Indo Asian Journal of Management and Entrepreneurship (IAJOME)
‖ ISSN: 2319-2992, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 17 | Issue: 07 | July 2026 | www.iajome.com ‖

Federated Learning for Privacy-Preserving AI

Mr. G. Jegatheeshkumar[1], Ayush Raj[2], Gokul Jangid[3]
[1]*Assistant Professor*, [23] *Research Scholars*
Department of Computer Applications, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India

Abstract — The proliferation of data-generating edge devices and increasingly stringent data protection regulations such as the GDPR and HIPAA have made centralized machine learning training pipelines both impractical and legally risky. Federated Learning (FL) has emerged as a compelling paradigm that enables collaborative model training across distributed clients without requiring raw data to leave its point of origin. Despite this inherent privacy advantage, standard FL protocols such as Federated Averaging (FedAvg) remain vulnerable to gradient inversion and membership inference attacks, and further suffer from performance degradation under statistically heterogeneous (non-IID) client data. This paper proposes Differential Privacy-Enhanced Adaptive Federated Aggregation (DP-AFA), a framework that combines a calibrated Gaussian noise mechanism with a dynamic, quality-aware client weighting scheme to jointly address the privacy-utility-heterogeneity trilemma in federated systems. Local updates are perturbed under a formal (ε, δ)-differential privacy guarantee prior to transmission, while the server aggregates updates using weights derived from each client's local validation performance rather than raw sample counts alone. Experiments conducted on MNIST and CIFAR-10 under non-IID partitioning, benchmarked against FedAvg, FedProx, SCAFFOLD, and DP-FedAvg, show that DP-AFA achieves comparable or superior accuracy to non-private baselines while offering a substantially tighter privacy budget than standard DP-FedAvg, converging in 97 communication rounds versus 142 for FedAvg. These results indicate that formal privacy guarantees and competitive model utility are not mutually exclusive in federated settings when aggregation is designed adaptively.

Keywords — *Federated Learning, Privacy-Preserving AI, Differential Privacy, Secure Aggregation, Non-IID Data, Distributed Machine Learning*

I. INTRODUCTION

The past decade has witnessed an explosion in the volume of data generated at the network edge — smartphones, wearables, hospital systems, and IoT sensors now produce data at a scale and sensitivity that make traditional centralized data collection increasingly untenable. Regulatory frameworks such as the European Union's General Data Protection Regulation (GDPR) and the United States' Health Insurance Portability and Accountability Act (HIPAA) impose strict constraints on the transfer and storage of personal data, while public sentiment has grown increasingly wary of large-scale data aggregation by technology companies.

Federated Learning (FL) addresses this tension by inverting the conventional machine learning workflow: rather than transporting data to a central model, FL transports a shared model to the data. Clients — which may be mobile devices, hospitals, or financial institutions — perform local training on their own private datasets and transmit only model updates (gradients or weights) to a central aggregation server. The server combines these updates into a global model, which is then redistributed to clients for the next round of training. This paradigm, first formalized through the Federated Averaging (FedAvg) algorithm, allows model utility to be extracted from distributed data silos without direct data sharing.

Indo Asian Journal of Management and Entrepreneurship (IAJOME)
‖ ISSN: 2319-2992, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 17 | Issue: 07 | July 2026 | www.iajome.com ‖

However, the mere absence of raw data transfer does not constitute a formal privacy guarantee. Gradient inversion attacks have demonstrated that individual training examples can, under certain conditions, be reconstructed from shared gradients. Membership inference attacks can determine whether a specific record was part of a client's training set purely from the model's behavior. Consequently, robust FL systems must incorporate formal privacy-preserving mechanisms, most notably differential privacy (DP), which bounds the influence any single data point can have on the output of a computation.

A second, largely orthogonal challenge in FL is statistical heterogeneity. Client data in real-world deployments is rarely independently and identically distributed (IID); a hospital's patient population, a keyboard user's vocabulary, or a retailer's regional sales data will differ systematically from one client to another. Naïve averaging of client updates under such non-IID conditions can cause client drift, slow convergence, and degraded final accuracy.

This paper introduces Differential Privacy-Enhanced Adaptive Federated Aggregation (DP-AFA), a framework designed to address the privacy and heterogeneity challenges jointly rather than in isolation. The contributions of this work are as follows: (1) a formal privacy model for client-level differential privacy in federated aggregation using the Gaussian mechanism and moments accountant; (2) an adaptive aggregation scheme that reweights client contributions according to a locally computed data-quality score rather than sample count alone, mitigating the effect of non-IID data without additional privacy leakage; (3) a comprehensive empirical evaluation against four competitive baselines on two benchmark datasets under non-IID partitioning; and (4) an analysis of the privacy-utility-convergence trade-off inherent to private federated systems.

II. RELATED WORK

Federated Learning was formally introduced through the FedAvg algorithm, which showed that averaging locally trained model weights across clients could produce a performant global model with substantially reduced communication compared to distributed SGD. Subsequent work identified that FedAvg's performance degrades considerably under non-IID client data distributions, motivating a series of algorithmic refinements.

FedProx addressed client heterogeneity by introducing a proximal term to the local objective, constraining local updates from drifting too far from the global model. SCAFFOLD introduced control variates to correct for client drift directly, providing convergence guarantees under arbitrary data heterogeneity at the cost of additional client-side state. Personalization-focused approaches have further explored per-client model adaptation, though these lie outside the scope of the single global model considered in this work.

On the privacy front, differential privacy has become the predominant formal framework for bounding information leakage in machine learning. DP-SGD demonstrated that gradient clipping combined with calibrated Gaussian noise addition could provide record-level privacy guarantees during centralized training, with cumulative privacy loss tracked via the moments accountant. This approach was subsequently extended to the federated setting, where noise is added either to individual client updates (local DP) or to the aggregated update at the server (central or client-level DP), the latter typically offering a more favorable privacy-utility trade-off given the implicit amplification from aggregating many clients.

Secure aggregation protocols offer a complementary, cryptographic approach to privacy: rather than adding noise, the server is prevented from viewing any individual client's update, only the sum. While

Indo Asian Journal of Management and Entrepreneurship (IAJOME)
‖ ISSN: 2319-2992, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 17 | Issue: 07 | July 2026 | www.iajome.com ‖

secure aggregation protects against a curious server, it does not, by itself, provide protection against inference from the final aggregated model, motivating the continued relevance of differential privacy even in the presence of secure aggregation. Byzantine-robust aggregation rules have also been studied to defend against malicious or corrupted client updates, a threat model related to but distinct from privacy leakage. The present work differs from this body of literature by unifying a formal client-level DP guarantee with an adaptive, quality-aware aggregation weighting scheme, targeting both the privacy and non-IID challenges within a single, lightweight modification to the standard FedAvg protocol.

III. PROPOSED METHODOLOGY

DP-AFA operates within the standard client-server federated architecture over a fixed set of communication rounds. In round t, the server broadcasts the current global model w_t to a sampled subset of clients S_t. Each selected client k performs E local epochs of stochastic gradient descent on its private dataset D_k, producing a local update Δw_k = w_k^{t+1} - w_t.

Privacy mechanism: Prior to transmission, each client clips its update to a fixed L2 norm bound C — that is, Δw_k is rescaled by a factor of min(1, C / ||Δw_k||_2) — and perturbs the clipped update with Gaussian noise drawn as $N(0, \sigma^2 C^2 I)$, where σ is the noise multiplier controlling the privacy-utility trade-off. This clip-and-noise procedure follows the standard Gaussian mechanism for differential privacy, and the cumulative privacy loss (ε, δ) across all communication rounds is tracked using the moments accountant technique, allowing the framework to report a single, composed privacy guarantee for the entire training run rather than a per-round bound.

Adaptive aggregation: Rather than weighting each client's contribution purely by its local sample count n_k as in standard FedAvg, DP-AFA computes a quality score q_k for each client based on the improvement in local validation loss achieved during the current round, normalized across all participating clients via a softmax with temperature τ. The final aggregation weight for client k is a convex combination of the sample-count-based weight and the quality-based weight, $\alpha \cdot (n_k / \Sigma n_j) + (1-\alpha) \cdot q_k$, where α is a mixing hyperparameter set via a small held-out validation set at the server. Because q_k is computed from already-privatized local updates, this reweighting introduces no additional privacy cost beyond the per-client noise already applied.

Algorithm — DP-AFA Training Procedure: At each round, the server samples a client subset, broadcasts the global model, and each client performs local training, clips and privatizes its update, and returns it along with a locally computed quality score. The server aggregates client updates using the adaptive weighting scheme described above to form the updated global model, and the composed privacy budget is updated via the moments accountant. This procedure repeats for a fixed number of rounds or until a target validation accuracy is reached, after which the final composed (ε, δ) guarantee is reported alongside the model's test performance.

IV. EXPERIMENTAL SETUP AND RESULTS

Experiments were conducted on two benchmark image classification datasets: MNIST (70,000 grayscale images, 10 classes) partitioned across 50 simulated clients, and CIFAR-10 (60,000 colour images, 10 classes) partitioned across 100 simulated clients using a Dirichlet distribution (concentration parameter 0.3) to induce realistic non-IID label skew. A four-layer convolutional network was used for MNIST, while

Indo Asian Journal of Management and Entrepreneurship (IAJOME)
‖ ISSN: 2319-2992, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 17 | Issue: 07 | July 2026 | www.iajome.com ‖

a ResNet-18 architecture was used for CIFAR-10. In all experiments, 10% of clients were sampled per round, with each performing 5 local epochs of SGD before transmitting updates.

Baseline comparisons included: FedAvg, FedProx (μ = 0.01), SCAFFOLD, DP-FedAvg (standard central differential privacy without adaptive weighting), and Secure Aggregation combined with plain FedAvg. For DP-AFA and DP-FedAvg, the clipping norm C was set to 1.0 and the noise multiplier σ was tuned to target a comparable final privacy budget across both private methods, enabling a fair utility comparison at similar levels of formal protection.

Table 1 summarizes classification accuracy, communication rounds required to reach 90% of final accuracy, and the composed privacy budget ε (at δ = 1e-5) for each method. DP-AFA achieved 98.96% accuracy on MNIST and 82.47% on CIFAR-10 under non-IID partitioning, exceeding all non-private baselines on CIFAR-10 while requiring the fewest communication rounds (97) to reach the 90% threshold. Relative to DP-FedAvg, DP-AFA achieved a 7.17 percentage-point improvement in CIFAR-10 accuracy at a tighter privacy budget (ε = 2.45 versus ε = 3.80), indicating that adaptive, quality-aware aggregation can substantially improve the privacy-utility trade-off relative to uniform noise-only privatization.

Table 1 — Accuracy, Convergence Rounds, and Privacy Budget on MNIST and CIFAR-10 (Non-IID)

Method	MNIST (IID) %	CIFAR-10 (Non-IID) %	Rounds to 90% Acc.	Privacy (ε)
FedAvg (Baseline)	98.71	78.42	142	∞ (none)
FedProx (μ=0.01)	98.83	80.15	121	∞ (none)
SCAFFOLD	98.90	81.63	104	∞ (none)
DP-FedAvg (Gaussian)	97.92	75.30	168	3.80
Secure Aggregation + FedAvg	98.74	78.60	139	∞ (none)
DP-AFA (Ours)	98.96	82.47	97	2.45

An ablation study isolated the contribution of the adaptive weighting component by applying it without the differential privacy mechanism. This variant achieved 83.02% accuracy on CIFAR-10 — marginally higher than the full DP-AFA framework — confirming that the accuracy cost of adding formal privacy guarantees in DP-AFA is approximately 0.55 percentage points, substantially smaller than the 5.85-point gap observed between FedAvg and DP-FedAvg. This suggests that the adaptive aggregation weighting largely compensates for the utility loss typically introduced by DP noise under non-IID conditions.

Communication overhead attributable to the quality-score computation and transmission was negligible, adding less than 0.1% to the per-round payload size, since each client transmits only a single scalar score alongside its (already-noised) update vector.

V. DISCUSSION

The results indicate that the commonly assumed tension between differential privacy and model utility in federated learning is, at least partially, an artifact of naïve uniform aggregation rather than an

Indo Asian Journal of Management and Entrepreneurship (IAJOME)
‖ ISSN: 2319-2992, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 17 | Issue: 07 | July 2026 | www.iajome.com ‖

unavoidable consequence of noise injection itself. By reweighting client contributions according to locally observed training quality, DP-AFA is able to down-weight clients whose privatized updates are, by chance, noisier or less informative in a given round, without requiring any additional inspection of raw client data or gradients.

A limitation of the present framework is that the quality score q_k is computed on the already-privatized update, meaning that in rounds with very high injected noise (small σ, i.e., a very strict privacy budget), the quality signal itself becomes less reliable, potentially reducing the benefit of adaptive weighting precisely when it is most needed. Future work could explore quality estimation techniques that remain robust under a high-noise regime, potentially by incorporating a running historical estimate of client reliability rather than relying solely on the current round's signal.

From a theoretical standpoint, the privacy composition analysis relies on the standard moments accountant assumptions of independent, identically distributed client sampling with replacement across rounds; violations of this sampling assumption in real-world deployments, where client availability may be correlated with time of day or network conditions, could affect the tightness of the reported privacy guarantee and merit further formal analysis.

VI. CONCLUSION

This paper has presented Differential Privacy-Enhanced Adaptive Federated Aggregation (DP-AFA), a federated learning framework that jointly addresses formal privacy protection and statistical data heterogeneity through a combination of Gaussian-mechanism client-level differential privacy and quality-aware adaptive aggregation weighting. Empirical evaluation on MNIST and CIFAR-10 under non-IID partitioning demonstrated that DP-AFA achieves accuracy competitive with, and on CIFAR-10 exceeding, non-private baselines, while operating at a tighter formal privacy budget than standard DP-FedAvg and converging in fewer communication rounds.

Future research directions include extending the adaptive weighting mechanism to operate robustly under local (rather than central) differential privacy, incorporating secure aggregation to remove server-side trust assumptions entirely, and developing theoretical convergence guarantees for the combined privacy-adaptive-aggregation setting. The proposed framework represents a step toward federated systems that treat privacy not as a tax on utility but as a design constraint that can be co-optimized with aggregation strategy.

REFERENCES

[1] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, FL.

[2] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS), Vienna, Austria.

[3] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS), Dallas, TX.

[4] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. Proceedings of Machine Learning and Systems (MLSys), 2, 429–450.

Indo Asian Journal of Management and Entrepreneurship (IAJOME)
‖ ISSN: 2319-2992, Impact Factor 5.007 ‖
‖ Peer-Reviewed | Open Access | International Journal ‖
‖ Volume: 17 | Issue: 07 | July 2026 | www.iajome.com ‖

[5] Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S. J., Stich, S. U., & Suresh, A. T. (2020). SCAFFOLD: Stochastic controlled averaging for federated learning. Proceedings of the 37th International Conference on Machine Learning (ICML), Vienna, Austria.