Ibn Al-Haitham Journal for Pure and Applied Sciences
Journal homepage: jih.uobaghdad.edu.iq
PISSN: 1609-4042, EISSN: 2521-3407
IHJPAS. 2026; 39(3)

Standardized Learning for Applicable Local Medical Systems: A Pilot Assessment

Ahmed Abdulaali Abdulsayed[1*]
[1]Software Computing and Artificial Intelligence, Computer Science, Islamic Azad University, Qazvin Branch, Iran.
* Corresponding Author.

Received: 21/ January /2026
Accepted: 25/April/2026
Published: 20/July/2026
doi.org/10.30526/39.3.4430

Abstract

Federated learning (FL) has become a feasible approach to developing medical prediction models by using distributed institutions without transferring the patient's raw data to the central server. The pilot experiment presented herewith confirms the usability of FL in local medical systems using the Wisconsin Breast Cancer Diagnostic dataset which is divided into five clients simulating independent medical institutions under IID and Non-IID settings. A lightweight multilayer perceptron was trained with the Federated Averaging (FedAvg) algorithm and compared to a centralized baseline under the same training conditions. Assessment Metrics for Model Performance: Accuracy, Precision, Recall, F1-Score, and Training Time. The accuracy of the centralized model was 96.49% while the federated model's accuracy was 94.74% under IID partitioning and 93.86% under Non-IID partitioning. This shows that the performance of the models dropped by an overall 2.6 percentage points in the heterogeneous setting. Notably, recall remained stable at 98.61% across all configurations, suggesting consistent sensitivity in picking up malignant cases. These findings suggest that FL is able to achieve reliable predictive performance while preserving the locality of data and can thus be employed as a method for privacy-aware learning in a collaborative local medical system.

Keywords: Federated Learning, Medical Systems, Data Privacy, Machine Learning, Experimental Study.

1. Introduction

Medical Systems Rely on Machine Learning for Their Accuracy of Functions and Effectiveness at Clinical Usage and Predictive Analytics. Nonetheless, the data-driven medical model is not effective due to the fragmented healthcare data in which data is held by multiple healthcare institutions as these are governed by privacy, security, and regulatory requirements[1,2]. There is a fundamental degree of tension between the need for large diverse datasets that models can learn from and the obligation to protect patient confidentiality and institutional data sovereignty.

Federated learning (FL) is a decentralized learning method that enables model training without using raw data from data holders. It allows several data holders to collectively train a model. In this scenario, each institution that participates will train a local model using its private data and will communicate only the model parameters or updates to the server coordinating the process and aggregating the parameters to produce the global model. This method limits direct exposure of data yet maintains collective learning. One significant milestone is the work of McMahan *et al.*, who propose the Federated Averaging (FedAvg) algorithm and show that decentralized training is communication-efficient despite the fact that client data can be statistically heterogenous, and communication resources are limited [3].

FL has been attracting considerable interest in medical applications because clinical data are highly sensitive and compliance, ethics and governance considerations play a key role in healthcare-oriented ML. Prior reviews and studies report that federated solutions are able to

achieve a similar performance to centralized training in performing tasks such as medical image segmentation, disease classification, and clinical risk prediction, while users maintain full local control over their institutional data sets[1,4]. According to the findings, FL presents a feasible solution for collaborative healthcare analytics where data access is limited due to legal, ethical, or operational issues.

Even though there have been advances in FL applied experimental studies, the studies evaluating the operational feasibility of FL in local medical systems remain limited[5-7]. Most importantly, these studies are reproducible and deployment evaluation oriented. This study gives a pilot evaluation of federated learning structure of local medical systems using open-access medical dataset to create a multi-institutional environment. The performance of our framework is compared to a centralized baseline using standard predictive performance measures and training-efficiency indicators. In this way, the study intends to evaluate the practical feasibility of federated learning as a data-locality-preserving strategy for collaborative medical prediction in local healthcare infrastructures.

2. Previous Studies

Federated learning (FL) has been previously shown to be an effective framework for training an ML model when the sharing of raw data is restricted for reasons of privacy, regulatory and/or operational. The literature, however, is quite varied. Some papers focus on large-scale system design, while others tackle medical applications, and others yet on privacy-enhanced or heterogeneous-data settings [6,7].

Another work on FL architecture is designed at the system level particularly for production deployments large-scale communications as well as fault tolerance and secure aggregation [8]. Their paper showed that federated training can be orchestrated across a very large number of distributed clients, which constitutes an important building block toward the scalable implementation of FL. Nonetheless, these large industrial systems are not totally representative of smaller localised medical settings where resource limitations, simpler model structures, and deployment feasibility may be more relevant.

Federated predictive modeling across institutions can achieve performance comparable to centralized approaches while preserving data locality[9]. Likewise, it was reported that the generalization of models in disease-classification tasks can be improved by the use of federated learning; especially when the institutional datasets differ in their demographic and clinical composition[10]. The aforementioned studies reveal that FL is practically useful in collaborative medical analysis. However, they primarily feature predictive feasibility rather than controlled experimental comparisons under explicitly simulated local-system conditions.

Other studies have investigated FL through lenses of privacy enhancement and medical governance. Ti was highlighter the regulatory, ethical and technical importance of privacy-preserving machine learning to be used in medical imaging, whereby FL can be an important enabler of inter-institutional collaboration[11]. Another researchers additionally studied FL in medical settings using differential privacy mechanisms. They concluded that useful predictive or segmentation performance could still be achieved under stricter privacy guarantees, although often with some loss in model accuracy[12]. The studies suggest that privacy protection in FL can be seen as a spectrum, ranging from the local presence of data to formal privacy mechanisms like differential privacy. Because performance comparisons in the literature may depend not only on the model architecture and dataset type but also the level of privacy protection.

The major determinant of federated performance is data heterogeneity. It was. explored the effects of non-IID distributions and communication constraints on convergence behavior and prediction accuracy in federated situations[13]. Their research suggests that performance deterioration in FL is often due to client-level statistical imbalance instead of just the federated paradigm. This is particularly valid for medical applications, where institutional datasets often differ in class balance, feature distribution, and patient demographic[14-16].

Based on previous research findings, FL facilitates collaborative medical learning by averting raw data centralization. Much of the literature, however, targets either industrial-scale systems, specialized imaging tasks, or privacy-enhanced frameworks that are quite different in objectives and complexity. By contrast, the present work implements a reproducible pilot evaluation of FL in a local medical-system setting leveraging an open-access tabular medical dataset, a lightweight multilayer perceptron model, and an explicit comparison between centralized, IID federated, and Non-IID federated training conditions. The study's positioning avoids proposing a new federated optimization algorithm, as it does not guarantee mathematical tractability in a generalized setting.

3. Materials and Methods

3.1. Research Design Overview

Using experimental and comparative research, this study investigates the practical feasibility of the federated learning in local medical systems. First, a centralized learning framework is implemented to serve as a performance baseline. Then, a federated learning framework is simulated to address independent medical institutions. Experimental conditions enable use of benchmarks and performance metrics to compare differing approaches.

3.2. Dataset Selection and Source

The data set has 30 real-valued features that have been computed from digitised fine needle aspirate (FNA) image digitised breast mass cell nuclei. Similarly, these features are statistical measures of nuclear morphology. Mean, standard error, and worst values for each feature are all included.

This particular dataset was chosen for three key reasons. To begin with, it complies with ethical requirements and is open access. It does not contain any identifiable patient information, making it appropriate for privacy-preserving research. Second, it offers a binary classification (benign versus malignant) benchmark task which would suit well for testing the predictive performance of both centralized and far federated learning models. Furthermore, the dataset has been utilized in previous peer-reviewed medical and machine learning experiments, thus lending possibility for comparison to already published results[17-19].

To mimic a distributed medical setting, the data was split into a number of subsets that represented separate local medical facilities. This setting allows us to examine federated learning with a statistically balanced and heterogeneous distribution of data, similar to real-life scenarios in multi-center healthcare. The description of the data is displayed in **Table 1**.

Table 1. Description of Dataset – Wisconsin Breast Cancer (Diagnostic).

Attribute	Description
Dataset Name	Breast Cancer Wisconsin (Diagnostic)
Source Repository [17]	UCI Machine Learning Repository
Original Donors	Dr. William H. Wolberg, University of Wisconsin
Total Samples	569
Number of Features	30 numerical features
Target Variable	Diagnosis (Benign = 0, Malignant = 1)
Data Type	Structured numerical data
Missing Values	None
Task Type	Binary classification
Ethical Status	Open-access, anonymized, non-identifiable data

3.3. Data Preprocessing and Normalization

To achieve numerical stability, ensure consistency over clients, and facilitate a fair comparison between centralized and federated learning settings, the dataset was processed before model training. Initially, all feature values were analysed for any missing cells and other data-quality issues. No missing values were found in the dataset that was selected. Hence, the complete dataset was retained for study.

To lessen the effect of scale difference between numerical values of attributes, each of the features was min–max normalised, by which they all fall within the range 0–1. This way, it prevents the features with large numerical value from dominating the learning and improves the optimisation stability of gradient-based learning. The normalized feature value x' was computed as **Equation (1).**

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad \textbf{(1)}$$

In the given formula, x indicates the original feature value while min and max denote minimum and maximum value of the respective feature are in training data.

To provide a good basis for testing the classifier, the data set was normalized, shuffled randomly, and divided into two data sets keeping 80% of the data for training and 20% of the data for testing. The test set is kept fixed and out of both centralized and federated training for the unbiased evaluation of model generalization.

For the federated learning setting, the training subset was further divided into multiple disjoint client datasets representing digital independent medical institutions. Two settings relating to data distribution were considered. In IID setting, the samples were divided randomly and nearly equally to the clients such that the clients have identical class composition and identical feature distribution. Unlike the IID setting, the Non-IID setting aims to achieve statistical heterogeneity across institutions by giving clients unequal local data characteristics. The heterogeneity was achieved through class-imbalance patterns, and in some cases feature-skewed partitions. Specifically, some clients were skewed towards benign samples, some towards malignant samples and some had more uneven distributions of features.

With this design, federated learning can be assessed under idealized local-data and heterogeneous local-data. It enables a comparison of the convergence, robustness, and predictive performance of the various participating models when they differ from one another in the underlying data profiles, which is a feature of realism of distributed medical environments.

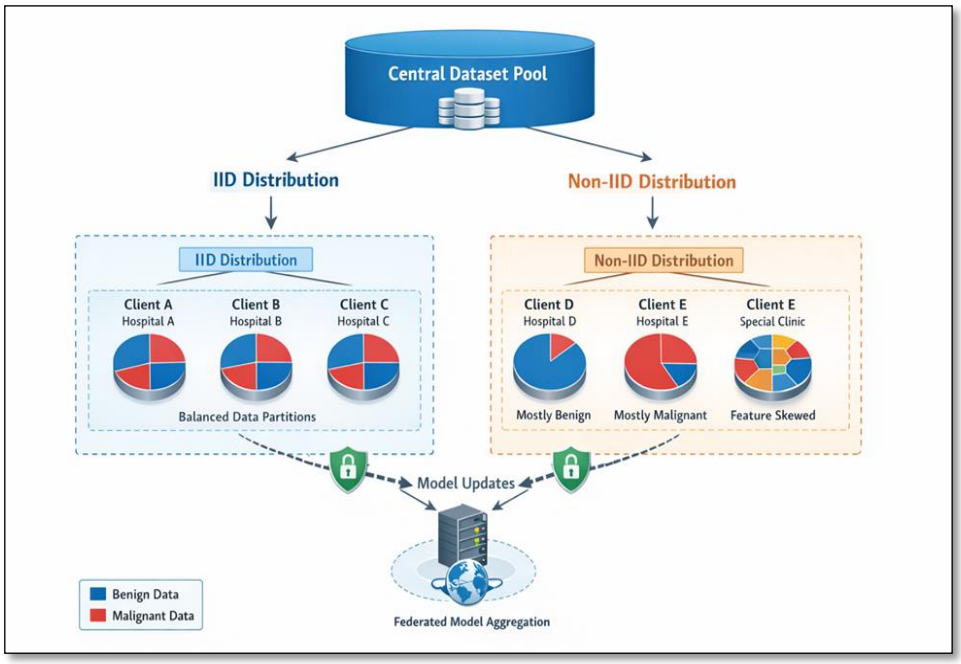

Figure 1. An illustration of partitioning IID and Non-IID client data for a federated learning situation.

Figure 1 shows a conceptual view of the design of experimental federated learning composed to simulate small local medical institutions. The central dataset pool is split into client-specific subsets under two different settings. In the IID setting, each client gets a balanced subset containing the same proportion of classes and similar feature distributions. In the Non-IID setting, local datasets are made heterogeneous on purpose, where some clients receive class-

biased partitions while others receive feature-skewed partitions. Each client locally trains the model using its dataset and communicates the model updates to the aggregation server. The server subsequently aggregates the client updates through the Federated Averaging (FedAvg) algorithm, yielding a new global model for the upcoming communication round. Local data can reside in one location while training together

3.4 Federated Learning Framework and Model Architecture

A lightweight multilayer perceptron classifier was chosen so that predictive performance was not sacrificed for one that could be easily run on localized medical systems that are resource constrained. The model input layer matches the 30-dimensional feature space of the selected dataset. After this, two hidden layers, which are fully connected and activate using a rectified linear unit, are present. Following that, an output layer, which has one neuron and a sigmoid activation, is present.

The federated training process was implemented using the Federated Averaging (FedAvg) algorithm. At the beginning of each global communication round t, the server broadcasts the current global model parameters $W_t$to all participating clients. Each client $k$then performs local training on its own private dataset for a fixed number of local epochs and produces updated model parameters W_t^k . These locally trained parameters are returned to the server and aggregated to construct the updated global model W_{t+1}according to **Equation (2)**:

$$W_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} W_t^k \quad \textbf{(2)}$$

where Kdenotes the total number of participating clients, $n_k$denotes the number of samples held by client k, and $n = \sum_{k=1}^{K} n_k$denotes the total number of training samples across all clients.

The Flower Federated Learning Framework Python was used for implementation [20]. The Flower framework managed client coordination, parameter exchanges, and server-side aggregations across communication rounds, while standard machine learning libraries were used for local model training at each client.

To compare learning paradigms in a fair manner, the configuration of the centralized baseline model took place with the same architecture, loss function, batch size, and optimization settings as the federated one. This design ensures that differences in performance are ascribed primarily to the training strategy and data-distribution setting, rather than changes in model. As the focus of the current work is on feasibility assessment, rather than privacy-mechanism design, the framework provides data locality but does not offer a formal privacy mechanism such as differential privacy and secure aggregation. Experimental setup is shown in **Table 2**.

Table 2. Experimental Setup and Training Settings

Parameter	Value
Number of Clients (K)	5
Global Rounds	20
Local Epochs per Round	3
Batch Size	32
Learning Rate	Adam
Optimizer	Diagnosis (Benign = 0, Malignant = 1)
Loss Function	Binary Cross-Entropy
Model Type	Multilayer Perceptron (MLP)
FL Framework	Flower (Python)
Evaluation Strategy	Centralized vs. Federated

3.5 Experimental Protocol

Experimental conditions were set to be controllable and reproducible, hence all the experiments were done. The systematically matching federated learning framework to a centralized baseline using the same model architecture, optimization setting and evaluation metrics. This design mitigates the effects of model-specific factors and hence viewpoint comparison here is on the impact of the training strategy and data-distribution setting.

Five simulated clients picked random to represent independent local medical institutions, all the experiments were run over 20 global communication rounds. During each round of communication, the clients performed local training for three epochs before sending the updated model parameters to the server for aggregation. To mitigate the effect of stochastic variation resulting from data partitioning and model initialization, each experiment was repeated three times, and the results reported are the mean values from the three runs.

Two scenarios of data distribution were analyzed. In the IID setting, the training samples were instantiated at clients distributed randomly and almost uniformly. Thus, the local data of each client is statistically similar. In the Non-IID setting, we assign clients' partitions that are class-imbalanced and feature-skewed to simulate realistic institutional heterogeneity, e.g., as may happen across multi-center medical environments. Experiments in the Proximal-AC settings were evaluated on a held-out test set, which was the same as that in centralized training. At the conclusion of every global round, the assessment of how well each method converged and how well it performed was measured. Furthermore, the meantime per round for training and the overall number of rounds for communication were taken as practical indicators of computational efficiency. This protocol allows for a balanced and reproducible assessment of related predictive performance and operational feasibility. It furthermore allows centralized and federated learning to be directly compared under both idealized homogeneous and more realistic heterogeneous client-data assumptions.

4. Results and Discussion

The quantitative results of the proposed framework in comparison to the centralized baseline under the IID and non-IID data distributions are reported in this section. To avoid applying random variability to the results due to data partitioning and model initialization, all results are shown as mean values across three independent experiment runs.

4.1 Performance Comparison Between Centralized and Federated Models

In **Table 3** the aggregated performance of the centralized model and the federated model in three runs. The best overall accuracy, 96.49%, its precision is 95.95%, its recall is 98.61%, and its F1 score is 97.26%. In the IID federated setting, the model achieved competitive prediction performance with 94.74% accuracy and 95.95% F1-score, although these results are slightly lower than when training is centralized. Under the Non-IID setting, the performance dropped further to an accuracy of 93.86% and F1 score of 95.30%, showing that statistical heterogeneity across clients harms federated optimization.

Table 3. Performance Measurement of Centralized and Federated Learning Models.

Model Type	Data Distribution	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Avg. Training Time per Round (s)
Centralized	—	96.49	95.95	98.61	97.26	0.0055
Federated	IID	94.74	93.42	98.61	95.95	0.0250
Federated	Non-IID	93.86	92.21	98.61	95.30	0.0249

Interestingly, the recall value recorded at 98.61% remained the same, in a centralized, IID federated, and Non-IID federated setting. This stability indicates that despite decentralizing the training and using heterogeneous local data distributions, the framework maintained high sensitivity for malignant-case detection. By contrast, the model appears to make less accurate and less precise predictions in the federated setting. It indicates that any impact from heterogeneity influenced the global balance in classification more than the model's ability to correctly identify the positive ones.

In relative terms, the accuracy of the Non-IID federated setting drops approximately 2.63 percentage points from the centralized baseline and 0.88 percentage point from the IID federated setting. The expected effect of statistical imbalance at the client level on convergence and global model consistency, on the other hand, aligns with such discrepancies. In terms of efficiency, the

average training time per round increases from 0.0055 s in the centralized setting to around 0.025 s in the federated settings, which are largely due to the added coordination and aggregation overhead. Nonetheless, the total training time was low, thus supporting this framework's practical applicability for localized medical-system deployment.

4.2 Convergence Behavior Across Communication Rounds

As depicted in **Figure 2**, during the global communication rounds, we assess the models learned using centralized and federated learning techniques on IID and Non-IID data distributions. The figure contrasts the learning stability, convergence speed, and the client-level statistical heterogeneity effect in optimizing model performance. The centralized baseline was reached early on by the federated model in the IID setting, as illustrated in **Figure 2**, due to a swift convergence. This behavior suggests that when local client datasets are statistically similar, local update aggregation continues to work and the global model can learn in a stable and effective manner. In such an environment, the federated algorithm retains much of the predictive power of centralized training while preserving the locality of data.

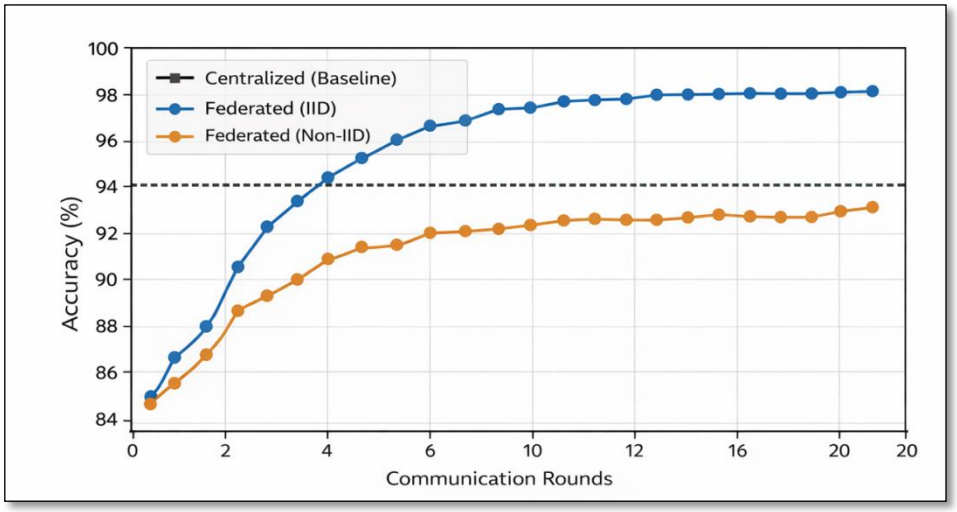

Figure 2. Convergence in accuracy of centralized and federated learning models across communication rounds for IID and Non-IID distributions

In comparison, the Non-IID setting showed slower convergence with larger round-to-round fluctuation. This behavior occurs due to statistical heterogeneity across clients, where class imbalance and feature skewness create asymmetric local gradient updates [14-16]. To approach convergence, the global model requires more communication rounds, as a result of that the aggregation process is less stable. Based on the above reasons, the Non-IID federated model shows oscillation in the early and middle rounds. However, it gradually becomes stable and converges to a similar level.

These observations corroborate the quantitative results of **Table 3**, in which the IID federated setup remained closer to the centralized baseline in terms of performance than the Non-IID setup. The analysis suggests that FedAvg exhibits good performance when used in localized medical-system settings, albeit the convergence performance of FedAvg is impacted more by the heterogeneous distributions of clients' data as compared to homogeneous distributions.

4.3 Comparison with Published Federated Learning Benchmarks

Table 4 contextualizes the reported results by comparing the current framework with selected federated learning studies on the WDBC dataset under different modeling and system assumptions. While these studies tackle the same benchmark task, direct numerical comparisons should be interpreted with caution. Indeed, the performance reported depends on the dataset itself, but also on differences in model complexity, privacy constraints, explainability requirements, system architecture, and evaluation design.

Table 4. Federated Learning Benchmarks Published for the WDBC Dataset.

Study	Model/ Setting	Accuracy (%)	Precision (%)	Recall/ Sensitivity (%)	F1-score (%)	Notes
[21]	FL + DNN (Feature Selection)	97.54	96.49	98.00	97.26	Collaborative FL framework
[22]	FL + Differential Privacy (ε = 1.9)	96.10	Not reported	Not reported	Not reported	Privacy–accuracy trade-off
[23]	FL + Explainable AI (XAI)	97.59	Not reported	Not reported	98.39	Trust-aware FL model
[24]	FL on Edge + Server (arXiv)	~99.95 (Edge) / ~98.00 (Server)	Not reported	Not reported	Not reported	Supporting preprint evidence

Their framework likely benefited from a model design that was more expressive than the lightweight MLP that has been employed in this study. In their findings, [23] demonstrate that there is a competitive performance for an explainable federated setting. This suggests that, under certain circumstances, it is possible to add some transparency to models without inducing a significant decrease in predictive performance. While [22] study, assessed federated learning with differential privacy constraints. The reported accuracy of 96.10% is consistent with the established trade-off between stronger privacy protection and predictive performance. According to[24], who present values that are far higher in the edge–server setting. Results presented as a preprint will vary due to differences in implementation scope, deployment assumptions, or evaluation conditions.

In comparison with these studies[21, 22, 24], the current federated framework was slightly less accurate under the Non-IID condition. It is scientifically expected rather than problematic. First, the present research purposefully utilized a lightweight MLP architecture to facilitate computational simplicity and local application rather than ultimate predictive complexity. The study intentionally focused on and tested IID and Non-IID client distributions in the second place. Furthermore, the heterogeneous setting introduces issues like class imbalance and feature skewness, making optimization more challenging. In addition, our work did not consider enhancement mechanisms that might change the performance profiles studied previously, such as differential privacy calibration, feature-selection pipelines, explainability modules or edge-specialized optimizers.

The comparative evidence presented in **Table 4** yields two findings. Initially, the accomplished performance of the proposed framework is within a practically plausible range when compared with the available literature on WDBC-based federated learning. Also, differences in performance across studies must be considered in light of design choices made, not simply rankings we obtain. The current findings lend support to the framework's ability to serve as a reproducible pilot assessment of federated learning feasibility in localized medical-system settings, rather than an attempt to outperform more specialized or privacy-enhanced versions of federated learning.

4.4 Summary of Key Findings

The experimental results indicate that the performance of federated learning is reasonably close to centralized training as long as client data is statistically balanced. The accuracy of the federated framework in the IID setting was not too far behind the conventional baseline accuracy. Under the Non-IID setting, the performance decreases slightly; client-level statistical heterogeneity does impact federated optimization. However, the framework remained stable and showed high recall, indicating consistent malignant-case detection sensitivity. Despite the addition of communication and aggregation delays in federated training, the observed cost was still operationally reasonable. As a whole, the results lend support to the viability of federated learning as a reproducible data-locality preservation collaborative prediction strategy for localized medical systems.

5. Conclusion

During the Wisconsin Breast Cancer Diagnostic study, a pilot study of federated learning for localized medical system was presented. The experiment comparison indicates that federated training can reach performance near that of centralized learning by meeting IID while it also remains sufficiently robust under Non-IID. The federated model obtained 94.74% accuracy in IID setting and 93.86% in the Non-IID setting, compared with 96.49% for the centralized baseline, for an overall degradation of about 2.6 percentage points over the heterogeneous setting. Recall remained consistent at 98.61% across all configurations, indicating that the framework maintained a strong capacity to identify malignant cases.

These findings suggest that federated learning can be a practical strategy to cooperatively make medical predictions when raw-data sharing is limited. The study findings indicate that the main performance differences stem less from decentralization per se than from client-level statistical heterogeneity, model simplicity, and distributed optimization constraints. This result is important from an applied perspective since it is greater realistic for confined healthcare infrastructures to keep (local) data at a single place than to transfer institutional datasets to a central repository.

Simultaneously, the present study has a number of limitations. The framework was tested on one open-source tabular data set using a lightweight MLP architecture, and the reported results are based on the mean performance over three runs without sophisticated statistical significance tests. Moreover, the research does not offer a formal privacy mechanism in the form of differential privacy and secure aggregation but does maintain data locality. For that reason, the present contribution is to be understood as a reproducible feasibility-oriented pilot assessment, not as a privacy-guarantee or algorithmically novel federated learning system.

Henceforth, adding formal privacy-preserving efforts, broader statistical analyses, more complex clinical datasets as well as stronger models would complete this framework. By extending the federated learning setting to more sophisticated problems, the results may direct a broader evaluation of its reliability and scaling.

Acknowledgment

Many thanks to Ibn Al-Haitham Journal for considering my paper to be published at their esteemed platform.

Conflict of Interest

There is no conflict to disclose.

Funding

This research did not receive any external funding.

References

1. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, Bakas S, Galtier MN, Landman BA, Maier-Hein KH, Ourselin S, Sheller MJ, Summers RM, Trask A, Xu D, Baust M, Cardoso MJ. The future of digital health with federated learning. npj Digit Med. 2020;3(1):119. https://doi.org/10.1038/s41746-020-00323-1
2. Kairouz P, McMahan HB, Avent B, Bellet A, Bennis M, Bhagoji AN, Bonawitz K, Charles Z, Cormode G, Cummings R, D'Oliveira R, Eichner H, El Rouayheb S, Evans D, Gardner J, Garrett Z, Gascón A, Ghazi B, Gibbons PB, Gruteser M, Harchaoui Z, He C, He L, Huo Z, Hutchinson B, Hsu J, Jaggi M, Javidi T, Joshi G, Khodak M, Konecny J, Korolova A, Koushanfar F, Koyejo S, Lepoint T, Liu Y, Mittal P, Mohri M, Nock R, Özgür A, Pagh R, Qi H, Ramage D, Raskar R, Raykova M, Song D, Song W, Stich SU, Sun Z, Suresh AT, Tramèr F, Vepakomma P, Wang J, Xiong L, Xu Z, Yang Q, Yu FX, Yu H, Zhao S. Advances and Open Problems in Federated Learning. Found Trends Mach Learn. 2021;14(1–2):1–210. https://doi.org/10.1561/2200000083

3. McMahan HB, Moore E, Ramage D, Hampson S, Agüera y Arcas B. Communication-efficient learning of deep networks from decentralized data. In: Proc 20th Int Conf Artif Intell Stat (AISTATS). PMLR; 2017. p. 1273–1282. URL: https://proceedings.mlr.press/v54/mcmahan17a.html
4. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, Milchenko M, Xu W, Marcus D, Colen RR, Bakas S. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Sci Rep. 2020;10:12598. https://doi.org/10.1038/s41598-020-69250-1
5. Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat Med. 2021;27:1735–1743. https://doi.org/10.1038/s41591-021-01506-3
6. Antunes RS, da Costa CA, Küderle A, Yari IA, Eskofier BM. Federated Learning for Healthcare: Systematic Review and Architecture Proposal. ACM Trans Intell Syst Technol. 2022;13(4):54:1–54:23. https://doi.org/10.1145/3501813
7. Teo ZL, Jin L, Liu N, Li S, Miao D, Zhang X. Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. Cell Rep Med. 2024;5(2):101419. https://doi.org/10.1016/j.xcrm.2024.101419
8. Bonawitz K, Eichner H, Grieskamp W, Huba D, Ingerman A, Ivanov V, Kiddon C, Konečný J, Mazzocchi S, McMahan HB, Van Overveldt T. Towards federated learning at scale: System design. Proc Mach Learn Syst (MLSys). 2019;1:374–388. URL: https://proceedings.mlsys.org/paper/2019/file/7b8b965ad4bca0e41ab51de7b31363a1-Paper.pdf
9. Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. Federated learning of predictive models from federated electronic health records. Int J Med Inform. 2018;112:59–67. https://doi.org/10.1016/j.ijmedinf.2018.01.007
10. Liu D, Chen J, Liu Q, Zhou B, Feng J. Federated learning for disease classification using multi-center clinical datasets. IEEE Access. 2020;8:113173–113181. https://doi.org/10.1109/ACCESS.2020.3003723
11. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nat Mach Intell. 2020;2(6):305–311. https://doi.org/10.1038/s42256-020-0186-1
12. Pfitzner B, Steckhan N, Arnrich B. Federated learning in a medical context: A systematic literature review. ACM Trans Internet Technol. 2021;21(2):1–31. https://doi.org/10.1145/3449390
13. Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: Concept and applications. ACM Trans Intell Syst Technol. 2019;10(2):1–19. https://doi.org/10.1145/3298981
14. Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated Optimization in Heterogeneous Networks. Proc Mach Learn Syst. 2020;2.
15. Karimireddy SP, Kale S, Mohri M, Reddi S, Stich S, Suresh AT. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. In: Proc 37th Int Conf Mach Learn (ICML). PMLR. 2020;119:5132–5143.
16. Li X, Jiang M, Zhang X, Kamp M, Dou Q. FedBN: Federated Learning on Non-IID Features via Local Batch Normalization. In: Int Conf Learn Represent (ICLR). 2021.
17. Dua D, Graff C. UCI Machine Learning Repository. Irvine (CA): University of California, School of Information and Computer Sciences; 2019. https://archive.ics.uci.edu/ml
18. Wolberg WH, Street WN, Mangasarian OL. Machine learning techniques to diagnose breast cancer from fine-needle aspirates. Cancer Lett. 1994;77(2–3):163–171. https://doi.org/10.1016/0304-3835(94)90099-X
19. Oh W, Nadkarni GN. Federated Learning in Health care Using Structured Medical Data. Adv Kidney Dis Health. 2023;30(1):4–16. https://doi.org/10.1053/j.akdh.2022.11.007
20. Beutel DJ, Topal T, Mathur A, Qiu X, Fernandez-Marques J, Gao Y. Flower: A Friendly Federated Learning Research Framework. arXiv. 2020;arXiv:2007.14390.
21. Almufareh MF, Tariq N, Humayun M, Almas B. A federated learning approach to breast cancer prediction in a collaborative learning framework. Healthcare. 2023;11(24):3185. https://doi.org/10.3390/healthcare11243185
22. Shukla S, Rajkumar S, Sinha A, Esha M, Elango K, Sampath V. Federated learning with differential privacy for breast cancer diagnosis enabling secure data sharing and model integrity. Sci Rep. 2025;15:13061. https://doi.org/10.1038/s41598-025-95858-2

23. Briola E, Nikolaidis CC, Perifanis V, Pavlidis N, Efraimidis P. A federated explainable AI model for breast cancer classification. In: Proc Eur Interdiscip Cybersecur Conf. 2024. https://doi.org/10.1145/3655693.3660255
24. Jha M, Maitri S, Lohithdakshan M, Duela SJ, Raja K. PrivFED—A framework for privacy-preserving federated learning in enhanced breast cancer diagnosis. arXiv. 2024;2405.08084. https://arxiv.org/abs/2405.08084