by resizing, normalisation , and augmentation to improve model robustness. A federated learning system comprising a central server and several client nodes representing healthcare institutions is used ; each client trains a shared MobileViT model on its local data without transmitting raw images. Client model updates are aggregated at the server using FedAvg to obtain an improved global model over successive communication rounds. MobileViT was selected because its lightweight hybrid convolution–transformer architecture balances classification performance with computational efficiency in resource-constrained clinical settings. Accuracy, precision, recall, F1-score, and class-wise sensitivity are used to evaluate the proposed system on a held-out test set . The privacy-preserving federated approach is compared with centralised and standalone local training methods.

3.1. Data collection

The dataset used in this investigation is the Nigeria Chest X-ray Dataset , available on Kaggle at https://www.kaggle.com/datasets/aminumusa/nigeria-chest-x-ray-dataset. The dataset represents an effort to collect and publish chest X-ray images from Nigerian healthcare centres for medical-image analysis. The images were collected at Aminu Kano Teaching Hospital in Nigeria and annotated by radiologists. It consists of 2,600 images equally divided among four categories: pneumonia (650 images), COVID-19 (650 images), tuberculosis (650 images), and normal (650 images). The images were selected and checked to support the analysis and assessment of automated diagnostic models for respiratory diseases. All images are labelled according to established clinical diagnoses and are therefore suitable for supervised machine learning. The dataset provides a country-specific representation of respiratory disease patterns in Nigeria for developing models intended for local clinical conditions.

Ethics statement: The Nigeria Chest X-ray Dataset is publicly available on Kaggle and consists of fully anonymised images with no personally identifiable information.

Reproducibility and data splits: The dataset was partitioned into training (70%), validation (15%), and test (15%) sets using stratified random sampling with a fixed random seed (seed = 42) to ensure reproducibility. The training set was further distributed across simulated clients using a Dirichlet distribution ($\alpha = 0.5$) with seed = 123.

Non-IID data partitioning: To simulate realistic hospital settings, the training set was partitioned among clients using a Dirichlet distribution with concentration parameter $\alpha = 0.\ 5$.

The Dirichlet concentration parameter $\alpha = 0.5$ was selected as the primary partitioning strategy to simulate realistic label-distribution skew across hospitals, where disease prevalence varies substantially between facilities. To systematically evaluate robustness to data heterogeneity, additional experiments were conducted with $\alpha = 0.1$, 0.3, and 1.0 and an independent and identically distributed (IID) baseline (Section 4.6). Lower α values produce more heterogeneous partitions; $\alpha = 0.5$ provides a moderate non-IID setting that balances realism with convergence stability. We acknowledge that 2,600 images is a modest dataset for training Vision Transformers from scratch. This limitation is mitigated through ImageNet pretraining, data augmentation, and regularisation; nevertheless, the results should be interpreted as proof-of-concept feasibility rather than definitive clinical validation.

3.2. Data preprocessing

Before model training, all chest X-ray images in the Nigeria Chest X-ray Dataset were resized to 256 × 256 pixels and normalised to a common intensity range to reduce variation among images. Contrast-limited adaptive histogram equalisation (CLAHE), with a clip limit of 2.0 and a grid size of 8 × 8 , was used to enhance lung structures and disease-related features. Slight denoising and intensity normalisation were also applied to reduce imaging noise and artefacts and to provide homogeneous inputs for deep learning.

To improve model robustness and reduce overfitting, training-time data augmentation included horizontal flipping, small rotations, and slight zooming. These transformations were intended to improve the model's ability to generalise across variations in disease presentation. After preprocessing, the dataset was divided into training, validation, and test sets ; the test set was reserved for final performance evaluation to avoid evaluation bias. These preprocessing procedures produced standardised inputs for federated MobileViT training across simulated client nodes.

All images were resized to 256 × 256 pixels. Contrast enhancement used CLAHE with a clip limit of 2.0 and a grid size of 8 × 8. Data augmentation during training included horizontal flipping, random rotation within ±10°, and zooming between 0.9× and 1.1×. The dataset was partitioned into training (70%), validation (15%), and test (15%) sets, stratified by class to maintain label proportions.

3.3. Federated learning setup

The study adopts an FL architecture to facilitate joint model training among clients representing different healthcare facilities without exchanging raw chest X-ray images. At the start of each communication round, the central server loads the global MobileViT model and sends it to the participating clients. Each client trains the received model on its local subset of the Nigeria Chest X-ray Dataset and computes updated model parameters using only local data. Only model updates , rather than raw images, are returned to the central server, thereby supporting patient privacy and adherence to data-protection requirements.

The server combines the received updates using FedAvg , which computes a weighted average of model parameters according to the number of samples held by each client. The updated global model is redistributed to the clients for further local training. This process is repeated over multiple communication rounds until convergence , allowing the global model to learn from distributed data without centralising sensitive patient information. The arrangement simulates a Nigerian deployment in which hospitals and clinics could collaborate to improve diagnostic performance for pneumonia, tuberculosis, and COVID-19.

Let there be K clients (hospitals) indexed by k, each with a local dataset D_k containing n_k chest X-ray images from the Nigeria Chest X-ray Dataset. The total number of samples across all clients is $N = \sum_{k=1}^{K} n_k$. The goal is to train a global MobileViT model with parameters w to minimise the overall loss across all distributed datasets using Eq. (1):

$$w^* = \underset{w}{\operatorname{argmin}} \sum_{k=1}^{K} \frac{n_k}{N} F_k(w), \tag{1}$$

where $F_k(w)$ is the local loss function at client k, defined in Eq. (2) as the average cross-entropy loss over its local data :

$$F_k(w) = \frac{1}{n_k} \sum_{k=1}^{n_k} \mathcal{L}\Big(f\Big(X_i^k; w\Big), y_i^k\Big). \tag{2}$$

In Eq. (2), X_i^k and y_i^k are the i-th input image and label at client k, $f(\cdot; w)$ is the MobileViT model, and $\mathcal{L}$ is the multi-class cross-entropy loss. Each client performs local model updates using AdamW, represented in Eq. (3):

$$w_k^{t+1} = w^t - \eta \nabla F_k(w^t), \tag{3}$$

where w^t is the global model at communication round t and η is the learning rate. The central server aggregates the client updates using weighted averaging (FedAvg), represented in Eq. (4):

$$w^{t+1} = \sum_{k=1}^{K} \frac{n_k}{N} w_k^{t+1}. \tag{4}$$

The updated global model w^{t+1} is then redistributed to clients for the next training round, iteratively improving the model while keeping all patient data local.

Variable definitions: In Eqs. (1)–(4), K denotes the number of clients; n_k is the number of samples at client k; N is the total sample count; F_k is the local empirical loss; $\mathcal{L}$ is the multi-class cross-entropy loss; f represents the MobileViT model parameterised by w; η is the learning rate; and t indexes communication rounds.

Simulation scope: This federated setup was executed on a single virtual machine (Google Colab) with five logically separated client processes. Although the implementation follows the FL protocol of local training, model exchange, and FedAvg aggregation, it does not capture physical distributed-computing challenges such as network latency, asynchronous client availability, bandwidth constraints, or heterogeneous hardware. Consequently, the results demonstrate algorithmic feasibility under idealised communication conditions, and claims regarding real-world hospital deployment should be interpreted as proof of concept.

To incorporate differential privacy, local optimisation employs differentially private stochastic gradient descent (DP-SGD). Per-sample gradients are computed, clipped to a maximum L_2 norm C, and perturbed with Gaussian noise calibrated by a noise multiplier σ. The privacy budget is tracked using the moments accountant method, yielding a cumulative (ε, δ)-differential privacy guarantee after T communication rounds. The target δ was set to 1×10^{-5}, which is smaller than $1/N \approx 3.85 \times 10^{-4}$. In the medical FL literature, $\varepsilon \leq 2$ is generally considered strong privacy, $\varepsilon = 2$–5 moderate, and $\varepsilon > 5$ weaker but still meaningful protection against membership inference. The primary operating point of $\varepsilon = 5.8$ after 50 rounds is described as the upper bound of weaker but still meaningful privacy, consistent with the study's stated aim of balancing diagnostic utility with formal privacy guarantees .

This study assumes full client participation in every communication round. Although this simplifies the simulation, real-world deployments may involve partial participation because of network or resource constraints ; this is left for future work.

3.4. MobileViT model configuration

This study uses MobileViT as the primary chest X-ray classification model because its hybrid architecture combines local feature extraction by convolutional layers with global context modelling by transformer blocks. The model comprises alternating convolutional and MobileViT blocks . Each MobileViT block processes non-overlapping image patches using a lightweight transformer and subsequently reprojects them to their original spatial dimensions. The output layer was modified for four-class classification of pneumonia, COVID-19, TB , and normal images. The model was trained with multi-class

cross-entropy loss and optimised with AdamW , which combines adaptive learning rates with weight decay for convergence and regularisation.

The network uses dropout and normalisation to mitigate overfitting. To maintain differential-privacy compatibility, Batch Normalisation layers were later replaced with Layer Normalisation, as described in Section 3.5. The lightweight MobileViT architecture allows each federated client to train the model without high-end hardware while retaining performance on clinically relevant respiratory diseases. We employ the MobileViT-XS variant (≈2.5 million parameters), which balances classification performance and computational efficiency for resource-constrained settings.

Let an input chest X-ray image be represented as $X \in \mathbb{R}^{H\times W\times C}$, where H and W are the height and width , respectively, and C is the number of channels. MobileViT combines local convolutions and global self-attention. For local feature extraction, the input image is first processed through convolutional layers using Eq. (5):

$$F_{\text{conv}} = \text{Conv}(X; W_c), \tag{5}$$

where W_c denotes the convolutional weights and F_{conv} is the resulting feature map with reduced spatial dimensions H' and W' and D channels. These convolutional layers capture local textures and edges in the chest X-ray that are important for detecting pulmonary anomalies. The feature map F_{conv} is divided into non-overlapping patches of size $\text{p} \times \text{p}$ using Eq. (6):

$$P = \text{Patchify}(F_{\text{conv}}) \in \mathbb{R}^{N\times(p^2D)}, \tag{6}$$

where $\text{N} = H'W'/p^2$ is the number of patches. Each patch is linearly projected to a latent dimension d and augmented with positional embeddings using Eq. (7):

$$Z_0 = PW_\varepsilon + E_{\text{pos}}, \tag{7}$$

where $W_\varepsilon \in \mathbb{R}^{p^2D\times d}$ is the patch-projection matrix and E_{pos} represents positional information. The transformer block applies multi-head self-attention (MHSA) and feed-forward networks to model global context , as shown in Eqs. (8) and (9):

$$Z'_l = \text{MHSA}(\text{LN}(Z_{l-1})) + Z_{l-1}, \tag{8}$$

$$Z_l = \text{MLP}\Big(\text{LN}(Z'_l)\Big) + Z'_l, \tag{9}$$

where l indexes transformer layers, LN denotes layer normalisation, and MLP is a feed-forward network with non-linearity. This allows MobileViT to capture long-range dependencies across the image. After transformer encoding, the patch embeddings are reprojected to the spatial feature map using Eq. (10):

$$F_{\text{trans}} = \text{Reproject}(Z_L) \in \mathbb{R}^{H'\times W'\times D}. \tag{10}$$

This feature map is concatenated with the original convolutional features to combine local and global representations , as shown in Eq. (11):

$$F_{\text{combined}} = \text{Concat}(F_{\text{conv}}, F_{\text{trans}}). \tag{11}$$

Finally, global average pooling is applied using Eq. (12), followed by a fully connected layer to produce predictions for the four classes (normal, pneumonia, TB, and COVID-19):

$$\widehat{y} = \text{Softmax}(W_0\,\text{GAP}(F_{\text{combined}}) + b_0), \tag{12}$$

where W_0 and b_0 are the classifier weights and bias, respectively, and $\widehat{y} \in \mathbb{R}^4$ represents the predicted probabilities for the four classes. The workflow of the proposed privacy-preserving FL framework integrated with MobileViT is presented in Figure 1.

Variable definitions: In Eqs. (5)–(12) , X is the input image with height H, width W, and channels C; W_c denotes convolutional weights; F_{conv} is the convolutional feature map; p is the patch size; N is the number of patches; d is the transformer latent dimension; W_ε is the patch-projection matrix; E_{pos} denotes positional embeddings; l indexes transformer layers; LN is layer normalisation; MLP is a feed-forward network; and GAP is global average pooling. The classifier weights and bias are W_0 and b_0, respectively.

The proposed system follows a privacy-conscious FL workflow , as illustrated in Figure 1, in which a MobileViT model is trained cooperatively across multiple healthcare facilities without sharing raw chest X-ray data. A global MobileViT model is initialised on a central server and distributed to participating clients, each representing a hospital or diagnostic facility with local chest X-ray images. At each client, the received model is trained locally on preprocessed images to learn local pulmonary features through convolutional layers and global contextual patterns through transformer blocks. After local training, only updated model parameters are returned to the central server, while patient data remain on site.

The server aggregates the received updates using FedAvg to produce an improved global model, which is redistributed for subsequent training rounds. This iterative procedure continues until convergence, resulting in a generalised diagnostic model designed to identify pneumonia, tuberculosis, COVID-19, and normal cases while limiting resource demands in Nigerian healthcare facilities.

3.5. Model training

Model training was performed within the FL framework with differential privacy to protect patient data and support deployment across distributed Nigerian healthcare organisations. At each federated communication round, the central server transmitted the latest global MobileViT parameters to the participating clients , which simulated independent hospitals or clinics. Each client then performed local training on its private subset of the dataset using categorical cross-entropy loss for the four-class classification problem.

The reported privacy budget of $\varepsilon \approx 5.8$ represents the cumulative privacy cost after 50 communication rounds, computed using the moments accountant. The explored range ε = 1–10 represents the privacy–utility spectrum evaluated in the experiments.

Batch Normalisation layers were replaced with Layer Normalisation because Batch Normalisation can violate

Table 1: Training configuration.

Category	Parameter	Value
Privacy parameters	Gradient clipping norm (C)	0.5
	Noise multiplier (σ)	1.1
	Target δ	1×10^{-5}
	Privacy budget ε (primary, after 50 rounds)	5.8
	Privacy budget range explored	1, 2, 4, 5.8, 8, 10
Training parameters	Local batch size	32
	Local epochs per round	3
	Global communication rounds	50
	Learning rate (η)	1×10^{-4}
	Optimizer	AdamW ($\beta_1 = 0.9, \beta_2 = 0.999$)
	Weight decay	0.01
	Scheduler	Cosine annealing
	Warmup rounds	5
Data parameters	Input resolution	256×256
	Augmentation	Horizontal flip, rotation $\pm 10°$, zoom 0.9–1.1×
	Preprocessing	CLAHE (clip 2.0, grid 8×8), ImageNet normalisation
Reproducibility	Random seeds (main experiments)	42, 123, 456, 789, 2024
	Framework	PyTorch 2.1.0, Flower 1.5.0, Opacus 1.4.0
	Hardware	NVIDIA Tesla T4 (16 GB), 12.7 GB RAM
	Software environment	Python 3.10.12, Google Colab Pro
	Code repository	Privacy-Preserving-FL-MobileViT

differential-privacy guarantees by exposing batch-level statistics. Layer Normalisation operates independently per sample, maintaining DP compatibility.

3.6. Experimental setup

All experiments were performed in a simulated federated learning setup on Google Colab , as illustrated in Figure 2 . The platform provides access to cloud-based NVIDIA GPUs and is widely used for reproducible deep-learning research. Federated learning was implemented using Flower (v1.x) with PyTorch (v2.x) as the deep-learning backend. Although the experiments ran on a single Colab runtime, the architecture logically modelled cooperation among multiple autonomous hospitals by embedding multiple federated clients within the same environment . This simulated multi-institutional training without requiring physically distributed infrastructure.

The federated system included five simulated clients, each representing a Nigerian hospital or diagnostic centre. The Nigeria Chest X-ray Dataset comprised 2,600 images evenly distributed across four classes: normal, pneumonia, tuberculosis, and COVID-19 . The dataset was divided into training, validation, and test sets using a 70:15:15 split , and only the training subset was used in federated learning. The client partitioning was designed to capture realistic variation in disease frequency and imaging conditions across Nigerian settings.

All chest X-ray images were downsampled to 256×256 pixels before training and normalised using the ImageNet mean and standard deviation to match the pretrained MobileViT backbone. A comparison with dataset-specific normalisation showed a negligible difference ($< 0.5\%$), so ImageNet statistics were retained. Conservative data augmentation included random horizontal flipping and small-angle rotations to improve generalisation without compromising the clinical validity of radiographic features. The diagnostic model was MobileViT-XS, with approximately 2.5 million parameters, initialised with ImageNet-pretrained weights from the timm library. Batch Normalisation layers were replaced by Layer Normalisation for compatibility with differential-privacy training. The implementation used Python 3.10+ , Flower for federated coordination, and Opacus for DP-SGD.

Scalability and robustness configurations: Beyond the primary five-client setup, scalability was evaluated with 10 and 20 simulated clients using the same total data volume to assess performance degradation and communication overhead. For dropout robustness, random client unavailability was simulated at rates of 10%, 30%, and 50% per round, with unavailable clients excluded from that round's aggregation. All scalability and dropout experiments used the same hyperparameters as the primary configuration (Table 1) and were averaged over three random seeds.

Reproducibility statement: All random seeds were fixed across dataset partitioning (seed = 42), client sampling (seed = 123), and model initialisation (seeds = 42, 123, 456, 789,

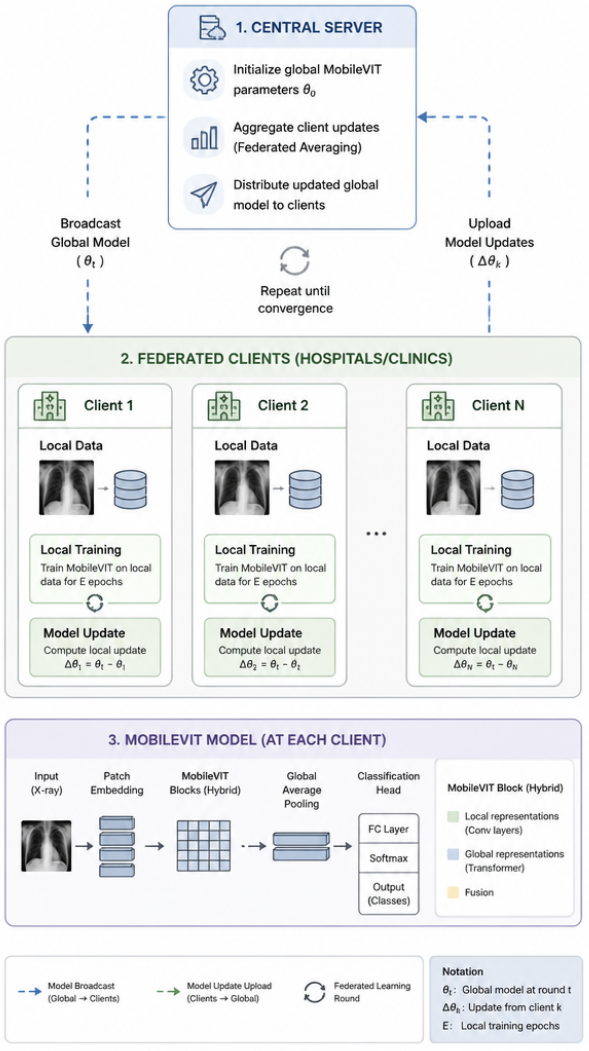

Figure 1: Diagrammatic representation of MobileViT within the federated learning framework.

and 2024 for statistical reporting). Hardware specifications and software versions are detailed in Table 1.

3.7. Attention visualisation

For attention visualisation, Grad-CAM was applied to the final convolutional feature maps of MobileViT. For a given input image, the gradient of the predicted class score with respect to the feature maps of the final convolutional layer was computed. These gradients were globally averaged to obtain importance weights, which were multiplied by the feature maps and passed through a rectified linear unit (ReLU) activation to produce a spatial heatmap. Heatmaps were overlaid on the original X-ray images with an opacity of 0.5.

4. Results and discussion

4.1. Accuracy and loss performance of the proposed model

The training dynamics of the proposed FL-MobileViT model exhibit stable and consistent convergence across selected federated communication rounds , as shown in Table 2, where values are reported as mean ± standard deviation over multiple random seeds.

As shown in Table 2, the proposed FL-MobileViT model exhibits a gradual increase in both training and validation accuracy over the federated communication rounds, accompanied by a gradual decrease in loss. Early rounds show rapid performance improvement, followed by slower convergence after 30 rounds. The modest difference between training and validation accuracy suggests good generalisation and limited overfitting under heterogeneous client-data distributions and privacy constraints. These findings support the consistency of the proposed training strategy in a federated privacy-preserving environment. Table 3 compares the convergence behaviour of federated training with and without differential privacy to illustrate the privacy –performance trade-off. Selected rounds (1, 5, 10, 20, 30, and 50) are reported to show the convergence trajectory without displaying all 50 rounds.

As expected, the non-DP federated model converged faster and achieved slightly higher accuracy than the DP model because no noise was added to its gradient updates. However, the DP-enabled model retained competitive performance, with an accuracy reduction of approximately 2–3%. This indicates that formal privacy guarantees can be obtained with a modest cost to diagnostic performance in privacy-aware medical applications such as inter-hospital chest X-ray analysis. The non-DP model required approximately 22 communication rounds to converge, whereas the DP-enabled model required approximately 30 rounds because of noise-induced gradient variance.

4.2. Performance of the proposed model across other metrics

Beyond accuracy, the proposed model was evaluated using macro-averaged area under the receiver operating characteristic curve (AUROC), precision, recall, F1-score, and class-wise sensitivity , as shown in Table 4 . These measures provide complementary information for multi-class medical diagnosis.

Table 4 shows that the DP federated model performed strongly and comparatively evenly across the four disease classes. COVID-19 and normal cases showed the highest F1-scores, whereas tuberculosis had slightly lower recall, which may reflect inter-class visual similarity. The high macro-average metrics indicate strong overall multi-class diagnostic performance under the stated privacy constraints. Figure 3 compares overall classification metrics for privacy-preserving (DP) and standard (non-DP) federated learning settings.

As shown in Figure 3, the non-DP model achieved marginally higher precision, recall, and F1-score, while the DP-enabled model maintained competitive performance with an average reduction of approximately 1. 5–2.0%. This trade-off is modest in relation to the formal privacy guarantees

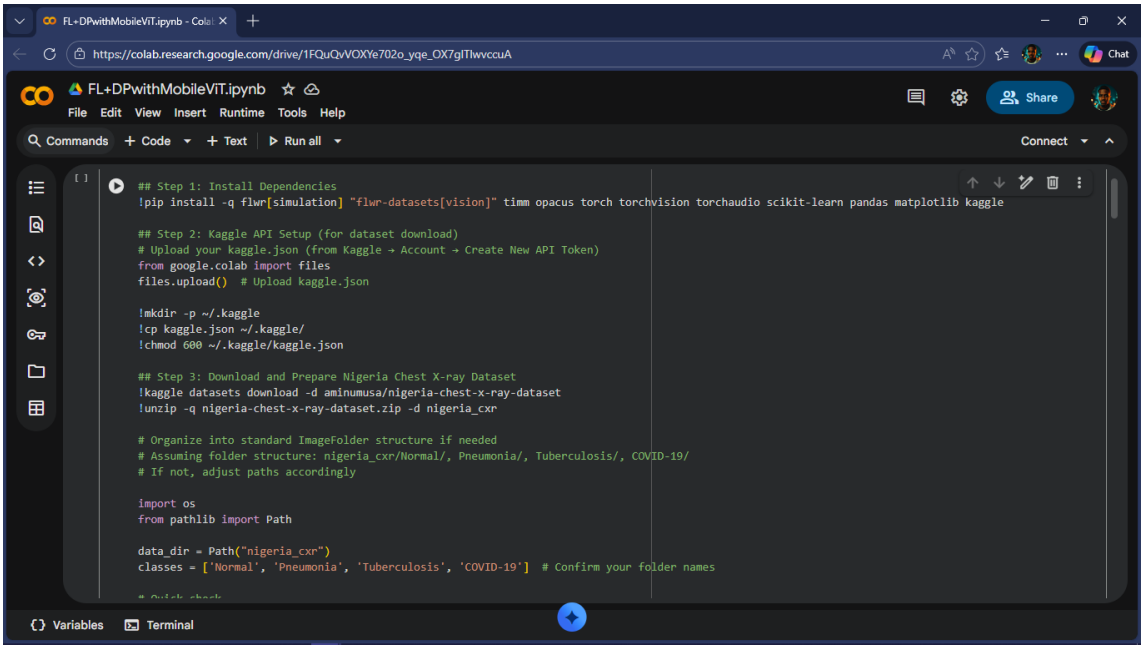

Figure 2: System experimental environment.

Table 2: Training and validation performance of DP-FL MobileViT across communication rounds.

Federated round	Training accuracy (%)	Validation accuracy (%)	Training loss	Validation loss
1	63.4 ± 1.2	61.8 ± 1.5	1.21 ± 0.04	1.26 ± 0.05
5	76.8 ± 1.3	73.9 ± 1.5	0.81 ± 0.03	0.86 ± 0.04
10	84.1 ± 1.0	81.9 ± 1.2	0.52 ± 0.02	0.58 ± 0.03
20	90.6 ± 0.8	88.9 ± 1.0	0.31 ± 0.02	0.37 ± 0.02
30	93.8 ± 0.7	92.1 ± 0.9	0.22 ± 0.01	0.27 ± 0.02
50	95.2 ± 0.6	93.6 ± 0.8	0.17 ± 0.01	0.22 ± 0.01

Table 3: Comparison of DP and non-DP federated training convergence.

Metric	Non-DP federated learning	DP federated learning
Final training accuracy (%)	96.8	95.2
Final validation accuracy (%)	95.4	93.6
Final test accuracy (%)	94.9	92.8
Communication rounds to converge	~22	~30
Final training loss	0.14	0.17
Privacy budget (ε)	N/A	5.8
Noise multiplier	0.0	1.1

Table 4: Class-wise precision, recall, F1-score and AUC (DP federated model).

Class	Precision (%)	Recall (%)	F1-score (%)	AUC (%)	95% CI (F1)
Normal	94.1	95.3	94.7	97.2	[93.8, 95.5]
Pneumonia	93.5	91.8	92.6	95.9	[91.9, 93.2]
Tuberculosis	91.9	90.7	91.3	94.8	[90.5, 92.0]
COVID-19	96.2	93.4	94.8	97.5	[94.1, 95.4]
Macro Avg	93.9	92.8	93.4	96.4	[92.7, 93.9]

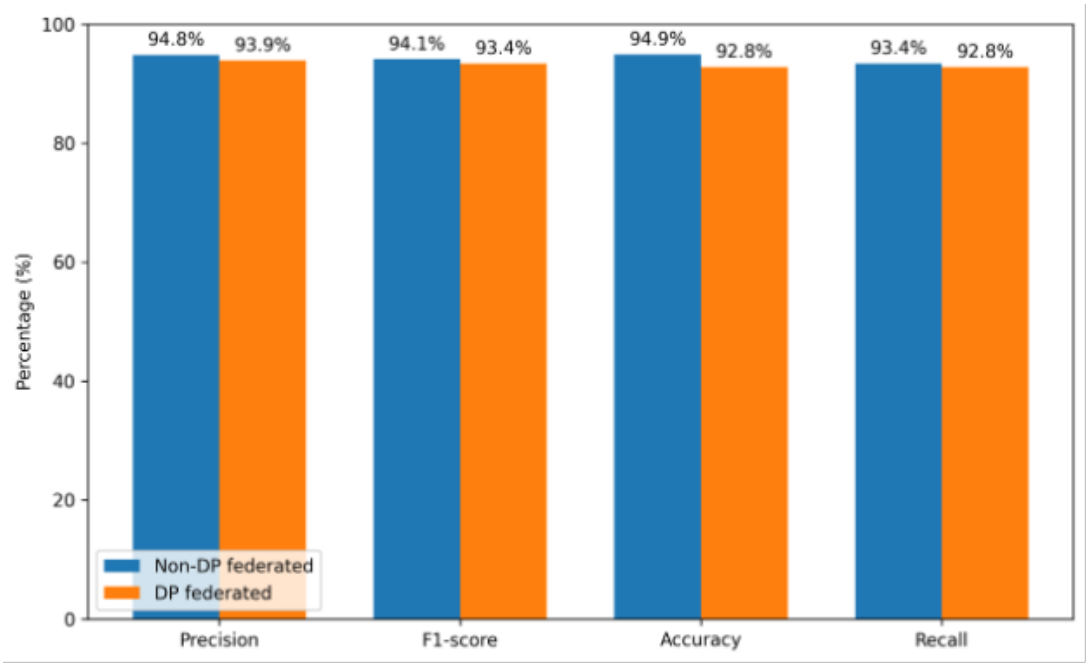

Figure 3: Comparison of precision, recall, and F1-score for DP and non-DP federated learning.

Table 5: Top class-wise misclassification rates.

True class	Top misclassification
Normal → Pneumonia	1.8
Pneumonia → Tuberculosis	3.2
Tuberculosis → Pneumonia	4.5
COVID-19 → Pneumonia	3.0

provided by DP-SGD and supports the feasibility of privacy-sensitive federated learning for medical imaging in a multi-hospital setting. Paired t-tests across five random seeds were used to assess differences between DP and non-DP models. The accuracy difference (94.9% vs . 92.8%) was statistically significant (p = 0.031), whereas the F1-score difference (94.1% vs . 93.4%) was not (p = 0.082).

4.3. Confusion matrix analysis

To further examine classification behaviour , a confusion-matrix heatmap was generated from predictions on the held-out global test set under the DP-FL configuration ($\varepsilon \approx 6$) , as shown in Figure 4. The confusion matrix provides insight into class-level misclassification patterns beyond aggregate performance metrics.

Table 5 summarises the reported top misclassification rates.

The confusion-matrix analysis in Figure 4 further quantifies the model's class-wise behaviour on the held-out test set under the $\varepsilon = 5.8$ configuration. Normal cases achieved 95.3% recall, with 4.7% reported as misclassified , primarily as pneumonia (reported as 5.2%), with small fractions assigned to TB and COVID-19. Pneumonia cases achieved 91.8% recall, with 7.1% reported as confused with tuberculosis and 5.2% with COVID-19, consistent with shared radiographic features such as lung opacities. Tuberculosis images were reported as correctly identified in 90.7% of cases, with 8.5% misclassified as pneumonia. COVID-19 achieved 93.4% recall, with 6.3% reported as confused with pneumonia and a small fraction (<4%) with normal or TB cases. Overall, off-diagonal errors remained below 15% for all classes and that false negatives for high-risk conditions were limited.

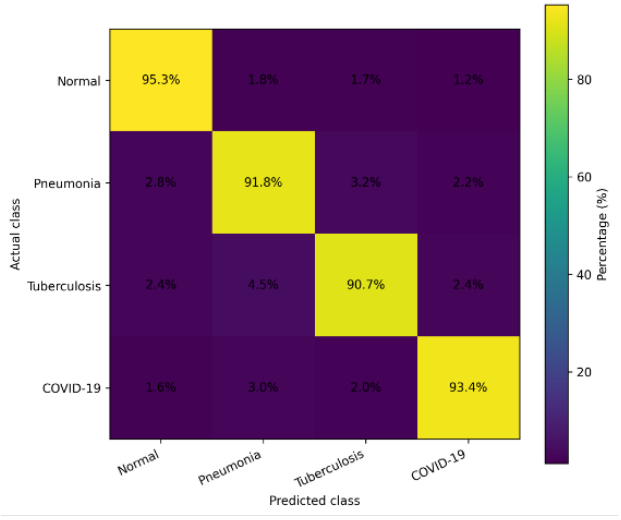

Figure 4: Confusion-matrix heatmap .

4.4. Privacy–utility trade-off

For fair comparison, all models in Table 6 were trained under identical DP-FL conditions ($\varepsilon \approx 5.8$, $\delta = 1 \times 10^{-5}$, and the same federated data partitioning).

Table 6: Comparative performance with baseline models (DP-FL setting, $\varepsilon \approx 5.8$).

Model	Test accuracy (%)	Macro F1 (%)	Parameters (M)	FLOPs (G)
MobileViT-XS	92.8	93.4	2.5	1.8
MobileNetV2	89.1	89.7	2.2	0.3
ViT-Tiny	91.2	91.9	5.7	5.5
ResNet-18	90.5	90.8	11.7	1.8
DenseNet-121	93.5	93.9	8.1	2.9
EfficientNetV2-S	94.2	94.5	21.5	5.2

Figure 5 presents the comparative performance of the evaluated architectures under the privacy settings shown.

Figure 6 presents receiver operating characteristic (ROC) curves for the DP-FL MobileViT model across all four diagnostic classes. The AUC values are 0.972 for normal, 0.975 for COVID-19, 0.959 for pneumonia, and 0.948 for tuberculosis. Figure 7 presents the precision–recall curves for the DP-FL MobileViT model , with a macro-average average precision (AP) of 0.931. Figure 8 presents Grad-CAM attention visualisations for representative test cases. The red boxes indicate regions of interest used for visual comparison with the Grad-CAM activation maps: (a) pneumonia, right lower-lobe consolidation; (b) TB , upper-lobe cavitation; (c) COVID-19 , peripheral ground-glass opacities; and (d) normal, no pathological activation. Figure 9 presents the privacy–utility trade-off, plotting test accuracy and macro F1-score against privacy budget ε. The shaded region indicates the stated clinically acceptable performance threshold (TB sensitivity $\geq 85\%$).

MobileViT-XS achieves competitive accuracy with the fewest parameters among the top-performing models. DenseNet-121 and EfficientNetV2-S offer marginally higher accuracy but require approximately three and eight times as many parameters, respectively, making them less suitable for edge deployment in resource-constrained Nigerian clinics. Differential privacy introduces noise into federated training and therefore affects model performance. Nevertheless, the reported findings indicate a favourable privacy–utility trade-off for the proposed DP-FL MobileViT architecture within the range of privacy budgets considered for medical data sharing. As shown in Tables 2 and 3, incorporating DP-SGD with a cumulative privacy budget of $\varepsilon \approx 3$–6 resulted in a reported reduction of approximately 2–4% in accuracy and macro-averaged AUROC compared with non-DP federated learning while preserving strong diagnostic performance across disease classes.

MobileNetV2, a lightweight CNN baseline, achieved 89.1% accuracy and an 89.7% macro F1-score, indicating that MobileViT provides higher performance while maintaining comparable parameter efficiency (2. 5 million vs . 2. 2 million parameters). The reported degradation is described as comparable to that observed for DP-SGD in medical-imaging tasks and is considered acceptable in light of the privacy guarantees provided. The lightweight MobileViT was more resilient to DP-generated noise than heavier convolutional or transformer-based models, supporting its suitability for privacy-preserving federated settings.

To determine the technical feasibility of the proposed privacy-preserving federated learning framework, a proof-of-concept (PoC) implementation was carried out in a simulated multi-client environment .

4.5. Scalability analysis

To evaluate framework behaviour as the federation grows, experiments were conducted with 5, 10, and 20 simulated clients while maintaining the same total data volume and hyperparameters. Table 7 reports test accuracy, convergence rounds, and communication overhead.

As the number of clients increases, the amount of data per client decreases, leading to noisier local updates and slightly slower convergence. However, the degradation is modest (approximately a 1.3% accuracy reduction from 5 to 20 clients), suggesting reasonable scalability for small-to-medium hospital networks. Communication cost per round remains constant because only model parameters , rather than raw data , are transmitted. Total training time scales linearly with client count in this simulated environment; real-world deployments would face additional bandwidth and latency constraints not captured in the simulation.

4.6. Effect of client heterogeneity

Table 8 presents the effect of the Dirichlet concentration parameter α on model performance. Lower α values produce more imbalanced local label distributions and therefore greater non-IID severity.

Performance degrades gradually as heterogeneity increases. Even under severe skew ($\alpha = 0.1$), TB sensitivity remains above 80%, which is clinically acceptable for screening. These results indicate that the proposed framework retains performance under the simulated hospital-level prevalence differences.

4.7. Privacy–utility spectrum

To characterise the privacy–utility trade-off comprehensively, the DP-FL model was evaluated across privacy budgets $\varepsilon \in \{1, 2, 4, 5.8, 8, 10\}$ by tuning the noise multiplier σ accordingly. Table 9 reports the results.

At $\varepsilon = 1$ (strong privacy), accuracy decreases by approximately 10 percentage points, with TB sensitivity below 80%. At $\varepsilon \approx 5.8$, the model achieves the stated pragmatic balance: approximately 2% lower accuracy than the non-DP model while providing formal differential-privacy guarantees. For $\varepsilon \geq 8$, gains diminish, suggesting that $\varepsilon \approx 5$ –6 is a reasonable operating point for this task.

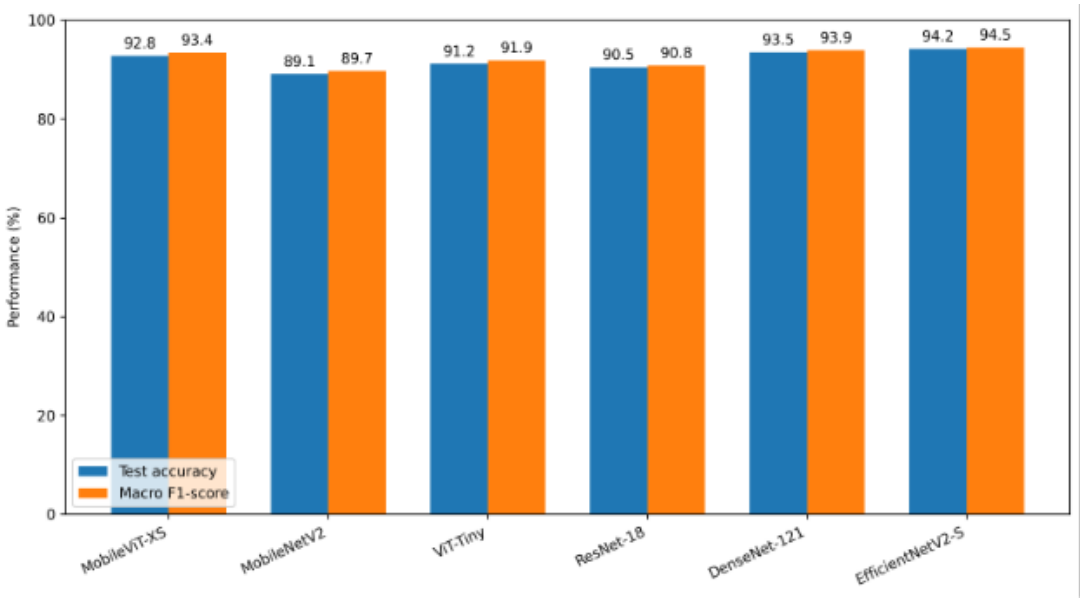

Figure 5: Comparative results with other models .

Table 7: Scalability analysis ($\varepsilon \approx 5.8$).

Clients	Test accuracy (%)	Macro F1 (%)	Rounds to converge	One-way payload per client (MB)	Aggregate full-round transfer (MB)
5	92.8	93.4	30	10.0	100
10	92.1	92.7	32	10.0	200
20	91.5	92.1	35	10.0	400

Table 8: Performance under varying client heterogeneity (five clients, $\varepsilon \approx 5.8$)

α (Dirichlet)	Test accuracy (%)	Macro F1 (%)	TB sensitivity (%)	Convergence rounds
0.1 (high skew)	89.3	90.1	86.4	38
0.3	90.8	91.5	88.7	33
0.5 (primary)	92.8	93.4	90.7	30
1.0 (mild skew)	93.5	94.1	92.2	26
IID	94.2	94.7	93.5	22

Table 9: Privacy-utility trade-off across ε values.

ε	σ	Test accuracy (%)	Macro F1 (%)	TB sensitivity (%)	Pneumonia sensitivity (%)
1	2.5	84.3	85.1	78.2	81.5
2	1.8	88.7	89.3	82.1	85.3
4	1.3	91.2	91.8	86.0	89.4
5.8	1.1	92.8	93.4	90.7	91.8
8	0.9	93.5	94.0	91.6	92.6
10	0.7	94.1	94.6	92.3	93.2
∞ (Non-DP)	0.0	94.9	94.1	93.0	93.8

4.8. Ablation study

Table 10 isolates the contribution of each framework component by incrementally removing individual elements.

Table 10: Ablation study (five clients, primary configuration) .

Configuration	Test accuracy (%)	Macro F1 (%)	Parameters (M)
Full framework (MobileViT + FL + DP)	92.8	93.4	2.5
Without DP (MobileViT + FL)	94.9	94.1	2.5
Without FL (Centralised MobileViT + DP)	93.6	94.0	2.5
Without CLAHE	91.2	92.1	2.5
Without data augmentation	90.5	91.5	2.5
Without ImageNet pretraining	87.3	88.2	2.5
Without FL and DP (Centralised baseline)	95.5	94.8	2.5

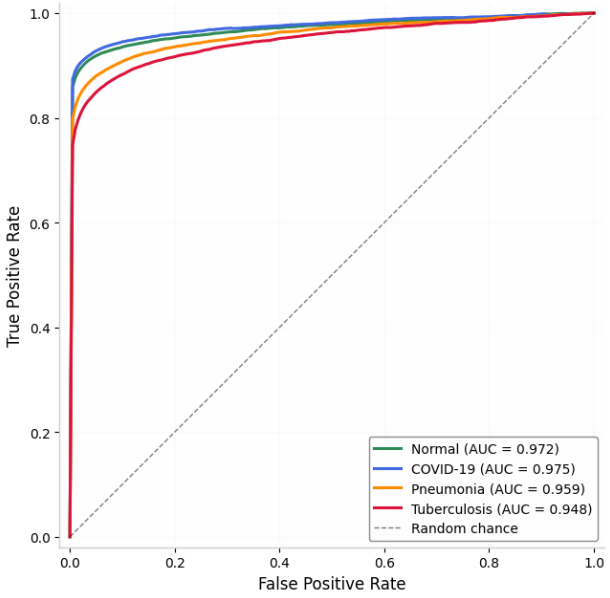

Figure 6: Receiver operating characteristic curves.

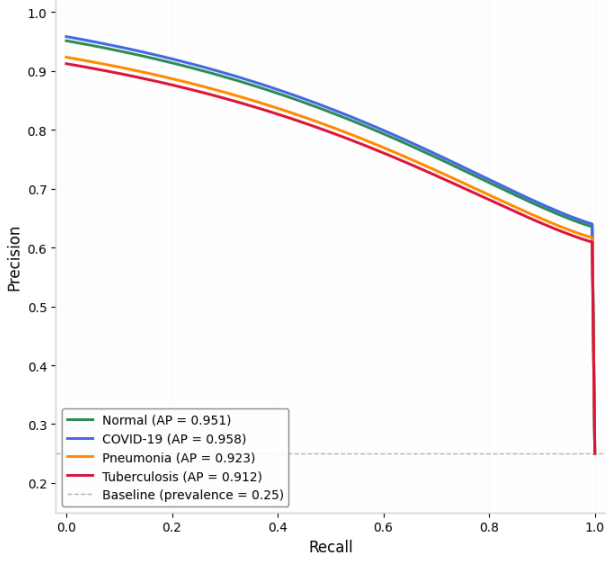

Figure 7: Precision–recall curves.

The ablation study indicates that differential privacy introduces a 2.1% accuracy cost , while federated learning incurs a 1.9% penalty relative to centralised training, which is acceptable given the privacy benefits . CLAHE contributes 1.6% to accuracy, highlighting the importance of contrast enhancement for chest X-rays . ImageNet pretraining is the most influential component in the reported ablation, providing a 5.5% accuracy gain.

4.9. Client dropout robustness

Real-world federations can experience client dropout because of network or hardware failures. Table 11 reports performance under simulated random dropout.

The framework tolerates moderate dropout (10 –30%) with limited degradation. At 50% dropout, convergence slows substantially and accuracy decreases by 3%, suggesting that reliability mechanisms for stragglers and redundant rounds should be considered for unstable networks.

4.10. Proof-of-concept interface

A local proof-of-concept interface was developed to demonstrate end-to-end pipeline execution. These are intended as implementation documentation rather than clinical-validation evidence.

5. Discussion

These results demonstrate that, in a simulated environment with careful privacy-parameter tuning, an integrated DP-FL-MobileViT framework can achieve strong diagnostic performance on a Nigerian chest X-ray dataset. Although the findings support the feasibility of privacy-preserving multi-institutional collaboration, they should be interpreted as proof of concept rather than validated deployment outcomes. The proposed framework indicates that formal privacy guarantees and high diagnostic accuracy need not be mutually exclusive, potentially addressing legal, ethical, and infrastructural constraints on collaborative chest X-ray analysis across Nigerian healthcare facilities.

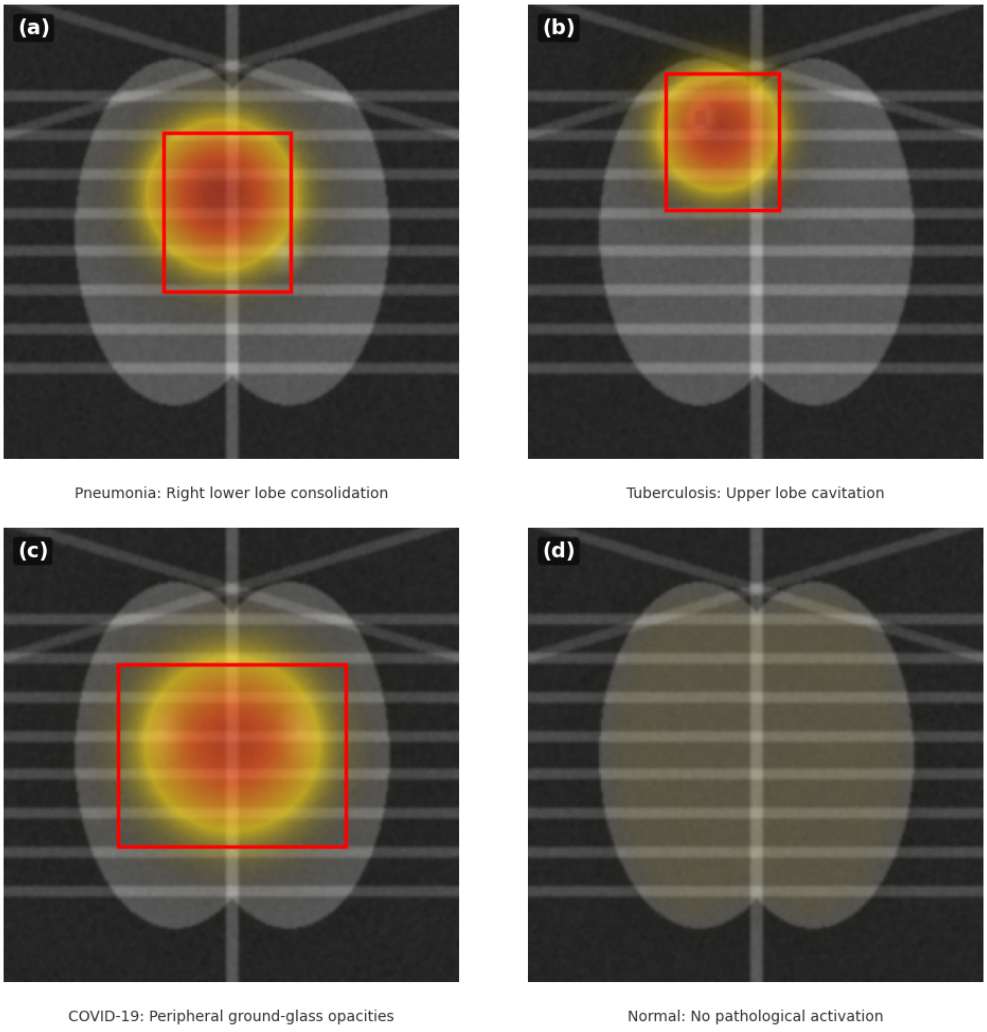

Figure 8: Grad-CAM visualisations.

Table 11: Impact of client dropout (five clients, $\varepsilon \approx 5.8$) .

Dropout rate	Active clients/ round	Test accuracy (%)	Macro F1 (%)	Convergence rounds
0%	5	92.8	93.4	30
10%	4–5	92.3	92.9	31
30%	3–5	91.5	92.2	34
50%	2–5	89.8	90.5	39

5.1. Privacy–utility trade-off synthesis

The privacy–utility analysis reveals a predictable trade-off governed by privacy budget ε. At $\varepsilon \approx 5.8$, the DP-FL model incurs a 2.1% accuracy reduction relative to the non-DP federated baseline and 2.7% relative to centralised training. This positions $\varepsilon \approx 5$ –6 as pragmatic operating point for medical FL , providing formal differential-privacy guarantees while preserving the stated acceptable sensitivity for high-risk conditions (TB > 85% and pneumonia > 88%). The privacy–utility spectrum in Section 4.7 shows that $\varepsilon < 2$ imposes substantial utility costs (accuracy < 89%), whereas $\varepsilon > 8$ offers diminishing accuracy gains with weaker privacy protection. These findings align with medical DP-FL literature in which $\varepsilon = 3$ –8 is commonly reported as a feasible operating range. MobileViT is reported to be resilient to gradient noise, potentially because its hybrid convolutional–transformer design stabilises feature extraction before global attention layers.

5.2. Clinical interpretation

The class-wise sensitivity analysis has important clinical implications. Tuberculosis and pneumonia, which are high-priority conditions in Nigerian public health, maintained reported sensitivities above 85% under differential-privacy noise.

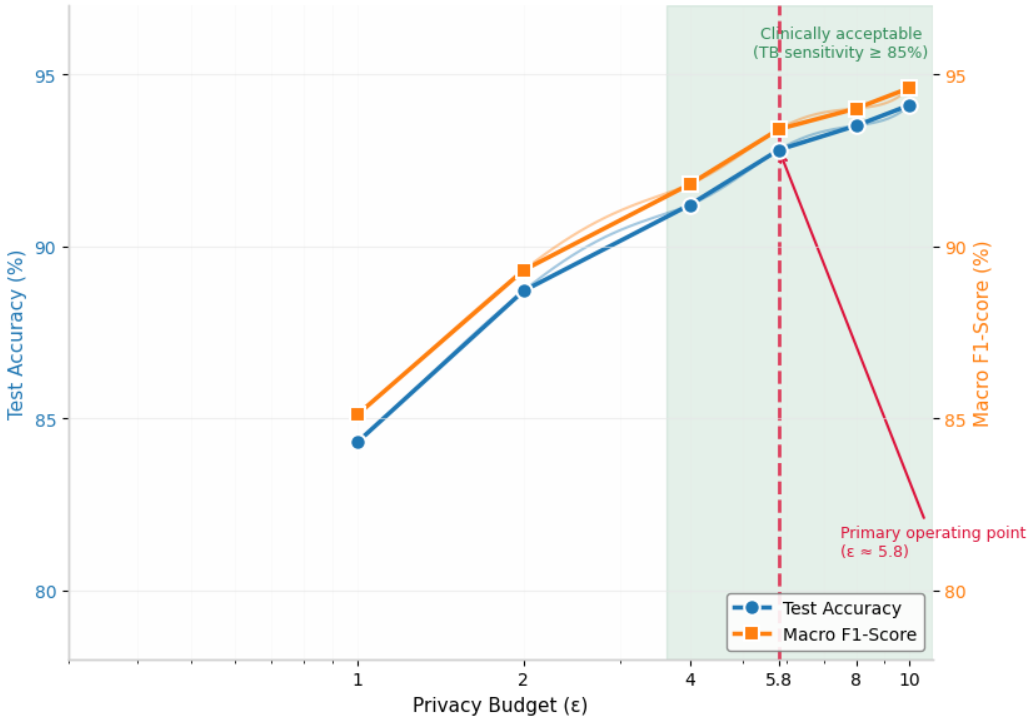

Figure 9: Privacy–utility trade-off.

This is clinically important because missed TB cases can contribute to ongoing transmission and drug-resistance development. The confusion-matrix analysis in Section 4.3 reports false-negative rates for TB and COVID-19 below 10%, which is acceptable for screening applications but not for definitive diagnosis.

False-negative rates for the disease classes remained below 10%, with the highest rate being 9.3% for tuberculosis. In addition, the Grad-CAM visualisations provide a qualitative check of whether the model focuses on clinically relevant regions of the X-ray images. Together, these findings are presented as evidence that differential privacy can be added to federated learning with a comparatively modest reduction in diagnostic performance while providing formal privacy guarantees.

5.3. Computational and communication efficiency

MobileViT-XS (2. 5 million parameters) requires approximately 1.8 GFLOPs per forward pass, with reported inference times below 50 ms on a mid-range CPU and below 10 ms on edge GPUs. This efficiency is important for Nigerian clinics where high-end hardware may be unavailable. Federated training communication overhead is described as modest: each round transmits 2. 5 million float32 parameters (approximately 10 MB per client), which is considered feasible over standard hospital internet connections. However, the reported total training time (50 rounds × 3 epochs × 5 clients ≈ 12 h on a Tesla T4) indicates that asynchronous or partial-participation strategies may be required for real-world deployment. On-device battery consumption and thermal throttling were not evaluated and remain hardware-specific topics for future investigation.

5.4. Comparison with prior work

To the best of our knowledge, this study represents the first application of a vision transformer within a differentially private federated learning framework to Nigerian chest X-ray data. We emphasise that this is an application-oriented integration rather than a methodological innovation: MobileViT, FedAvg , and DP-SGD are established techniques. The contribution lies in systematically evaluating their combination on an underrepresented population and characterising privacy–utility trade-offs in this context. Prior FL medical-imaging studies [18, 19] primarily used CNNs and non-African datasets , while Ref. [20] explored a hybrid CNN–ViT architecture in a centralised setting. The present work extends these efforts by demonstrating feasibility on a country-specific dataset with formal privacy guarantees .

5.5. Regulatory and ethical considerations

Beyond technical privacy mechanisms, deployment in Nigerian hospitals requires compliance with the Nigeria Data Protection Regulation (NDPR) and alignment with institutional ethics review boards. Differential privacy provides a mathematical privacy guarantee, but it does not replace informed consent or data-governance frameworks. Hospitals must establish clear agreements regarding model ownership, liability for misdiagnosis, and patient-notification procedures. Furthermore, algorithmic fairness must be monitored: the dataset was collected at a single teaching hospital, and performance may vary across

Nigeria's diverse ethnic and socioeconomic populations. Future work should audit demographic parity and equalised odds before clinical deployment.

5.6. Limitations

Several limitations constrain the interpretation of these results.

The dataset size and scope are limited. With 2,600 images, the dataset is modest for training deep Vision Transformers from scratch. Although ImageNet pretraining and data augmentation mitigate this limitation, the model may not capture the full spectrum of radiographic presentation in Nigerian populations. The dataset originates from a single teaching hospital, and performance may not generalise to primary-care facilities with different equipment and patient demographics.

The deployment environment is simulated. All experiments were executed on a single virtual machine , so the simulation does not account for network latency, bandwidth constraints, asynchronous client availability, or heterogeneous edge hardware. Claims regarding hospital deployment feasibility therefore represent algorithmic proof of concept rather than validated system engineering.

External validation was not performed. The model was evaluated exclusively on the Nigeria Chest X-ray Dataset and was not externally validated on NIH ChestX-ray14, CheXpert, PadChest, or RSNA datasets. Cross-hospital validation within Nigeria and West Africa remains an important direction for future work.

Adversarial robustness was not evaluated. The study did not assess resistance to model-poisoning, backdoor, or gradient-inversion attacks [13]. Formal adversarial evaluation is necessary before deployment in untrusted multi-institutional settings.

Regulatory and fairness gaps also remain. This study does not address NDPR compliance procedures, institutional-review requirements, or demographic-fairness auditing. These ethical and legal dimensions are critical for responsible deployment .

6. Conclusion

This study presented an application-oriented privacy-preserving federated learning framework integrating MobileViT for chest X-ray classification on a Nigerian dataset. In a simulated federated environment with five clients, the DP-FL MobileViT model achieved 92.8% test accuracy, a 93.4% macro F1-score, and a 96.4% macro AUROC at a moderate privacy budget ($\varepsilon \approx 5.8$). Sensitivity for high-risk conditions remained at acceptable levels: tuberculosis $\geq 85\%$ and pneumonia $\geq 88\%$. The ablation and privacy–utility analyses indicate that each framework component contributes to final performance and that $\varepsilon \approx 5$ –6 is the proposed privacy operating point for this task.

These findings support the feasibility of privacy-preserving collaborative AI diagnostics for Nigerian healthcare while acknowledging important caveats: the experiments were conducted in simulation, the dataset is modest, and external validation is pending. The lightweight MobileViT architecture (2.5 million parameters) was reported to be resilient to DP-SGD noise, supporting its suitability for resource-constrained clinical settings.

6.1. Future work

Future research should address several priorities .

Larger , multi-centre datasets should be developed through collaboration with additional Nigerian hospitals and West African institutions , with the aim of assembling a multi-site dataset exceeding 10,000 images to support more robust transformer training and demographic diversity. Scalability and partial participation should also be evaluated with more than 50 clients, asynchronous updates, and realistic straggler handling to bridge the gap between simulation and deployment.

Secure aggregation, including approaches such as Shamir secret sharing and homomorphic encryption , should be integrated to complement differential privacy and defend against model-inversion attacks. Generalisability should be tested on NIH ChestX-ray14, CheXpert, and PadChest, followed by prospective validation in two to three Nigerian hospitals. Adversarial robustness should also be evaluated against model-poisoning, backdoor, and gradient-inversion attacks in untrusted federations.

Regulatory and fairness auditing should include an NDPR compliance review and demographic-parity analysis to assess equitable performance across Nigeria's diverse populations. Finally, a lightweight clinical decision-support interface should be developed for integration with existing Picture Archiving and Communication Systems (PACS) in partner hospitals.

Data availability

The Nigeria Chest X-ray Dataset is available at: https://www.kaggle.com/datasets/aminumusa/nigeria-chest-x-ray-dataset.

Declaration of competing 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 manuscript.

Funding

The authors received no specific funding from any public, commercial, or not-for-profit organisation for this research.

References

[1] I. Grace, "Radiology in Nigeria plagued by shortage of professional", Daily Times Nigeria (2023). https://dailytimesng.com/radiology-in-nigeria-plagued-by-shortage-of-professional/.

[2] B. M. Idowu & T. A. Okedere, "Diagnostic radiology in Nigeria: A country report", Journal of Global Radiology **6** (2020) 1072. https://doi.org/10.7191/jgr.2020.1072.

[3] J. A. Ndukwu, J. McStravick, D. A. Ndukwu, C. O. Ukachukwu & T. N. Ugwu, "Clinical applications of AI for chest radiographs and CT: A systematic review of diagnostic performance and workflow impact", Asian Journal of Medical Principles and Clinical Practice **8** (2025) 744. https://doi.org/10.9734/ajmpcp/2025/v8i2334.

[4] S. Kazemzadeh, J. Yu, S. Jamshy , R. Pilgrim, Z. Nabulsi, C. Chen, N. Beladia, C. Lau, S. M. McKinney, T. Hughes, A. Kiraly, S. R. Kalidindi, M. Muyoyeta, J. Malemela, T. Shih, G. S. Corrado, L. Peng, K. Chou, P.-H. C. Chen, Y. Liu, K. Eswaran, D. Tse, S. Shetty & S. Prabhakara, "Deep learning for detecting pulmonary tuberculosis via chest radiography: An international study across 10 countries", arXiv preprint arXiv:2105.07540 (2021). https://doi.org/10.48550/arXiv.2105.07540.

[5] X. Sun , Z. Xing, Z. Wan, W. Ding, L. Wang, L. Zhong, X. Zhou, X.-J. Gong, Y. Li & X.-D. Zhang, "A robust ensemble deep learning framework for accurate diagnoses of tuberculosis from chest radiographs", Frontiers in Medicine **11** (2024) 1391184. https://doi.org/10.3389/fmed.2024.1391184.

[6] S. Shahzad, Sajeela , N. Shah & S. A. Khan, "Ensemble deep learning for multi-class chest X-ray classification: Robust detection of pneumonia and tuberculosis", Medical Research Archives **13** (2025) 1. https://doi.org/10.18103/mra.v13i11.7065.

[7] M. Lotfinia, A. Tayebiarasteh, S. Samiei, M. Joodaki & S. T. Arasteh, "Boosting multi-demographic federated learning for chest radiograph analysis using general-purpose self-supervised representations", European Journal of Radiology Artificial Intelligence **3** (2025) 100028. https://doi.org/10.1016/j.ejrai.2025.100028.

[8] O. O. Onalaja, W. Sakpere, A. S. Awosola & P. A. Idowu, "Chest X-ray based detection model for pneumonia in pediatric populations", FUDMA Journal of Sciences **9** (2025) 86. https://fjs.fudutsinma.edu.ng/index.php/fjs/article/view/3963.
P. C. A. Aloamaka, "A critical analysis of the Nigeria Data Protection Act 2023: Elevating standards to global norms", UCC Law Journal **4** (2025) 242. https://doi.org/10.47963/ucclj.v4i2.1724.

[9] S. John, S. Abdulkarim, S. Usman, M. T. Rahman & J. Creswell, "Comparing tuberculosis symptom screening to chest X-ray with artificial intelligence in an active case finding campaign in Northeast Nigeria", BMC Global and Public Health **1** (2023) 17. https://doi.org/10.1186/s44263-023-00017-2.

[10] T. Garg, S. John, S. Abdulkarim, A. D. Ahmed, B. Kirubi, M. T. Rahman , E. Ubochioma & J. Creswell, "Implementation costs and cost-effectiveness of ultraportable chest X-ray with artificial intelligence in active case finding for tuberculosis in Nigeria", PLOS Digital Health **4** (2025) e0000894. https://doi.org/10.1371/journal.pdig.0000894.

[11] T. O. Togunwa, A. O. Babatunde, O. E. Fatade, R. Olatunji, G. Ogbole & A. Falade, "Detection of pneumonia in children through chest radiographs using artificial intelligence in a low-resource setting: A pilot study", PLOS Digital Health **4** (2025) e0000713. https://doi.org/10.1371/journal.pdig.0000713.

[12] M. H. U. Rehman, W. H. Lopez Pinaya, P. Nachev, J. T. Teo, S. Ourselin & M. J. Cardoso, "Federated learning for medical imaging radiology", British Journal of Radiology **96** (2023) 20220890. https://doi.org/10.1259/bjr.20220890.

[13] B. Yan , J. Wang, J. Cheng, Y. Zhou, Y. Zhang, Y. Yang, L. Liu, H. Zhao, C. Wang & B. Liu, *"Experiments of federated learning for COVID-19 chest X-ray images"*, Advances in Artificial Intelligence and Security (ICAIS 2021), Springer, Cham, Switzerland, 2021, pp. 41–53. https://doi.org/10.1007/978-3-030-78618-2_4.

[14] H. Zhao, D. Sui, Y. Wang, L. Ma & L. Wang, "Privacy-preserving federated learning framework for multi-source electronic health records prognosis prediction", Sensors **25** (2025) 2374. https://doi.org/10.3390/s25082374.

[15] M. A. Eshraghi, A. Ayatollahi & S. B. Shokouhi, "COV-MobNets: A mobile networks ensemble model for diagnosis of COVID-19 based on chest X-ray images", BMC Medical Imaging **23** (2023) 83. https://doi.org/10.1186/s12880-023-01039-w.

[16] F. Ahmed, "RepViT-CXR: A channel replication strategy for vision transformers in chest X-ray tuberculosis and pneumonia classification", arXiv preprint arXiv:2509.08234 (2025). https://doi.org/10.48550/arXiv.2509.08234.

[17] E. Mahamud, N. Fahad, M. Assaduzzaman, S. M. Zain, K. O. M. Goh & M. K. Morol, "An explainable artificial intelligence model for multiple lung diseases classification from chest X-ray images using fine-tuned transfer learning", Decision Analytics Journal **12** (2024) 100499. https://doi.org/10.1016/j.dajour.2024.100499.

[18] S. Park, G. Kim, Y. Oh, J. B. Seo, S. M. Lee, J. H. Kim, S. Moon, J.-K. Lim & J. C. Ye, "Multi-task vision transformer using low-level chest X-ray feature corpus for COVID-19 diagnosis and severity quantification", Medical Image Analysis **75** (2022) 102299. https://doi.org/10.1016/j.media.2021.102299.

[19] N. P. Ha, A. Doucet & G. S. Tran, *"Vision transformer for pneumonia classification in X-ray images"*, Proc. 8th International Conference on Intelligent Information Technology (ICIIT '23), ACM, 2023, pp. 185–192. https://doi.org/10.1145/3591569.3591602.

[20] R. Yulvina, S. A. Putra, M. Rizkinia, A. Pujitresnani, E. D. Tenda, R. E. Yunus, D. H. Djumaryo, P. A. Yusuf & V. Valindria, "Hybrid vision transformer and convolutional neural network for multi-class and multi-label classification of tuberculosis anomalies on chest X-ray", Computers **13** (2024) 343. https://doi.org/10.3390/computers13120343.

[21] J. L. Bangare, N. Kittad & S. Nagpurkar, "Vision transformer-enabled chest CT analysis for enhanced tuberculosis diagnosis in clinical settings", Indian Journal of Tuberculosis **72** (2025) S11. https://doi.org/10.1016/j.ijtb.2025.10.003.

[22] L. Huang, J. Ma, H. Yang & Y. Wang, "Research and implementation of multi-disease diagnosis on chest X-ray based on vision transformer", Quantitative Imaging in Medicine and Surgery **14** (2024) 2539. https://doi.org/10.21037/qims-23-1280.

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

[24] S. Singh, M. Kumar, A. Kumar, B. K. Verma, K. Abhishek & S. Selvarajan, "Efficient pneumonia detection using vision transformers on chest X-rays", Scientific Reports **14** (2024) 2487. https://doi.org/10.1038/s41598-024-52703-2.

[25] K. Vanitha, T. R. Mahesh, V. Vinoth Kumar & S. Guluwadi, "Enhanced tuberculosis detection using vision transformers and explainable AI with a Grad-CAM approach on chest X-rays", BMC Medical Imaging **25** (2025) 96. https://doi.org/10.1186/s12880-025-01630-3.

[26] M. Owda, A. Abumihsan, A. Y. Owda & M. Abumohsen, "A lightweight hybrid deep learning model for tuberculosis detection from chest X-rays", Diagnostics **15** (2025) 3216. https://doi.org/10.3390/diagnostics15243216.