Analyzing Privacy Trade-offs of Federated Learning in Medical Data Processing

Shahedul Islam Shahed
ID: 0112230274
Dept. of Computer Science and Engineering United International University
Dhaka, Bangladesh
sshahed223274@bscse.uiu.ac.bd

Meherab Hossain Bhueyan
ID: 011221466
Dept. of Computer Science and Engineering United International University
Dhaka, Bangladesh
mbhuyian221466@bscse.uiu.ac.bd

Sharmin Sultana LizaID: 011221331
Dept. of Computer Science and Engineering United International University
Dhaka, Bangladesh
sliza221331@bscse.uiu.ac.bd

***Abstract*—The growing integration of artificial intelligence into healthcare has introduced powerful opportunities for early disease detection and personalized medicine. However, training medical AI models centrally often violates privacy, security, and data-sharing regulations. Federated Learning has emerged as a paradigm that allows multiple institutions to collaboratively train a global model while keeping patient data local. Despite its promise, FL faces fundamental trade-offs between privacy and utility—particularly when combined with Differential Privacy. In this study, we examine these trade-offs through extensive experiments on the Pima Indians Diabetes dataset. We compare four FL optimization strategies, such as FedAvg, FedProx, FedAvgM, and FedAdam, under varying conditions of data heterogeneity, client participation, and privacy noise levels. Our findings show that FL can reach nearly centralized performance up to 78.86% accuracy while preserving privacy. We also show that small DP noise $\sigma = 0.01$ has minimal impact, but higher levels $\sigma \geq 0.05$ lead to measurable performance degradation. The results offer practical recommendations for privacy-preserving FL deployment in medical contexts and shed light on the sensitivity of federated optimizers to noise and non-IID data.**

***Index Terms*—Federated Learning, Differential Privacy, Medical Data, Privacy-Utility Trade-off, Federated Optimization, Healthcare AI.**

I. Introduction

Artificial intelligence has transformed healthcare analytics by enabling the automatic prediction of diseases from medical records, imaging, and genomic data. Centralized learning approaches, however, require the aggregation of sensitive patient data across hospitals and laboratories, raising major concerns about privacy, consent, and data breaches. Regulations such as HIPAA and GDPR explicitly restrict unrestricted sharing of health records across institutions. Consequently, most hospitals operate with small, isolated datasets that limit the generalization of AI models.

Federated Learning offers an alternative paradigm for distributed training [1]. It allows multiple clients—such as hospitals or diagnostic centers—to train a global model collaboratively without transferring raw data. Each client computes local updates on its data, and a central server aggregates them to update the shared model. While this structure reduces direct privacy risks, model updates can still reveal sensitive patterns or enable membership inference attacks.

To strengthen privacy guarantees, FL can be combined with Differential Privacy [2]. DP introduces random noise and gradient clipping to mask the contribution of individual participants, quantified by parameters (ε, δ). However, noise inevitably reduces model accuracy, leading to a core trade-off between privacy protection and predictive performance. Understanding this trade-off under realistic settings—especially in the medical domain—is critical for the adoption of FL in clinical environments.

This work systematically studies how optimization choices, data heterogeneity, and DP noise interact to affect model utility in a medical prediction task. By comparing four server-side optimizers and exploring diverse client and privacy settings, we provide empirical insights into designing privacy-aware federated systems for healthcare.

II. Literature Review

A. Federated Learning and its Baselines

The foundational FL algorithm, *Federated Averaging* (FedAvg), was introduced by McMahan *et al.* [1]. FedAvg periodically averages local model updates from participating clients. Although simple, it effectively reproduces centralized performance on many tasks. However, non-IID data across clients often slows convergence and causes model divergence.

B. Optimizer Variants

To overcome heterogeneity issues, Li *et al.* proposed FedProx [3], adding a proximal term that penalizes deviations from the global model, mitigating "client drift." Reddi *et al.* [4] later introduced adaptive server optimizers such as FedAdam, FedYogi, and FedAdagrad, which apply adaptive learning rates to global aggregation. FedAvgM extends FedAvg by integrating server-side momentum to smooth weight updates.

C. Non-IID Data and Client Drift

In federated environments, clients often have data drawn from distinct distributions. Hsu *et al.* [5] characterized this using a Dirichlet distribution with concentration α, where smaller α values increase heterogeneity. Such non-IID conditions degrade FL performance by misaligning update directions, making optimization unstable. Methods such as FedProx and FedAvgM explicitly address this issue through regularization and momentum.

D. Differential Privacy in FL

Differential Privacy (DP) ensures that the inclusion or exclusion of a single user does not significantly affect the model outcome [2]. It operates by clipping gradient norms and injecting Gaussian noise. The Re´nyi DP accountant improves privacy estimation for iterative algorithms. In FL, DP can be applied at the client level, protecting entire institutions rather than individual samples.

E. FL in Medical Applications

Healthcare presents strong motivations for FL deployment. Cross-institutional studies in imaging, pathology, and EHR analysis have demonstrated FL's potential to train robust models without sharing raw data. However, few works have performed controlled comparisons across optimizers and privacy parameters on structured medical datasets. Our work fills this gap by providing a unified evaluation of FL optimizers and DP effects in a realistic healthcare simulation.

III. Materials and Methods

A. Dataset

The experiments used the Pima Indians Diabetes dataset, a benchmark for medical prediction containing 768 samples with eight numerical features such as Glucose, BMI, Blood Pressure, and Age. The binary target variable indicates diabetes presence (1) or absence (0). The dataset contains no missing values, simplifying preprocessing.

B. Preprocessing

We performed an 80/20 stratified train–test split to preserve class balance. Standardization used Z-score scaling:

$$X' = \frac{X - \mu}{\sigma}, \tag{1}$$

C. Federated Simulation Setup

We simulated $K \in \{10, 20, 50\}$ virtual clients. Each communication round included:

1) Server broadcasts the current global model w_t;
2) A random subset (fraction $f \in \{1.0, 0.5\}$) of clients trains locally for one epoch;
3) Clients return local updates Δw_i^t to the server, which aggregates them.

Data were partitioned in two modes:

- **IID:** Uniform random distribution of samples across clients.
- **Non-IID:** Dirichlet distribution ($\alpha = 0.3$) to induce heterogeneous label distributions.

D. Server-Side Optimizers

FedAvg:

$$w^{t+1} = w^t + \eta \sum_{i=1}^{K} p_i \Delta w_i^t \tag{2}$$

where p_i is the proportion of client i's data.

FedProx: Each client minimizes

$$\min_{w_i} f_i(w_i) + \frac{\mu}{2} \| w_i - w^t \|^2 \tag{3}$$

to control deviation from the global model.

FedAvgM:

$$v^{t+1} = \beta v^t + (1 - \beta) \sum_i p_i \Delta w_i^t, \tag{4}$$

$$w^{t+1} = w^t + \eta v^{t+1}. \tag{5}$$

FedAdam: FedAdam maintains first and second moment estimates similar to Adam optimizer but applied to aggregated updates.

E. Differential Privacy Mechanism

We applied client-level Gaussian DP to each update:

$$\tilde{\Delta w_i} = \frac{\Delta w_i}{\max(1, \|\Delta w_i\|_2 / C)}, \tag{6}$$

$$\bar{w} = \frac{1}{K} \sum_i \tilde{\Delta w_i} + \mathcal{N}(0, \sigma^2 C^2 I), \tag{7}$$

where μ and σ are computed from training data only to prevent data leakage.

where $C = 1.0$ bounds sensitivity and $\sigma \in \{0, 0.01, 0.05, 0.10\}$ controls noise magnitude.

F. Evaluation Protocol

Each configuration was trained for 50 global rounds with $E = 1$ local epoch. We report test accuracy (mean $\pm$ standard deviation) across two random seeds. Baselines include centralized training and basic FedAvg without DP.

IV. Results and Discussion

A. Overall Performance

Centralized training achieved 77.92% accuracy. In contrast, basic FL configurations reached accuracies up to 78.86% (FedAvgM, $f = 0.5$), matching or even slightly exceeding the centralized results. This demonstrates the potential of FL to achieve high utility without data sharing.

B. Optimizer Comparison

FedAvg and FedProx consistently performed best across IID and non-IID settings, averaging 76.4% accuracy. FedAvgM produced similar results, while FedAdam underperformed (74.0%) and showed higher sensitivity to noise. These results confirm that adaptive optimizers may require careful fine-tuning under DP perturbations.

C. Impact of Differential Privacy

Table II summarizes performance across privacy noise levels. Small noise ($\sigma = 0.01$) incurred negligible cost, whereas strong noise ($\sigma = 0.10$) reduced accuracy by 2–3%. This quantifies the privacy–utility balance achievable in FL for healthcare data.

TABLE I: Performance Comparison of Different Training Strategies

Model	Accuracy	F1-Score	ROC-AUC
Centralized	77.92%	0.658	0.835
Basic FL (FedAvg, K=10, IID)	72.08%	0.592	0.791
Optimized FL (FedProx, K=20, Non-IID, frac=0.5)	78.86%	0.681	0.841
Optimized FL + DP (FedProx, σ=0.01)	78.51%	0.675	0.839

TABLE II: Effect of Differential Privacy Noise on Accuracy

Noise Level (σ)	Mean Accuracy (%)	Std.
0.00	76.42	0.87
0.01	76.44	0.85
0.05	75.58	1.76
0.10	74.67	3.01

D. Effect of Client Scaling and Participation

As the number of clients increased from 10 to 50, performance dropped by approximately 1.5 percentage points due to higher aggregation variance. Partial participation ($f = 0.5$) sometimes improved generalization by adding stochastic sampling noise, acting as a regularizer.

E. Non-IID Heterogeneity

Non-IID distributions reduced average accuracy by roughly one percentage point compared to IID. FedProx and FedAvgM showed better stability due to their regularization and momentum mechanisms, making them suitable choices for real-world heterogeneous data.

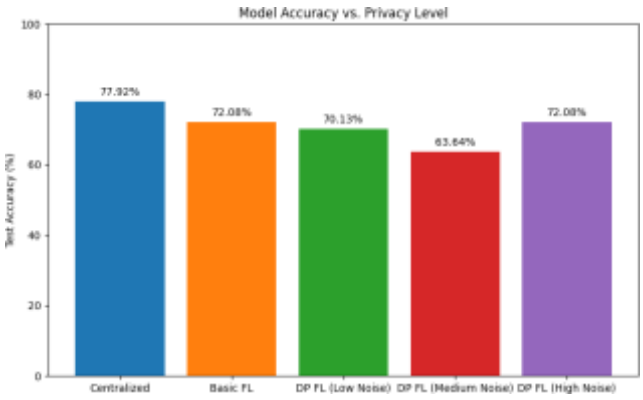

Fig. 1: Performance comparison of FL optimizers under varying noise and client settings.

F. Practical Insights

In medical FL deployments:

- FedProx and FedAvgM are robust defaults for moderate heterogeneity.
- Small DP noise ($\sigma = 0.01$) offers strong privacy with minimal cost.
- 10–20 clients and partial participation balance communication and accuracy.
- FedAdam should be used with caution, especially under strong DP noise.

V. Future Work

Future research will focus on several key areas to advance the practical application of privacy-preserving FL in healthcare:

- **Formal Privacy Accounting:** Conduct a formal privacy analysis using Re´nyi DP to compute tighter (ε, δ) bounds for each experimental configuration.
- **Personalized FL:** Investigate personalized FL techniques (e.g., FedPer, FedBN) to improve model performance for individual clients and reduce the negative impact of non-IID data.
- **Robustness and Security:** Explore Byzantine-resilient aggregation methods and anomaly detection to protect the global model from malicious or faulty clients.
- **Communication Efficiency:** Implement and evaluate communication-saving techniques such as model quantization and structured updates to reduce the overhead in resource-constrained environments.
- **Expanded Evaluation Metrics:** Move beyond accuracy to evaluate model performance using metrics like ROC-AUC, calibration, and fairness across different demographic subgroups present in medical data.

VI. Conclusion

This study conducted a comprehensive empirical analysis of the privacy–utility trade-offs in federated learning on medical

data. Through comparisons of multiple optimizers and DP configurations, we found that FedAvg, FedProx, and FedAvgM deliver stable performance close to centralized baselines, even under partial participation and data heterogeneity. Small amounts of DP noise can preserve privacy with negligible performance degradation. These findings underscore that privacy-preserving FL is a viable and powerful paradigm for sensitive healthcare applications when configured thoughtfully, offering a path to building robust, collaborative AI models without compromising patient confidentiality.

REFERENCES

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, vol. 54. PMLR, 2017, pp. 1273–1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html

[2] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318.

[3] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proceedings of Machine Learning and Systems*, vol. 2, 2020, pp. 429–450.

[4] S. J. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Konec˘ny`, S. Kumar, and H. B. McMahan, "Adaptive federated optimization," in *International Conference on Learning Representations*, 2021. [Online]. Available: https://openreview.net/forum?id=LkFG3lB13U5

[5] T.-M. Hsu, H. Qi, and M. Brown, "Measuring the effects of non- identical data distribution for federated visual classification," *arXiv preprint arXiv:1909.06335*, 2019.