Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 6 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i06.59653

PAPER

From Federated Learning to Real-Time Pneumonia Prediction: Deploying an AdaptiveMesh-Based Global Model on Resource-Constrained Devices

Lamir Shkurti[1], Arsim Susuri[2] (✉), Vehebi Sofiu[1], Besnik Qehaja[1]

[1]UBT Higher Education Institution, Prishtina, Kosovo

[2]University "Ukshin Hoti", Prizren, Kosovo

arsim.susuri@uni-prizren.com

ABSTRACT

Federated learning (FL) enhances data privacy but faces significant deployment challenges on resource-constrained edge devices. Building on our previous development of the AdaptiveMesh algorithm, this study presents an end-to-end implementation of a resource-aware FL framework for pneumonia detection from chest X-ray images. The system is optimized for heterogeneous hardware and unstable network conditions. Validated on two independent datasets, the proposed global model achieved high diagnostic reliability, reaching up to 95% accuracy and critical recall performance for clinical use. A key contribution of this work is the practical transition from decentralized training to real-time prediction; we integrated the trained model into a lightweight Flask-based web application deployed on devices such as Raspberry Pi 4 and Intel Atom MiniPCs. Our results confirm that the AdaptiveMesh-based system provides a stable, privacy-preserving, and feasible solution for medical diagnostics in low-resource environments, bridging the gap between algorithmic optimization and real-world application.

KEYWORDS

adaptive federated learning (FL), wireless mesh networks, edge artificial intelligence (AI) deployment, medical image classification, resource-constrained devices

1 INTRODUCTION

In today's world, massive volumes of information are continuously generated by billions of embedded, mobile, and wearable devices. In the healthcare domain, medical sensors, imaging systems, and home monitoring devices produce sensitive patient data that is critical for building accurate predictive models. While this data offers enormous potential for improving diagnostics and clinical decision-making, it is subject to strict privacy, ethical, and regulatory constraints, such as the General

Shkurti, L., Susuri, A., Sofiu, V., Qehaja, B. (2026). From Federated Learning to Real-Time Pneumonia Prediction: Deploying an AdaptiveMesh-Based Global Model on Resource-Constrained Devices. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(6), pp. 162–182. https://doi.org/10.3991/ijim.v20i06.59653

Article submitted 2025-11-15. Revision uploaded 2026-02-03. Final acceptance 2026-02-06.

Data Protection Regulation (GDPR), making centralized data collection increasingly impractical.

Chest radiography (X-ray) remains one of the most widely used and cost-effective imaging modalities for diagnosing pneumonia. However, accurate interpretation of X-ray images requires experienced radiologists, a resource that is often scarce in low-resource and remote healthcare environments. Although artificial intelligence (AI) and machine learning (ML) techniques have demonstrated strong potential for automated medical image analysis, their deployment is frequently limited by privacy regulations that restrict the sharing of sensitive patient data across institutions.

Federated learning (FL) has emerged as a promising paradigm to address these challenges by enabling decentralized model training while keeping data localized on edge devices. By sharing only model updates rather than raw data, FL enhances privacy preservation and regulatory compliance. However, deploying FL in real-world healthcare settings introduces substantial challenges, particularly in heterogeneous and resource-constrained environments. Edge devices participating in FL often exhibit significant variability in CPU capacity, memory availability, thermal constraints, and network stability. Conventional FL approaches typically treat all clients uniformly, an assumption that leads to inefficiencies and system failures in such environments. Addressing these issues requires adaptive mechanisms that dynamically tailor training behavior to the real-time resource state of each participating device.

Our prior research [1–4] has progressively addressed these challenges through a series of studies on adaptive federated learning for medical image classification on low-power edge devices. Specifically, we introduced FederatedMesh [1] for FL over wireless mesh networks, BACA [2] for CPU and bandwidth-aware workload adaptation, and AdaptiveMesh [3], which extends adaptivity by incorporating CPU load, memory usage, temperature, and bandwidth constraints. These approaches were further validated across both physical edge platforms and virtualized cloud environments in [4], demonstrating improved stability and energy efficiency during federated training.

Building on this foundation, the present work shifts the focus from optimizing federated training alone to enabling a complete, end-to-end deployment pipeline. In contrast to our earlier studies, this work demonstrates the practical transition from decentralized training to real-time inference on resource-constrained edge devices.

Distinct from our earlier works, this study extends AdaptiveMesh beyond adaptive federated training by enabling an end-to-end pipeline from decentralized learning to real-time deployment on resource-constrained edge devices. The model is trained using a balanced subset of the COVID-19 Radiography Database [5], focusing on the NORMAL and Viral Pneumonia classes, and its generalization capability is further validated on the independent COVID-19 X-ray Dataset (Train & Test Sets) [6]. This two-dataset validation strategy ensures that the system's performance is not limited to a single data distribution and supports its applicability in diverse clinical environments. In contrast to prior centralized approaches and FL studies without deployment [7], [8], the trained global model is integrated into a lightweight Flask-based web application, enabling real-time pneumonia prediction on low-power devices such as Raspberry Pi 4 and Intel Atom MiniPCs.

1.1 Contributions

The key contributions of this study, which distinguish it from our previous works, are as follows:

End-to-End Transition: We demonstrate the practical transition from decentralized adaptive training to real-time inference on resource-constrained hardware.

Real-Time Deployment: We integrate the AdaptiveMesh-trained global model into a lightweight, Flask-based web application, enabling immediate pneumonia detection on devices such as Raspberry Pi 4 and Intel Atom MiniPCs.

Cross-Dataset Validation: To ensure clinical robustness, the system is evaluated using two independent datasets: the COVID-19 Radiography Database and the COVID-19 X-ray Dataset.

Quantification of Inference Performance: Unlike studies focused only on training accuracy, this work provides a quantitative analysis of resource consumption and latency during the prediction phase at the edge.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 describes the AdaptiveMesh algorithm, dataset preparation, and system architecture. Section 4 presents the Flask-based deployment for real-time inference. Section 5 reports experimental results and performance evaluation. Section 6 discusses the findings, and Section 7 concludes the paper.

2 RELATED WORK

Federated learning has increasing attention in the medical domain as a privacy-preserving paradigm for training machine learning models on sensitive patient data. Prior work spans a wide range of applications, including pneumonia from chest X-ray images, secure federated architectures, and edge-based learning systems. Nevertheless, most existing studies address these aspects in isolation, and only a limited number consider the combined challenges of adaptive training, device heterogeneity, and real-time deployment on resource-constrained edge devices.

2.1 Federated learning for medical image analysis

Several studies have explored FL for medical image classification, particularly for pneumonia detection. Rana and Marwaha [9] propose an unsupervised FL framework based on a modified LeNet-5 architecture and fuzzy c-means clustering, demonstrating competitive performance without labeled data. However, their work does not consider real-time deployment or adaptation to device-level constraints.

Similarly, Siddique et al. [10] present an FL-based COVID-19 classification system using decentralized chest X-ray datasets. While their study models client-server communication, it is limited to simulated environments and does not address adaptive workload management or deployment on resource-constrained hardware.

Chamarti et al. [11] introduce PneumoNet, a horizontal FL framework for pneumonia detection across healthcare institutions. Although their system achieves strong classification accuracy and demonstrates practical usability through a

graphical interface, it relies on standard federated averaging and does not incorporate adaptive mechanisms to handle heterogeneous device capabilities or fluctuating resource conditions.

Overall, these works demonstrate the feasibility of FL for privacy-preserving medical imaging but largely assume homogeneous clients and stable computational resources. Our work explicitly addresses device heterogeneity and real-time resource fluctuations through adaptive workload management.

2.2 Secure and heterogeneous federated learning frameworks

Other studies focus on privacy, security, and scalability in heterogeneous federated environments. Hoang et al. [12] propose APPFLx, a secure FL framework designed for biomedical applications across heterogeneous infrastructures. While their system ensures privacy and scalability, it primarily targets cloud-based deployments rather than real-time inference on edge devices.

Roy et al. [13] introduce a semi-synchronous FL framework with chaos-based encryption to enhance secure communication. Their approach reduces synchronization overhead and preserves accuracy in asynchronous settings; however, it does not incorporate adaptive training strategies based on device-level metrics such as CPU usage, memory limitations, or thermal conditions. Moreover, evaluation is performed in simulated environments rather than on physical edge hardware.

Sáinz-Pardo Díaz and López García [14] analyze FL performance under intermittent client participation and network heterogeneity, highlighting the impact of non-IID data and client dropout on convergence. While their analysis provides valuable insights into robustness, it remains simulation-based and does not include adaptive mechanisms that respond to real-time system constraints or support real-time inference deployment.

Tabassum et al. [15] demonstrated a practical Android application for depression detection using smartphone sensors, highlighting the feasibility of FL in mHealth while ensuring data privacy. Similarly, Ankayarkanni et al. [16] proposed P2FLF, a fog computing-based FL framework that integrates differential privacy and encryption, achieving competitive accuracy with reduced training time.

These frameworks advance security and scalability but remain focused on cloud or simulated environments. Crucially, they lack adaptive mechanisms that respond to real-time edge-device constraints during training, which is essential for stable operation in real-world, resource-constrained healthcare settings.

2.3 Edge and real-time medical AI systems

Beyond federated learning, several works investigate real-time or edge-based medical image analysis. Sun et al. [17] propose a robust centralized deep learning model for pneumoconiosis staging in the presence of noisy labels, improving generalization but without incorporating FL or edge-aware learning.

Patel et al. [18] present an explainable transfer learning framework using EfficientNet-B4 and Grad-CAM for multiclass lung disease classification. While their approach achieves high accuracy and improves clinical interpretability, it operates

in a centralized setting and does not consider federated training or deployment on resource-constrained devices.

Nair et al. [19] introduce a real-time COVID-19 detection system combining CNNs and extreme learning machines, optimized via a metaheuristic algorithm. Although the system demonstrates real-time prediction capability, it is fully centralized and does not address data privacy or device heterogeneity.

Qayyum et al. [20] propose a multi-modal federated learning framework integrating chest X-ray images and cough audio data for COVID-19 diagnosis at the edge. While their system explores edge deployment and modality fusion, it lacks adaptive mechanisms for dynamic resource management, such as monitoring CPU load, memory usage, or temperature during operation.

These systems achieve real-time performance and great accuracy but operate in centralized or non-adaptive federated environments. Our work bridges the gap between adaptive, resource-aware, federated training and deployment on low-power edge devices.

2.4 Adaptive and resource-aware learning approaches

Kareem et al. [21] propose a GDPR-compliant FL framework for pneumonia detection using pre-trained CNN models. Although their work preserves privacy and evaluates performance across different architectures, it does not address adaptivity to heterogeneous device constraints or real-world deployment on low-resource hardware.

Alshareef et al. [22] introduce a federated reinforcement learning (FRL) approach for kidney disease classification in IoT environments, incorporating energy optimization. While their work demonstrates adaptive learning behavior, it targets a different medical domain and does not explicitly control training behavior based on real-time system metrics such as CPU temperature or bandwidth availability.

While these approaches introduce adaptability, they either target different domains or lack real-time deployment validation. Our work extends adaptive FL to pneumonia detection and demonstrates practical feasibility in clinical edge environments.

2.5 Summary and research gap

In summary, existing studies demonstrate that FL can effectively preserve privacy in medical image analysis and that edge-based systems can support real-time inference. However, most prior work addresses these challenges independently. Few approaches jointly consider (i) adaptive training based on real-time device constraints, (ii) deployment on heterogeneous, resource-limited edge hardware, and (iii) end-to-end integration from federated training to real-time inference.

In contrast, our work addresses this gap by combining adaptive federated learning with a fully deployed, real-time diagnostic pipeline. The proposed AdaptiveMesh-based system dynamically adjusts training workloads based on device-level metrics and demonstrates practical feasibility through deployment on physical edge devices such as Raspberry Pi 4 and Intel Atom MiniPCs. This integrated approach advances the state of the art by bridging the gap between adaptive

federated training and real-world medical AI deployment in resource-constrained environments.

3 ADAPTIVEMESH TRAINING PROCEDURE

The AdaptiveMesh training procedure is designed to optimize FL performance in resource-constrained environments by adapting the training behavior of each client in real time. Unlike conventional FL approaches that treat all clients uniformly, AdaptiveMesh introduces a dynamic training mechanism tailored to each device's current computational state and network condition.

3.1 Client profiling and metrics monitoring

Each client device continuously monitors key hardware and network metrics, including:

- CPU usage (% utilization),
- CPU temperature (°C),
- Available RAM (MB),
- Training time per epoch (seconds),
- Bandwidth (Mbps).

These metrics are collected both before and during local training to inform decisions on the number of images, the number of local epochs, batch size, and the communication frequency.

3.2 Dynamic parameter adjustment

Based on real-time device metrics, the client-side training procedure adjusts workload to prevent overheating, memory exhaustion, and long training delays. Specifically:

- If CPU usage exceeds 80% or temperature exceeds 85°C, the number of local epochs and images is reduced.
- If bandwidth is below 2 Mbps, the client may skip or delay communication rounds.
- Devices with low training time and idle resources are allocated more data and training epochs.

3.3 Federated training cycle

Each FL round follows the standard federated workflow, enhanced with adaptive mechanisms:

1. The server distributes the current global model to all selected clients.
2. Clients locally train the model on their private data, applying adaptive workload settings.

3. After training, each client sends only model updates, not raw data, to the central server.
4. The server performs Federated Averaging (FedAvg) to aggregate updates and improve the global model.

The dataset used for training is dynamically sampled per round. This dynamic allocation ensures non-IID data distribution across clients and mimics real-world heterogeneity in data availability.

3.4 Training setup

The model was trained for 50 adaptive FL rounds using a subset of the COVID-19 Radiography Database, distributed among four heterogeneous edge clients:

- Raspberry Pi 4 (2x),
- Intel Atom MiniPC, and
- Standard laptop.

AdaptiveMesh successfully managed device heterogeneity by ensuring balanced training loads. Devices under high load trained for fewer epochs or smaller batches, while more capable nodes contributed proportionally more. This adaptivity resulted in improved stability, reduced overheating, and efficient training convergence.

4 FEDERATED LEARNING WORKFLOW AND SYSTEM ARCHITECTURE

Our FL system follows a client-server paradigm where edge devices participate in model training without sharing raw data. The workflow consists of four main steps: (1) Data remains local at the client; (2) The global model is distributed to clients; (3) Clients train locally and send model updates; (4) The server aggregates updates to refine the global model. Figure 1 illustrates this process along with deployment.

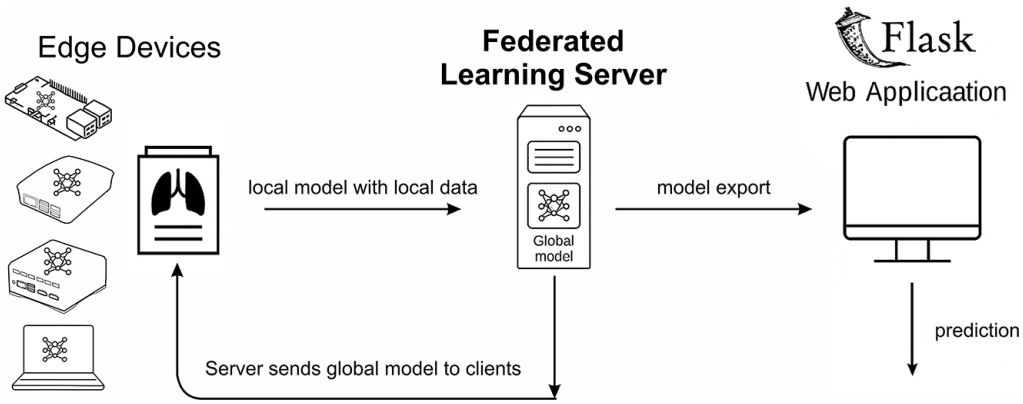

Fig. 1. System architecture: Federated learning across edge devices, followed by model deployment via Flask for real-time inference

4.1 Proposed system architecture

The proposed system targets privacy-preserving pneumonia detection from pediatric chest X-ray images, optimized for execution on heterogeneous and resource-constrained edge devices. The implementation is developed in Python using TensorFlow and Keras.

Model architecture. The classification model is based on EfficientNetB0, pre-trained on ImageNet due to its favorable balance between performance and computational efficiency. The network consists of convolutional and pooling layers followed by a global average pooling layer and a fully connected softmax output layer for binary classification (NORMAL vs. PNEUMONIA). A uniform weight initializer is used to ensure consistency across distributed training sessions.

Data preprocessing. All images are resized to 224 × 224 pixels to match the model's input requirements.

To optimize storage efficiency on resource-constrained edge devices, the dataset is stored on disk in uint 8 format (0–255). During both training and inference, images are dynamically converted to floating-point representation and preprocessed on-the-fly using the standard ImageNet preprocessing pipeline. This operation includes scaling and channel-wise normalization, ensuring compatibility with pre-trained CNN weights while minimizing disk and memory overhead on edge hardware. During training, real-time data augmentation is applied using the ImageDataGenerator API, including random rotations, horizontal flips, and zoom operations to improve generalization.

Training configuration. Local training is performed using the Adam optimizer with a learning rate of 1×10^{-4} and categorical cross-entropy as the loss function. Batch size, number of epochs, and the number of training images are dynamically adjusted by *AdaptiveMesh* based on real-time device metrics.

Each client runs a dedicated monitoring thread that records CPU usage, temperature, memory availability, bandwidth, and training duration. These metrics are logged locally and periodically reported to the central server to support adaptive decision-making.

4.2 Demonstration Flask-based deployment

To demonstrate real-world deployability, the final global model trained using AdaptiveMesh is integrated into a lightweight Flask-based web application. The application enables real-time pneumonia detection on edge devices.

During inference, the system performs the following steps:

1. Load the trained model into memory at runtime.
2. Resize uploaded X-ray images to 224 × 224 pixels.
3. Apply the same preprocessing pipeline used during training.
4. Perform classification and return predictions with confidence scores.

The web interface allows users to upload chest X-ray images and instantly receive predictions labeled as NORMAL or PNEUMONIA. This deployment demonstrates the complete pipeline from adaptive federated training to privacy-preserving, low-latency inference on resource-constrained hardware.

Figures 2 and 3 illustrate the user interface and prediction output, respectively.

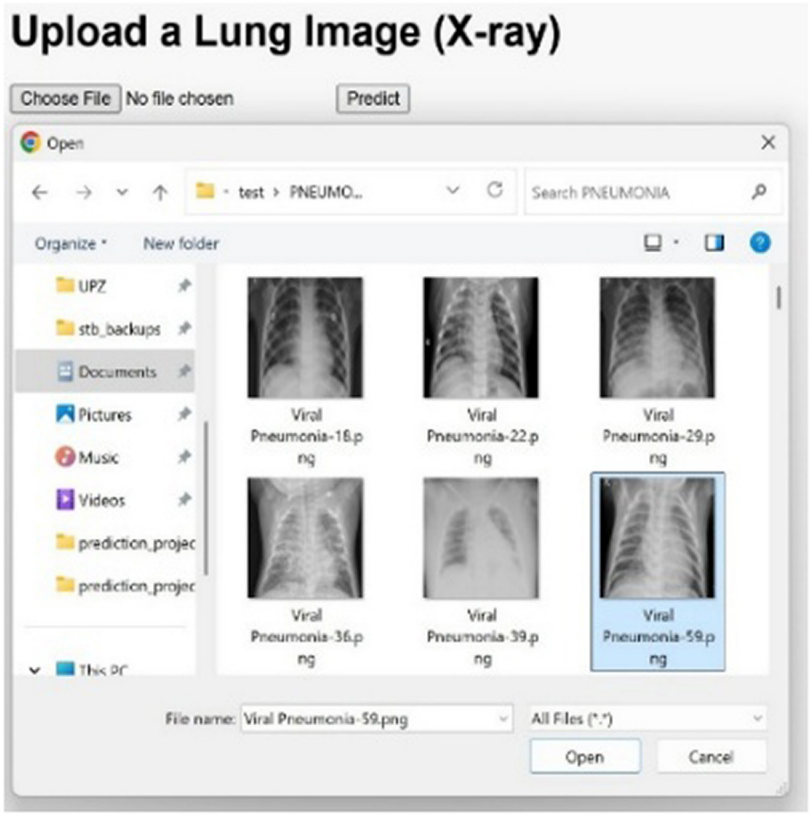

Fig. 2. User interface for uploading chest X-ray image

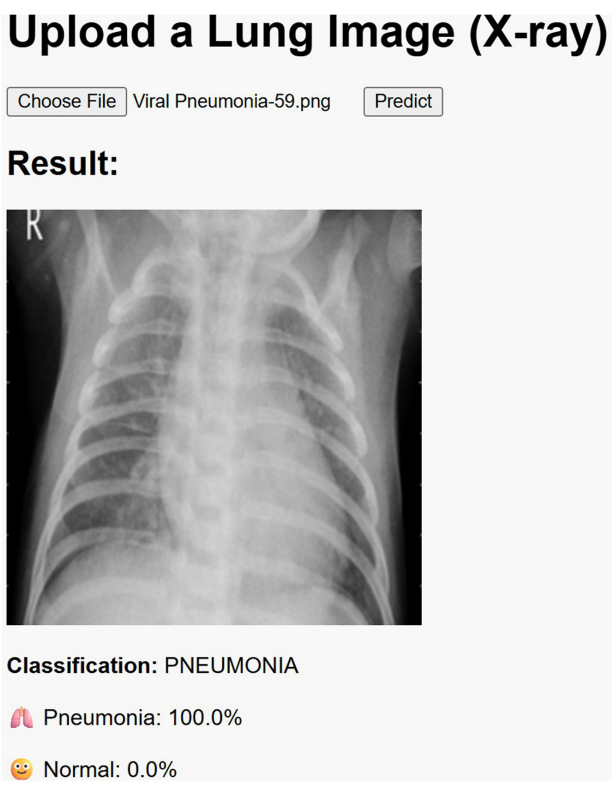

Fig. 3. Prediction result interface showing classification and confidence scores

5 RESULTS AND EVALUATION

To evaluate the performance of our deployed federated model for pneumonia detection, we conducted a detailed analysis using multiple classification metrics, including accuracy, precision, recall, F1 score, and confusion matrix. The global model trained with AdaptiveMesh was first tested on a non-IID test subset of the COVID-19 Radiography Database, considering two classes: PNEUMONIA and NORMAL. This evaluation provided insights into both generalization and potential overfitting behavior. Furthermore, to verify robustness across datasets, the model was also evaluated on the independent COVID-19 X-ray Dataset (Train and Test Sets), confirming its ability to generalize effectively beyond the primary training distribution.

5.1 Results on the COVID-19 radiography database

To evaluate the model trained with AdaptiveMesh, we used the balanced subset of the COVID-19 Radiography Database. From the original set of 1,345 PNEUMONIA and 10,192 NORMAL images, we randomly sampled 1,250 images per class to form a balanced dataset of 2,500 images. Using stratified sampling, the data was split into 1,000 images per class (2,000 total) for training and 250 images per class (500 total) for testing, followed by resizing to 224 × 224 pixels, normalization to [0, 1], and on-the-fly augmentation. Balancing was performed to avoid bias toward the majority NORMAL class, though this does not reflect the true clinical distribution where normal cases are far more frequent.

Figure 4 illustrates sample chest X-ray images from the Chest X-Ray Images (Pneumonia) dataset, highlighting clear visual differences between healthy lungs and those affected by pneumonia.

For evaluation, we generated and analyzed several metrics: the confusion matrix, classification report (including precision, recall, and F1 score per class), average precision scores, and precision-recall curves for each class. These metrics provided a comprehensive understanding of the model's performance and its ability to distinguish between normal and pneumonia cases.

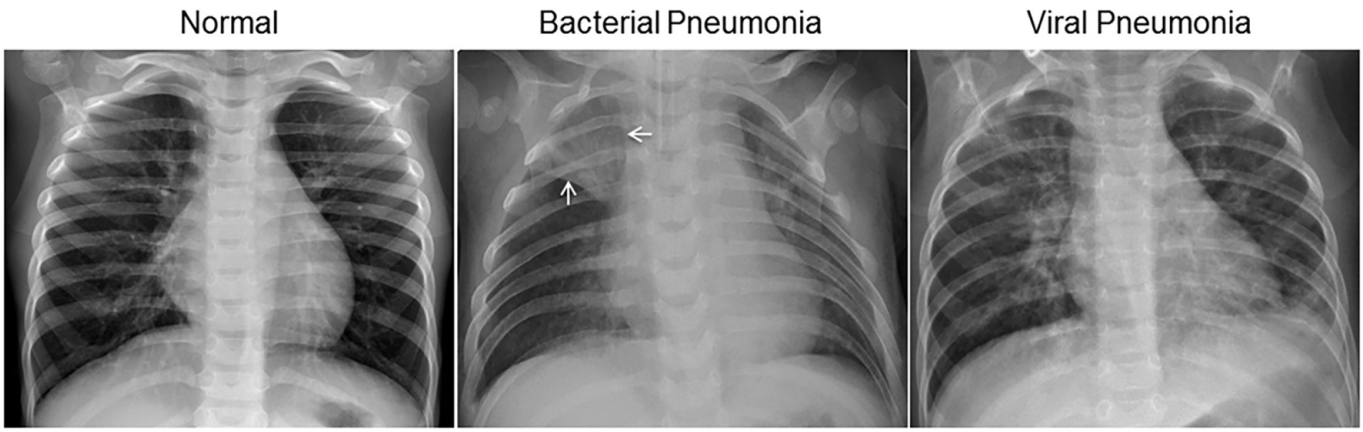

Fig. 4. Examples of chest X-rays in patients with pneumonia as highlighted in [23]

Performance on test set. Table 1 presents the evaluation of metrics on the test set using the trained EfficientNetB0 model. The model achieved an overall accuracy of 91%. Both classes obtained high F1 scores, with PNEUMONIA achieving 0.90 and NORMAL 0.91, showing a balanced performance across classes.

Table 1. Classification report on test set

Class	Precision	Recall	F1 Score	Support
PNEUMONIA	0.93	0.88	0.90	250
NORMAL	0.89	0.93	0.91	250
Accuracy	0.91	0.91	0.91	0.91
Macro Avg	0.91	0.91	0.91	500
Weighted Avg	0.91	0.91	0.91	500

Confusion matrix. The confusion matrix in Figure 5 shows that out of 250 pneumonia cases, 220 were correctly identified while 30 were misclassified as normal. For the normal class, 233 out of 250 were correctly predicted, with 17 misclassified as pneumonia.

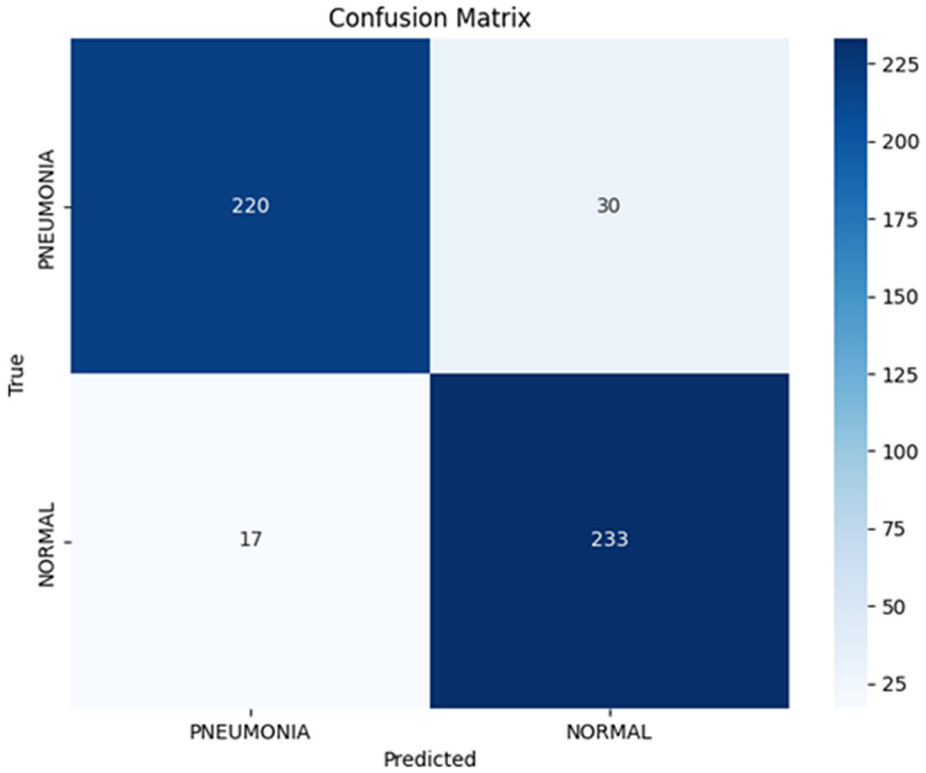

Fig. 5. Confusion matrix for pneumonia classification on the test set

Precision, recall, and F1 score per class. Figure 6 illustrates the precision, recall, and F1 score for each class, confirming the model's balanced ability to both detect pneumonia and correctly identify normal cases.

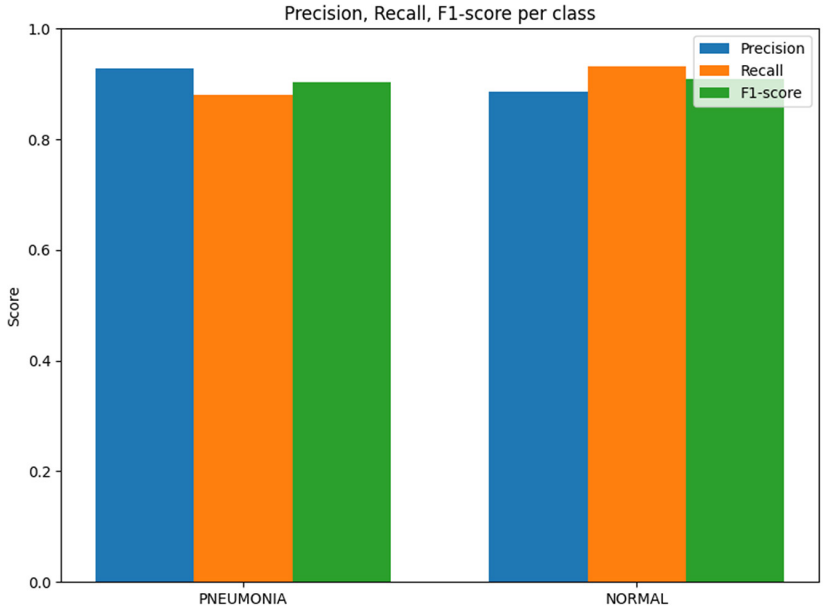

Fig. 6. Precision, recall, and F1 score per class

Precision–recall curve. Figure 7 displays the precision–recall curves for both classes. The curves remain close to the top-right corner, indicating high precision and recall values across thresholds, with an average precision score exceeding 0.91 for both classes.

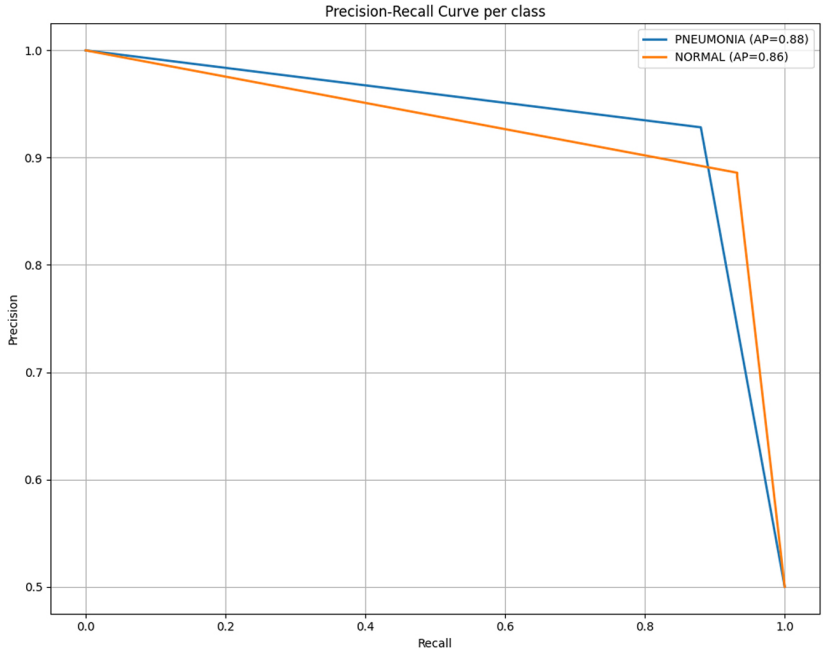

Fig. 7. Precision–recall curve for both classes

ROC curve and AUC. The receiver operating characteristic (ROC) curves for both classes are shown in Figure 8. The model achieved a high area under the curve (AUC) score of 0.91 for PNEUMONIA and 0.91 for NORMAL, indicating strong discriminative capability. The curves are consistently close to the top-left corner, reflecting the model's ability to maintain a high true positive rate while minimizing false positives across decision thresholds.

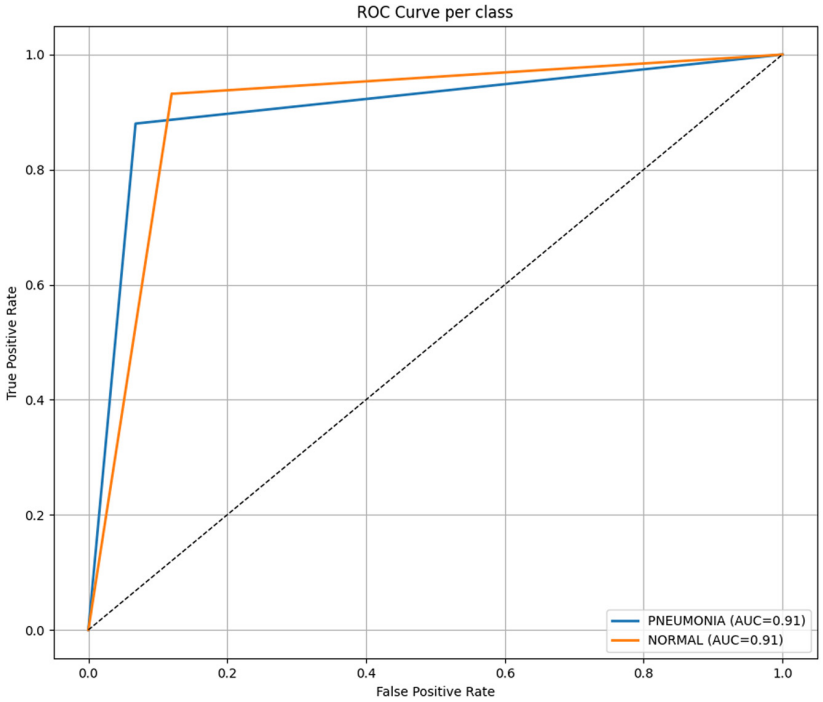

Fig. 8. ROC curve and AUC for both classes

Confidence intervals for performance metrics. Table 2 reports the 95% confidence intervals (CI) for the key evaluation metrics of the EfficientNetB0 model on the test set, obtained via bootstrapping with 2000 resamples. The results indicate consistently high performance for both classes, with narrow confidence intervals, highlighting the model's stability.

Table 2. 95% confidence intervals (Bootstrap, n = 2000) for AUC, F1 score, precision, and recall

Class	AUC	F1 Score	Precision	Recall
PNEUMONIA	0.906 (0.881–.931)	0.904 (0.876–0.930)	0.929 (0.893–0.962)	0.880 (0.840–0.919)
NORMAL	0.906 (0.881–0.931)	0.908 (0.883–0.934)	0.886 (0.847–0.923)	0.932 (0.899–0.963)

The AUC values of 0.906 for both classes reflect high discriminative power, with narrow CIs (±0.025) confirming the robustness of the model across resamples. F1 scores above 0.90, combined with high precision and recall, indicate a balanced tradeoff between false positives and false negatives. Such consistency suggests that the model is well-suited for deployment in real-world diagnostic applications, where reliability across varying data samples is essential.

5.2 Results on the COVID-19 X-ray Dataset (Train and Test Sets)

To further evaluate the robustness and generalization ability of the proposed AdaptiveMesh-based global model, we conducted an additional experiment using the COVID-19 X-ray Dataset (Train and Test Sets) [6] from Kaggle. Similar to the previous setup, we considered a binary classification problem (PNEUMONIA vs. NORMAL).

The preprocessing pipeline was identical: resizing all images to 224 × 224 pixels, normalizing to the [0, 1] range, and on-the-fly data augmentation.

Performance on test set. Table 3 presents the evaluation metrics on the Kaggle test set. The model achieved an overall accuracy of 95%. Both classes obtained high F1 scores, with PNEUMONIA achieving 0.95 and NORMAL also 0.95, demonstrating balanced performance across classes.

Table 3. Classification report on Kaggle test set

Class	Precision	Recall	F1 Score	Support
PNEUMONIA	1.00	0.90	0.95	20
NORMAL	0.91	1.00	0.95	20
Accuracy	0.95	0.95	0.95	0.95
Macro Avg	0.95	0.95	0.95	40
Weighted Avg	0.95	0.95	0.95	40

Confusion matrix. The confusion matrix in Figure 9 shows that out of 20 pneumonia cases, 18 were correctly identified while two were misclassified as normal. For the normal class, all 20 cases were correctly predicted with no misclassifications.

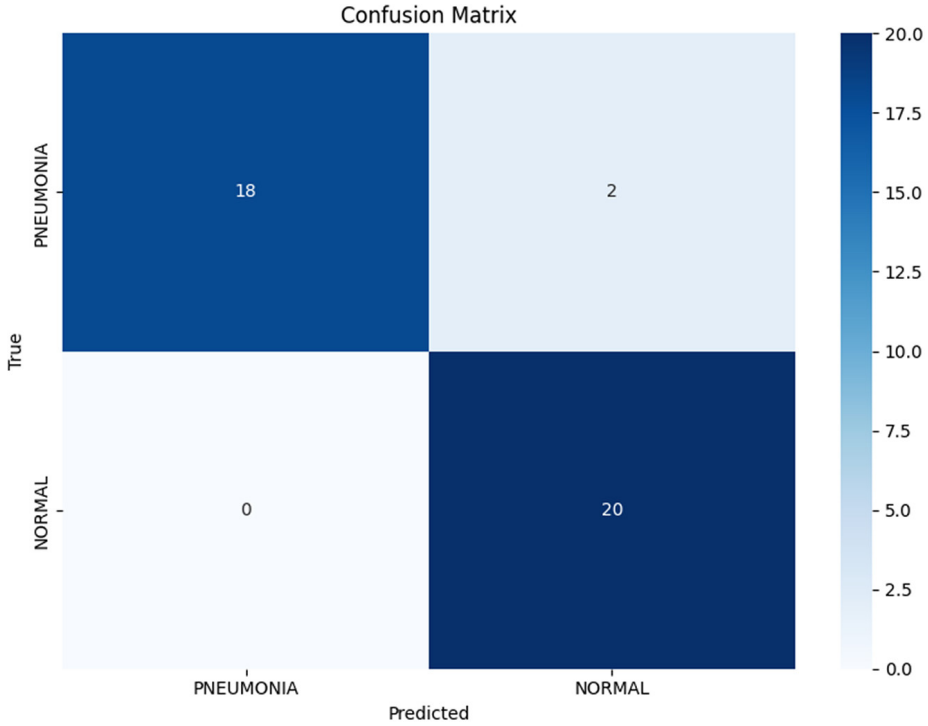

Fig. 9. Confusion matrix for pneumonia classification on the Kaggle test set

Precision, recall, and F1 score per class. Figure 10 illustrates the precision, recall, and F1 score for each class, confirming the model's balanced ability to both detect pneumonia and correctly identify normal cases.

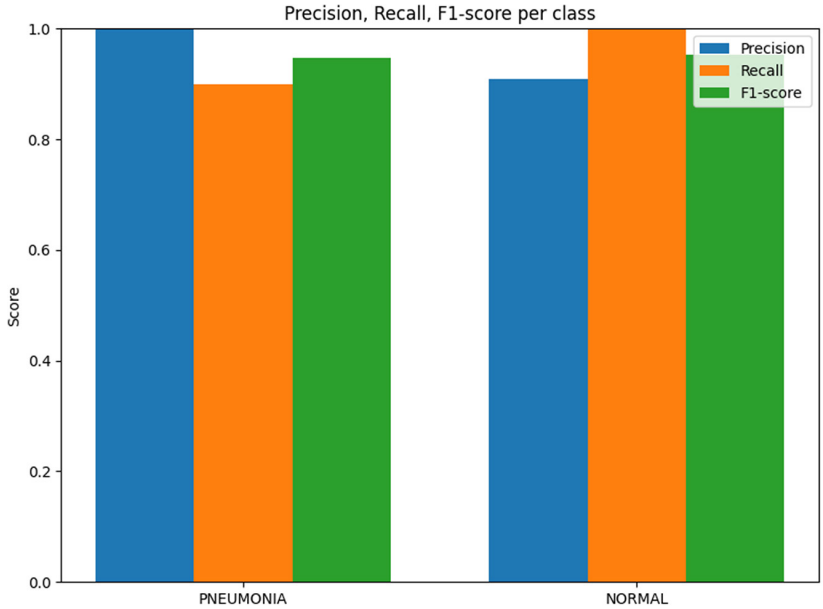

Fig. 10. Precision, recall, and F1 score per class (Kaggle dataset)

Precision–recall curve. Figure 11 displays the precision–recall curves for both classes. The curves remain close to the top-right corner, indicating high precision and recall values across thresholds. The average precision score exceeded 0.95 for both classes.

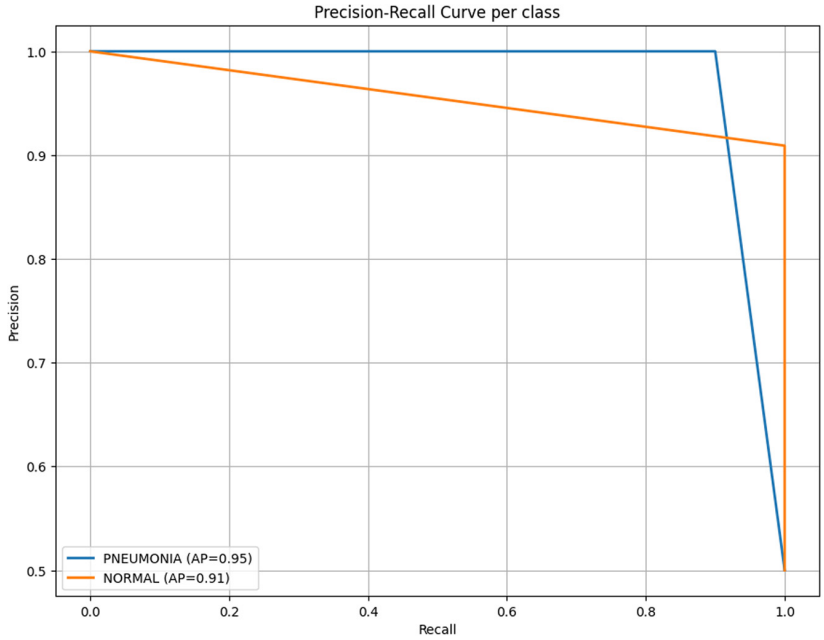

Fig. 11. Precision–recall curve for both classes (Kaggle dataset)

ROC curve and AUC. The ROC curves for both classes are shown in Figure 12. The model achieved an AUC score of 0.951 for PNEUMONIA and 0.951 for NORMAL, indicating strong discriminative capability. The curves are consistently close to the

top-left corner, reflecting the model's ability to maintain a high true positive rate while minimizing false positives across decision thresholds.

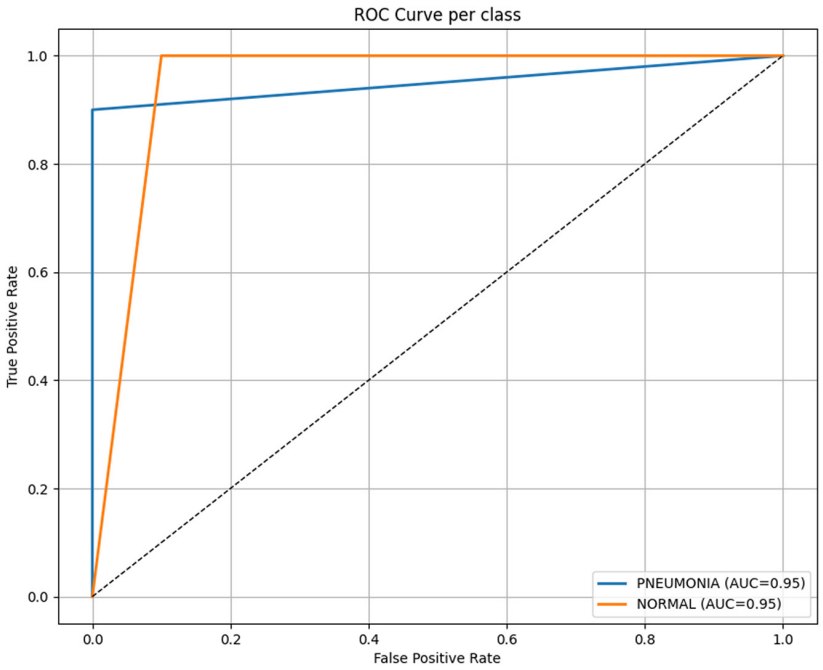

Fig. 12. ROC Curve and AUC for both classes (Kaggle dataset)

Confidence intervals for performance metrics. Table 4 reports the 95% CI for the key evaluation metrics of the model on the Kaggle test set, obtained via bootstrapping with 2000 resamples. The results indicate consistently high performance with narrow confidence intervals, highlighting the model's stability.

Table 4. 95% Confidence intervals (Bootstrap, n = 2000) for AUC, F1 score, precision, and recall (Kaggle dataset)

Class	AUC	F1-score	Precision	Recall
PNEUMONIA	0.951 (0.875–1.000)	0.947 (0.857–1.000)	1.000 (1.000–1.000)	0.902 (0.750–1.000)
NORMAL	0.951 (0.875–1.000)	0.952 (0.872–1.000)	0.910 (0.773–1.000)	1.000 (1.000–1.000)

5.3 Comparison with a standard FL baseline (FedAvg)

To assess the practical benefits of AdaptiveMesh, we compare it with the standard FedAvg algorithm, which is widely used as a baseline in federated learning. Unlike FedAvg, which assigns a uniform training workload to all participating clients, AdaptiveMesh dynamically adapts the local training configuration based on real-time device constraints. In our experiments on heterogeneous edge devices, FedAvg frequently led to thermal instability on resource-constrained clients such as Raspberry Pi 4, with CPU temperatures exceeding 80°C and occasional training interruptions. AdaptiveMesh mitigates this issue by proactively reducing the number of local epochs and training images when thermal or

utilization thresholds are reached, thereby maintaining system stability and preventing node dropouts (see Table 5).

This comparison focuses on deployment feasibility and operational stability rather than accuracy alone, which is particularly critical for real-world edge-based medical applications.

Table 5. Qualitative comparison between AdaptiveMesh and FedAvg in resource-constrained edge environments

Algorithm	Thermal Awareness	Bandwidth Adaptation	Resource Optimization	Training Stability
FedAvg	No	No	No	Low (on edge)
AdaptiveMesh	Yes	Yes	Yes	High

5.4 Inference performance on edge devices

After completing federated training, the final global model was deployed in a Flask-based web application for real-time pneumonia prediction. During inference, users upload a single chest X-ray image, and the system immediately returns the classification result (PNEUMONIA or NORMAL) along with a confidence score.

On a Raspberry Pi 4, the observed single-image inference latency averaged 1.18 seconds, resulting in an instant response from the user's perspective. CPU utilization during inference remained below 65%, and memory usage stayed under 420 MB. On the Intel Atom MiniPC, inference latency was consistently below 0.96 seconds, with CPU usage under 55% and memory usage below 380 MB. These results confirm that the proposed system supports real-time, low-latency medical image inference on resource-constrained edge devices.

6 DISCUSSION

The results presented in Section 5 demonstrate that the EfficientNetB0-based model, fine-tuned on a balanced subset of the COVID-19 Radiography Database, achieves strong classification performance in distinguishing between NORMAL and PNEUMONIA cases. The balanced sampling of 1,250 images per class mitigated the class imbalance present in the original dataset, contributing to stable precision and recall values across classes. This approach highlights the importance of careful dataset preparation in improving model reliability, particularly for minority classes such as PNEUMONIA. This balancing was necessary to prevent bias toward the majority class and to ensure fair evaluation of the model's discriminative ability. However, it does not reflect the true clinical distribution, where normal cases are far more prevalent. As a result, performance metrics on the balanced dataset may appear inflated compared to real-world prevalence.

The model reached a test accuracy of approximately 91%, with both classes achieving F1 scores above 0.90. Notably, the recall for the PNEUMONIA class was 0.88, indicating strong sensitivity in detecting critical cases with minimal false negatives. The precision values (PNEUMONIA: 0.93, NORMAL: 0.89) suggest that the model also maintains a low rate of false positives, supporting its potential integration into real-world screening workflows.

The performance can be partly attributed to the extensive data augmentation strategies applied during training, which enhanced the model's generalization to unseen images.

Evaluating the model on an independent COVID-19 X-ray Dataset further confirmed its robustness, achieving an overall accuracy of 95% and balanced F1 scores of 0.95 for both classes. These consistent results suggest that the model is not only dataset-specific but also generalizable.

While these results are promising, several limitations warrant discussion. The model's performance may vary when applied to populations or imaging conditions not represented in the training datasets, such as pediatric cases, different X-ray machines, or low-quality images. Additionally, inference on resource-constrained edge devices may be limited by computational capacity, requiring careful optimization of model size and execution speed. Future studies should also consider evaluating model interpretability and explainability to increase trust among clinicians and facilitate adoption in practice.

Looking forward, the integration of more advanced architectures, larger and more diverse datasets, and additional adaptive mechanisms in federated learning environments could further enhance model robustness and scalability. Extending this work to embedded IoT devices (e.g., TinyML applications) may enable effective deployment in resource-limited environments, allowing real-time, privacy-preserving pneumonia detection at the edge. Beyond the numerical results, a critical perspective highlights the practical implications of deploying AdaptiveMesh in real healthcare environments. While accuracy and recall values are encouraging, the true impact lies in the system's ability to operate reliably under heterogeneous device constraints and unstable networks. This emphasizes that federated learning solutions must be evaluated not only by statistical performance but also by their robustness, scalability, and clinical usability in resource-limited settings.

7 CONCLUSION

This study presented a practical, end-to-end pipeline for pneumonia detection using chest X-ray images. Two publicly available datasets were employed: the COVID-19 Radiography Database and the COVID-19 X-ray Dataset (Train and Test Sets). For the Radiography Database, a balanced subset was constructed by focusing on the NORMAL and PNEUMONIA classes, and an EfficientNetB0 model pre-trained on ImageNet was fine-tuned within a federated learning framework.

The resulting system achieved a test accuracy of 91%, an AUC of 0.906 for both classes, and high F1 scores (0.904 for PNEUMONIA, 0.908 for NORMAL) with narrow 95% confidence intervals, highlighting the model's robustness. Furthermore, evaluation on the independent COVID-19 X-ray dataset confirmed consistent performance with an overall accuracy of 95% and F1 scores of 0.95 for both classes. The balanced dataset design, combined with on-the-fly data augmentation, contributed to strong generalization performance without overfitting.

Given its compact architecture and high accuracy across multiple datasets, the model is well-suited for integration into federated learning frameworks, enabling privacy-preserving diagnostics in low-resource healthcare environments. This aligns with the broader vision of deploying AI-assisted screening tools directly on edge devices, reducing reliance on centralized computation while preserving patient data privacy.

7.1 Limitations and future work

Although the results are promising, several limitations remain. The model has not yet been validated in real clinical settings, where image quality, acquisition conditions, and patient demographics may differ from the datasets used. Additionally, while our dataset subsets were balanced for this study, the original COVID-19 Radiography Database exhibits significant class imbalance, which could impact performance in broader applications.

Future research directions include:

- Extending validation to diverse, real-world clinical environments.
- Investigating adaptive thresholding strategies to further optimize the sensitivity–specificity trade-off.
- Incorporating additional classes such as COVID-19 and lung opacity to broaden diagnostic coverage.
- Applying model compression techniques (e.g., pruning, quantization) to enhance deployment efficiency on ultra-low-power devices.
- Exploring cross-modality learning with CT scans or ultrasound for multimodal diagnostic support.

8 REFERENCES

[1] L. Shkurti, M. Selimi, and A. Besimi, "Federatedmesh: Collaborative federated learning for medical data sharing in mesh networks," in *Collaborative Computing: Networking, Applications and Worksharing*, H. Gao, X. Wang, and N. Voros, Eds., Springer, Cham, 2024, pp. 154–169. https://doi.org/10.1007/978-3-031-54531-3_9

[2] L. Shkurti and M. Selimi, "BACA: Bandwidth and CPU-aware adaptive federated learning for wireless environments," in *2024 13th Mediterranean Conference on Embedded Computing (MECO)*, 2024, pp. 1–5. https://doi.org/10.1109/MECO62516.2024.10577810

[3] L. Shkurti and M. Selimi, "Adaptivemesh: Adaptive federate learning for resource-constrained wireless environments," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 20, no. 14, pp. 22–37, 2024. https://doi.org/10.3991/ijoe.v20i14.50559

[4] L. Shkurti and M. Selimi, "Enhancing adaptive behavior in federated learning: The adaptivemesh algorithm for interactive mobile and bandwidth-limited resource-constrained wireless environments," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 19, no. 10, pp. 32–55, 2025. https://doi.org/10.3991/ijim.v19i10.54067

[5] T. Rahman *et al.*, "COVID-19 radiography database," *kaggle*, 2022. https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

[6] J. P. Cohen, "COVID-19 image data collection," *GitHub*, 2020. https://github.com/ieee8023/covid-chestxray-dataset

[7] F. Chollet *et al.*, Keras, 2015. https://github.com/keras-team/keras

[8] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning (ICML)*, 2020.

[9] N. Rana and H. Marwaha, "Pneumonia detection from X-ray images using federated learning–an unsupervised learning approach," *Measurement: Sensors*, vol. 37, p. 101410, 2025. https://doi.org/10.1016/j.measen.2024.101410

[10] A. A. Siddique, S. M. U. Talha, M. Aamir, A. D. Algarni, N. F. Soliman, and W. El-Shafai, "Covid-19 classification from X-ray images: An approach to implement federated learning on decentralized dataset," *Computers, Materials and Continua*, vol. 75, no. 2, pp. 3883–3901, 2023. https://doi.org/10.32604/cmc.2023.037413

[11] C. H. Kumar, E. Jangam, C. H. Kumar, and T. Philip Paul Arunodhayam, "Horizontal federated learning for advancing pneumonia detection with distributed medical imaging data," in *2025 6th International Conference on Recent Advances in Information Technology, (RAIT)*, 2025, pp. 1–8. https://doi.org/10.1109/RAIT65068.2025.11088951

[12] T.-H. Hoang *et al.*, "Enabling end-to-end secure federated learning in biomedical research on heterogeneous computing environments with APPFLx," *Computational and Structural Biotechnology Journal*, vol. 28, pp. 29–39, 2025. https://doi.org/10.1016/j.csbj.2024.12.001

[13] A. Roy, D. R. Mahanta, and L. B. Mahanta, "A semi-synchronous federated learning framework with chaos-based encryption for enhanced security in medical image sharing," *Results in Engineering*, vol. 25, p. 103886, 2025. https://doi.org/10.1016/j.rineng.2024.103886

[14] J. S.-P. Díaz and Á. L. García, "Study of the performance and scalability of federated learning for medical imaging with intermittent clients," *Neurocomputing*, vol. 518, pp. 142–154, 2023. https://doi.org/10.1016/j.neucom.2022.11.011

[15] N. Tabassum, M. Ahmed, N. J. Shorna, M. M. U. R. Sowad, and H. M. Z. Haque, "Depression detection through smartphone sensing: A federated learning approach," *Int. J. Interact. Mob. Technol.*, vol. 17, no. 1, pp. 40–56, 2023. https://doi.org/10.3991/ijim.v17i01.35131

[16] B. Ankayarkanni, N. K. Pani, M. Anand, V. Malathy, and Bhupati, "P2FLF: Privacy-preserving federated learning framework based on mobile fog computing," *Int. J. Interact. Mob. Technol.*, vol. 17, no. 17, pp. 72–81, 2023. https://doi.org/10.3991/ijim.v17i17.42835

[17] W. Sun *et al.*, "A deep learning method for pneumoconiosis staging on chest X-ray under label noise," *IEEE Access*, vol. 13, pp. 134031–134040, 2025. https://doi.org/10.1109/ACCESS.2025.3590783

[18] A. N. Patel *et al.*, "An explainable transfer learning framework for multi-classification of lung diseases in chest X-rays," *Alexandria Engineering Journal*, vol. 98, pp. 328–343, 2024. https://doi.org/10.1016/j.aej.2024.04.072

[19] T. Hu *et al.*, "Real-time covid-19 diagnosis from X-ray images using deep CNN and extreme learning machines stabilized by chimp optimization algorithm," *Biomedical Signal Processing and Control*, vol. 68, p. 102764, 2021. https://doi.org/10.1016/j.bspc.2021.102764

[20] A. Qayyum, K. Ahmad, M. A. Ahsan, A. Al-Fuqaha, and J. Qadir, "Collaborative federated learning for healthcare: Multi-modal COVID-19 diagnosis at the edge," *IEEE Open Journal of the Computer Society*, vol. 3, pp. 172–184, 2022. https://doi.org/10.1109/OJCS.2022.3206407

[21] A. Kareem, H. Liu, and V. Velisavljevic, "A federated learning framework for pneumonia image detection using distributed data," *Healthcare Analytics*, vol. 4, no. 100204, 2023. https://doi.org/10.1016/j.health.2023.100204

[22] M. A. Mohammed *et al.*, "Federated-reinforcement learning-assisted IoT consumers system for kidney disease images," *IEEE Transactions on Consumer Electronics*, vol. 70, no. 4, pp. 7163–7173, 2024. https://doi.org/10.1109/TCE.2024.3384455

[23] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122.e9–1131.e9, 2018. https://doi.org/10.1016/j.cell.2018.02.010

9 AUTHORS

Lamir Shkurti holds a Ph.D. degree from the Faculty of Contemporary Sciences and Technologies at Southeast European University in North Macedonia and is currently employed as an Assistant Professor at the University for Business and Technology (UBT) in Kosovo. His research interests include federated learning for resource-constrained platforms, embedded machine learning, the Internet of Things (IoT), wireless mesh networks, cloud computing, virtualization technologies, software development, and algorithm optimization.

Arsim Susuri holds a Ph.D. in Computer Science, an MSc in Electrical Engineering, and a BSc in Computer Engineering and Telecommunication. His research focuses on natural language processing, operating systems, artificial intelligence, machine learning, and computer architecture. He has published numerous papers. Dr. Arsim Susuri is an Associate Professor at the University Ukshin Hoti Prizren (E-mail: arsim.susuri@uni-prizren.com).

Vehebi Sofiu is an Associate Professor at the Faculty of Energy Engineering and Management and the Faculty of Computer Science and Engineering at UBT, Prishtina. He holds a Ph.D. in Energy and Computer Engineering, with a focus on mathematical modeling of the effects of solar panels and climatic conditions on alternative energy utilization. His professional experience includes project management in energy auditing, energy efficiency, and optimization. His research interests also include hybrid vehicles and platform-based energy models.

Besnik Qehaja holds a Ph.D. degree in Information and Communication Technologies from Corvinus University of Budapest and is currently an Associate Professor and Dean of the Department of Computer Science and Engineering at the University for Business and Technology (UBT) in Kosovo. His academic background is rooted in Software Engineering and Telecommunications. His research interests include digital transformation, educational technologies, telemedicine, artificial intelligence, Internet of Things (IoT), smart city infrastructures, learning management systems, and immersive technologies (VR/AR). He has led national and international projects in eHealth, AI-driven education, and smart digital ecosystems, contributing significantly to innovation in higher education and healthcare systems.