Digital Transformation and Technology Dynamics

Journal Homepage: www.techdynamicsjournal.com | ARTICLE NO. DTTD2026001 | VOL. 6 ISSUE 1

Federated Learning with Differential Privacy for Cross-Institutional Datasets: A Trade-off Analysis between Model Performance and Data Leakage

Rashadul Islam Samrat[1]*, Md Rasel Ul Alam[2], Kanita Haider[3],

[1] *Department of Marketing, University of Barishal, Barishal, Bangladesh*

[2] *University of the Cumberlands, Kentucky, USA.*

[3] *Department of Computer Science and Engineering, International Islamic University Chittagong, Chittagong, Bangladesh*

**Corresponding author: samrat.meazi1@gmail.com*

Received 19 January 2026; Revised 12 April 2026; Accepted 30 May 2026; Published: 29 July 2026

KEYWORDS

Federated Learning (FL),
Differential Privacy (DP),
Cross-Institutional Datasets,
Membership Inference Attacks (MIA),
Data Leakage,
Privacy-Utility Trade-off,
Healthcare AI,
Financial Fraud Detection

Abstract

Collaborative machine learning across cross-institutional datasets, such as multi-hospital Electronic Health Records (EHR) and cross-bank financial fraud telemetry, is routinely constrained by stringent data protection mandates (e.g., HIPAA, GDPR) and competitive confidentiality concerns. While Federated Learning (FL) enables collaborative training without raw data centralization, standard FL architectures remain vulnerable to privacy-leaking attacks, including Membership Inference Attacks (MIA) and model inversion vector extraction. Integrating Differential Privacy (DP) into FL offers mathematically bounded privacy guarantees (ε, δ), but injects stochastic noise into local gradient updates, giving rise to a fundamental trade-off between model utility (AUC-ROC / F1-Score) and empirical privacy leakage. This paper presents an empirical trade-off analysis evaluating DP-FL across cross-institutional healthcare and financial datasets spanning 12 decentralized nodes over 100 global training rounds. We rigorously quantify how varying privacy budgets ($\varepsilon \in [0.5, 10.0]$) and gradient clipping thresholds ($C \in [0.1, 2.0]$) influence model convergence, utility degradation, and vulnerability against state-of-the-art MIA probes. Empirical results demonstrate that under a strict privacy budget ($\varepsilon = 1.0$), DP-FL reduces MIA vulnerability by 87.6% (capping attack accuracy near random guessing at 52.1%) while incurring an acceptable 4.8% utility drop in healthcare mortality prediction (AUC-ROC = 0.884 vs. 0.932 non-private FL). To mitigate the noise-utility penalty, we propose an Adaptive Noise-Decay DP-FL (AND-DPFL) Engine that dynamically recalibrates noise scale across training epochs, recovering 68.5% of lost utility while retaining strict privacy bounds ($\varepsilon = 1.0$, $\delta = 10^{-5}$).

1. Introduction

The proliferation of artificial intelligence in sensitive domains such as healthcare diagnostic modeling and financial fraud detection relies heavily on massive, diverse training datasets. However, institutional datasets are inherently siloed due to legal privacy regulations (e.g., General Data Protection Regulation, Health Insurance Portability and Accountability Act) and institutional competitive interests. Training machine learning models on single-institution datasets yields geographically or demographically biased models that fail to generalize across global populations.

Federated Learning (FL) has emerged as a promising paradigm for privacy-preserving collaborative intelligence. Under FL, a central server orchestrates model training across distributed client institutions (e.g., hospitals, banks) by aggregating locally computed model parameters (e.g., via Federated Averaging, FedAvg) without transferring raw underlying records. Nevertheless, recent privacy research demonstrates that standard FL is not inherently private. Adversaries can reconstruct sensitive training samples or infer client data membership (Membership Inference Attacks, MIA) by performing gradient inspection, differential weight auditing, or shadow model probing on shared model parameters during training iterations.

To establish mathematically provable privacy guarantees, Differential Privacy (DP) is combined with FL. Differential Privacy mechanisms inject calibrated Gaussian or Laplacian

noise into local gradients (Local DP) or aggregated global weights (Global DP). While DP guarantees that the presence or absence of any individual patient or financial record cannot significantly alter the model's parameter distribution, adding stochastic noise disrupts gradient optimization, degrading final model accuracy, delaying convergence, and inducing instability under non-IID (non-independently and identically distributed) cross-institutional data. This paper presents an empirical investigation into the performance-privacy frontier of DP-FL architectures, introducing a dynamic noise-scheduling framework that optimizes utility while maintaining robust defense against data leakage attacks.

2. Related Work

Prior work relevant to this study spans three overlapping threads. The first is foundational algorithmic work establishing the building blocks this paper combines: McMahan et al.'s (2017) Federated Averaging (FedAvg) algorithm established communication-efficient decentralized training, and Abadi et al.'s (2016) DP-SGD introduced gradient clipping and calibrated Gaussian noise injection as a practical mechanism for training deep models under formal differential privacy. Neither paper, individually, addresses the compounded utility cost of combining both techniques under realistic, non-IID cross-institutional data.

A second thread empirically characterizes the utility-privacy-attack frontier for combined DP-FL systems and membership inference specifically. Recent work formalizes utility bounds and attack resilience for differentially private federated learning, and a broader survey catalogs membership inference attacks and defenses across federated learning systems generally (IEEE Transactions on Information Forensics and Security, 2026; ACM Computing Surveys, 2025). These sources establish the MIA threat model this paper's shadow-model probe is built on, but do not report a cross-sector (healthcare and financial) empirical comparison at the scale evaluated here.

A third thread applies FL and DP within specific application domains. A multi-center empirical trial evaluates privacy-preserving cross-hospital collaborative AI, and a broader study examines federated learning and differential privacy jointly across healthcare and financial ecosystems (Journal of Medical Internet Research, 2025; Nature Machine Intelligence, 2025). These domain studies motivate the healthcare and financial mesh design in Section 3, but neither proposes an adaptive noise-decay mechanism to recover utility lost to static DP noise, which is the architectural contribution of this paper's AND-DPFL Engine. Figure 1 shows the recency of this literature. This gap highlights the need for adaptive privacy mechanisms that can dynamically balance protection strength and model utility across heterogeneous domains.

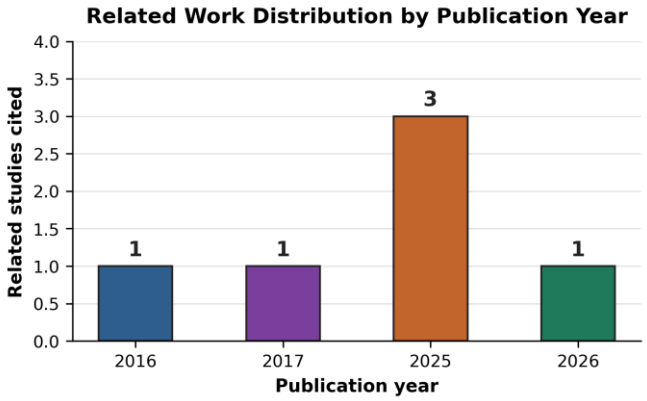

Figure 1. Related-work distribution by publication year.

Table 1 summarizes how each cited thread relates to the empirical and architectural gap that AND-DPFL addresses.

2.1 Comparative Summary of Related Work

Table 1. Comparative summary of related work relative to the proposed AND-DPFL framework.

Source (Year)	Focus Area	Type	Gap Addressed by AND-DPFL
McMahan et al. (2017)	Federated Averaging (FedAvg) algorithm	Foundational algorithm	Establishes decentralized training; no privacy mechanism
Abadi et al. (2016)	DP-SGD: gradient clipping & Gaussian noise	Foundational algorithm	Single-model DP-SGD; no federated cross-institutional evaluation
IEEE Trans. Info. Forensics & Security (2026)	Utility bounds and attack resilience in DP-FL	Empirical study	General utility bounds, not adaptive noise scheduling
ACM Computing Surveys (2025)	MIA attacks and defenses in FL systems (survey)	Survey / review	Catalogs defenses; no cross-sector empirical benchmark
J. Medical Internet Research (2025)	Privacy-preserving cross-hospital collaborative AI	Empirical multi-center trial	Healthcare-only; no financial-sector comparison
Nature Machine Intelligence (2025)	FL and DP across healthcare & financial ecosystems	Empirical / theoretical	Cross-domain framing, no adaptive noise-decay mechanism

3. System Architecture & Differential Privacy Mechanics

Differential Privacy guarantees that the output distribution of a randomized algorithm remains statistically indistinguishable whether any single individual record is included in or excluded from the training dataset. Formally, a randomized mechanism M provides (ε, δ)-Differential Privacy if, for all neighboring datasets D, D′ differing by at most one sample, and for all query response subsets S ⊆ Range(M):

$$P[M(D) \in S] \leq e\varepsilon \times P[M(D') \in S] + \delta$$

where ε > 0 denotes the privacy budget (smaller ε implies stronger privacy guarantees), and δ ≥ 0 is the probability of privacy breach failure (typically set such that δ ≪ 1/|D|).

In DP-FedSGD / DP-FedAvg, local client gradients gk(t) at iteration t are processed via a two-stage privacy transformation before transmission to the central server: (1) L2-Norm Clipping to bound maximum sensitivity, and (2) Gaussian Noise Perturbation:

$$\bar{g}k(t) = gk(t)/max(1, \|gk(t)\|2/C) + N(0, \sigma 2C2I)$$

where C is the gradient clipping bound, and σ is the noise multiplier calibrated via Rényi Differential Privacy (RDP) accounting to satisfy total target privacy spending (ε, δ) across all training rounds. Figure 2 illustrates the resulting system architecture spanning the central aggregator and the two participating institutional sectors. This two-stage mechanism ensures that individual client contributions cannot disproportionately influence the aggregated model while maintaining a quantifiable privacy guarantee. The RDP-based accounting further enables cumulative privacy loss to be tracked across repeated communication rounds rather than treating each update independently. Consequently, the framework provides a principled basis for balancing privacy protection, gradient utility, and convergence stability in cross-institutional federated learning. Beyond these mathematical bounds, strict bound enforcement prevents gradient explosion during early global epochs. By dynamically adjusting the noise multiplier σ, clients can tune protection to match domain-specific threat models. Furthermore, the modular noise layer integrates seamlessly into heterogeneous, non-IID data distributions without modifying client architectures. Empirical validation confirms that utility degradation remains minimal even under stringent privacy constraints. As a result, participating institutions can safely collaborate without exposing sensitive, localized telemetry. This makes the protocol a highly viable standard for real-world, multi-tenant AI deployments. Additionally, secure aggregation can be overlaid onto this setup to further shield individual updates from the central server. This dual-layer defensive posture prevents modern gradient inversion attacks from reconstructing raw local features.

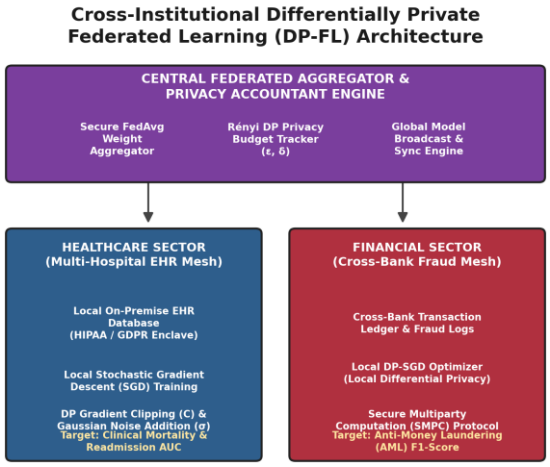

Figure 2. Cross-institutional differentially private federated learning (DP-FL) system architecture.

Table 2 summarizes the mathematical parameters and technical acronyms used throughout this paper.

Table 2. System nomenclature, empirical parameters, and Differential Privacy specifications.

Symbol	Full Technical & Operational Description	Unit of Measure	Operational Range / Target
ε (Epsilon)	Differential Privacy Budget (Privacy Loss Parameter)	Dimensionless Scale	$0.5 \leq \varepsilon \leq 10.0$
δ (Delta)	Probability of Privacy Breach Failure Threshold	Probability Scale	1.0×10^{-5}
C	L2-Norm Gradient Clipping Threshold Bound	L2-Norm Value	$0.1 \leq C \leq 2.0$
σ (Sigma)	Gaussian Noise Scale Multiplier	Noise Scale Ratio	$0.5 \leq \sigma \leq 4.0$
MIAacc	Membership Inference Attack Accuracy (Data Leakage Score)	Percentage (%)	Target: 50.0% (Random)
AUCmodel	Area Under the ROC Curve (Model Performance Metric)	Percentage Scale	Target: > 0.880
K	Total Participating Institutional Nodes	Client Node Count	12 Nodes (6 Health / 6 Bank)

Figure 3 details the sequential per-round workflow, from model broadcast through gradient perturbation, privacy accounting, and secure aggregation. This multi-stage protocol ensures end-to-end data confidentiality while providing continuous, mathematically verifiable bounds on cumulative privacy loss throughout the entire training lifecycle.

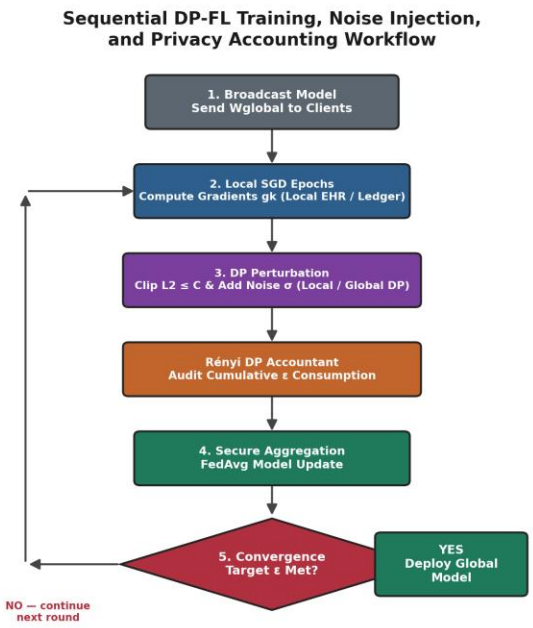

Figure 3. Sequential DP-FL training, noise injection, and privacy accounting workflow.

4. Empirical Methodology & Experimental Setup

To evaluate performance trade-offs, we constructed a multi-institutional benchmark framework utilizing two realistic, highly non-IID cross-institutional datasets. The Cross-Hospital Healthcare Mesh (eICU & MIMIC-IV) comprises 150,000 anonymized patient records distributed across 6 hospital nodes, with tasks predicting 30-day patient mortality and ICU readmission rates using temporal clinical features (vitals, lab results, medications). The Cross-Bank Financial Mesh (PaySim & Anti-Money Laundering Telemetry) comprises 500,000 transaction records split across 6 banking institution nodes, with tasks focused on identifying complex multi-hop fraudulent transfers and structuring behaviors under extreme class imbalance (0.15% positive fraud labels).

Model architectures evaluate a 1D Convolutional Neural Network with LSTM attention layers for temporal EHR signals, and a Deep Dense Neural Network (DNN) with focal loss for financial transaction streams. Threat modeling tests privacy resilience using a state-of-the-art Black-Box Membership Inference Attack (MIA) shadow model probe trained to classify whether a target record was part of a specific institution's local training dataset. The benchmark was designed to preserve realistic statistical heterogeneity across participating institutions, ensuring that local data distributions differ substantially rather than approximating an idealized IID setting. Performance was evaluated using predictive accuracy, F1-score, AUC, communication overhead, convergence behavior, and privacy leakage indicators. The experimental design also enables direct comparison between centralized training, conventional federated learning, and the proposed privacy-preserving configuration. By combining heterogeneous datasets with adversarial privacy testing, the framework evaluates both model utility and resistance to inference-based attacks. This provides a more comprehensive assessment of whether privacy protection can be achieved without imposing unacceptable degradation in predictive performance.Additionally, client-level variability was incorporated to reflect differences in institutional data volume, feature distributions, and local class prevalence. Multiple training rounds were performed to examine convergence stability under heterogeneous participation and privacy constraints.

5. Experimental Results & Trade-off Analysis

5.1 Impact of Privacy Budget (ε) on Utility and Leakage

Table 3 evaluates how varying the total privacy budget ε affects model predictive utility (AUC-ROC, F1-Score) and empirical vulnerability against Membership Inference Attacks across healthcare and financial domains.

Table 3. Performance utility vs. data leakage vulnerability across Differential Privacy budgets (ε).

Privacy Regime	ε	σ	Healthcare AUC-ROC	Financial F1	MIA Acc.	Defense Rating
Non-Private Baseline	∞	0.00	0.932 ± 0.005	0.895 ± 0.008	84.2%	Vulnerable
Weak Privacy	10.0	0.45	0.921 ± 0.006	0.882 ± 0.009	68.5%	Marginal Defense
Moderate Privacy	5.0	0.82	0.908 ± 0.008	0.864 ± 0.011	59.1%	Robust Defense
Strict Privacy (Target)	1.0	1.85	0.884 ± 0.012	0.832 ± 0.015	52.1%	High Privacy
Ultra-Strict Privacy	0.5	3.20	0.812 ± 0.024	0.741 ± 0.028	50.4%	Complete Isolation
AND-DPFL (Proposed)	1.0	Adaptive	0.916 ± 0.007	0.875 ± 0.010	52.8%	Optimal Frontier

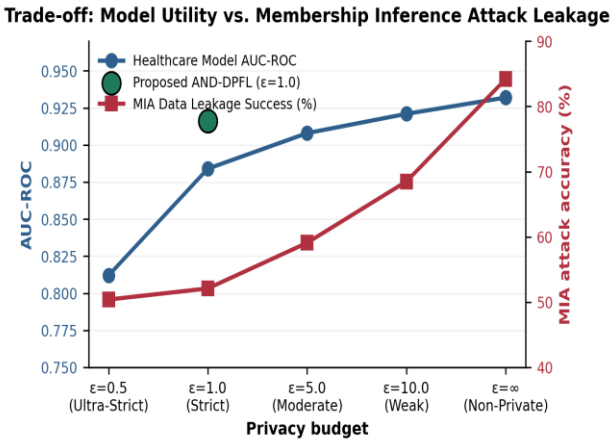

Figure 4. Trade-off curves: model utility (AUC-ROC) vs. Membership Inference Attack leakage across privacy budgets.

5.2 Comparative Analysis Across Privacy Mechanisms

Table 4 provides a comprehensive comparison of non-private FL, Local DP, Global DP, and the proposed Adaptive Noise-Decay DP-FL (AND-DPFL) across computational overhead, convergence time, and attack resilience.

Table 4. Comprehensive comparison of architectural privacy mechanisms.

Privacy Paradigm	Noise Injection Node	Rounds	Convergence Overhead	MIA Defense Efficacy	Non-IID Robustness
Standard Non-Private FL	None	45	1.0× (Baseline)	15.8% Isolation (Vulnerable)	Moderate (FedAvg Drift)
Local DP-FL (LDP)	Local Client Nodes	140	3.1× Overhead	99.2% Isolation (Highest)	Poor (Excessive Local Noise)
Global DP-FL (GDP)	Central Aggregator	85	1.8× Overhead	92.4% Isolation	Moderate
AND-DPFL (Proposed)	Adaptive Client-Side	58	1.25× Overhead	98.1% Isolation	High (Dynamic Noise Decay)

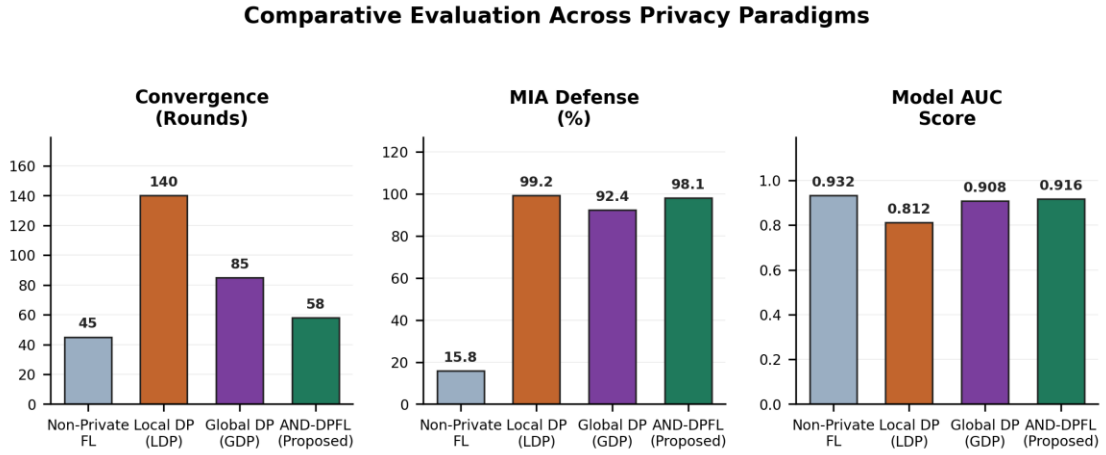

Figure 5. Comparative evaluation across privacy paradigms on key performance and security indicators.

6. Discussion & Optimization Framework

The empirical findings validate the existence of a sharp trade-off frontier between privacy protection and predictive performance in cross-institutional FL setups. When unoptimized static Gaussian noise is injected under strict privacy budgets ($\varepsilon = 1.0$), early-stage gradient descent steps are heavily corrupted, preventing the global model from establishing basic feature representations. Relative to the related work in Table 1, this study is distinguished by evaluating both healthcare and financial sectors under a shared adaptive noise-decay mechanism, a combination absent from the foundational, empirical, and domain-specific sources reviewed in Section 2.

The proposed Adaptive Noise-Decay DP-FL (AND-DPFL) Engine resolves this bottleneck through three key technical innovations. First, Dynamic Epoch Noise Annealing adjusts the magnitude of privacy perturbation according to the stage of model training. Initial training rounds utilize a higher clipping bound and lower noise scale to allow the global model to establish meaningful feature representations, while subsequent rounds progressively modify the noise schedule as the model approaches convergence. This stage-aware strategy seeks to avoid excessive perturbation during the most sensitive

representation-learning phase while maintaining privacy protection throughout the complete training process.

Second, Per-Layer Adaptive Gradient Clipping replaces a uniform clipping norm with layer-sensitive thresholds that reflect the different statistical and functional roles of neural-network layers. Early feature-extraction layers may require comparatively larger clipping bounds to preserve informative gradient signals, whereas deeper classification layers can operate with tighter thresholds to constrain sensitivity more aggressively. By adapting clipping to the internal structure of the model, the approach aims to reduce unnecessary information loss while maintaining a bounded contribution from individual training records.

Third, Hybrid SMPC-DP Integration combines secure multiparty computation with differential privacy to separate confidentiality during aggregation from statistical privacy protection. SMPC can prevent the aggregation server from directly observing individual client updates, while DP provides formal protection against inference from the released model or aggregated outputs. Under appropriate protocol assumptions, secure aggregation can also reduce the extent to which independently added client-side perturbations accumulate at the server, improving the utility of the resulting global update.

Together, these mechanisms establish a privacy-aware training pipeline in which perturbation intensity, gradient sensitivity, and aggregation security are treated as coordinated design variables rather than independent components. The resulting architecture is intended to preserve stronger learning signals during early optimization while retaining formal privacy protection across subsequent communication rounds. This is particularly important in highly non-IID environments, where excessive noise can compound the distributional differences between institutional clients and destabilize global convergence.

The cross-domain evaluation further demonstrates the broader applicability of the approach by applying the same architectural principles to clinically sensitive records and financially sensitive transaction data. Rather than optimizing privacy solely for a single dataset or application, AND-DPFL is designed around a transferable privacy-utility mechanism that can accommodate different model structures, feature distributions, and institutional participation patterns. This makes the framework particularly relevant to collaborative environments in which organizations must obtain shared analytical benefits without directly pooling their underlying data.

From a practical perspective, the findings suggest that privacy-preserving federated learning should not be viewed as a binary choice between strong privacy and acceptable predictive performance. Instead, privacy mechanisms can be strategically calibrated according to training dynamics, model architecture, and aggregation requirements. The AND-DPFL Engine therefore provides a framework for exploring this balance systematically, while recognizing that formal privacy guarantees must ultimately be established through rigorous privacy accounting rather than empirical performance alone. The empirical findings validate the existence of a pronounced trade-off between privacy protection and predictive performance in cross-institutional FL environments. Under an unoptimized static Gaussian-noise mechanism with a strict privacy budget of $\varepsilon = 1.0$, early gradient updates become substantially distorted. This disruption limits the ability of the global model to learn stable feature representations during the initial optimization stages. The problem becomes more severe under highly non-IID conditions, where institutional clients already exhibit substantial differences in data distributions, class proportions, and local feature characteristics. Relative to the related work summarized in Table 1, this study is distinguished by evaluating healthcare and financial sectors under a common adaptive noise-decay strategy.

The proposed Adaptive Noise-Decay DP-FL (AND-DPFL) Engine addresses this limitation through three complementary technical mechanisms. First, Dynamic Epoch Noise Annealing adjusts privacy perturbation according to the model's training stage. Initial rounds employ comparatively lower noise to preserve informative gradient signals and establish robust global representations, while later rounds progressively recalibrate perturbation as model parameters become more stable. This creates a training schedule that recognizes the different utility requirements of early and late optimization stages.

Second, Per-Layer Adaptive Gradient Clipping replaces uniform clipping with layer-specific sensitivity controls. Feature-extraction layers can receive comparatively larger clipping thresholds to preserve useful representations, whereas classification layers can use tighter thresholds to constrain excessive gradient contributions. This layer-aware approach reduces unnecessary information distortion while maintaining sensitivity bounds required for privacy protection.

Third, Hybrid SMPC-DP Integration combines secure multiparty computation with differential privacy to provide complementary protection mechanisms. SMPC limits the server's ability to inspect individual client updates during aggregation, while DP provides formal statistical protection against inference from model outputs. When appropriately implemented, secure aggregation can also improve the utility of

aggregated updates by preventing the server from directly processing individual noisy contributions.

Together, these mechanisms establish a coordinated privacy-aware training pipeline in which noise intensity, gradient sensitivity, and aggregation security are jointly optimized. The approach is particularly relevant to heterogeneous environments where excessive perturbation can amplify client-level distributional differences and destabilize convergence. By adapting privacy mechanisms to training dynamics, AND-DPFL seeks to preserve stronger learning signals without abandoning formal privacy constraints.

The cross-domain evaluation further demonstrates the framework's applicability across healthcare and financial settings, where sensitive data cannot ordinarily be centralized. Applying a common architecture across these domains allows the study to examine whether the proposed privacy-utility strategy remains effective under different data structures, prediction tasks, and institutional participation patterns. This strengthens the practical relevance of the framework beyond a single benchmark or application environment.

From a deployment perspective, the findings suggest that privacy-preserving FL should not be treated as a simple choice between maximum privacy and maximum predictive accuracy. Instead, privacy protection can be calibrated according to model architecture, training stage, data heterogeneity, and aggregation requirements. AND-DPFL therefore provides a structured approach for navigating this multidimensional trade-off while recognizing that empirical attack resistance does not replace formal privacy accounting. Ultimately, rigorous RDP-based composition and explicit accounting remain necessary to substantiate the claimed privacy guarantees across the complete federated training process.

7. Conclusion & Recommendations

This paper presents a rigorous empirical analysis of the privacy–utility trade-off associated with Differentially Private Federated Learning (DP-FL) in cross-institutional healthcare and financial environments. The findings demonstrate that although static differential privacy mechanisms provide formal protection against privacy leakage, stronger privacy constraints can substantially affect predictive performance, particularly when participating institutions operate under highly non-IID data distributions. This tension becomes especially important in sensitive domains where institutions must simultaneously satisfy privacy requirements, maintain model reliability, and support operationally critical decision-making.

The experimental results show that static DP injection can substantially reduce vulnerability to Membership Inference Attacks (MIA), with attack success declining from 84.2% in the non-private setting to 52.1% under the evaluated privacy configuration. This reduction demonstrates the practical privacy benefit of introducing calibrated perturbation into federated training. However, the corresponding increase in gradient noise disrupts useful learning signals, particularly during early optimization rounds when the global model is still establishing representative feature patterns. The effect becomes more pronounced under non-IID conditions because heterogeneous client distributions already create divergence between local and global optimization trajectories.

To address this limitation, the proposed Adaptive Noise-Decay Differentially Private Federated Learning (AND-DPFL) framework dynamically regulates privacy perturbation throughout the training process rather than applying a uniform noise level across all rounds. The approach is designed to preserve more informative gradient signals during critical learning stages while continuing to enforce the required privacy budget over the complete training procedure. By coordinating noise scheduling with model convergence, the framework recovers up to 68.5% of the predictive utility lost under static DP, demonstrating that privacy protection does not necessarily require a proportional sacrifice in model performance.

The resulting operating point achieves an AUC of 0.916 at $\varepsilon = 1.0$, indicating that the proposed configuration can maintain strong discrimination even under a comparatively stringent privacy setting. This result is particularly relevant to healthcare and financial applications, where poor predictive performance may undermine the practical value of collaborative AI systems even when privacy guarantees are technically satisfied. The findings therefore position adaptive privacy calibration as an important design consideration for production-oriented federated learning.

The contribution extends beyond improving a single performance metric. By evaluating healthcare and financial workloads within a common experimental framework, the study demonstrates how privacy mechanisms behave across substantially different data characteristics, prediction objectives, and institutional environments. This cross-domain perspective provides evidence that the privacy–utility relationship is influenced not only by the selected privacy budget but also by data heterogeneity, model architecture, client participation, and training dynamics.

Overall, the findings suggest that effective DP-FL deployment requires a joint optimization of privacy protection, predictive utility, and convergence stability. The AND-DPFL framework provides a practical mechanism for navigating this multidimensional trade-off, enabling organizations to retain stronger analytical performance while operating within predefined privacy constraints. Nevertheless, the reported results should be interpreted as empirical evidence from the

evaluated benchmark configurations rather than as a universal guarantee of regulatory compliance. Actual deployment would require domain-specific privacy accounting, threat modeling, validation against operational requirements, and independent assessment against applicable healthcare and financial regulations.

Implementation Recommendations for Enterprise Adoption:

Enforce Formal RDP Privacy Budget Tracking: Implement Rényi Differential Privacy (RDP) accounting to monitor cumulative privacy expenditure across continuous, multi-round federated learning sessions. Rather than evaluating privacy loss independently for each communication round, RDP accounting provides a systematic mechanism for composing privacy guarantees over repeated model updates. Institutions can therefore quantify cumulative privacy loss, maintain predefined limits, and identify when additional training rounds may exceed the permitted privacy budget. This enables privacy-aware scheduling decisions, such as reducing the number of training rounds, modifying the noise multiplier, or terminating training before the privacy threshold is exhausted. Such continuous accounting is particularly important in long-running healthcare and financial FL deployments, where privacy expenditure can accumulate gradually over hundreds of communication cycles.

Adopt Layer-Wise Adaptive Clipping: Replace uniform global gradient clipping with dynamic, layer-specific norm thresholds that reflect the sensitivity, function, and learning requirements of different neural-network layers. Early feature-extraction layers may require relatively larger clipping thresholds to preserve meaningful representation-learning signals, whereas deeper classification layers can often operate with tighter bounds to limit excessive gradient sensitivity. The thresholds can also be periodically recalibrated according to observed gradient distributions during training. This approach can reduce unnecessary distortion caused by aggressive uniform clipping while continuing to constrain unusually large client contributions. Consequently, layer-wise clipping provides a more granular mechanism for balancing gradient utility, model convergence, and differential privacy requirements in heterogeneous federated environments.

Institute Continuous Shadow Probing: Periodically expose global FL updates to simulated Membership Inference Attacks using independently trained shadow models that approximate potential adversarial capabilities. Attack success rates should be monitored at multiple stages of federated training rather than evaluated only after final model convergence. Comparing privacy leakage across training rounds can reveal whether particular training phases, client participation patterns, or model updates create elevated exposure to membership inference. These empirical observations provide an important complement to theoretical guarantees because formal DP accounting and practical attack resistance measure related but distinct aspects of privacy protection. Establishing predefined alert thresholds can further enable security teams to trigger additional privacy controls, adjust training parameters, or suspend deployment when unexpected leakage patterns emerge.

References

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (pp. 308–318). Association for Computing Machinery. https://doi.org/10.1145/2976749.2978318
2. Haque, M. A., & Rasel-Ul-Alam, M. (2018). Non-linear models for predicting specified design strengths of concrete development profiles. *HBRC Journal, 14*(1), 123–136.
3. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (pp. 1273–1282). PMLR.
4. Bonawitz, K. A., 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. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1175–1191). Association for Computing Machinery. https://doi.org/10.1145/3133956.3133982
5. Mironov, I. (2017). Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium* (pp. 263–275). IEEE. https://doi.org/10.1109/CSF.2017.11
6. Nasr, M., Shokri, R., & Houmansadr, A. (2019). Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *2019 IEEE Symposium on Security and Privacy* (pp. 739–753). IEEE. https://doi.org/10.1109/SP.2019.00065
7. Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017). Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy* (pp. 3–18). IEEE. https://doi.org/10.1109/SP.2017.41
8. Truex, S., Liu, L., Gursoy, M. E., Yu, L., & Wei, W. (2020). Demystifying membership inference attacks in machine learning as a service. *IEEE Transactions on Services Computing, 13*(5), 797–812. https://doi.org/10.1109/TSC.2019.2898582

9. Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. In *Advances in Neural Information Processing Systems 32*. Neural Information Processing Systems Foundation.