Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 14 May 2026
Revised 18 June 2026
Accepted 02 July 2026

Natural Resources for Human Health 2026, Vol. 6 (4s): 343–362

KEYWORDS:
Lorem ipsum; dolor sit amet; consectetur; adipiscing elit; natural resources; phytochemistry.

An Empirical Study of Federated Brain Tumor Classification Under IID and Non-IID Client Distributions with DP-Inspired Local Perturbation

[1]Aswathy R Nair, [2]N.Sivakumar

[1]Marwdai University,India
[2] Marwadi University,India

ABSTRACT: Federated learning (FL) is an approach to protect privacy during medical image analysis due to the possibility to jointly train the model without sharing data between institutions. In this study, we provide a comparative analysis of federated brain tumor classification performance in case of IID and non-IID clients with and without differential privacy (DP)-based local perturbation. The performance metrics are calculated according to the accuracy, F1-score, and AUC measures for the multi-client FL convolutional neural network (CNN) based binary classifier. We demonstrate that the best results were obtained for the IID client configuration without DP application, with 97.30% test accuracy, 0.9754 F1-score, and 0.9947 AUC values, while the best configuration among the non-IID ones resulted in 93.30% test accuracy, 0.9401 F1-score, and 0.9761 AUC values. The application of DP caused significant degradation of classification performance with IID and non-IID configurations leading to 84.25% and 78.79% test accuracies, respectively. This demonstrates the presence of a privacy-utility tradeoff. Overall, the results suggest that federated brain tumor classification is achievable and works effectively; however, the effectiveness of federated brain tumor classification depends greatly on the client data heterogeneity and DP application.

1. INTRODUCTION

The problem of classifying brain tumors using CT and MRI scans is still relevant today because an early and accurate diagnosis is critical for effective treatment. It is now widely known that deep learning models achieve good results in the domain of image processing; however, they require access to large datasets for proper training. Unfortunately, medical images are stored separately at multiple institutions due to legal, moral, and organizational restrictions associated with sharing private information. For this reason, Federated Learning (FL) has been proposed as a new technique in machine learning that allows multiple parties to cooperatively train models by only exchanging models' parameters and not sharing raw data [1], [2], [3]. FL is gaining increasing importance in the field of medical image processing and AI-assisted medicine [4], [5], [6].

However, practical implementation of FL in healthcare remains challenging due to several key factors. For example, the first one is the problem of non-independent and identically distributed (non-IID) client data distribution, whereby the sample sizes, imaging equipment, demographics, and class imbalance may vary between different hospitals, leading to an unstable convergence and decreased prediction accuracy [7], [8], [9]. Another issue to consider when applying FL to healthcare data involves the adoption of privacy protection techniques, such as differential privacy (DP), which guarantees the privacy of individual records by clipping gradients and adding noise, although at the cost of model prediction capability [10], [11], [12]. Previous research has established that the issues of heterogeneity and privacy are fundamental when applying FL to medical imaging tasks, such as the classification of brain tumors [13], [14], [15].

1.1. Problem Statement

It is imperative to diagnose and plan the treatment of brain tumors accurately through CT and MRI imaging for which accurate brain tumor classification is necessary since these imaging modalities provide anatomical and pathological information about the brain. However, training effective deep learning models for such tasks requires huge and varied datasets, which are challenging to collect as medical images are distributed among several hospitals and are constrained by the issues of data security, ethical considerations, and regulations regarding data exchange. Federated Learning (FL) can help overcome this limitation by allowing collaborative learning using decentralized client machines with no requirement of transmitting raw data from the machines to a central server. FL, however, faces certain challenges related to non-IID client data distributions due to imbalance, heterogeneity, different data sizes and modalities at each hospital client, along with privacy-preserving perturbations through differential privacy-like noise injection techniques that could affect its efficiency in practice. Though several studies have investigated federated learning and privacy-preserving methods individually, medical image classification using multimodal CT-MRI data, along with non-IID setting, has not received much attention. Hence, the problem that will be considered in this paper is to assess the effectiveness of federated learning algorithms in achieving reliable predictive performance, efficient model convergence, and good system efficiency in the scenario of non-IID clients, with privacy-preserving local perturbations.

1.2 Research Gap

Even though several research efforts have proved the utility of federated learning approaches in medical image classification, especially brain tumor segmentation, most previous works consider single-modal image databases, advanced optimization techniques, or differential privacy-based schemes alone. Few studies have considered the possibility of federated brain tumor classification in a distributed manner based on an actual multimodal dataset having CT and MRI images. Moreover, not many studies provide an empirical investigation regarding the relationship between different distribution scenarios of clients (e.g., IID vs. non-IID) and the role of local perturbation inspired by differential privacy (DP) in terms of accuracy, convergence, communication complexity, computational burden, and overall privacy-utility trade-off. Most previous works have also prioritized achieving high accuracy levels without much consideration of the overall feasibility issues associated with practical applications. This is another major gap that needs to be filled through our research study.

1.3. Novelty of the Research

The novelty of the research is in the proposal of an empirical benchmark that combines the utility of federated brain tumor classification on a real multimodal dataset of CT and MRI images by leveraging two relevant imaging sources at once in the same decentralization setting. Unlike several studies, which use MRI-only datasets, assume IID distribution, or analyze privacy alone, this study examines both IID versus non-IID distribution and with or without

differential privacy-inspired local perturbation in several federated learning settings. Furthermore, while most existing studies focus solely on predictive accuracy, this study also analyzes other aspects such as convergence behavior, communication costs, and overall system performance.

1.4 Contributions

- Created a federated learning approach for the task of binary brain tumor classification utilizing a unique multimodal dataset with both CT and MRI scans.
- Modeled practical decentralized medical environments via five clients divided into IID and non-IID data splits.
- Investigated eight experimental setups with different combinations of IID/Non-IID distributions, FedAvg/FedProx models, and local perturbations with noise levels equal to 0 and 0.1.
- Measured performance using the Accuracy, Precision, Recall, F1-score, Specificity, AUC, and Loss metrics.
- Obtained excellent classification results, where the best IID no-noise condition yielded an Accuracy score of 97.29%, F1-score of 0.9754, and AUC of 0.9947.
- Studied round-based convergence dynamics and systems' properties such as communication cost, training time, CPU consumption, DP overhead, and the approximate privacy measurement.
- Successfully identified that the IID no-noise configuration outperformed other experimental conditions, whereas the non-IID noisy configuration demonstrated the most significant drop in performance due to client heterogeneity and privacy-utility trade-offs in federated health applications.

- Supplied a thorough performance benchmark considering accuracy, privacy, heterogeneity, and scalability issues in federated learning, providing an exemplary basis for further studies.

2. LITERATURE REVIEW

2.1. IID vs. Non-IID Performance: A Critical Gap

An important finding from the above literature review is that the explicit separation of IID and non-IID accuracies is not always observed and sometimes missing for brain tumor FL studies [7], [11], [12], [13]. Specifically, Ay et al. [7] assessed several FL algorithms only using IID-like data split settings, and FedAvg achieved an accuracy rate of 85.55% after 10 rounds while Ft-FedAvg gained an accuracy rate of 85.80% after 30 rounds without considering any non-IID configurations [7]. Furthermore, Mahlool and Abed [11], [12] presented their classification results as accuracies of 98% and 97.14% for two brain tumor datasets and 0.82 and 0.96 for two more datasets, yet there was no investigation into any non-IID data configuration [11], [12]. In addition, Nandan et al. [13] achieved 99.7% accuracy without considering any non-IID configurations [13]. The above pattern has also been reflected in other studies on FL in the healthcare sector. Rauniyar et al. [3] found out that there was potential for DP-FL in brain tumor segmentation by studying studies involving the BraTS 2018 dataset, but they failed to establish comparisons between IID and non-IID data performance [3]. Prayitno et al. [4] confirmed the reality of non-IID data in the medical field, noting the negative effects of non-IID data distribution in models, but performance degradation from non-IID data was inconsistently quantified across studies [4]. The performance issues arising from non-IID data have been affirmed by Madathil et al. [5] that FedAvg models using non-IID data show poor accuracies and slow convergence rates as explained by Zhao et al. [5].

2.2. Studies Addressing Non-IID Performance

The studies examined herein revealed Narmadha and Varalakshmi [9] as the only study with clear information on how non-IID influences the performance metrics with reference to the number of communication rounds needed to attain target Dice scores being 40–70% higher for non-IID compared to IID configurations although the exact Dice score coefficient for the former configuration was not provided separately [9]. The overall Dice score of 0.94 obtained is with regard to IID configuration of the Kaggle LGG-MRI dataset [9]. In their experiment, Raheem et al. [10] observed consistent supremacy of FUSION in relation to SOTA FL frameworks when non-IID configuration is considered among TCGA-LGG brain MRI dataset, Kvasir-SEG, and UDIAT datasets despite the numerical value of the accuracy being expressed relatively to SOTA [10]. In their research, Alsaleh et al. [8] simulated client distribution in both IID and non-IID configurations with an overall accuracy of 99.07% being achieved in IID-like conditions while observing convergence instability in the non-IID condition though the accuracy figure for non-IID condition could not be determined separately [8]. Rehman et al. [1] analyzed some studies which had revealed FedDis superiority over baseline FL approaches by 11% in terms of non-IID configuration in relation to brain MRI segmentation [1], [6].

2.3. Impact of Differential Privacy on IID and Non-IID Accuracy

The connection between differential privacy and the IID and non-IID impact is a further layer of complication. It was shown by Ay et al. [7] that Dp-FedAvg incurred accuracy penalties in comparison with the classic FedAvg in the IID setting, which follows the well-known trade-off between privacy and accuracy found throughout the FL literature [14], [15], [16]. In particular, it was found by Shukla et al. [16] in research on breast cancer FL using DP that the combination of FL and DP resulted in a 96.1% accuracy for $\varepsilon = 1.9$, versus 97.7% in the absence of DP; this confirms that there is an inherent accuracy penalty associated with the use of DP [16]. Similarly, it was established by Ziegler et al. [17] that DP FL applied to chest X-ray classification using DenseNet121 attained an AUC of 0.94 for $\varepsilon = 6$, with performance decreasing for stricter privacy levels [17]. This has particular implications in terms of brain tumor FL, as non-IID and DP have not been considered simultaneously in the brain tumor FL literature [1], [3].

2.4. Scalability and Its Relationship to IID/ Non-IID Performance

The scalability challenges directly affect the ability to analyze the effectiveness of non-IID performance. The works by Ay et al. [7] have involved only three federated clients, while Alsaleh et al. [8] simulated only five clients, and the same is true about the work done by Mahlool and Abed [11], [12]. All of these are well under the scale of realistic multi-institutional deployments [1], [3]. Rauniyar et al. [3] stated that scalable Federated Learning for brain tumor applications involves overcoming not only

statistical but also system heterogeneity, but very few researchers take on this challenge together [3]. In their paper, Madathil et al. [5] have proven that there are two main forms of heterogeneity in federated healthcare settings, namely statistical and system heterogeneity, and the latter should be considered simultaneously for achieving scalable deployments [5].

2.5. Multimodal FL and IID/ Non-IID Considerations

In contrast to conventional non-IID label or quantity skew, multimodal FL adds further layers of heterogeneity. Specifically, according to Che et al. [18], multimodal federated learning adds modality heterogeneity as an extra challenge where different clients might have different modality combinations [18]. In brain tumor FL studies, full integration of multimodal inputs, such as T1, T2, and FLAIR MRI sequences, is still at a conceptual level [1], [9], and none of the papers under review provided separate results on whether multimodal brain tumor FL was IID or non-IID in terms of its accuracy rates [1], [9]. According to Jabbar et al. [19], in vertical FL for glaucoma detection, their proposed QAVFL architecture performed well with non-IID data, reaching an accuracy rate of 98.6% [19].

2.6. Consolidated Research Gaps

Summarizing upon the above review, the following research gaps are identified regarding IID and non-IID performance reporting and evaluation :

- Lack of explicit IID and non-IID accuracy reporting: A large proportion of brain tumor FL studies only provide one accuracy value without breaking down the performance metrics based on the data distribution condition [7], [11], [12], [13]; thus, it is hard to make comparisons among studies.
- Non-IID robustness: Most brain tumor FL studies do not perform tests with non-IID distributions [7], [11], [12], [13], or they show performance deterioration without solving convergence problems [5], [8], [9].
- Differential Privacy (DP)-non-IID impact: There is no quantitative analysis on how DP and non-IID distribution affect the accuracy loss of brain tumor FL studies [1], [3], [17].
- Multimodal non-IID analysis: To date, there is no study that investigates IID and non-IID performance for multimodal brain tumor FL systems [1], [9], [18].
- Scalability issues: The vast majority of brain tumor FL studies are conducted with a small number of clients (typically 3–5 clients), which falls far short of practical deployment [3], [5], [7], [8].

Table 1. Comparative summary of selected literature related to federated medical imaging and brain tumor analysis Part-1

Ref. No.	Author(s)	Year	Dataset Used	Multimodal	IID	Non-IID	FL Algorithm Used	Differential Privacy Applied	Scalability Addressed?
[6]	Ay et al.	2024	Brain MRI datasets	No	Yes	No	FedAvg variants	Yes	Partial
[2]	Alsaleh et al.	2025	CT + synthetic MRI	Yes	Yes	Yes	FedAvg, FedProx	No	Partial
[4]	Narmadha & Varalakshmi	2025	LGG-MRI	Partial	Yes	Yes	FedAvg	Mentioned	Partial
[5]	Raheem et al.	2025	TCGA-LGG + others	No	Yes	Yes	Fusion FL	No	Yes
[1]	Rehman et al.	2023	Multi-modal brain scans	Yes	Yes	Yes	FedAvg, FLOP, FedDis	Yes	Partial

[14]	Mahlool & Abed	2022	MRI datasets	No	Yes	No	FedAvg	No	Partial
[16]	Nandan et al.	2024	Brain tumor MRI	No	Yes	No	FedAvg	No	No
Proposed	**Present Study**	**2026**	**Single multimodal CT + MRI dataset**	**Yes**	**Yes**	**Yes**	**FedAvg / FedProx**	**Yes**	**Yes**

Table 2. Comparative summary of selected literature related to federated medical imaging and brain tumor analysis Part-2

Ref. No.	Author(s)	Performance	Accuracy (IID)	Accuracy (Non-IID)	Main Limitation / Gap
[6]	Ay et al.	85.80%	85.80%	Not reported	No non-IID, no multimodal
[2]	Alsaleh et al.	99.07%	99.07%	Not separately reported	Synthetic MRI, no formal DP
[4]	Narmadha & Varalakshmi	Dice 0.94	Dice 0.94	Not separate	Multimodal not implemented
[5]	Raheem et al.	Strong segmentation results	Competitive	Strong	No formal DP
[1]	Rehman et al.	FedDis +11%	Comparable	11%	Review-based
[14]	Mahlool & Abed	98.00%	98.00%	Not reported	No DP, no non-IID
[16]	Nandan et al.	99.70%	99.70%	Not reported	No DP, no non-IID
Proposed	**Present Study**	**97.29% Acc, F1=0.9754, AUC=0.9947**	**97.29%**	**93.30% (best no-noise), 79.26% (noisy)**	**Strong unified benchmark with multimodal + non-IID + privacy + system metrics**

3. METHODOLOGY

3.1 Overall Federated Framework

The paper adopted a controlled federated learning model to classify binary brain tumors to focus on the impacts of data distribution to clients and the local perturbation that is inspired by DP. The procedure included the extraction, preprocessing, and stratified splitting, the partitioning of the dataset into IID and Non-IID clients, the round-based federated training, the equal-weight server aggregation, and the assessment conducted by predictive and system-level metrics. A total of eight

combinations of two data modes (IID, Non-IID), two algorithm labels (FedAvg, FedProx) and two noise levels (0.0, 0.1) were investigated.

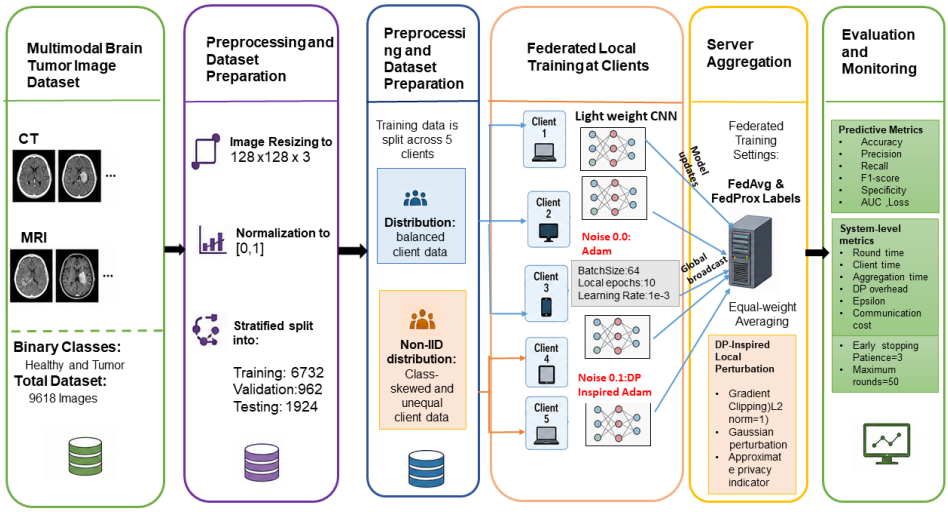

Figure 1. Overview of the proposed federated brain tumor classification framework under IID and Non-IID client distributions with DP-inspired local perturbation.

3.2 Dataset Preparation and Experimental Setup

The dataset contained 9,618 CT and MRI brain images, including 4,300 healthy and 5,318 tumor samples [45]. Labels were assigned from folder names, with healthy coded as 0 and tumor as 1. All images were resized to $128 \times 128 \times 3$ and normalized to [0,1]. Stratified sampling produced 6,732 training, 962 validation, and 1,924 test samples.

Let the dataset be $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i$is an image and $y_i \in \{0,1\}$. The split was $\mathcal{D} = \mathcal{D}_{tr} \cup \mathcal{D}_{val} \cup \mathcal{D}_{te}$, with $| \mathcal{D}_{tr} | = 6732$, $| \mathcal{D}_{val} | = 962$, and $| \mathcal{D}_{te} | = 1924$. Table 2 summarizes the fixed experimental settings used throughout the federated framework.

Table 3. Experimental setup of the proposed federated brain tumor classification framework

Parameter	Value
Task	Binary brain tumor classification
Input image size	128 × 128 × 3
Total dataset size	9,618 images
Training / Validation / Test	6,732 / 962 / 1,924
Number of clients	5
Data distribution modes	IID, Non-IID
Algorithm labels	FedAvg, FedProx
Local model	CNN
Convolution filters	32, 64, 128
Hidden dense layer	64 neurons
Output layer	1 sigmoid neuron

Batch size	64
Local epochs	10
Maximum communication rounds	50
Learning rate	1×10^{-3}
Optimizer (no-noise)	Adam
Optimizer (privacy-aware)	DP-inspired Adam
Noise multipliers	0.0, 0.1
L_2 clipping norm	1
δ	1×10^{-5}
Aggregation rule	Equal-weight averaging
Early stopping patience	3

Table 3 presents the fixed configuration applied across all eight experimental settings. These shared settings ensured that the observed differences in performance were examined under a common framework while varying only the data-distribution mode, algorithm label, and perturbation setting.

3.3 Client Partitioning Under IID and Non-IID Settings

The training subset was distributed across $K = 5$ clients. Under IID partitioning, shuffled training indices were divided approximately equally:

$$\mathcal{D}_{tr} = \bigcup_{k=1}^{K} \mathcal{D}_k^{IID}, \mathcal{D}_i^{IID} \cap \mathcal{D}_j^{IID} = \emptyset \ (i \neq j) \quad (1)$$

This produced client sizes of 1347, 1347, 1346, 1346, and 1346. Under Non-IID partitioning, class-wise indices were first separated and then one class was made dominant per client; each client retained 80% of its dominant-class subset and 20% of its minority-class subset:

$$\mathcal{D}_k^{Non\text{-}IID} = \mathcal{S}_{k,dom} \cup \mathcal{S}_{k,min} \quad (2)$$

This created unequal local sizes of 716, 630, 715, 629, and 715 and class-skewed client data. Table 4 summarizes the client-wise sample distribution under the two partitioning schemes.

Table 4. Client data distribution under IID and Non-IID settings

Setting	Client 1	Client 2	Client 3	Client 4	Client 5
IID	1347	1347	1346	1346	1346
Non-IID	716	630	715	629	715

Table 4 shows that the IID setting preserved an almost uniform allocation across all five clients, whereas the Non-IID setting produced unequal client sizes. This distributional contrast provides the basis for examining how balanced and heterogeneous client partitions affect federated model behavior in the later results.

3.4 Local Classification Model

A lightweight CNN was used as the local model on both client and server sides. The architecture contained three convolutional stages with 32, 64, and 128 filters, each followed by max-pooling, then flattening, a dense layer with 64 neurons, and a sigmoid output neuron. For an input image x, the model predicted $\hat{y} = f(x; \theta)$, and local training minimized binary cross-entropy:

$$\mathcal{L}_{BCE} = -\frac{1}{m}\sum_{i=1}^{m}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \quad (3)$$

3.5 DP-Inspired Local Optimization and Privacy Indicator

When the noise multiplier was 0.0, standard Adam was used. When $\sigma > 0$, a DP-inspired Adam variant applied gradient clipping and Gaussian perturbation. For each gradient tensor g,

$$\tilde{g} = g \cdot \min\left(1, \frac{C}{\|g\|_2 + \varepsilon_s}\right), \hat{g} = \tilde{g} + \mathcal{N}(0, \sigma^2 C^2) \quad (4)$$

where $C = 1$and $\sigma \in \{0.0, 0.1\}$. An approximate privacy indicator was tracked as

$$\epsilon = \begin{cases} 0, & \sigma = 0, \\ \frac{\sqrt{2r\ln(1/\delta)}}{\sigma}, & \sigma > 0, \end{cases} \quad (5)$$

with round index $r$and $\delta = 10^{-5}$. This $\epsilon$was treated as a comparative indicator, not a formal accountant-based differential privacy guarantee.

3.6 Federated Training and Global Aggregation

For each configuration, the server initialized a global CNN and broadcast it to all clients each communication round. Every client trained a local copy on its own partition for 10 epochs and returned updated parameters. The server aggregated client models by equal-weight averaging:

$$\theta^{(r+1)} = \frac{1}{K}\sum_{k=1}^{K} \theta_k^{(r+1)} \quad (6)$$

The same averaging rule was used under both FedAvg and FedProx labels. However, the implementation did not add a proximal term to the local FedProx objective, so the FedAvg/FedProx comparison should be interpreted as a label-level empirical contrast within the same averaging-based loop. Table 5 lists the eight evaluated experimental configurations defined by the combination of data mode, algorithm label, and noise multiplier.

Table 5. Evaluated experimental configurations

Case	Data Mode	Algorithm Label	Noise Multiplier
1	IID	FedAvg	0
2	IID	FedAvg	0.1
3	IID	FedProx	0
4	IID	FedProx	0.1
5	Non-IID	FedAvg	0
6	Non-IID	FedAvg	0.1
7	Non-IID	FedProx	0
8	Non-IID	FedProx	0.1

Table 5 shows the full experimental matrix used in the study. This design enabled direct comparison across balanced and heterogeneous client distributions, as well as between unperturbed and perturbation-enabled local training settings within the same federated framework.

Algorithm 1. Federated brain tumor classification under IID/Non-IID partitioning with DP-inspired local perturbation

1. Extract and preprocess all images.

2. Split the dataset into training, validation, and test subsets using stratified sampling.
3. Partition the training set into IID and Non-IID client subsets.
4. For each combination of data mode, algorithm label, and noise setting, initialize the global CNN.
5. For each communication round, broadcast the global model to all clients.
6. Train each client locally for 10 epochs using batch size 64.
7. If $\sigma > 0$, apply clipping and Gaussian perturbation during local optimization.
8. Aggregate returned client parameters by equal-weight averaging.
9. Evaluate the global model and apply early stopping based on validation loss.

3.7 Evaluation Metrics and Early Stopping

The global model was tested after every round on training, validation and on test sets. The desired predictive measures were accuracy, precision, recall, F1-score, specificity, AUC, and loss. Round time, client training time, aggregation time, DP overhead, approximate epsilon, communication cost and CPU usage were system level metrics. Hard predictions were based on a threshold of 0.5. On the transmission of parameters at float32 precision, communication cost was computed, and on the transmission of parameters at float64 precision, DP overhead has also been computed against the equivalent no-noise client-training base. Early termination restarted the patience parameter in case the validation loss raised and finished training on three sequential non-enhancing cycles. Table 6 summarizes the evaluation metrics used in the study.

Table 6. Evaluation metrics used in the study

Category	Metrics
Predictive metrics	Accuracy, Precision, Recall, F1-score, Specificity, AUC, Loss
System-level metrics	Round time, Client training time, Aggregation time, DP overhead, Approximate epsilon, Communication cost, CPU usage

Table 6 shows that the evaluation framework combined predictive and computational measures. This allowed the study to assess not only classification quality, but also the round-level system behavior associated with different data-distribution and perturbation settings.

4. RESULTS

4.1 Experimental Overview

This is where the results of the eight experimental settings that are created by the combination of two client-distribution modes (IID and Non-IID), two labels of the algorithms (FedAvg and FedProx) and two local-noise environments (0 and 0.1) are reported. The obtained final results are provided in four complementary perspectives: final predictive performance of all environments, degradation compared to the highest possible baseline, representative round-wise behaviour, and representative system-level properties. As per the result package approved, the body of the manuscript refers to the approved final tables only, and the figures are obtained out of the already approved final result tables and the reported round-wise traces only. To give a visual summarization instantly of the predictive ranking in all settings, Figure 2 compares the end test accuracy, F1-score, and AUC processes of the eight experimental settings.

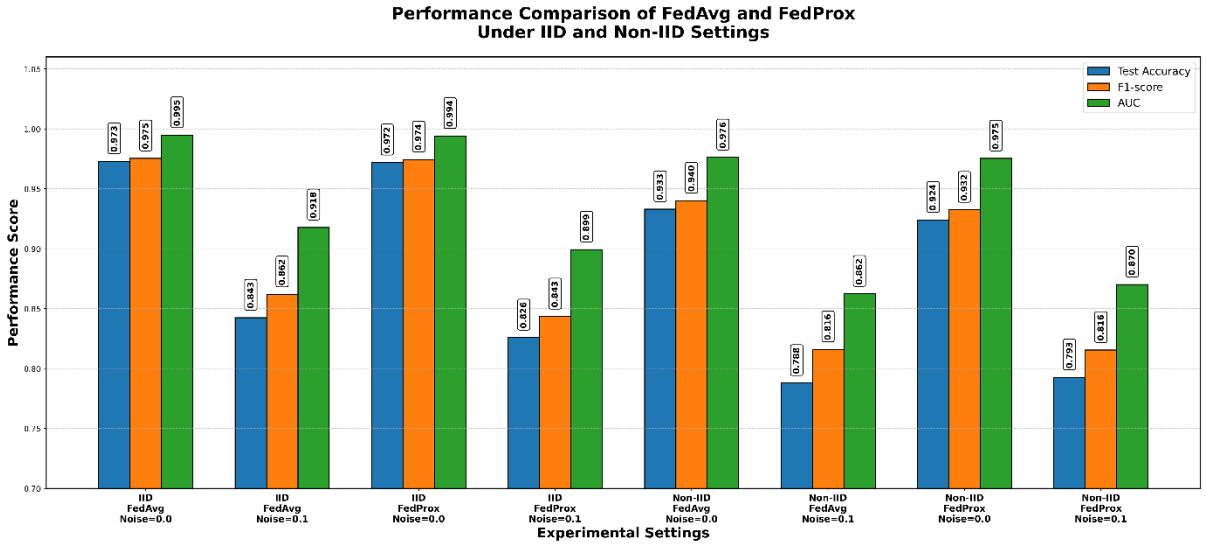

Figure 2. Overall comparison of final predictive performance across the eight experimental settings.

Figure 2 is a strong indicator of the high and low-performing IID no-noise settings and the noisy settings, respectively. The two no-noise cases of IID dominate the performance range in all the three measures but the noisy Non-IID settings are at the lower end of the comparison. The figure thus gives a brief visual account of the same ranking that is found in Table 6 as a numerical account.

4.2 Final Performance Comparison Across All Settings

The final performance table approved gives the principal summary of the end-point predictive performance of the eight configurations. The best ultimate score is provided in the IID, FedAvg, noise 0 condition with test accuracy is 0.972973, F1-score is 0.975402 and AUC is 0.994667. Corresponding IID FedProx, noise 0 is numerically similar and thus with an accuracy of 0.971933, F1-score of 0.974383 and AUC 0.994044. The weakest final results are registered at the bottom of the range under the noisy Non-IID settings. Specifically, Non-IID, FedAvg, noise 0.1 yields accuracy 0.787942, F1-score 0.816050, and AUC 0.862315, while Non-IID, FedProx, noise 0.1 yields accuracy 0.792620, F1-score 0.815534, and AUC 0.870125. Table 7 gives the comparative final.

Table 7. Final performance comparison across all experimental settings

Data Setting	Algorithm Label	Noise	Epsilon (ϵ)	Test Accuracy	F1-score	AUC
IID	FedAvg	0	0	0.972973	0.975402	0.994667
IID	FedAvg	0.1	126.957062	0.842516	0.861833	0.917774
IID	FedProx	0	0	0.971933	0.974383	0.994044
IID	FedProx	0.1	135.722808	0.825884	0.843385	0.899297
Non-IID	FedAvg	0	0	0.932952	0.940084	0.976132
Non-IID	FedAvg	0.1	143.955777	0.787942	0.816050	0.862315
Non-IID	FedProx	0	0	0.924116	0.932470	0.975355
Non-IID	FedProx	0.1	143.955777	0.792620	0.815534	0.870125

Table 7 indicates a similar ranking among metrics. At matched no-noise conditions the IID settings are greater than the Non-IID settings of both labels. The noisy setting is smaller than the no-noise setting in the matched conditions of data-distribution.

The two no-noise rows of IID are hardly different in their final predictive performance, and the noisy environments are shown in a distinctly different performance band.

4.3 Performance Degradation Relative to the Best Baseline

The amount of degradation with respect to the strongest observed configuration is measured by the table of approved results as a result of degradation with respect to IID, FedAvg, noise 0 baseline. This presentation renders the relative decrease in predictive performance clear and not to be deduced throughout Table 6. In Table 8, the degradation values have been summarized.

Table 8. Performance degradation relative to the best baseline (IID, FedAvg, no noise)

Comparison Case	Accuracy Drop	F1-score Drop	AUC Drop
IID, FedAvg, Noise 0.1	0.130457	0.113569	0.076893
IID, FedProx, Noise 0.0	0.001040	0.001019	0.000623
IID, FedProx, Noise 0.1	0.147089	0.132017	0.095370
Non-IID, FedAvg, Noise 0.0	0.040021	0.035318	0.018535
Non-IID, FedAvg, Noise 0.1	0.185031	0.159352	0.132352
Non-IID, FedProx, Noise 0.0	0.048857	0.042932	0.019312
Non-IID, FedProx, Noise 0.1	0.180353	0.159868	0.124542

The least degradation is exhibited by IID, FedProx, noise 0 with 0.001040, 0.001019, 0.000623 reduction in accuracy, F1-score, and AUC respectively. Maximum degradation is experienced in Non-IID, FedAvg, noise 0.1, where accuracy, F1-score and AUC are reduced by 0.185031, 0.159352 and 0.132352. The two noisy Non-IID settings thus demonstrate the highest distance as compared to the best performing baseline. The degradation of the test accuracy, F1-score, and AUC in comparison to the IID, FedAvg, and no-noise baseline is plotted in Figure 3 in order to provide a better visual comparison of the deterioration of the performance under the best-performing configuration.

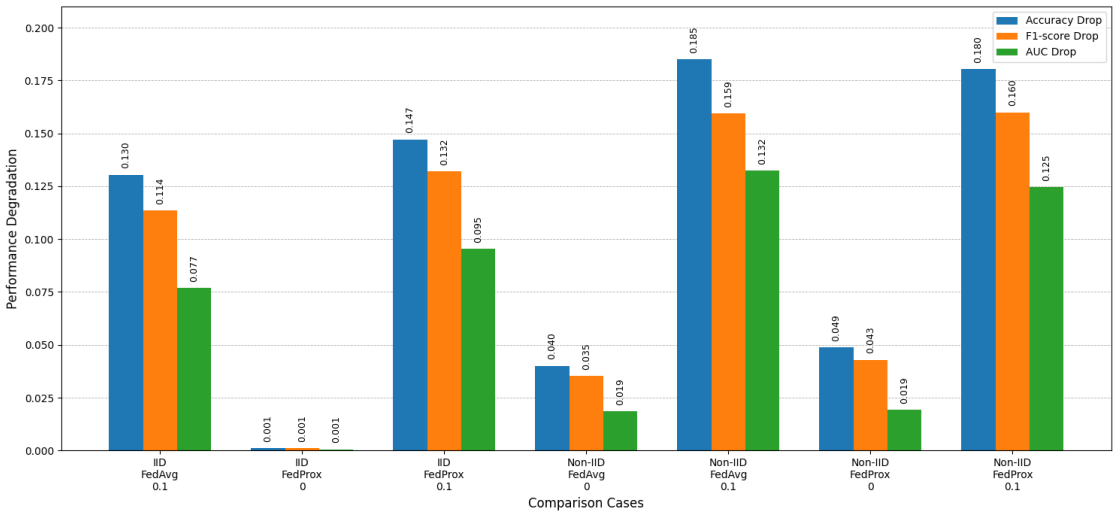

Figure 3. Degradation in test accuracy, F1-score, and AUC relative to the IID, FedAvg, no-noise baseline.

Figure 3 indicates the proportional relationship between the magnitude of performance loss in the 7 non-baseline cases. The least amount of degradation can be realized in the IID, FedProx, no-noise setups, and the most amount of degradation can be realised in the noisy Non-IID sets. This graphical comparison is helpful in supplementing Table 7 since the difference in performance relative to the best baseline is much easier to absorb visually.

4.4 Representative Round-Wise Behavior

Representative round-wise snapshots are also part of the approved final result set to demonstrate the development of the selected configurations throughout the communication rounds. These exemplary records take note of both positive and challenging environments and fill in the end-point comparison. Table 9 shows the representative round regressional values.

Table 9. Representative round-wise behavior for selected configurations

Configuration	Round	Test Accuracy	F1-score	AUC	Round Time (s)	Client Time (s)	Aggregation Time (s)	Comm. Cost (MB)
IID, FedAvg, Noise 0.0	1	0.8851	0.8924	0.9547	65.81	58.14	0.01	33.98
IID, FedAvg, Noise 0.0	10 (final)	0.9730	0.9754	0.9947	61.67	54.71	0.01	339.80
Non-IID, FedAvg, Noise 0.0	7 (final)	0.9330	0.9401	0.9761	36.94	30.06	0.01	237.86
Non-IID, FedAvg, Noise 0.1	1	0.6195	0.7244	0.6807	38.10	31.12	0.01	33.98
Non-IID, FedProx, Noise 0.1	9 (final)	0.7926	0.8155	0.8701	37.98	31.10	0.01	305.82

In Table 9, there is a clear separation in the chosen clean trajectory of IID compared to the non-IID trajectories, which are noisy. The given IID, FedAvg, no-noise set starts with the 0.8851 test accuracy in the Round 1 and peaks at the last reported round at 0.9730. In comparison, the Non-IID, FedAvg, noise 0.1 configuration begins at a significantly lower accident threshold (0.6195 test accuracy and 0.6807 AUC in Round 1) and the last representative row for Non-IID, FedProx, noise 0.1 is still below the no-noise settings at 0.7926 test accuracy and 0.870 In order to complement these representative round-wise snapshots, Figure 4 shows the convergence behavior of the selected configurations in terms of accuracy of test across communication rounds.

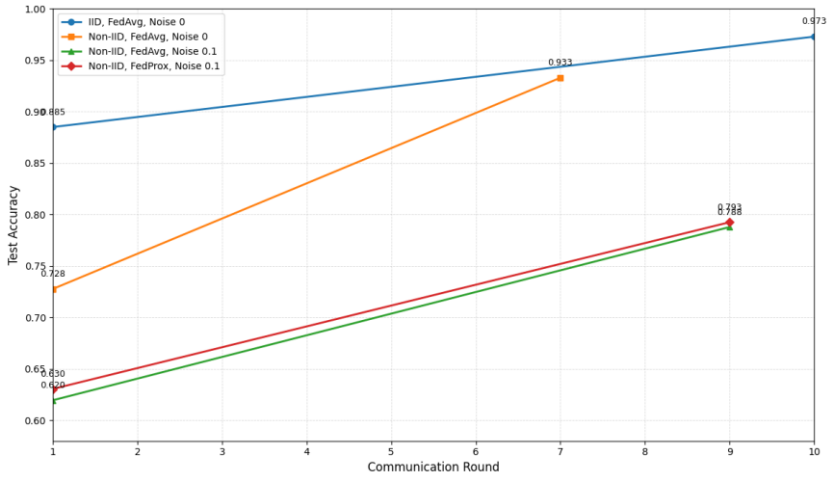

Figure 4. Representative convergence curves of test accuracy across communication rounds for selected configurations.

Figure 4 shows there is a much differentiation between the cleaner IID path and harder Non-IID conditions. No lack of accuracy The IID, FedAvg and no-noise versions reach the ultimate accuracy value maximum, and the noisy Non-IID versions are found in a much worse performance domain when training. This value is therefore a complement of Table 8 since it illustrates both favorable and challenging federated environment as time goes by as a relative behaves maximally.

4.5 Effect of Client Distribution

The impact of the distribution of the data can be seen in the end result performance table as well as the representative round-wise table and it aligns with the approved client-partition summary applied in the methodological section. With IID partitioning, the client datasets are fairly balanced with 13461347 samples each. In Non-IID partitioning, each client is between 629 and 716 samples, and the results log indicates strongly disproportional class distributions one client with 120 healthy and 596 tumor samples and another with 481 healthy and 149 tumor samples.

This difference in distribution can be observed in the last predictive outcomes with equal privacy. For FedAvg with noise 0, test accuracy changes from 0.972973 in IID to 0.932952 in Non-IID, while AUC changes from 0.994667 to 0.976132. For FedProx with noise 0, test accuracy changes from 0.971933 to 0.924116, while AUC changes from 0.994044 to 0.975355. The same direction is maintained in the noisy scenarios, in which the IID, FedAvg, noise 0. 1 setting outperforms the Non-IID, FedAvg, noise 0.1 setting, in terms of accuracy, F1-score, and AUC. In order to make a more specific comparison on effects of distribution, Figure 5 compares the final model operation with IID and Non-IID client distributions in matched noise settings.

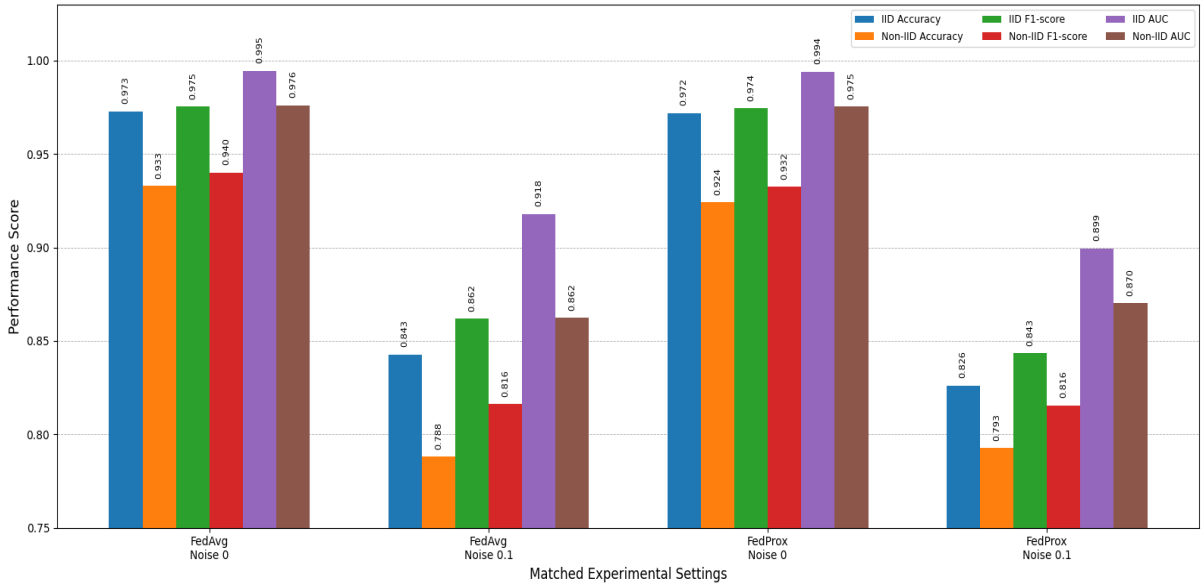

Figure 5. Effect of IID and Non-IID client distributions on final model performance under matched noise settings.

Figure 5 indicates that IID systems are always more accurate and have higher F1-score and AUC as compared to their Non-IID counterparts at similar noise conditions. It can be seen that the separation is present both in the noise-free environment and the noisy one, so the distribution of clients has a quantifiable impact on the ultimate predictive performance.

4.6 System-Level Characteristics and Privacy Indicator

The accepted system level table is a summary of representative system behavior of chosen configurations. Such values are round time, client time, aggregation time, DP overhead, approximation epsilon, cost of communication and CPU usage. The system-level statistics represented are provided in Table 10.

Natural Resources for Human Health
eISSN: 2583-1194

Table 10. System-level characteristics of representative settings

Configuration	Round Considered	Round Time (s)	Client Time (s)	Aggregation Time (s)	DP Overhead (%)	Epsilon (ϵ)	Comm. Cost (MB)	CPU Usage (%)
IID, FedAvg, Noise 0.0	1	65.81	58.14	0.01	0.00	0.000	33.98	47.0
IID, FedAvg, Noise 0.0	10 (final)	61.67	54.71	0.01	0.00	0.000	339.80	67.5
IID, FedAvg, Noise 0.1	1	62.76	55.83	0.01	-3.98	47.985	33.98	67.6
Non-IID, FedAvg, Noise 0.1	1	38.10	31.12	0.01	3.33	47.985	33.98	60.4
Non-IID, FedProx, Noise 0.1	9 (final)	37.98	31.10	0.01	2.88	143.956	305.82	60.6

According to Table 10, the time of aggregation in all exemplary rows is insignificant, but the part of the client training time denotes the vast majority of the round time. There is a rise in communication cost as the training increases, as the means are 339.80 MB in the last round of IID, FedAvg 305.82 MB in the last rounds of Non-IID, FedProx noise 0.1 used row and FedAvg noise 0.3 used row, respectively. Approximate epsilon is only recorded in noisy setting and is more with the rounds in reported representative rows. The negative value of DP overhead in the IID, FedAvg, noise 0.1 row is reinstated as it is in the final table approved. In order to enhance the legibility of informative computation patterns, Figure 6 demonstrates a system-level comparison of round time, client time, communication cost, and approximate epsilon in a variety of settings of choice.

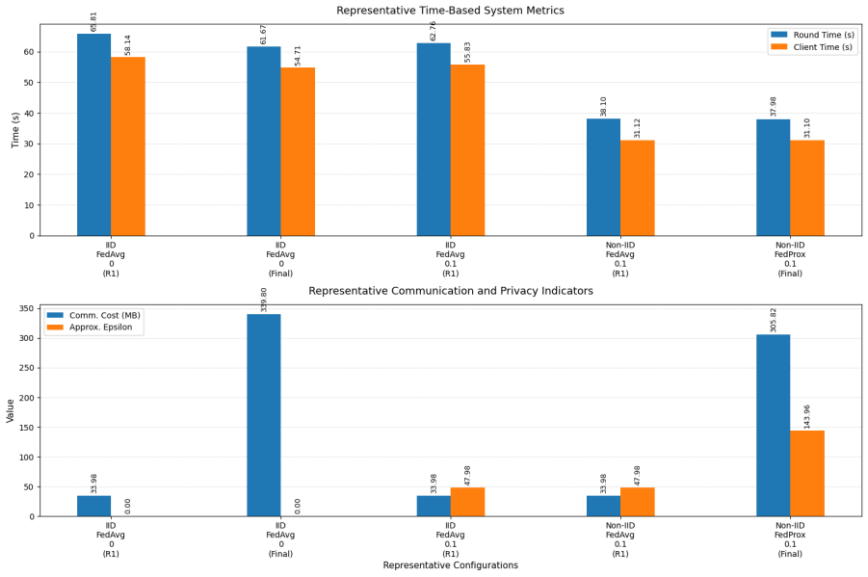

Figure 6. Representative system-level comparison of round time, client time, communication cost, and approximate epsilon across selected settings.

As indicated by Figure 6, the most significant portion of the round duration is credited to the client training time, whereas the aggregation time is insignificant in the chosen representative cases. It also depicts how the communication cost goes as a function of further training, and that only in perturbation-enabled settings do nonzero values of approximate epsilon occur. This number is integrated with Table 9, thus giving a concise graphical representation of the calculated and privacy-associated properties of the tested federated designs.

5. DISCUSSION

The findings develop an evident empirical ranking of the considered settings. Maximum predictive accuracy was observed in the IID partitioning without local perturbation with IID FedAvg, no-noise configuration with 0.972973 test accuracy, 0.975402 F1-score and 0.994667 AUC which is numerically very similar to the corresponding IID, FedProx, no-noise setting. The non-noise Non-IID configurations performed the worst with accuracy of 0.787942, FedAvg, noise 0.1 and 0.792620 with Non-IID, FedProx, noise 0.1. The trend is confirmed by the degradation analysis: the least significant fall in comparison with the best baseline was observed in IID, FedProx, no noise, the greatest decreases were observed in the case of the noisy Non-IID cases. The current structure in practice was most effective when the data of all clients were balanced, and local updates were not contaminated, and the combination between heterogeneity and DP-inspired perturbation had the highest utility loss. The trend is connected with the current evidence that the notion of privacy-preserving medical deep learning has to be assessed based on explicit privacy-utility trade-offs instead of privacy language as such [56].

These results are consistent with the literature on federated medical-imaging more generally that the heterogeneity of the clients is still one of the key obstacles to stable decentralized training. The survey of federated medical image analysis highlights the fact that non-IID distributions, local sample composition imbalance, and site-specific variation may significantly affect convergence and top quality of final models [57]. The same is claimed in meta-survey studies in radiomics and medical image research that federated approaches should be evaluated based on realistic data fragmentation and not just with positive end-point studies [58]. The current work substantiates that point: under matched no-noise operations, the labels of both algorithms experienced negative final accuracy and AUC in Non-IID training scenarios and such a pattern was present in IID training. The findings thus confirm the opinion that distributional imbalance is not an incidental instance of implementation, but the fundamental determinant of federated medical-image performance.

This interpretation is reinforced by results in the representative round-wise. The clean IID setup exhibited a reasonably good initial round and reached to the highest final performance, also the noisy Non-IID settings were randomly starting at significantly lower levels and remaining in a lower performance space during their final reported rounds. This can be explained by the fact that the work on decentralized healthcare benchmark revealed that federated difficulty can not be perceived with reference to terminal measures but should also be evaluated in terms of training behavior in a realistic institutional fragmentation [59]. Similar studies on brain tumor involving modality-conscious federated designs also indicate that decentralized neuro-oncology models are also very sensitive to local data structure and compatibility of updates across sites [60]. Though the current experiment is binary classification and not segmentation the same high-level principles can be applied: not only architecture or optimizer selection, but also statistical structure of local client data, can influence federated brain tumor analysis.

The research paper is also consistent with the current body of work of task-specific research on federated brain tumor analysis. The reviews of deep learning and federated learning in brain tumors segmentation always claim that preserving privacy, cross-site-resistance, and heterogeneity are issues yet to be addressed [61]. Later diagnosis-focused literature that adopts peer-to-peer federated learning shows that decentralized brain tumor models are becoming more warped to the demands of applications instead of being seen as generic healthcare classifiers [42]. Secure MRI brain tumor classification studies also indicate that it is possible to achieve privacy-conscious federation pipelines, though they also suggest that security-driven design decisions also should be evaluated in the context of predictive utility [43]. The current work adds to this task-related literature in demonstrating that in a single controlled setting, there is an increased practical cost of perturbation in cases when local data are already heterogeneous.

Simultaneously, the algorithm-label comparison presented in the given research must be viewed with restraint. Even though both FedAvg and FedProx settings are also reported, the framework in use employed the same equal-weight aggregation rule and no proximal term was included in the local FedProx objective. As such, the almost identicality of the two clean IIDs settings, as well as the minor variation in the more uncomfortable configurations, cannot be considered as a strong optimizer-

level finding. The contribution more defendable is the existence of documentation of interaction between the client distribution, local perturbation, convergence behavior and system-level properties in a bounded federated brain tumor classification framework. This read is in line with pioneering privacy preserving federated brain tumor research that already identified feasibility, yet with significant space to improved refinements of optimization and privacy modeling in the future [44].

These results imply several things. First, it is possible to have privacy-conscious federated brain tumor classification, but its practicality is highly implemented by the structure of client data. Second, local training based on perturbation must be evaluated together with heterogeneity, as the minimal levels of performance achieved in this paper were obtained in the case of both. Third, at system level the results indicate that the training time of clients was the biggest total round time, whereas in representable scenarios, the role of aggregation time was negligible. The experiments are limited in that it uses only one dataset, a lightweight CNN, five clients, and equal-weight aggregation, and tolerance of an approximate privacy indicator, but not an actual accountant-based guarantee. Future effort must hence build this base by formalized privacy accounting, more actual heterogeneity-conscious optimization, wider benchmarking, and repeated test on more diverse client-partition models.

Table 11 .Comparative analysis

Ref. No.	Dataset Used	FL Algorithm Used	Accuracy (IID)	Accuracy (Non-IID)
[6]	Brain MRI datasets	FedAvg variants	85.80%	Not reported
[2]	CT + synthetic MRI	FedAvg, FedProx	99.07%	Not separately reported
[4]	LGG-MRI	FedAvg	Dice 0.94	Not separate
[5]	TCGA-LGG + others	Fusion FL	Competitive	Strong
[1]	Multi-modal brain scans	FedAvg, FLOP, FedDis	Comparable	11%
[14]	MRI datasets	FedAvg	98.00%	Not reported
[16]	Brain tumor MRI	FedAvg	99.70%	Not reported
Proposed	**Single multimodal CT + MRI dataset**	**FedAvg / FedProx**	**97.29%**	**93.30% (best no-noise), 79.26% (noisy)**

Table 11 shows that although several existing federated learning methods achieved high IID accuracy, most did not properly evaluate Non-IID scenarios or report robustness under heterogeneous data distributions. The proposed method, using a single multimodal CT + MRI dataset with FedAvg/FedProx, achieves strong performance with 97.29% IID accuracy and 93.30% Non-IID accuracy, while still maintaining 79.26% accuracy in noisy conditions. Compared to previous studies, the proposed framework provides more consistent, reliable, and practical performance across both IID and Non-IID environments, making it the best overall approach for real-world medical imaging applications.

6. LIMITATIONS

This study was conducted using a single multimodal CT + MRI dataset within a controlled federated learning environment consisting of five simulated clients. The framework focused on CNN-based binary brain tumor classification under selected IID and Non-IID client distributions using equal-weight aggregation. Advanced federated optimization strategies, adaptive aggregation methods, and extensive client imbalance scenarios were not fully explored. In addition, cross-dataset validation, large-scale scalability analysis, and evaluation in real-world multi-hospital clinical environments were not included in the current study.

7. FUTURE WORK

Future work can extend the proposed framework by evaluating larger and more diverse medical imaging datasets with increased numbers of federated clients and realistic multi-hospital deployment settings. More advanced deep learning architectures such as ResNet, EfficientNet, and Vision Transformers can also be investigated alongside personalized and heterogeneity-aware federated optimization methods and adaptive aggregation strategies to improve robustness under complex Non-IID and client

imbalance conditions. Furthermore, the framework can be expanded toward multi-class brain tumor classification, segmentation, and broader clinical decision-support applications.

6. CONCLUSION

In this research, the controlled empirical study of federated brain tumor classification under IID and Non-IID clients distributions, with or without DP-inspired local perturbation were evaluated. In all the tested environments, the findings revealed an apparent and consistent trend, where the best predictive performance was obtained under the IID, no-noise, settings and worsening performance when Non-ID partitioning was implemented, and the worst results were when local perturbation was added. All the comparison tables, the study of the degradation, and the representative round-wise behavior, as well as the system-level observations, validated the identical key finding that federated brain tumor classification can be practiced in the privacy-aware context, however, the usability of the technique is highly contingent on the heterogeneity of the client data and the perturbation during local optimization. The practical implications of these findings are that high performance in balanced federated settings cannot be expected to directly generalize to non-homogenous institutional settings, and that privacy-aware perturbation must be a quantifiable utility trade-off instead of an uncompromisingly neutral designed addition in such already adversarial Non-IID settings. In the framework of the introduced framework, the study is more of an empirical reference than the contribution to privacy-accounting or the development of optimizers. The experiment was restricted to one dataset, a light CNN, five clients, and combining uniformly, aggregation and a rough indicator of privacy. Subsequent research ought thus to build on this foundation with formal privacy accounting, actual heterogeneity-conscious optimization, more extensive cross-dataset benchmarking and cross-architecture benchmarking and re-evaluation in more varied scenarios involving clients and distributions.

AUTHOR CONTRIBUTIONS

A.N. led the conceptualization and methodology of the study, implemented the federated learning framework, conducted experimental investigation, performed formal analysis, interpreted the results, and prepared the original manuscript draft. A.N. also contributed to the design and evaluation of the IID and non-IID experimental settings, including privacy-oriented perturbation analysis. S.N. contributed to data curation, validation, visualization, supervision, and provided critical review and editing of the manuscript. Both authors contributed to refining the study, discussing the findings, and approved the final version of the manuscript.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

Ethical approval was not required for this study.

DATA AVAILABILITY STATEMENT

The original dataset used in this study is publicly available from the Kaggle repository (*Brain Tumor Multimodal Image (CT & MRI)* by Murtozalikhon). Client-level partitions and derived data splits generated for the federated learning experiments are available from the corresponding author upon reasonable request.

REFERENCES

[1] M. H. ur Rehman, W. H. L. Pinaya, P. Nachev, J. Teo, S. Ourselin, and M. J. Cardoso, "Federated learning for medical imaging radiology," *British Journal of Radiology*, vol. 96, no. 1150, 2023, doi: 10.1259/bjr.20220890.

[2] Alsaleh, G. G. Tejani, S. Mishra, P. Sharma, and S. J. Mousavirad, "A federated learning-based privacy-preserving image processing framework for brain tumor detection from CT scans," Scientific Reports, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-07807-8.

[3] K. ElBedoui, W. Barhoumi, and J. Cho, "SoK: Federated Learning and Unlearning for Medical Image Analysis," *Expert Systems*, vol. 42, no. 6, 2025, doi: 10.1111/exsy.70063.

[4] K. Narmadha and P. Varalakshmi, "Enhanced brain tumour segmentation using a hybrid dual encoder–decoder model in federated learning," *Scientific Reports,* vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-17432-0.

A. Raheem, Z. Yang, H. Yu, M. A. Manan, F. Sabah, and S. Ahmed, "FUSION: Uncertainty-Guided Federated Semi-Supervised Learning for Medical Image Segmentation," *Iet Image Processing,* vol. 19, no. 1, 2025, doi: 10.1049/ipr2.70147.

[5] Ş. Ay, E. Ekıncı, and Z. Garip, "A brain tumour classification on the magnetic resonance images using convolutional neural network based privacy-preserving federated learning," *International Journal of Imaging Systems and Technology*, vol. 34, no. 1, 2024, doi: 10.1002/ima.23018.

[7] L. Che, J. Wang, Y. Zhou, and F. Ma, "Multimodal Federated Learning: A Survey," 2023, doi: 10.20944/preprints202307.1420.v1.

[6] G. Raj, "A Privacy-Preserving and Interpretable Federated Learning Framework With Clinician Feedback for Multi-Hospital Clinical Decision Support," *Concurrency and Computation Practice and Experience*, vol. 37, no. 27–28, 2025, doi: 10.1002/cpe.70365.

[7] R. Torkzadehmahani, "Privacy-preserving *Artificial Intelligence Techniques in Biomedicine,*" 2020, doi: 10.48550/arxiv.2007.11621.

[8] Rauniyar et al., "Federated Learning for Medical Applications: A Taxonomy, Current Trends, Challenges, and Future Research Directions," Ieee Internet of Things Journal, vol. 11, no. 5, pp. 7374–7398, 2024, doi: 10.1109/jiot.2023.3329061.

[9] P. Prayitno et al., "A Systematic Review of Federated Learning in the Healthcare Area: From the Perspective of Data Properties and Applications," *Applied Sciences,* vol. 11, no. 23, p. 11191, 2021, doi: 10.3390/app112311191.

[10] N. T. Madathil, F. K. Dankar, M. Gergely, A. N. Belkacem, and S. Alrabaee, "Revolutionizing healthcare data analytics with federated learning: A comprehensive survey of applications, systems, and future directions," *Computational and Structural Biotechnology Journal,* vol. 28, pp. 217–238, 2025, doi:10.1016/j.csbj.2025.06.009.

[11] M. H. ur Rehman, W. H. L. Pinaya, P. Nachev, J. Teo, S. Ourselin, and M. J. Cardoso, "Federated learning for medical imaging radiology*," British Journal of Radiology,* vol. 96, no. 1150, 2023, doi: 10.1259/bjr.20220890.

[12] D. H. Mahlool and M. H. Abed, "Optimize Weight sharing for Aggregation Model in Federated Learning Environment of Brain Tumor classification," *Journal of Al-Qadisiyah for Computer Science and Mathematics,* vol. 14, no. 3, 2022, doi: 10.29304/jqcm.2022.14.3.989.

[15] D. H. Mahlool and M. H. Abed, "Distributed brain tumor diagnosis using a federated learning environment," *Bulletin of Electrical Engineering and Informatics,* vol. 11, no. 6, pp. 3313–3321, 2022, doi: 10.11591/eei.v11i6.4131.

[13] U. Nandan, C. V. Sai, N. R. Sai, and A. Viswanadapalli, "Enhanced Brain Tumor Detection and Privacy Preserving Using Federated Learning," *International Journal of Scientific Research in Science and Technology,* vol. 11, no. 6, pp. 131–144, 2024, doi: 10.32628/ijsrst24116166.

[14] S. Shukla, S. Rajkumar, A. Sinha, M. Esha, E. Konguvel, and S. Vidhya, "Federated learning with differential privacy for breast cancer diagnosis enabling secure data sharing and model integrity," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-95858-2.

[15] J. Ziegler, B. Pfitzner, H. Schulz, A. Saalbach, and B. Arnrich, "Defending against Reconstruction Attacks through Differentially Private Federated Learning for Classification of Heterogeneous Chest X-Ray Data," 2022, doi: 10.48550/arxiv.2205.03168.

[16] Jabbar, J. Huang, M. K. Jabbar, and A. Ali, "Adaptive Multimodal Fusion in Vertical Federated Learning for Decentralized Glaucoma Screening," Brain Sciences, vol. 15, no. 9, p. 990, 2025, doi: 10.3390/brainsci15090990.

[17] S. Soltanieh, "Federated Learning in Neurology: Bridging Data Privacy and Artificial Intelligence for Brain Health," *Seminars in Neurology*, vol. 46, no. 01, pp. 038–048, 2025, doi: 10.1055/a-2769-6752.

[18] [21] M. Adnan, S. Kalra, J. C. Cresswell, G. W. Taylor, and H. R. Tizhoosh, "Federated learning and differential privacy for medical image analysis," *Scientific Reports*, vol. 12, no. 1, Art. no. 1953, 2022.

[19] M. G. Crowson, D. Moukheiber, A. R. Arévalo, B. D. Lam, S. Mantena, A. Rana, *et al.*, "A systematic review of federated learning applications for biomedical data," *PLOS Digital Health*, vol. 1, no. 5, Art. no. e0000033, 2022.

[20] S. Nazir and M. Kaleem, "Federated learning for medical image analysis with deep neural networks," *Diagnostics*, vol. 13, no. 9, Art. no. 1532, 2023.

[21] M. Jiang, Y. Zhong, A. Le, X. Li, and Q. Dou, "Client-level differential privacy via adaptive intermediary in federated medical imaging," in *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Cham, Switzerland: Springer Nature Switzerland, Oct. 2023, pp. 500–510.

[22] H. Guan, P.-T. Yap, A. Bozoki, and M. Liu, "Federated learning for medical image analysis: A survey," *Pattern Recognition*, vol. 151, Art. no. 110424, 2024.
[23] Z. L. Teo, L. Jin, N. Liu, S. Li, D. Miao, X. Zhang, *et al.*, "Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture," *Cell Reports Medicine*, vol. 5, no. 2, 2024.
[24] F. Zhang, D. Kreuter, Y. Chen, S. Dittmer, S. Tull, T. Shadbahr, *et al.*, "Recent methodological advances in federated learning for healthcare," *Patterns*, vol. 5, no. 6, 2024.
[25] M. Li, P. Xu, J. Hu, Z. Tang, and G. Yang, "From challenges and pitfalls to recommendations and opportunities: Implementing federated learning in healthcare," *Medical Image Analysis*, vol. 101, Art. no. 103497, 2025.
[26] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artificial Intelligence and Statistics (AISTATS)*, 2017, pp. 1273–1282.
[27] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proc. Machine Learning and Systems*, vol. 2, pp. 429–450, 2020.
[28] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," in *Proc. Int. Conf. Machine Learning (ICML)*, 2020, pp. 5132–5143.
[29] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conf. Computer and Communications Security (CCS)*, 2016, pp. 308–318.
[30] M. J. Sheller *et al.*, "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," *Scientific Reports*, vol. 10, p. 12598, 2020.
[31] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated learning for healthcare informatics," *J. Healthcare Informatics Research*, vol. 5, no. 1, pp. 1–19, 2021.
[32] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020.
[33] M. Manthe, S. Duffner, and C. Lartizien, "Federated brain tumor segmentation: An extensive benchmark," *Medical Image Analysis*, vol. 97, Art. no. 103270, 2024.
[34] E. H. Lee, M. Han, J. Wright, M. Kuwabara, J. Mevorach, G. Fu, *et al.*, "An international study presenting a federated learning AI platform for pediatric brain tumors," *Nature Communications*, vol. 15, no. 1, Art. no. 7615, 2024.
[35] Al-Saleh, G. G. Tejani, S. Mishra, S. K. Sharma, and S. J. Mousavirad, "A federated learning-based privacy-preserving image processing framework for brain tumor detection from CT scans," Scientific Reports, vol. 15, no. 1, Art. no. 23578, 2025.
[36] M. Z. Muntaqim and T. A. Smrity, "Federated learning framework for brain tumor detection using MRI images in non-IID data distributions," *Journal of Imaging Informatics in Medicine*, vol. 38, no. 6, pp. 3909–3927, 2025.
[37] R. Haripriya, N. Khare, and M. Pandey, "Privacy-preserving federated learning for collaborative medical data mining in multi-institutional settings," *Scientific Reports*, vol. 15, no. 1, Art. no. 12482, 2025.
[38] J. Wen, X. Li, X. Ye, X. Li, and H. Mao, "A highly generalized federated learning algorithm for brain tumor segmentation," *Scientific Reports*, vol. 15, no. 1, Art. no. 21053, 2025.
[39] L. Zheng *et al.*, "Sensitivity-aware differential privacy for federated medical imaging," *Sensors*, vol. 25, no. 9, Art. no. 2847, 2025.
[40] C. Chen, H. Pan, K. Zhang, Z. Li, and F. Yu, "Prototype-based personalized federated learning for medical image classification," *Knowledge-Based Systems*, vol. 326, Art. no. 114021, 2025.
[41] E. Albalawi, M. T. R., A. Thakur, V. V. Kumar, M. Gupta, S. B. Khan, and A. Almusharraf, "Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor," *BMC Medical Imaging*, vol. 24, no. 1, Art. no. 110, 2024.
[42] G. Appasami and N. Savarimuthu, "Cross-dataset performance evaluation and secure federated learning with weighted model aggregation for MRI brain tumor image classification," *The Journal of Supercomputing*, vol. 81, no. 16, pp. 1–26, 2025.
[43] Z. Zhou, G. Luo, M. Chen, Z. Weng, and Y. Zhu, "Federated learning for medical image classification: A comprehensive benchmark," *IEEE Journal of Biomedical and Health Informatics*, 2025.
[44] P. N. Srinivasu, G. J. Lakshmi, S. C. Narahari, J. Shafi, J. Choi, and M. F. Ijaz, "Enhancing medical image classification via federated learning and pre-trained model," *Egyptian Informatics Journal*, vol. 27, Art. no. 100530, 2024.
[45] Q. Li, B. He, and D. Song, "Model-contrastive federated learning," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 10713–10722.

[46] S. Sharma and K. Guleria, "A comprehensive review on federated learning based models for healthcare applications," *Artificial Intelligence in Medicine*, vol. 146, p. 102691, 2023.
[47] D. Upreti, E. Yang, H. Kim, and C. Seo, "A comprehensive survey on federated learning in the healthcare area: Concept and applications," *Computer Modeling in Engineering & Sciences*, vol. 140, no. 3, p. 2239, 2024.
[48] R. Kumar *et al.*, "Privacy-preserving blockchain-based federated learning for brain tumor segmentation," *Computers in Biology and Medicine*, vol. 177, p. 108646, 2024.
[49] S. Alphonse, F. Mathew, K. Dhanush, and V. Dinesh, "Federated learning with integrated attention multiscale model for brain tumor segmentation," *Scientific Reports*, vol. 15, no. 1, p. 11889, 2025.
[50] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, *et al.*, "The future of digital health with federated learning," *npj Digital Medicine*, vol. 3, no. 1, art. no. 119, 2020.
[51] E. Darzidehkalani, M. Ghasemi-Rad, and P. M. A. van Ooijen, "Federated learning in medical imaging: Part I: Toward multicentral health care ecosystems," *Journal of the American College of Radiology*, vol. 19, no. 8, pp. 969–974, 2022.
[52] E. Darzidehkalani, M. Ghasemi-Rad, and P. M. A. van Ooijen, "Federated learning in medical imaging: Part II: Methods, challenges, and considerations," *Journal of the American College of Radiology*, vol. 19, no. 8, pp. 975–982, 2022.
[53] M. Mohammadi, M. Vejdanihemmat, M. Lotfinia, M. Rusu, D. Truhn, A. Maier, and S. Tayebi Arasteh, "Differential privacy for medical deep learning: Methods, tradeoffs, and deployment implications," *npj Digital Medicine*, 2026.
[54] F. R. da Silva, R. Camacho, and J. M. R. Tavares, "Federated learning in medical image analysis: A systematic survey," *Electronics*, vol. 13, no. 1, art. no. 47, 2023.
[55] Raza, A. Guzzo, M. Ianni, R. Lappano, A. Zanolini, M. Maggiolini, and G. Fortino, "Federated learning in radiomics: A comprehensive meta-survey on medical image analysis," Computer Methods and Programs in Biomedicine, vol. 267, art. no. 108768, 2025.
[56] M. Zenk, U. Baid, S. Pati, A. Linardos, B. Edwards, M. Sheller, *et al.*, "Towards fair decentralized benchmarking of healthcare AI algorithms with the Federated Tumor Segmentation (FeTS) challenge," *Nature Communications*, vol. 16, no. 1, art. no. 6274, 2025.
[57] H. Liu, D. Wei, Q. Dai, X. Wu, Y. Zheng, and L. Wang, "Federated modality-specific encoders and partially personalized fusion decoder for multimodal brain tumor segmentation," *Medical Image Analysis*, art. no. 103759, 2025.
[58] M. F. Ahamed, M. M. Hossain, M. Nahiduzzaman, M. R. Islam, M. R. Islam, M. Ahsan, and J. Haider, "A review on brain tumor segmentation based on deep learning methods with federated learning techniques," *Computerized Medical Imaging and Graphics*, vol. 110, art. no. 102313, 2023.
[59] N. Onaizah, Y. Xia, and K. Hussain, "FL-SiCNN: An improved brain tumor diagnosis using siamese convolutional neural network in a peer-to-peer federated learning approach," Alexandria Engineering Journal, vol. 114, pp. 1–11, 2025.
[60] G. Appasami and N. Savarimuthu, "Federated learning for secure medical MRI brain tumor image classification," *The European Physical Journal Special Topics*, vol. 234, no. 15, pp. 4813–4827, 2025.
[61] W. Li, F. Milletarì, D. Xu, N. Rieke, J. Hancox, W. Zhu, *et al.*, "Privacy-preserving federated brain tumour segmentation," in *International Workshop on Machine Learning in Medical Imaging*. Cham, Switzerland: Springer International Publishing, 2019, pp. 133–141.
[62] murtozalikhon, "Brain tumor multimodal image (CT & MRI)," *Kaggle*. [Online]. Available: https://www.kaggle.com/datasets/murtozalikhon/brain-tumor-multimodal-image-ct-and-mri.