SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Adaptive Federated Learning With Intelligent Client Selection For Automated Chest X-Ray Pneumonia Detection

Miss. Nutan M. Dhande[1] , Dr. Anupa Sinha[2]

[1]Research Scholar Computer Science & Engineering Kalinga University, Naya Raipur nutandhande@gmail.com
[2]Assistant Professor Computer Science & Engineering Kalinga University, Naya Raipuranupa.sinha@kalingauniversity.ac.in

Abstract- While federated learning is a promising approach for privacy-preserving medical image analysis, optimal participation of the clients is essential for achieving good global model convergence, classification accuracy and communication efficiency, but it is not always observed. In this study, an adaptive Federated Learning with intelligent client selection for automatic Chest X-ray pneumonia detection is proposed. The goal is to design the best selection of clients for a distributed healthcare system in terms of maximized diagnostic accuracy, and minimised communication overhead. The proposed framework combines three intelligent client selection strategies: Random, Performance-Based, and Diversity-Based, Federated Averaging aggregation, privacy-preserving collaborative learning and the ImageNet-pretrained MobileNetV2. The Chest X-Ray Pneumonia images are split into 10 simulated hospitals and local models are trained independently with each hospital without sharing patient information and are then iteratively aggregated. The experimental results show that Performance Based selection round with 3 clients gives the best accuracy of the test of 83.36% which is more than the Random selection round (80.32%) and Diversity Based selection round (76.80%). In the 2-client scenario, the Random selection is the best with a score of 80.64%, surpassing the Performance-based (77.28%) and Diversity-based (80.16%). The higher the round participation of the clients, the better the test accuracy becomes with a 2.72% increase when the participation is increased from two to three clients per round, but the communication cost rises from 16 to 24 client-rounds. This novelty of the work is that for a given participation budget, the intelligent selection of clients has been studied under the same architecture, thus allowing for a thorough communication–accuracy trade off analysis. The study provides practical suggestions for adaptive client scheduling in privacy-preserving medical imaging systems and proves that the intelligent participation of patients with quality-aware substantially improves federated pneumonia detection, with retaining efficiency of communication and privacy of patient's data.

Keywords Federated Learning; Client Selection Strategies; Chest X-Ray Pneumonia Detection; MobileNetV2 Transfer Learning; Federated Averaging (FedAvg); Privacy-Preserving Medical Imaging

1. Introduction

Pneumonia is still one of the main causes of morbidity and mortality in the world and chest X-ray is still the most common imaging technique used to detect pneumonia in clinical practice [1]. Deep Convolutional Neural Networks (DCNNs) have been shown to have a great capacity to automate the interpretation of radiographs, and have been trained to perform at the level of an expert on large, centrally curated imaging datasets [2]. The use of these centralized models in clinic use is however restricted, as medical imaging data are scattered among many hospitals, and cannot be easily combined due to strict privacy regulations and patient confidentiality [3]. Federated learning is a promising paradigm that has recently appeared, in which the raw patient data remains locally with the institutions while the institutions collaboratively train a shared global model without compromising privacy while benefiting from having access to large datasets for collaborative training [4]. The

Federated Averaging algorithm is the "standard" approach for coordinating distributed training on heterogeneous clients from hospitals, with locally trained model updates being averaged over using a set of weights [5]. Although federated learning shows great potential in healthcare, it still suffers from a number of problems, such as statistical heterogeneity, frequent client withdrawals, and the high communication costs associated with transmitting the model updates over multiple communication rounds [6]. The question of client selection, that is which subset of the institutions participating in a given training round contributes to a given training round's convergence speed, and the final diagnostic accuracy, has increasingly become a topic of research. To go with the demands of federated deployment, lightweight architectures like MobileNetV2 are also deployed in order to ease the computational and bandwidth challenge at the resource-limited hospital nodes and to make transfer learning from a pretrained backbone, trained on the ImageNet dataset, an appealing approach for medical image classification tasks [8]. This work is inspired by these factors and examines the performance of the three different client selection strategies - random sampling, performance-based exploitation, and diversity-based round-robin rotation - for diagnostic accuracy and communication efficiency in a simulated cross-silo federated environment of ten hospital clients that perform automated pneumonia detection on paediatric chest radiographs. In contrast to designing model architectures, systematic variation of the selection strategy and the number of clients sampled per communication round, is a relatively under-researched aspect of adaptive client scheduling that this work aims to identify, and achieve practical, evidence-based guidelines for adaptive client scheduling that balance the diagnostic performance of the system with the communication cost incurred in a privacy-preserving medical imaging system. The results will feed into practitioners who deploy federated pipelines in hospital networks of limited communication bandwidth and participation reliability, where the quality of diagnostics achieved can be directly influenced by the quality of the communication.

The major points of this research are highlighted as follows:

- Proposes an adaptive federated learning framework that combines two models, MobileNetV2 and Federated Averaging, and applies it in the task of pneumonia detection in 10 hospitals, while ensuring privacy.
- Compares the different client selection strategies (Random, Performance Based, and Diversity Based) systematically across a range of participation budgets with clients, within the same architecture.
- Derives a trade-off analysis between the accuracy of the messages and establishing the optimal selection strategy for clients in a round depending on the budget for the client participation in each round.

2. Related Work

There have been a few studies that have looked at medical imaging applications in federated learning, targeting the specific problems encountered when collaborating across institutions and datasets across varying hospitals [9]. Initial federated solutions for chest X-ray classification showed potential for models trained on distributed hospital datasets to achieve near-central performance on data that is kept locally [10]. Follow-on work has investigated what happens when the data are not independent and identically distributed (iid) for each client, and found that the class imbalance and heterogeneity in the hospital populations can significantly impact the convergence behaviour of the global model [11]. To address these impacts, researchers proposed some weight schemes for the aggregation based on adaptable weights according to data quality, not just the size of the dataset [12]. Another approach for rapid convergence is to investigate client selection strategies: performance-aware strategies that select classes with the largest loss reduction in the most recent training round have been proven to greatly increase the convergence rate in the initial rounds of training [13]. On the other hand, fairness-based selection methods which ensure fair inclusion of all clients in the training process have been suggested to avoid systematic exclusion of the underrepresented institutions from the training process [14]. Concurrently, transfer learning from ImageNet-pretrained backbones (e.g., MobileNetV2, ResNet) has been extensively applied to medical imaging federated pipelines to minimize the number of trainable parameters and its communication load per round [15]. Additionally, communication-efficient federated learning methods, such as using gradient compression and quantized models to save bandwidth in the limited-bandwidth hospital networks have been explored [16].

In the context of pneumonia, multiple deep learning investigations have been conducted on the Kaggle Chest X-Ray dataset, and the accuracy has been high at the centralized level, but there is an open question on federated generalization across institutions [17]. Existing studies on federated medical imaging have requested empirical assessment of client scheduling policies with various participation budgets, with many previous ones

set the number of clients per round without any empirical evidence [18]. As a whole, it brings to light that the ability to aggregate across multiple clients using FedAvg, to transfer knowledge to individual clients, and to preserve privacy are now a well-documented part of federated medical image pipelines; however, a detailed comparison of client selection strategies for various participation budgets, and in particular when applied to the detection of pneumonia in chest X-ray images, remains poorly studied and is the focus of the present study through controlled experimentation under a fixed federated architecture with two different client-participation budgets, and thus directly compared with respect to the chest X-ray pneumonia detection problem.

Table 1. Summary of related work on federated learning for medical image analysis

Ref.	Study Focus	Model / Architecture	FL / Aggregation Technique	Dataset	Key Finding / Gap
[15]	Transfer-learning backbones in FL medical imaging	MobileNetV2 / ResNet	FedAvg	Multi-hospital chest X-ray	Fewer trainable params lower communication payload
[16]	Communication-efficient federated learning	CNN	Gradient quantization / compression	Radiograph data	Bandwidth reduced without major accuracy loss
[17]	Centralized pneumonia classification benchmark	CNN ensembles	Centralized (no FL)	Kaggle Chest X-Ray	High centralized accuracy; federated generalization unexplored
[18]	Client scheduling under participation budgets	Generic CNN	FedAvg, variable client count	Simulated multi-client imaging	Selection policy fixed without justification
[19]	Non-IID data in federated imaging	CNN	FedProx-style aggregation	Multi-site radiology	Heterogeneity degrades convergence speed
[20]	Fairness-aware client participation	CNN	Round-robin scheduling	Distributed medical images	Ensures equitable participation, prevents starvation
[21]	Performance-based client prioritization	CNN	Loss-ranked selection	Federated benchmark datasets	Faster early-round convergence via exploitation
[22]	Privacy-preserving aggregation	CNN	Differential-privacy FedAvg	Chest radiograph datasets	Improved privacy at modest accuracy trade-off
[23]	Lightweight CNN for edge FL deployment	MobileNet variants	FedAvg	Mobile health imaging	Reduced computation for constrained clients

[24]	Systematic FL client-selection survey	Multiple architectures	Comparative FL strategies	Cross-domain benchmarks	Highlights need for accuracy–communication trade-off study

3. Methodology

3.1 Research Objective

This study aims to investigate and compare different client selection strategies in a FL scheme to find out the best client selection approach which improves the overall performance of the model and minimizes the communication overhead for automated pneumonia detection from chest X-ray images.

3.2 Methodology Overview

It is a federated learning pipeline typically used for cross-silo federated learning where there are 10 simulated hospitals with private partitions of the training data. The architecture consists of a lightweight convolutional network which is trained locally at different clients and weights are federated averaged centrally (FedAvg). All experiments were conducted with the same clients and only the selection strategy of clients and the number of clients sampled per round were varied across experiments.

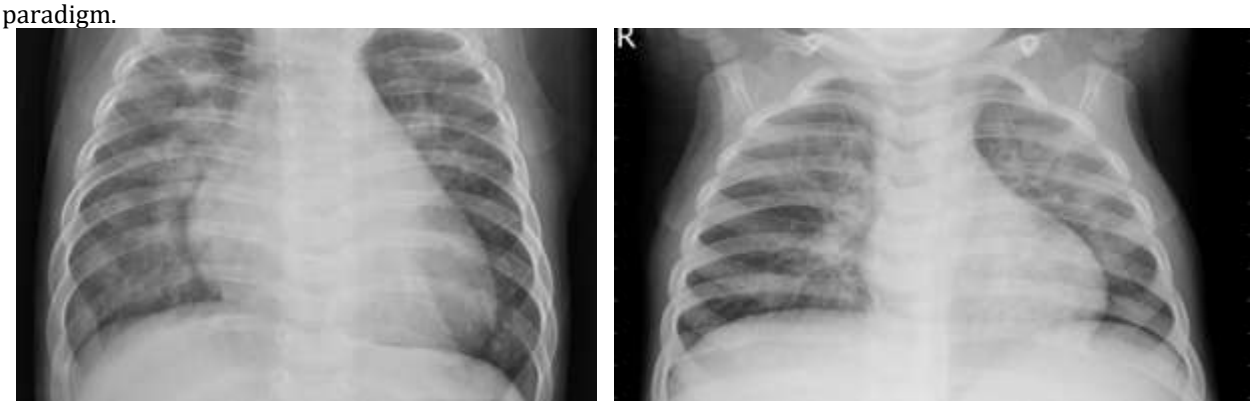
3.3 Dataset Details

The study is based on the publicly available Chest X Ray Pneumonia database (Kaggle, Paul Mooney) [25] which consists of paediatric anterior–posterior chest radiographs, where each image is classified as either NORMAL or PNEUMONIA, that is split into training, validation, and test folders. However, there were a significant number of pneumonia cases that affected the results, because they were in the training period, shaping the recall of shaping per class.

Table 2. Chest X-Ray Pneumonia dataset composition and role of each split.

Split	NORMAL	PNEUMONIA	Total	Role in Pipeline
Train	1,341	3,881	5,222	Partitioned across 10 clients
Validation			16	Per-round global evaluation
Test	234	391	625	Final global model evaluation
Classes			2	NORMAL, PNEUMONIA

Figure 1 shows some example chest X-ray images of the pneumonia dataset used to develop the models. The samples feature different levels of lung opacity, anatomical features, and disease phenotypes, offering rich visual diversity to facilitate effective feature learning, model training, and evaluation in the federated learning paradigm.



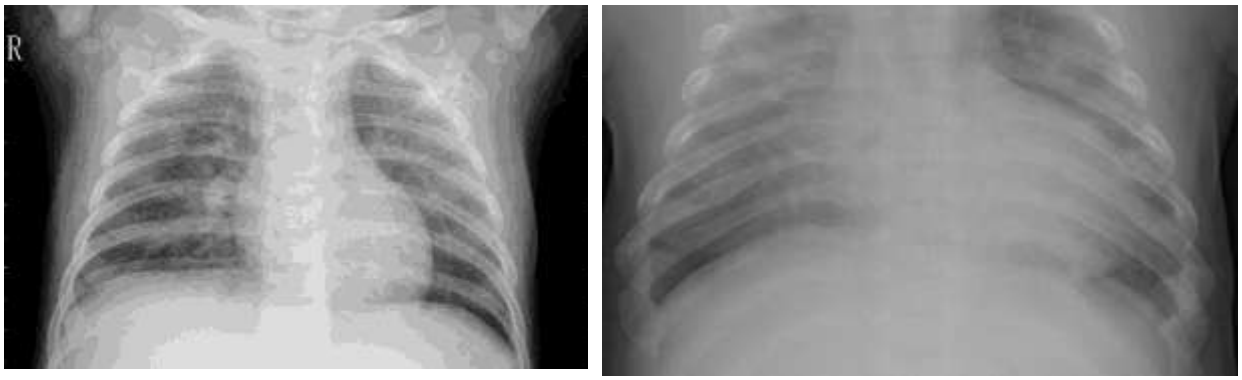


Figure 1: Input sample data images form the dataset

3.4 Pre-processing Techniques

The images are preprocessed in the same way before being put into the network. The input resolution of the images varies and the Pixel intensity of the image varies too, so each image is resized to a fixed square resolution and converted into a normalised tensor, which is then standardised using ImageNet channel statistics to match the pre-trained MobileNetv2 backbone. The six transformations are a way to formally express this pipeline.

$$I_r = \text{Resize}(I, H \times W), \text{where } (H, W) \in \{(128,128), (64,64)\} \quad (1)$$

The resizing can be formulated as described with (1) which maps the original radiograph of arbitrary dimension into a fixed resolution of the square, using bilinear interpolation, and ensures the uniformity of tensor shapes when batching with all the participating hospital clients.

$$x' = \frac{I_r}{255},\ x' \in [0,1]^{\{H \times W \times 3\}} \quad (2)$$

To be able to use channel-wise normalisation meaningfully with inputs, there must be a step of scaling the 8-bit intensities to a floating-point tensor that ranges between the unit interval [0, 1]. This step is performed using equation (2) prior to channel-wise normalisation.

$$x_{\text{norm}[c]} = \frac{x'[c] - \mu_c}{\sigma_c},\ c \in \{R, G, B\} \quad (3)$$

Equation (3) is a channel-wise z-score normalisation with fixed constants for mean and standard deviation of the ImageNet radiograph intensity distribution the pretrained MobileNetV2 backbone was trained with.

$$x_{\text{flip}} = \text{Flip}_{h(x_{\text{norm}})} \text{ with probability } p = 0.5 \quad (4)$$

During training, the image is flipped horizontally with probability $1/2$, as described in Equation (4); this can be regarded as a left-right anatomical variation that exposes the network to this type of variation.

$$x_{\text{rot}} = \text{Rotate}(x_{\text{norm}}, \theta), \theta \sim U(-10°, +10°) \quad (5)$$

The random rotation transform in Equation (5) is used to model the variation in patient positioning by rotating the radiograph with a random angle selected uniformly from a bounded interval, which represents realistic variation in acquisition of the radiograph.

$$x_{\text{jit}} = \alpha \cdot x_{\text{norm}} + \beta, \alpha \sim U(0.8,1.2), \beta \sim U(-0.2,0.2) \quad (6)$$

However, equation (6) represents the colour-jitter operation with randomly sampled linear rescaling of the brightness and contrast of each pixel to simulate a range of exposure times and imaging devices between hospitals, within a 20% range.

Only Equations (1)–(3) are used for validation and test images; no augmentation (Equations (4)–(6)) is applied to the images, thus making the comparison between the evaluation results of the different client selection strategies deterministic.

3.5 Data Augmentation

Each hospital has about 522 images then, augmentation is applied on the fly to each training partition (not applied to complete datasets), which adds diversity to the relatively small number of images available to each client and decreases overfitting.

- Random horizontal flip: Randomly mirrors the radiograph horizontally, introducing anatomical variation to the left–right side of the image, thus exposing the network to left–right variation.
- Random rotation: small random rotations up to ±10° are used, to mimic patient positioning differences.

- Colour jitter: brightness and contrast are randomly varied by up to ±20%, as would occur if the imaging equipment and exposure settings varied in different hospitals.

3.6 Distribute Data over Clients

There are 10 clients that represent 10 independent hospitals with a total of 5222 training images. An IID partition is used: the entire index list is randomly rearranged and partitioned into 10 parts and each client receives a similarly balaced class, eliminating the effects of confounds caused by distributing the data. To make comparisons for the future a non-IID mode has been implemented as well.

Table 3. IID partition of training data across ten simulated hospitals (≈522 images each).

Client	Samples	Client	Samples
Client 01	523	Client 06	522
Client 02	523	Client 07	522
Client 03	522	Client 08	522
Client 04	522	Client 09	522
Client 05	522	Client 10	522

Algorithm 1 Data Partitioning (IID / Non-IID)
Input: full training set D; number of clients N; flag non_iid
Output: client partitions {D1 .. Dn}
1: indices ← Shuffle([0 .. |D|-1])
2: if non_iid = False then // IID
3: splits ← SplitEqually(indices, N)
4: else // Non-IID (class-skewed)
5: class0 ← indices with label NORMAL
6: class1 ← indices with label PNEUMONIA
7: for k = 0 to N-1 do
8: take a per-client slice of class0 and class1
9: D_k ← Shuffle(slice0 ∪ slice1)
10: end for
11: end if
12: return {D1 .. Dn}

3.7 Federated Training Configuration

All experiments have the same training configuration to control the comparison of the strategies. All the differences between the two experiments lie in the degree of sampling (i.e. number of clients sampled per round), and in the size of the input image.

Table 4. Basic model and federated training configuration for both experiments

Parameter	Experiment 1 (2/round)	Experiment 2 (3/round)
Total clients (hospitals)	10	10
Clients selected per round	2	3
Communication rounds	8	8
Local epochs per round	1	1
Batch size	64	64
Learning rate	0.001	0.001
Optimizer	Adam	Adam
Loss function	Cross-Entropy	Cross-Entropy
Input image size	128 × 128	64 × 64
Data partition	IID	IID
Aggregation	FedAvg (weighted)	FedAvg (weighted)
Random seed	42	42

3.8 Classification Model

The same architecture (MobileNetV2) is always applied to all clients and strategies, and is pre-trained on ImageNet. It is also light weight, which is suited for a federated simulation in which the model is being copied

and aggregated over and over again. Transfer learning is used: The trainable parameters are limited (164,226), with the feature extractor frozen, just a new classification head is trained.

Table 5. MobileNetV2 architecture used for classification (164,226 trainable parameters)

Layer	Type	Configuration	Trainable
Backbone	MobileNetV2 features	ImageNet pretrained, frozen	No
Head-1	Dropout	p = 0.3	
Head-2	Linear	1280 → 128	Yes
Head-3	ReLU	activation	
Head-4	Dropout	p = 0.2	
Head-5	Linear	128 → 2 (output)	Yes

Algorithm 2 BuildModel: MobileNetV2 Transfer Learning
Input: ImageNet-pretrained MobileNetV2; number of classes = 2
Output: model ready for federated fine-tuning
1: model ← MobileNetV2(pretrained = True)
2: for each param in model.features do
3: param.requires_grad ← False // freeze backbone
4: end for
5: model.classifier ← Sequential(
Dropout(0.3), Linear(1280→128), ReLU(),
Dropout(0.2), Linear(128→2))
6: return model

3.9 Optimisation and Loss Functions

All local training uses a uniform optimization and loss configuration, to provide an unfair and impartial comparison of the client selection strategies. The Adam optimizer is used due to its adaptive learning ability and a learning rate of 0.001 and batch size of 64 is used to optimize only the trainable classification head while the backbone of MobileNetV2 is frozen. For binary classification of NORMAL and PNEUMONIA, Cross-Entropy loss has been used as the objective function. At training time, through standard backpropagation, only the gradients of the trainable parameters are calculated, which allows for stable convergence, efficient parameter updates, lower complexity and better classification performance when including all the federated clients.

3.10 Model Training and Evaluation

The proposed Federated learning framework optimizes the model using several rounds of communication with the central server, without it being allowed to see the raw patient information. The server chooses a subset of the hospital clients that will participate in the communication during each round and communicates the current global parameters of the model to the subset of clients. The model is then trained independently by each of the selected clients for an epoch with their respective data sets (chest X-ray images) with the Adam optimizer and Cross-Entropy loss, keeping the patient data local to the respective client throughout the training process. Local optimization results in new model parameters which are sent to the central server where the models from all clients are combined via a weighted averaging technique based on the number of samples used for training on the client. The weighted aggregation can make a greater contribution by clients with larger data sets, while maintaining client privacy. After every communication round, the new global model is tested on a held out validation set to check for convergence and stability in the training process. After the completion of all the rounds of communication, the final classification model for the global model is evaluated on an unseen 625-image test set with the help of Confusion Matrix, Precision, Recall, F1-Score and Accuracy metrics, to get a comprehensive analysis of classification performance. In addition, the cost of the communication is measured as a 'real' measure of communication overhead: the product of the number of clients involved in the communication and the number of rounds of communication. This allows a quantitative evaluation of the communication cost between the various client selection strategies, while maintaining the diagnostic accuracy of the communication.

3.11 Client Selection Strategies

There are three strategies for deciding which of the clients play each round – each are different rules for balancing exploration/exploitation/fairness among the hospitals.

- Random Selection: k clients are randomly selected from the 10 hospitals at each round, the baseline without any modifications; wide exploration but fluctuating composition round to round.
- Performance-Based Selection: If a client has already participated, the loss recorded in their local training is used to rank them, the k with the lowest loss are selected and if they have not yet participated, k are randomly selected (falling back to random in round one).
- Diversity Based Selection (Round Robin): With a round robin rotation, the eight clients are rotated around so that each hospital gets to be the client in each of the 8 rounds with maximum chance of coverage.

Table 6. Comparison of the three client selection strategies.

Strategy	Selection Principle	Strength	Weakness
Random	Uniform random sample	Broad exploration, simple	Unstable, high variance
Performance-Based	Lowest recent local loss	Exploits best data	May over-select few clients
Diversity-Based	Round-robin rotation	Fair, full data coverage	Ignores client quality

Algorithm 3 Random Selection
Input: client set of size N; count k
Output: k selected client indices
1: indices ← [0,1,...,N-1]
2: selected ← SampleWithoutReplacement(indices, k)
3: return selected

Algorithm 4 Performance-Based Selection
Input: client set of size N; count k; scores {l_1..l_N}
Output: k selected client indices
1: if scores is empty then
2: return RandomSelection(N, k) // cold start
3: end if
4: ranked ← argsort(scores) ascending
5: selected ← ranked[0:k]
6: return selected

Algorithm 5 Diversity-Based Selection (Round Robin)
Input: client set of size N; count k; round index r
Output: k selected client indices
1: start ← (r × k) mod N
2: selected ← [(start+i) mod N for i = 0..k-1]
3: return selected

3.12 Federated Training Loop and Aggregation

The server side orchestration algorithm initializes the global model and executes the fixed rounds, activates the selection strategy, activates the local model training, executes FedAvg aggregation, and evaluates the global model after each round.

Algorithm 6 Federated Training Loop (FedAvg Orchestration)
Input: N clients with data {D1..Dn}; rounds R; clients/round k;
selection strategy S; local epochs E; lr eta
Output: trained global model w

```
1: w ← BuildModel()
2: scores ← [0.5, 0.5, ..., 0.5]
3: for r = 1 to R do
4:   selected ← S(clients, k, scores, r)
5:   models ← []; sizes ← []
6:   for each c in selected do
7:     w_c, l_c ← LocalTrain(copy(w), D_c, E, eta)
8:     models.append(w_c); sizes.append(|D_c|)
9:     scores[c] ← l_c
10:  end for
11:  w ← FedAvg(models, sizes)
12:  val_acc ← Evaluate(w, D_val)
13: end for
14: test_acc ← Evaluate(w, D_test)
15: return w
```

Algorithm 6. Server-side federated training loop.

FedAvg averages the weight tensors returned by selected clients, layer by layer, over the model state dictionary (where the contribution from each client is proportional to the number of samples he/she trained on):

$$w_{global} = \Sigma_c \left(\frac{n_c}{n}\right) \cdot w_c,\ n = \Sigma_c n_c \quad (7)$$

From equation (7) it is seen that the local parameters of the hospital that provides more samples, have a more proportionate influence on the global aggregated model parameters in each round.

4. Results

For both experiments the results are reported in terms of the accuracy of the validation done per round, the accuracy of the test, the communication cost, and the confusion matrix of the best strategy.

4.1 Experiment 1 -Two Clients per Round

Random selection had the highest final accuracy in the test and two client selections were made in each round (a total of 8 client rounds). The performance-based selection led to a lower selection performance plateau due to the fact that the same two clients were selected over and over again and were not diverse enough to have a wide range of data to be used for the global model.

Table 7. Per-round validation accuracy by strategy- Experiment 1 (2 clients/round).

Round	Random (Val %)	Performance-Based (Val %)	Diversity-Based (Val %)
1	68.75	50.00	50.00
2	56.25	56.25	56.25
3	62.50	56.25	75.00
4	75.00	81.25	56.25
5	87.50	81.25	62.50
6	93.75	81.25	93.75
7	68.75	68.75	68.75
8	81.25	68.75	75.00

Table 7 shows the per-round validation accuracy for both the three strategies in Experiment 1 for the eight communication rounds with two clients chosen per round. Of course, the random selection begins at 68.75 per cent and then falls during the training period, but it does climb to a peak of 93.75 per cent at round six, and remains in the vicinity of 81.25 per cent thereafter. Performance Based selection starts at 50.00% with no loss scores available, and then steadily increases to 81.25% in round four, but levels off because they are using the same limited client pair. This volatility is echoed in Diversity-Based selection, with its rate rising to 93.75 percent with the Random selection and sometimes being able to match exploitation's performance with this smaller participation budget.

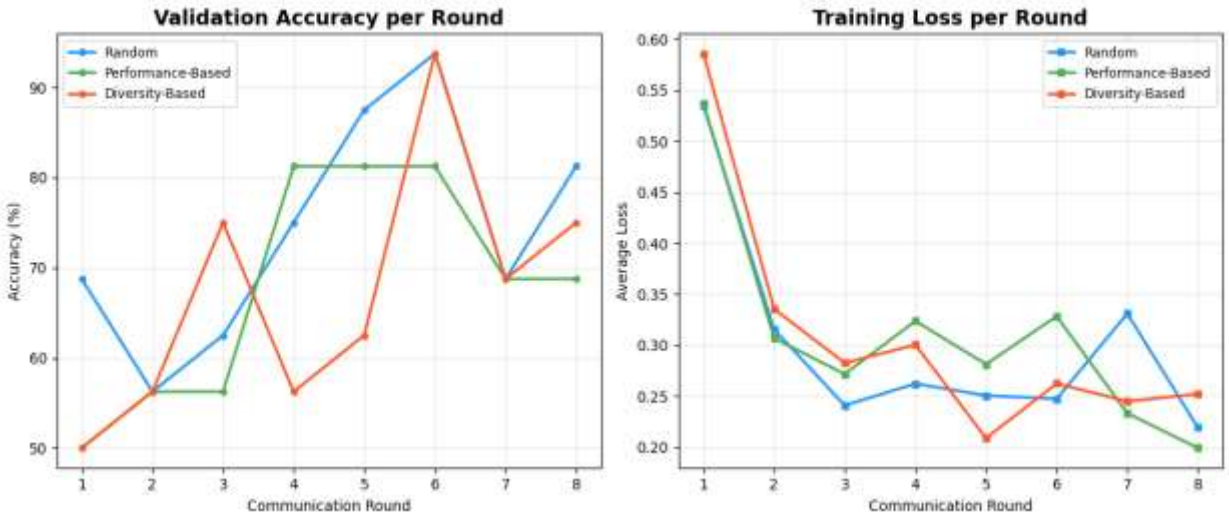


Figure 2. Validation accuracy (left) and training loss (right) per communication round Experiment 1.

In the case of Experiment 1, validation accuracy and training loss are plotted per round, as seen in Figure 2, where training loss of both other selection strategies has reached a plateau at a lower level than validation accuracy, and Random selection continues to steadily rise through the rounds and reach its best at round 6.

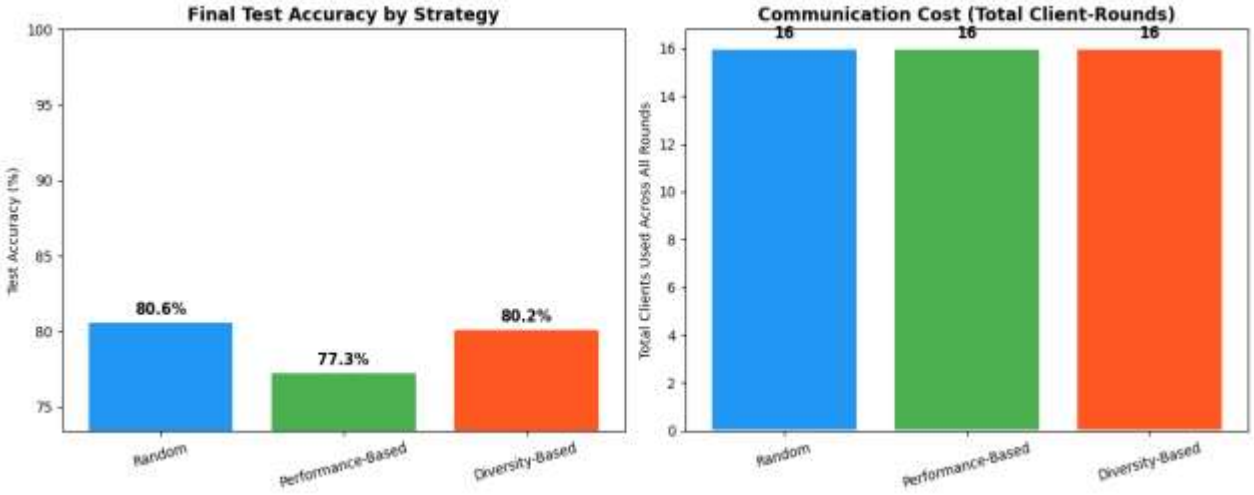


Figure 3. Final test accuracy and communication cost by strategy - Experiment 1.

The results in Figure 3 also show a visual comparison of the final test accuracy with the communication cost (sixteen rounds) for the three strategies used in Experiment 1.

Table 8. Strategy comparison summary- Experiment 1. Best strategy (Random) highlighted.

Strategy	Test Accuracy	Peak Val. Accuracy	Comm. Cost
Random	80.64%	93.75%	16
Performance-Based	77.28%	81.25%	16
Diversity-Based	80.16%	93.75%	16

The results from the Experiment 1 strategy comparison are summarised in Table 8, which shows the accuracy of the test, the best accuracy achieved during validation, and the number of communications needed. Under

this two-client settings, random selection outperforms other selection methods with the highest accuracy (80.64) and the highest peak validation accuracy (93.75). In terms of test accuracy, Diversity Based selection comes next with 80.16 percent accuracy and a comparable peak accuracy during validation, with fair rotation being a good runner up. Performance Based selection trials achieved a test accuracy of 77.28 percent, the lowest level of test accuracy in the series, and the highest of 81.25 percent was also repeated, in which the same pair of clients was oversampled. All three strategies have a cost of communication of 16 client-rounds, with the only difference in accuracy being due to the quality of the selections.

The confusion matrix and per-class metrics of the best model present higher pneumonia recall (0.96) while normal is lesser (0.54) which is a direct reflection of class imbalance present in the training set.

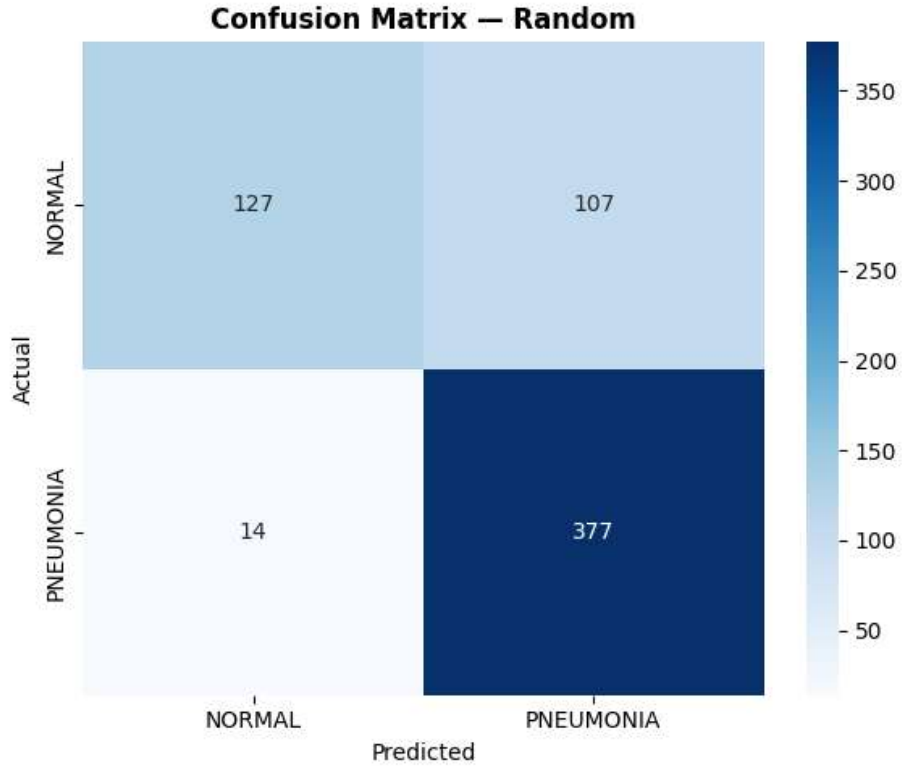


Figure 4. Confusion matrix of the best strategy (Random) - Experiment 1.

The best strategy of Experiment 1, Random selection, is shown on figure 4 where Pneumonia is detected with a high rate but there is a significant number of cases with "normal" label but incorrect diagnosis of "pneumonia" due to the class imbalance in the training set.

Table 9. Classification report for the best strategy (Random)- Experiment 1.

Class	Precision	Recall	F1-score	Support
NORMAL	0.90	0.54	0.68	234
PNEUMONIA	0.78	0.96	0.86	391
Accuracy			0.81	625
Macro avg	0.84	0.75	0.77	625
Weighted avg	0.82	0.81	0.79	625

The classification report for Random selection, the best model from Experiment 1, for the test set of size 625, is shown in Table 9. NORMAL class has high precision (0.90) and low recall (0.54) with F1 score of 0.68, meaning that there are numerous normal radiographs that are misclassified as pneumonia. The PNEUMONIA

class has the opposite trend, where precision is 0.78 and the recall is 0.96 resulting in an F1-score of 0.86, due to the imbalance of the training set (1341 normal vs. 3881 pneumonia). The overall accuracy is 0.81, the macro-averaged F1-score is 0.77 and the weighted average is 0.79, indicating that it is biased towards the majority class.

4.2 Experiment 2 - Three Clients per Round

Performance Based selection obtained the highest final test accuracy (83.36%) in the first experiment but in the second experiment it got the highest, when participation was increased to three clients per round (communication cost of 24 client-rounds). This strategy could exploit high quality clients and yet have averaged over a sufficient amount of data for a stable performance, three clients per round.

Table 10. Per-round validation accuracy by strategy - Experiment 2 (3 clients/round).

Round	Random (Val %)	Performance-Based (Val %)	Diversity-Based (Val %)
1	50.00	50.00	50.00
2	56.25	62.50	50.00
3	75.00	56.25	68.75
4	62.50	50.00	81.25
5	68.75	87.50	87.50
6	62.50	81.25	62.50
7	81.25	81.25	81.25
8	68.75	87.50	68.75

The results for Experiment 2 are presented in Table 10 in terms of the average accuracy for each round of the game, with three clients playing each round over eight rounds. The three strategies start at the same level at 50.00 percent with one of the strategies diverging. Performance Based selection rises rapidly, to 87.50 percent at round five and to the same level at round eight, when it's based on a larger number of clients, so that no one is overconcentrated on one client pair. In diversity-based selection, the throughput is 87.50 percent after round five too, which is due to full coverage of the selection process with the help of rotation. Random selection reaches its maximum of 81.25 percent, but it's more volatile and less likely to hit the 87.50 percent high, with quality-aware and fairness-aware selection both performing better than pure randomness when participation budget is expanded to three clients.

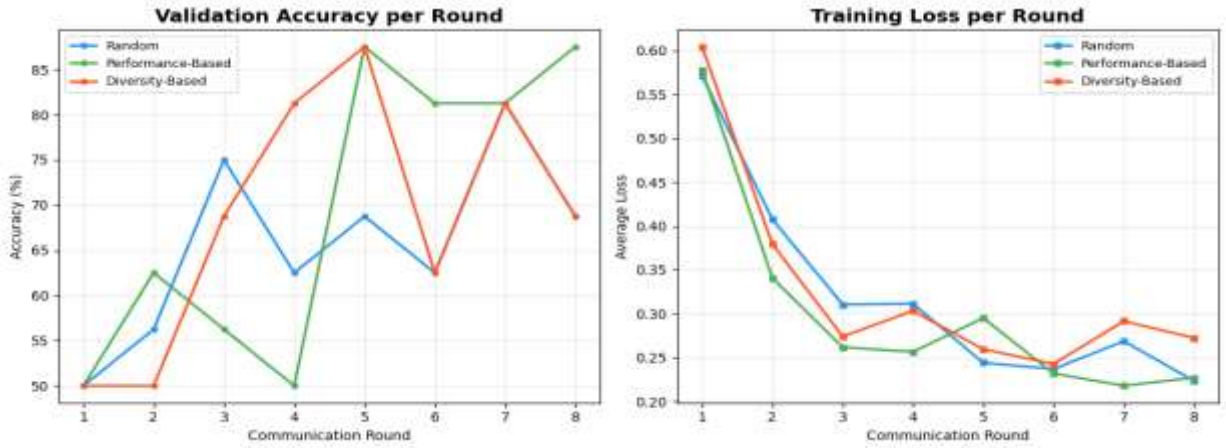


Figure 5. Validation accuracy (left) and training loss (right) per communication round- Experiment 2.

Figure 5 displays the accuracy validation curve and the training loss curve for the Experiment 2 with the highest peak validation accuracy of 87.50% over the eight communication rounds with PerformanceBased selection.

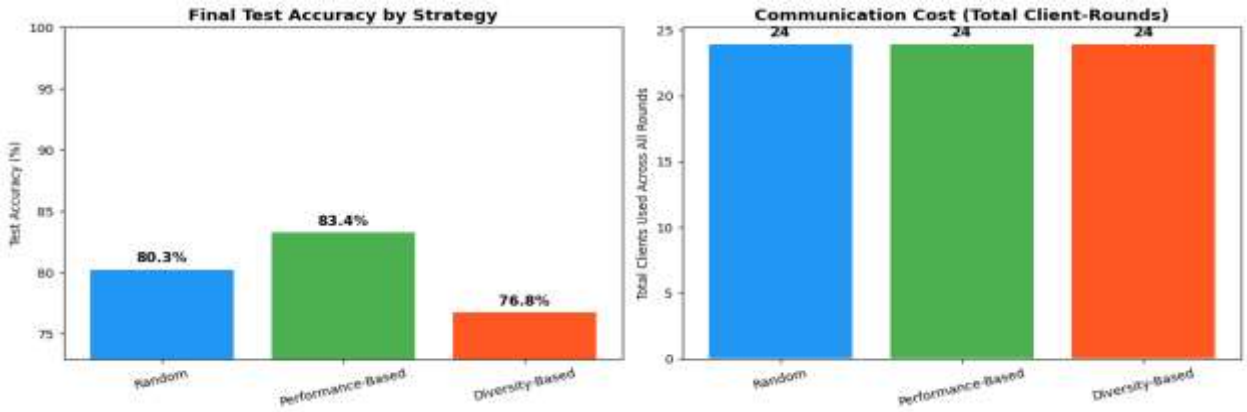


Figure 6. Final test accuracy and communication cost by strategy - Experiment 2.

The comparison between final test accuracy and communication cost for Experiment 2 is shown in Fig. 6 and the performance-based selection method is apparent to be the best performer at the common communication cost of twenty-four rounds.

Table 11. Strategy comparison summary - Experiment 2

Strategy	Test Accuracy	Peak Val. Accuracy	Comm. Cost
Random	80.32%	81.25%	24
Performance-Based	83.36%	87.50%	24

The results of Experiment 2 strategy comparison are summarised in Table 11, and include test accuracy, peak accuracy on validation, and communication cost. The best performance is obtained for both experiments in the case of Performance-Based selection (83.36 percent test accuracy and 87.50 percent peak validation accuracy), indicating that the selection over quality is effective after three clients are involved in a round. Random selection does not fare much better, with a test score accuracy of 80.32 percent and a lower peak of 81.25 percent, while Diversity-Based selection has the lowest accuracy of 76.80 percent with a peak accuracy of 87.50 percent. The gain that can be achieved with the three strategies is the same (twenty-four client-rounds), and the gains from the Performance-Based strategy are due to better scheduling only. Experiment 2, the best model, has a confusion matrix that is more balanced than Experiment 1; the Recall of the Normal class increased to 0.71 and the Recall of Pneumonia remained high at 0.91.

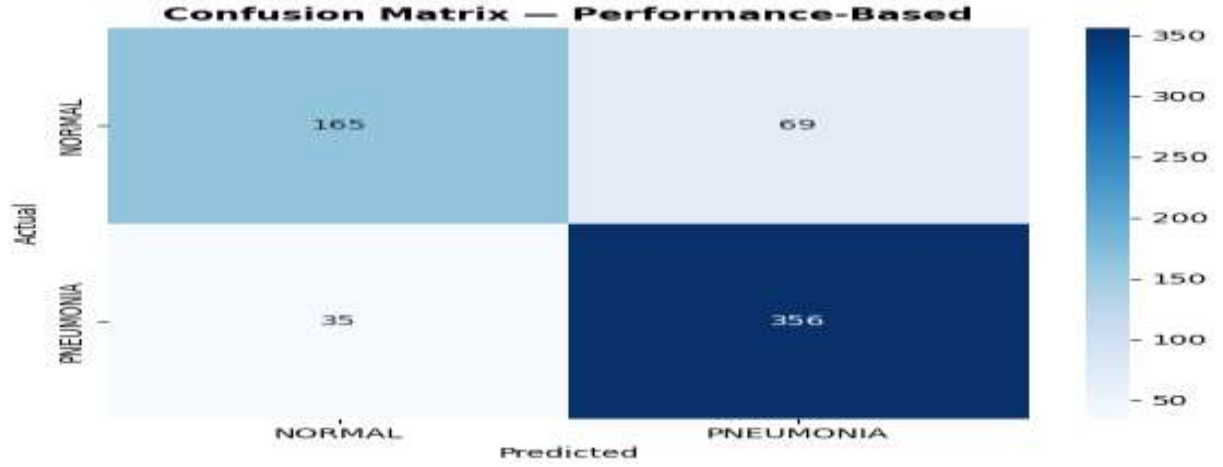


Figure 7. Confusion matrix of the best strategy (Performance-Based) - Experiment 2.

The confusion matrix for the Experiment 2's model (Performance-based selection) is shown in figure 7, where the model achieved a more balanced result, and the normal-class recall showed a slight improvement compared to the best model of Experiment 1.

Table 12. Classification report for the best strategy (Performance-Based)- Experiment 2.

Class	Precision	Recall	F1-score	Support
NORMAL	0.82	0.71	0.76	234
PNEUMONIA	0.84	0.91	0.87	391
Accuracy			0.83	625
Macro avg	0.83	0.81	0.82	625
Weighted avg	0.83	0.83	0.83	625

Table 12 shows the per-class classification report for the best model from Experiment 2, a model that is based on Performance based selection, on the same 625-image test set. The NORMAL class demonstrates a greater precision (0.82) and better recall (0.71) with an F1-score of 0.76 compared to the recall of 0.54 in Experiment 1. The results for the PNEUMONIA class are also good with values of 0.84, 0.91 and 0.87 for precision, recall and F1 score respectively. The overall accuracy is 0.83, which is higher than the accuracy of Experiment 1 which was at 0.81. The macro and weighted average F1-scores are 0.82 and 0.83 respectively, which represents a significantly more balanced classifier, benefiting neither the majority class of pneumonia nor the other classes, and has a significantly higher diagnostic accuracy for all classes combined.
A supplementary run, with a fixed subset of five randomly selected clients (Clients 1, 3, 6, 8 and 9) per round, achieved only 75.52% test accuracy at 40 client-rounds, showing that a fixed committee does not lead to higher test accuracy or lower overhead than does the adaptive strategies.

5. Comparative Analysis of Models

Combining both experiments helps explain the accuracy communication trade-off, and also reveals the necessity to implement different strategies based on how many rounds the client participates in:

Table 13. Consolidated comparison across both experiments

Experiment	Strategy	Test Acc.	Peak Val. Acc.	Comm. Cost	Rank
Exp 1 (2/round)	Random	80.64%	93.75%	16	1
Exp 1 (2/round)	Diversity-Based	80.16%	93.75%	16	2
Exp 1 (2/round)	Performance-Based	77.28%	81.25%	16	3
Exp 2 (3/round)	Performance-Based	83.36%	87.50%	24	1
Exp 2 (3/round)	Random	80.32%	81.25%	24	2
Exp 2 (3/round)	Diversity-Based	76.80%	87.50%	24	3

Table 13 combines all six ways of combining the strategies and experiments together and ranks them according to test accuracy. Performance-Based selection is the overall winner at the twenty-four client-rounds in Experiment 2 (test accuracy = 83.36 percent). In this smaller budget, the first two methods, Random selection and Diversity-Based, perform extremely well within themselves, both at 80.64 percent, followed by Performance-based at 77.28 percent. The order in Experiment 2 changes with Random coming in second (80.32

percent) and Diversity-Based coming in third (76.80 percent). This indicates that there is no single best strategy for all, but rather the optimal strategy will depend on the combination of strategy and the number of clients sampled per round.

5.1 Key Observations

The experimental results show that the selection strategy used by the clients and the number of selected clients in a communication round have significant impacts on the effectiveness of federated learning. The selected performance-based configuration that results in performance of 83.36% test accuracy is the best overall performance among all the evaluated configurations with 3 clients involved in every communication round. This setup was found to consistently outperform all the other strategies by being able to identify clients with high local training performance, and with a high level of diversity because of increased participation. The findings show that intelligent client scheduling based on the concept of quality and information can greatly enhance the convergence and generalization ability of the global federated model in pneumonia detection of chest radiography.

The comparison also shows that the success or failure of a client selection strategy is much a function of the client's participation budget. However, when only 2 clients were involved in each round of communication, Performance Based selection resulted in the lowest test accuracy (77.28%) as it kept selecting the same high performing clients, thereby not introducing diversity in the data, and also limiting the global model to the same distribution of patients. Random selection, on the other hand, showed the best performance in this limited context with test accuracy of 80.64%, and achieved a similar value of 80.16% for Diversity Based selection. These findings show that as the communication budget gets smaller, more diversity among clients becomes more significant.

Another remark is on the balance between the performance of the diagnosis and the overhead of communication. The higher number of clients led to a 50% increase in the communication cost from 16 to 24 client-rounds but also resulted in a higher highest test accuracy (+2.72 percentage points) of 80.64% to 83.36%. The modest additional communication burden led to a significant improvement in the performance of the model, suggesting that a small boost in participation costs can make a significant difference in global model performance. Further, the best model achieved a more balanced confusion matrix with a higher normal-class recall (0.54 to 0.71) but didn't compromise on the pneumonia recall (0.91). This reduction in prediction bias also demonstrates that the overall prediction accuracy is boosted with the proposed intelligent client selection, and class-balanced diagnosis is also improved, so the proposed framework would be more suitable for privacy-preserving clinical decision support systems.

6. Conclusion

This paper proposed an adaptive federated learning algorithm, intelligent client selection for automated pneumonia detection in chest X-ray images, in privacy-preserving distributed healthcare systems. The proposed framework was systematically evaluated using the test of the impact of different client selection strategies (Random, Performance-Based, and Diversity-Based) and client participation budgets (MobileNetV2 transfer learning) and enabled the study of various factors. The study showed experimentally that the efficiency of the communication and the diagnostic performance depend crucially on the scheduling of clients. The test accuracy for Performance Based selection was the highest at 83.36% (with 3 clients per communication round), while that of Random selection was more successful at 80.64% (with 2 clients per communication round). The results also indicated that the greater the involvement of client, the more accurate the diagnosis became by 2.72% with a moderate increase in the communication overhead from 16 to 24 client-rounds. In addition to performance improvement, the proposed framework successfully boosted the class balanced prediction performance by boosting the recall performance for normal cases without degrading pneumonia detection performance. Overall, the study shows that involving the client as an adaptive, quality-aware participant is a viable approach towards the enhancement of federated medical image analysis without compromising on data privacy and communication efficiency. The results of this research offer important design principles for distributed healthcare institutions to adopt federated learning systems. Future research will examine dynamic client scheduling for non-IID data distributions, optimizing for personalized data, communication-efficient aggregation methods, and large-scale evaluation with real datasets from multiple institutions for clinical data to further enhance robustness, scalability, and clinical utility.

References

[1] C. Ortiz-Toro, A. García-Pedrero, M. Lillo-Saavedra, and C. Gonzalo-Martín, "Automatic detection of pneumonia in chest X-ray images using textural features," Comput. Biol. Med., vol. 145, Art. no. 105466, 2022, doi: 10.1016/j.compbiomed.2022.105466.
[2] Z. Xu and H. Zhang, "Pneumonia detection from enhanced chest X-ray images based on Double SGAN model," Sci. Rep., vol. 16, Art. no. 9922, 2026, doi: 10.1038/s41598-026-39785-w.
[3] R. Kundu, R. Das, Z. W. Geem, G.-T. Han, and R. Sarkar, "Pneumonia detection in chest X-ray images using an ensemble of deep learning models," PLoS ONE, vol. 16, no. 9, Art. no. e0256630, 2021, doi: 10.1371/journal.pone.0256630.
[4] A. Y. Nageye, A. D. Jimale, M. O. Abdullahi, et al., "Enhancing deep learning for pneumonia detection: Developing web based solution for Dr. Sumait Hospital in Mogadishu Somalia," Discov. Appl. Sci., vol. 7, Art. no. 309, 2025, doi: 10.1007/s42452-025-06735-6.
[5] M. Fiszman, W. W. Chapman, D. Aronsky, R. S. Evans, and P. J. Haug, "Automatic detection of acute bacterial pneumonia from chest X-ray reports," J. Am. Med. Inform. Assoc., vol. 7, no. 6, pp. 593–604, 2000, doi: 10.1136/jamia.2000.0070593.
[6] Y. Xie, B. Zhu, Y. Jiang, B. Zhao, and H. Yu, "Diagnosis of pneumonia from chest X-ray images using YOLO deep learning," Front. Neurorobot., vol. 19, Art. no. 1576438, 2025, doi: 10.3389/fnbot.2025.1576438.
[7] M. F. Hashmi, S. Katiyar, A. W. Hashmi, and A. G. Keskar, "Pneumonia detection in chest X-ray images using compound scaled deep learning model," Automatika, vol. 62, nos. 3–4, pp. 397–406, 2021, doi: 10.1080/00051144.2021.1973297.
[8] F. Wahid, S. Azhar, S. Ali, M. S. Zia, F. Abdulaziz Almisned, and A. Gumaei, "Pneumonia detection in chest X-ray images using enhanced restricted Boltzmann machine," J. Healthc. Eng., vol. 2022, Art. no. 1678000, pp. 1–17, 2022, doi: 10.1155/2022/1678000.
[9] H. Slimi, A. Balti, S. Abid, et al., "Trustworthy pneumonia detection in chest X-ray imaging through attention-guided deep learning," Sci. Rep., vol. 15, Art. no. 40029, 2025, doi: 10.1038/s41598-025-23664-x.
[10] S. Potharaju, S. N. Tambe, K. Dasari, N. Srikanth, R. Venkatarao, and S. Tambe, "Enhanced X-ray image classification for pneumonia detection using deep learning based CBAM and SE mechanisms," Intell.-Based Med., vol. 12, Art. no. 100299, 2025, doi: 10.1016/j.ibmed.2025.100299.
[11] A. Khan, M. U. Akram, and S. Nazir, "Automated grading of chest X-ray images for viral pneumonia with convolutional neural networks ensemble and region of interest localization," PLoS ONE, vol. 18, no. 1, Art. no. e0280352, 2023, doi: 10.1371/journal.pone.0280352.
[12] Lutfunhar, H. (2026). Secure Wireless Sensor Communication Using Transformer-Based Threat Detection. International Journal on Advanced Computer Theory and Engineering, 15(2), 52–58. https://doi.org/10.65521/ijacte.v15i2.3305.
[13] C. Usman, S. U. Rehman, A. Ali, A. M. Khan, and B. Ahmad, "Pneumonia disease detection using chest X-rays and machine learning," Algorithms, vol. 18, no. 2, Art. no. 82, 2025, doi: 10.3390/a18020082.
[14] R. Siddiqi and S. Javaid, "Deep learning for pneumonia detection in chest X-ray images: A comprehensive survey," J. Imaging, vol. 10, no. 8, Art. no. 176, 2024, doi: 10.3390/jimaging10080176.
[15] A. Saber, A. Fateh, P. Parhami, A. Siahkarzadeh, M. Fateh, and S. Ferdowsi, "Efficient and accurate pneumonia detection using a novel multi-scale transformer approach," Sensors, vol. 25, no. 23, Art. no. 7233, 2025, doi: 10.3390/s25237233.
[16] M. S. A. Reshan, K. S. Gill, V. Anand, S. Gupta, H. Alshahrani, A. Sulaiman, and A. Shaikh, "Detection of pneumonia from chest X-ray images utilizing MobileNet model," Healthcare, vol. 11, no. 11, Art. no. 1561, 2023, doi: 10.3390/healthcare11111561.
[17] Uppalapati, O. (2026). Optimization-Driven Intelligent Diagnosis of Neuropsychiatric Conditions Using EEG Analytics. International Journal on Advanced Computer Engineering and Communication Technology, 15(2), 34–40. Retrieved from https://journals.mriindia.com/index.php/ijacect/article/view/3376.
[18] E. Yanar, F. Hardalaç, and K. Ayturan, "PELM: A deep learning model for early detection of pneumonia in chest radiography," Appl. Sci., vol. 15, no. 12, Art. no. 6487, 2025, doi: 10.3390/app15126487.
[19] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, "Efficient pneumonia detection in chest X-ray images using deep transfer learning," Diagnostics, vol. 10, no. 6, Art. no. 417, 2020, doi: 10.3390/diagnostics10060417.
[20] D. Zhang, F. Ren, Y. Li, L. Na, and Y. Ma, "Pneumonia detection from chest X-ray images based on convolutional neural network," Electronics, vol. 10, no. 13, Art. no. 1512, 2021, doi: 10.3390/electronics10131512.

[21] G. Parra-Cabrera, J. J. Jiménez-Delgado, and F. D. Pérez-Cano, "Artificial intelligence for pulmonary abnormality detection in chest X-ray imaging: A detailed review of methods, datasets and future directions," Technologies, vol. 14, no. 3, Art. no. 147, 2026, doi: 10.3390/technologies14030147.
[22] S. Batra, H. Sharma, W. Boulila, V. Arya, P. Srivastava, M. Z. Khan, and M. Krichen, "An intelligent sensor based decision support system for diagnosing pulmonary ailment through standardized chest X-ray scans," Sensors, vol. 22, no. 19, Art. no. 7474, 2022, doi: 10.3390/s22197474.
[23] Z. Mustafa and H. Nsour, "Using computer vision techniques to automatically detect abnormalities in chest X-rays," Diagnostics, vol. 13, no. 18, Art. no. 2979, 2023, doi: 10.3390/diagnostics13182979.
[24] J. Becker, J. A. Decker, C. Römmele, M. Kahn, H. Messmann, F. Schwarz, T. Kroencke, and C. Scheurig-Muenkler, "Artificial intelligence-based detection of pneumonia in chest radiographs," Diagnostics, vol. 12, no. 6, Art. no. 1465, 2022, doi: 10.3390/diagnostics12061465.
[25] Chest X-Ray Pneumonia dataset: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/code