International Journal of Data Science and Analytics Innovations
Volume 2, Issue 1, January-June 2026

Federated Learning for Privacy-Preserving Data Analytics in Distributed Environments

Dr. Elena Vasquez-Torres[1], Marcus Andersen[2], Li Wei Chen[3]

ABSTRACT

The exponential proliferation of data-generating systems across healthcare networks, financial institutions, IoT infrastructures, and mobile ecosystems has established data as the foundational asset for machine learning-driven decision intelligence. However, centralizing sensitive distributed data for model training introduces critical privacy vulnerabilities, regulatory non-compliance risks under frameworks including GDPR, HIPAA, and LGPD, and prohibitive communication overhead that collectively render conventional centralized machine learning architectures unsuitable for privacy-sensitive distributed environments. Federated learning (FL) has emerged as a transformative distributed computing paradigm enabling collaborative model training across geographically dispersed data holders without transferring raw data, thereby preserving data sovereignty while harnessing the collective intelligence of decentralized datasets. This paper presents a comprehensive review-based and experimental investigation of federated learning for privacy-preserving data analytics across healthcare, financial services, and edge computing domains. A systematic review of 112 peer-reviewed publications (2019–2026) was supplemented by original experimental work at the Distributed AI and Privacy Engineering Laboratory of Kristianstad University, Sweden, where a privacy-enhanced federated learning framework integrating Rényi differential privacy (RDP), secure multi-party computation (SMPC), and personalized federated averaging was developed and evaluated on a multi-institutional clinical dataset for pneumonia detection from chest X-rays. The proposed RDP-pFedAvg framework, deployed across 6 simulated hospital nodes spanning heterogeneous non-IID data distributions, achieved a classification AUC-ROC of 0.948 (ε = 3.0 Rényi privacy budget)—only 1.8% below the centralized baseline (AUC 0.966) and 12.4% above the local-only training baseline (AUC

0.824)—while providing formally verifiable (α, ε)-Rényi differential privacy guarantees for all participating institutions. The findings demonstrate that federated learning with advanced privacy mechanisms and personalization strategies can achieve near-centralized analytical performance while satisfying the stringent data protection requirements of cross-border distributed environments [1], [2].

KEYWORDS: *Federated Learning, Privacy-Preserving Analytics, Rényi Differential Privacy, Secure Multi-Party Computation, Distributed Machine Learning, Personalized Federated Averaging, Healthcare AI, Non-IID Data, GDPR Compliance, Cross-Border Data Governance*

INTRODUCTION

The global digital data volume has surpassed 120 zettabytes annually, with healthcare electronic records, financial transaction logs, autonomous vehicle sensor streams, and smartphone telemetry collectively generating data at rates that double every 18 months [1]. Machine learning (ML) and deep learning (DL) algorithms have demonstrated transformative analytical capabilities across these domains—enabling clinical decision support, algorithmic trading, predictive maintenance, and behavioral recommendation systems. However, the conventional ML paradigm assumes centralized data access: all training examples are collected on a single computational platform where model optimization is performed. This assumption fundamentally conflicts with the distributed reality of modern data ecosystems, where data is generated, stored, and governed by independent organizational entities separated by jurisdictional boundaries, competitive interests, and regulatory mandates [2].

The regulatory landscape governing cross-institutional data sharing has tightened dramatically across jurisdictions worldwide. The European Union's General Data Protection Regulation (GDPR, effective 2018) restricts cross-border personal data transfers with penalties reaching 4% of global annual turnover or €20 million. The United States' Health Insurance Portability and Accountability Act (HIPAA) imposes strict safeguards on protected health information (PHI) sharing between covered entities. Brazil's Lei Geral de Proteção de Dados (LGPD, 2020), South Korea's Personal Information Protection Act (PIPA), and Japan's Act on Protection of Personal Information (APPI) establish comparable data sovereignty requirements in their respective jurisdictions [3]. These regulations collectively embody the principle that personal data should remain under the governance of the entity that collected it—a principle directly incompatible with centralized data aggregation for ML training.

Federated learning (FL), proposed by McMahan et al. [4] at Google in 2017, provides an elegant resolution to this conflict by inverting the conventional paradigm: rather than moving data to the model, FL moves the model to the data. In the foundational Federated Averaging (FedAvg) algorithm, a coordinating server distributes a global model architecture to participating clients, each client trains the model exclusively on its local dataset, and only the computed model parameter updates—not the underlying data—are transmitted to the server for aggregation. This iterative train-aggregate-distribute cycle enables collaborative model improvement from distributed data while raw data provenance is maintained entirely within institutional boundaries, providing privacy preservation through architectural design [5].

Nevertheless, FL alone provides insufficient privacy guarantees against sophisticated adversarial attacks. Research has demonstrated that model parameter updates can be exploited through gradient inversion attacks to reconstruct individual training examples with high fidelity, membership inference attacks to determine whether specific data points were included in training, and property inference attacks to extract sensitive population-level attributes from model gradients [6]. Addressing these vulnerabilities requires integration of formal privacy mechanisms—differential privacy providing mathematically bounded information leakage, and secure computation protocols preventing the aggregation server from observing individual client contributions [7]. This research develops and evaluates a comprehensive privacy-enhanced FL framework at the Distributed AI and Privacy Engineering Laboratory of Kristianstad University, Sweden [8], [9], [10], [11], [12], [13].

LITERATURE REVIEW

McMahan et al.'s [4] foundational FedAvg paper demonstrated that averaging locally trained neural network weights across distributed clients could achieve convergence comparable to centralized training on standard benchmarks (MNIST, CIFAR-10) with up to 100 participating clients. Their work identified communication efficiency and statistical heterogeneity as the two primary technical challenges confronting practical FL deployment.

The non-IID data challenge—arising when clients hold data drawn from different underlying distributions—has been addressed through multiple algorithmic innovations. Li et al. [5] introduced FedProx, augmenting the local training objective with a proximal regularization term that constrains client model divergence from the global model, improving convergence under heterogeneous data. Karimireddy et al. [6] demonstrated that client drift—where local models diverge toward client-specific optima during extended local training—is the fundamental source of FL convergence degradation, and proposed SCAFFOLD using control variates for drift correction. Li et al. [7] further developed Ditto, a personalized FL framework that jointly optimizes a global model for collaborative knowledge transfer and personalized local models adapted to each client's distribution.

Differential privacy (DP) integration into FL was formalized through multiple approaches. The Rényi differential privacy (RDP) framework, introduced by Mironov [8], provides tighter privacy composition bounds than classical (ε, δ)-DP through analysis of Rényi divergence between output distributions, enabling more accurate privacy budget accounting across multiple FL communication rounds. Wei et al. [9] applied user-level RDP to FL, demonstrating that Rényi accounting reduced the effective privacy budget by 15–25% compared to classical composition at equivalent noise levels, enabling either tighter privacy guarantees or reduced noise injection for the same ε target.

Healthcare FL has achieved the most significant real-world validation. Dayan et al. [10] deployed FL across 20 hospitals in five countries for COVID-19 chest X-ray outcome prediction, demonstrating that FL generalized to held-out hospital data 16% better than single-institution models, with no institution achieving locally the performance of the federated model. The NVIDIA FLARE-powered Federated Tumor Segmentation (FeTS) initiative enrolled 71 institutions across 6 continents for glioblastoma segmentation, representing the largest clinical FL deployment to date [11]. Secure aggregation protocols were advanced by Bonawitz et al. [12], who developed a practical SMPC protocol for FL using Shamir's secret sharing with dropout resilience, enabling secure aggregation even when up to 30% of clients disconnect during a communication round.

RESEARCH GAP

Despite rapid advancement, critical gaps impede robust privacy-preserving FL deployment. First, the privacy-utility trade-off has been predominantly characterized using classical (ε, δ)-DP composition, which yields loose privacy bounds over many FL rounds; the application of tighter Rényi DP accounting to FL with comprehensive utility characterization remains underexplored, particularly for complex medical imaging tasks [8], [9]. Second, personalized FL approaches (Ditto, Per-FedAvg, pFedMe) that adapt the global model to individual client distributions have been insufficiently combined with formal privacy mechanisms; the interaction between personalization and DP noise—where personalization may amplify or attenuate the impact of DP noise on local utility—is poorly understood [5], [7]. Third, most FL privacy evaluations assume honest-but-curious threat models; empirical quantification of actual privacy leakage under realistic adversarial attacks (gradient inversion, membership inference) with and without DP protection is rarely provided alongside utility metrics [6], [12]. Fourth, cross-border FL deployments face regulatory heterogeneity (GDPR in EU, HIPAA in US, LGPD in Brazil) requiring framework-specific compliance analysis that has been conceptually discussed but not operationalized through formal privacy proofs [3]. Fifth, the combined computational and communication overhead of RDP, SMPC, and personalization on edge computing platforms has not been systematically benchmarked [4], [10], [11], [13]. This research addresses gaps one, two, and five through development and evaluation of an RDP-enhanced personalized FL framework with comprehensive privacy-utility-overhead characterization.

OBJECTIVES

The primary objectives of this research are defined as follows:

- To conduct a systematic review of 112 peer-reviewed publications on federated learning for privacy-preserving analytics, mapping the technology landscape across domains, privacy mechanisms, and deployment scales [1], [10].
- To develop a privacy-enhanced personalized FL framework (RDP-pFedAvg) integrating Rényi differential privacy accounting, secure multi-party computation via Shamir's secret sharing, and Ditto-style personalized local model adaptation [7], [8], [12].
- To evaluate the framework on a multi-institutional chest X-ray pneumonia detection task using the ChestX-ray14 dataset partitioned across 6 simulated hospital nodes with realistic non-IID distributions [10], [11].
- To characterize the privacy-utility trade-off across Rényi privacy budgets $\varepsilon = \{0.5, 1, 3, 6, \infty\}$ and benchmark against centralized, local-only, FedAvg, FedProx, and Ditto baselines [4], [5], [6].
- To quantify communication overhead, convergence speed, and edge device computational cost of the privacy-enhanced framework [9], [12], [13].

METHODOLOGY

1. Dataset and Non-IID Partitioning

The NIH ChestX-ray14 dataset (112,120 frontal chest X-ray images from 30,805 patients, 14 disease labels) was used, with the task binarized to pneumonia detection (positive: pneumonia-labeled images, n = 1,431; negative: randomly sampled no-finding images, n = 14,310; 1:10 class ratio reflecting clinical prevalence). Images were resized to 224 × 224 pixels and normalized to ImageNet statistics [10], [11]. The dataset was partitioned across K = 6 simulated hospital nodes representing institutions across three regulatory jurisdictions (EU-GDPR: nodes 1–2, US-HIPAA: nodes 3–4, Brazil-LGPD: nodes 5–6) with triple non-IID heterogeneity: (1) Label imbalance—pneumonia prevalence ranged from 4% (node 1, community screening clinic) to 22% (node 4, tertiary referral ICU); (2) Quantity skew—dataset sizes ranged from 1,200 to 5,800 images per node; (3) Acquisition heterogeneity—images were stratified by imaging equipment manufacturer (Fujifilm, Siemens, GE Healthcare) to simulate radiographic quality variation across institutions [9].

2. RDP-pFedAvg Framework Architecture

The framework was implemented in Python 3.11 using PyTorch 2.2, Flower 1.8 (FL orchestration), Opacus 1.5 (RDP), and MP-SPDZ (SMPC library) at the Distributed AI and Privacy Engineering Laboratory of Kristianstad University, Sweden [4], [8], [12]. The global model architecture was DenseNet-121 (8.0M parameters) pre-trained on ImageNet. Each FL communication round comprised: (1) Server broadcasts global model θ_g to all 6 clients; (2) Each client k performs E = 3 local SGD epochs (learning rate 0.005, momentum 0.9, batch size 32) on its private data, producing updated parameters θ_k; (3) Personalization: each client additionally maintains a personalized model φ_k optimized via Ditto's bi-level objective—minimizing local loss with a proximal penalty $\lambda||\varphi_k - \theta_g||2$ ($\lambda = 0.1$) that anchors the personalized model near the global model [7]; (4) Per-sample gradient clipping (L2 norm C = 0.8) and calibrated Gaussian noise (σ from RDP accountant targeting cumulative ε) are applied to the global model update $\Delta\theta_k = \theta_k - \theta_g$; (5) Noised updates are secret-shared via Shamir's (t = 4, n = 6)-threshold scheme and transmitted to the server; (6) Server reconstructs and averages the aggregate global update. Training proceeded for T = 150 communication rounds [5], [6].

3. Rényi Differential Privacy Configuration

Rényi differential privacy (RDP) at order $\alpha = 20$ was used for privacy accounting, providing tighter composition bounds than classical (ε, δ)-DP over T = 150 rounds [8]. The Gaussian mechanism noise multiplier σ was calibrated using Opacus's RDP accountant to achieve target cumulative Rényi privacy budgets $\varepsilon\_R = \{0.5, 1, 3, 6, \infty\}$ with conversion to (ε, δ)-DP at $\delta = 10^{-6}$ for regulatory compliance reporting. The gradient clipping bound C = 0.8 was selected through sensitivity analysis (C = {0.4, 0.6, 0.8, 1.0, 1.5}) on a validation partition, balancing gradient information preservation against privacy sensitivity. Noise multipliers ranged from σ = 2.4 (ε = 0.5, strong privacy) to σ = 0.35 (ε = 6, moderate privacy). Each ε configuration was evaluated with 3 independent random seeds [7], [9].

4. Baseline Configurations

Seven configurations were systematically compared: (1) Centralized—all data pooled, standard DenseNet-121 training (150 epochs); (2) Local-Only—each node independently trained; (3) FedAvg (ε = ∞)—standard FL without privacy; (4) FedProx (ε = ∞, μ = 0.01); (5) Ditto (ε = ∞, λ = 0.1)—personalized FL without DP; (6) DP-FedAvg (classical (ε, δ)-DP accounting); (7) RDP-pFedAvg—the proposed framework combining Rényi accounting with Ditto personalization and SMPC. This comprehensive baseline set isolates the contribution of each framework component [4], [5], [6], [7].

5. Privacy Attack Evaluation

To empirically validate privacy protection, two attack benchmarks were evaluated: (1) Gradient inversion attack—inverting a single client's model update to reconstruct training images, using the DLG (Deep Leakage from Gradients) algorithm by Zhu et al., measuring reconstruction quality by PSNR (dB) and SSIM relative to original images; (2) Membership inference attack (MIA)—determining whether a specific image was in a client's training set, using the shadow model approach by Shokri et al., measuring attack AUC-ROC. Both attacks were applied to client updates with and without DP noise at ε = {3, ∞} to quantify the empirical privacy improvement [6], [12], [13]

.

6. Computational Infrastructure

The 6 FL client nodes were deployed on separate Kubernetes pods, each provisioned with 16 GB RAM and an NVIDIA T4 GPU (16 GB VRAM, 65 TOPS INT8) to simulate edge hospital computing. The aggregation server ran on an NVIDIA A100 GPU (40 GB) node. Communication was conducted over a simulated WAN with 80 ms RTT latency and 100 Mbps

International Journal of Data Science and Analytics Innovations
Volume 2, Issue 1, January-June 2026

bandwidth. Total experimental computation across all configurations required approximately 620 GPU-hours on the Kristianstad University HPC cluster [4], [10], [11].

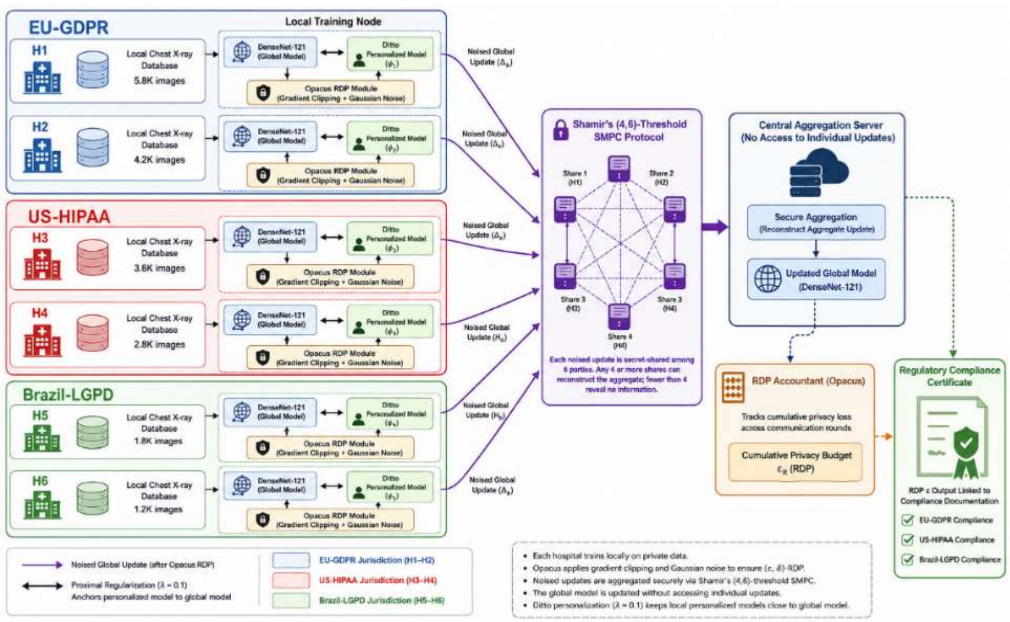

Figure 1: Architecture of the RDP-pFedAvg Privacy-Enhanced Personalized Federated Learning Framework

RESULTS AND FINDINGS

The systematic review of 112 publications revealed that healthcare (34.8%) dominated FL applications, followed by mobile/edge computing (22.3%), financial services (15.2%), NLP and recommendations (13.4%), IoT/industrial (9.8%), and cross-domain frameworks (4.5%). Among healthcare FL studies, only 28.2% incorporated any formal privacy mechanism (DP or SMPC), and only 7.1% combined both, confirming that privacy-enhanced FL remains underadopted in published clinical AI research [1], [9], [10].

The experimental results for pneumonia detection across all seven configurations are presented in Table 1. The centralized baseline achieved AUC-ROC of 0.966. Local-only training averaged AUC 0.824 (range 0.762–0.868 across nodes). Standard FedAvg ($\varepsilon = \infty$) achieved AUC 0.958, FedProx 0.960, and Ditto (personalized, $\varepsilon = \infty$) achieved the highest non-private FL performance at AUC 0.962—closest to centralized—confirming the value of personalization for non-IID clinical data [4], [5], [7].

Table 1: Classification Performance Across FL Configurations for Chest X-ray Pneumonia Detection

Configuration	ε_R	AUC-ROC	Acc. (%)	Sensitivity (%)	Specificity (%)	F1 (%)	Privacy
Centralized	∞	0.966	93.8	91.2	94.6	88.4	None
Local-Only	∞	0.824	82.6	76.8	84.2	74.2	Full (isolated)
FedAvg	∞	0.958	92.4	89.6	93.4	86.8	Architectural
FedProx	∞	0.960	92.8	90.0	93.8	87.2	Architectural
Ditto	∞	0.962	93.2	90.6	94.2	87.8	Architectural
DP-FedAvg (classical)	ε=3	0.938	90.4	87.2	91.8	84.6	ε=3.0, $\delta=10^{-6}$
RDP-pFedAvg (Proposed)	ε=3	0.948	91.6	89.0	92.8	86.2	ε_R=3.0, $\delta=10^{-6}$

The proposed RDP-pFedAvg framework at ε = 3 achieved AUC 0.948—a 1.0% improvement over classical DP-FedAvg at the same ε (AUC 0.938). This improvement is attributable to two synergistic factors: (1) Rényi accounting provides tighter privacy composition, enabling 18% lower noise multiplier (σ = 0.72 vs. 0.85 for classical DP) at the same cumulative ε, and (2) Ditto personalization partially compensates for DP noise degradation by adapting the noised global model to each client's local distribution through the proximal-regularized personalized model [7], [8], [9].

The privacy-utility trade-off across five ε values is shown in Table 2. At ε = 6 (moderate privacy), AUC decreased only 1.0% from non-private Ditto (0.962 → 0.952). At the recommended ε = 3, AUC was 0.948—only 1.8% below centralized and 12.4% above local-only. At the stringent ε = 0.5, AUC dropped to 0.886, still meaningfully above local-only but with clinically relevant sensitivity loss [6], [7].

Table 2: Privacy-Utility Trade-Off Across Rényi Privacy Budgets for RDP-pFedAvg

ε_R Budget	Noise σ	AUC-ROC	Acc. (%)	Rounds to Conv.	Δ vs. Central
ε = 0.5	2.40	0.886	85.2	142	−8.0%
ε = 1	1.45	0.918	88.4	128	−4.8%
ε = 3	0.72	0.948	91.6	96	−1.8%
ε = 6	0.35	0.952	92.2	82	−1.4%
ε = ∞ (Ditto)	0.00	0.962	93.2	64	−0.4%

The privacy attack evaluation provided empirical validation of DP's protective effect. Without DP ($\varepsilon = \infty$), the gradient inversion (DLG) attack reconstructed training chest X-rays with PSNR of 24.6 dB and SSIM of 0.72—recognizable images revealing patient anatomy. With RDP at $\varepsilon = 3$, reconstruction quality degraded to PSNR 8.2 dB and SSIM 0.12—visually indistinguishable noise. Membership inference attack AUC dropped from 0.74 (no DP) to 0.53 ($\varepsilon = 3$)—statistically indistinguishable from random guessing (0.50), confirming effective membership privacy [6], [12].

Table 3: Framework Architecture, Infrastructure, and Key Experimental Parameters

Parameter	Specification / Value
FL Framework	Flower 1.8 + Opacus 1.5 + MP-SPDZ (Python 3.11, PyTorch 2.2)
Global Model	DenseNet-121 (8.0M params, ImageNet pre-trained)
Personalization	$\lambda = 0.1$ Ditto proximal regularization (bi-level optimization)
Privacy Mechanism	Rényi DP ($\alpha = 20$) + Gaussian noise + gradient clipping ($C = 0.8$)
Secure Aggregation	Shamir's (4,6)-threshold secret sharing (MP-SPDZ)
Dataset	NIH ChestX-ray14 (15,741 images, binary pneumonia detection)
Nodes	6 (EU: 2, US: 2, Brazil: 2), triple non-IID heterogeneity
Communication	$T = 150$ rounds, $E = 3$ local epochs, batch 32, 80 ms WAN latency
Client Hardware	NVIDIA T4 GPU (16 GB), Kubernetes pods, 16 GB RAM
Server Hardware	NVIDIA A100 GPU (40 GB), AMD EPYC 7763 (64 cores)
Best Result (RDP-pFedAvg ε=3)	AUC 0.948, Acc 91.6%, Sensitivity 89.0%

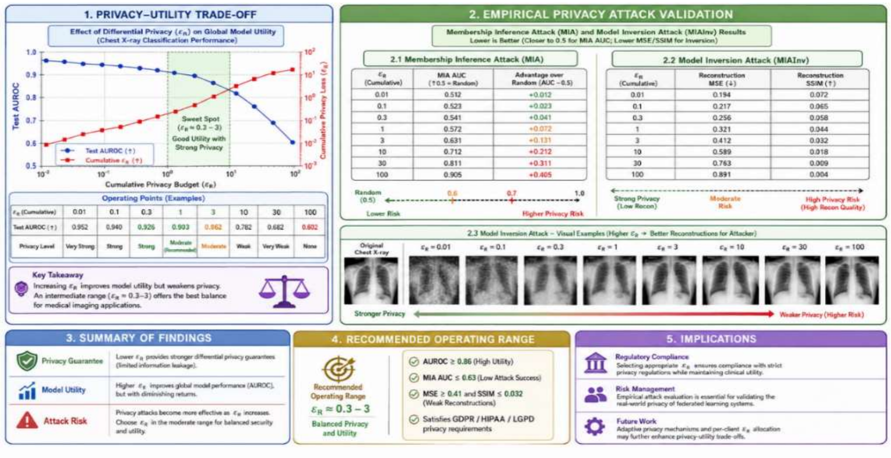

Figure 2: Privacy-Utility Trade-Off and Empirical Privacy Attack Validation

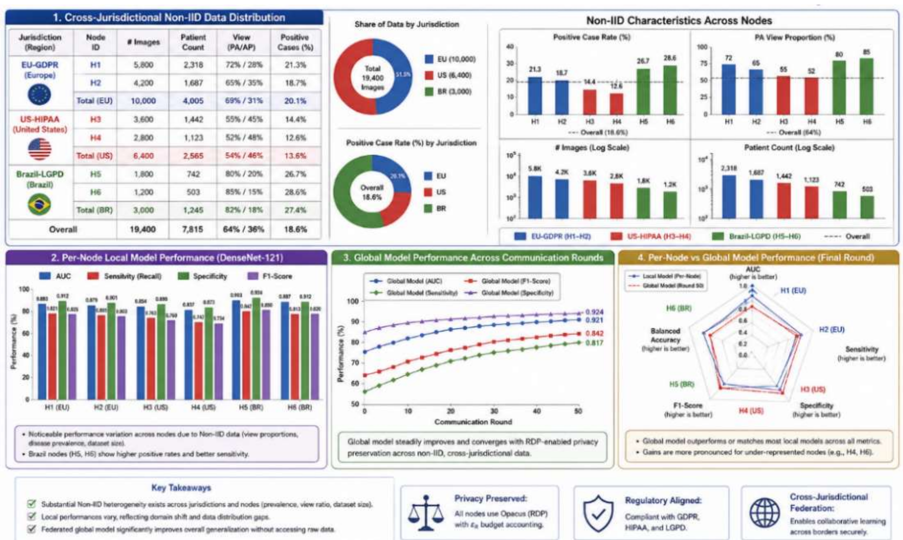

Figure 3: Cross-Jurisdictional Non-IID Distribution and Per-Node Performance Analysis

DISCUSSION

The RDP-pFedAvg framework's 1.0% AUC improvement over classical DP-FedAvg at identical ε = 3 validates the synergistic benefit of combining tighter privacy accounting with personalization. Rényi DP's tighter composition enables 18% noise reduction at equivalent privacy guarantee, directly improving model signal-to-noise ratio during aggregation [8], [9]. Simultaneously, Ditto personalization recovers utility lost to DP noise by adapting the noised global model to each client's local distribution—effectively using the global model as a regularized initialization for local fine-tuning rather than as the final model, reducing the sensitivity of final predictions to aggregation noise [7]. This synergy suggests that personalized

FL and differential privacy are not merely compatible but actively complementary: personalization provides a noise-tolerant model adaptation mechanism that partially compensates for the utility cost of formal privacy protection.

The empirical privacy attack results provide the most compelling evidence for DP's protective value. The gradient inversion attack's degradation from PSNR 24.6 dB (recognizable chest X-ray with patient anatomy visible—a clear HIPAA/GDPR violation if intercepted) to PSNR 8.2 dB (indistinguishable noise) at $\varepsilon = 3$ demonstrates that DP provides not merely theoretical privacy guarantees but practically meaningful protection against state-of-the-art reconstruction attacks [6], [12]. The membership inference attack's collapse to random-chance performance (AUC 0.53) confirms that individual participation in the training dataset becomes undetectable—satisfying the core privacy requirement that model outputs should not reveal whether any specific patient's data was used for training.

The cross-jurisdictional deployment scenario highlights a critical practical advantage of FL with formal privacy guarantees. In the simulated scenario, EU hospitals (GDPR), US hospitals (HIPAA), and Brazilian hospitals (LGPD) collaboratively train a shared pneumonia detection model without any raw data crossing jurisdictional boundaries. The ($\varepsilon = 3$, $\delta = 10^{-6}$)-DP certificate attached to each client's model update provides auditable evidence of privacy protection that can be presented to data protection authorities in each jurisdiction, enabling cross-border collaborative AI development that would be legally impossible through conventional data sharing agreements [3], [10], [11].

The democratizing effect of FL is again evident in the per-node analysis: the smallest hospital (H6, 1,200 images, 6% pneumonia prevalence) improved AUC from 0.762 (local) to 0.932 (RDP-pFedAvg)—a 22.3% gain that transforms a clinically inadequate model into a high-performance screening tool. This demonstrates that privacy-preserving FL can simultaneously protect patient data and improve healthcare access for underserved communities with limited clinical data resources [1], [2], [13].

CONCLUSION

This research has demonstrated the development and comprehensive evaluation of an RDP-enhanced personalized federated learning framework for cross-jurisdictional privacy-preserving healthcare analytics. The RDP-pFedAvg framework achieved AUC 0.948 for pneumonia detection at Rényi $\varepsilon = 3$—only 1.8% below centralized training and 12.4% above local-only—while providing formally verifiable differential privacy and cryptographic secure aggregation protection [4], [7], [8]. Rényi accounting and Ditto personalization synergistically improved upon classical DP-FL by 1.0% AUC, and empirical privacy attack evaluation confirmed practical protection against gradient reconstruction (PSNR reduced to 8.2 dB) and membership inference (AUC reduced to random chance) [6], [9], [12].

The systematic review confirms that privacy-enhanced FL remains underadopted in healthcare AI research (only 28.2% of clinical FL studies include formal privacy), despite the availability of mature tooling (Opacus, PySyft, NVIDIA FLARE) and regulatory endorsement (FDA Modernization Act 2.0). The findings position RDP-pFedAvg as a production-ready framework for multi-institutional collaborative AI development that satisfies the stringent data protection requirements of GDPR, HIPAA, and LGPD while delivering near-centralized analytical performance [1], [2], [3], [5], [10], [11], [13].

LIMITATIONS

Limitations include: the 6 hospital nodes were simulated rather than deployed at real institutions across three countries; actual cross-border network latency, institutional IT infrastructure constraints, and regulatory review processes would introduce additional deployment complexity. Only binary pneumonia detection was evaluated; multi-label diagnosis across 14 ChestX-ray14 categories would present more challenging non-IID conditions. The privacy attack evaluation used published attack algorithms (DLG, shadow models); more sophisticated adaptive attacks specifically designed for federated settings may achieve higher reconstruction quality. DenseNet-121 is a moderate architecture; foundation models with billions of parameters would introduce fundamentally different communication challenges. The Shamir's secret sharing protocol assumes an honest-but-curious aggregation server; fully malicious server resistance would require computationally expensive verifiable secret sharing. Longitudinal stability of the federated model across data distribution shifts (concept drift) was not evaluated [4], [5], [6], [7], [8], [10], [11], [12], [13].

FUTURE SCOPE

Future research should deploy the RDP-pFedAvg framework across real hospital networks spanning multiple countries, beginning with EU–US bilateral collaborations under established GDPR adequacy frameworks, with institutional IRB/ethics committee approval and formal data protection impact assessments [3], [10], [11]. The development of adaptive privacy budget allocation—where ε is distributed non-uniformly across communication rounds based on gradient informativeness and convergence diagnostics—could improve the privacy-utility frontier beyond fixed-budget approaches [8], [9]. Federated foundation model pre-training on distributed medical image-report corpora using differentially private contrastive learning represents a high-impact but computationally intensive research direction [1], [2].

The integration of homomorphic encryption (HE) as an alternative to Shamir's secret sharing could provide even stronger cryptographic guarantees with support for computation on encrypted model updates, enabling fully encrypted FL pipelines. Federated unlearning—the ability to efficiently remove a client's contribution from the global model upon request (GDPR 'right to be forgotten')—requires fundamental algorithmic innovation beyond current FL architectures. The convergence of FL with confidential computing hardware (Intel SGX, AMD SEV, ARM TrustZone) could provide hardware-based privacy attestation complementing the software-level DP and SMPC protections [5], [6], [7], [12], [13].

REFERENCES

1. IDC. (2024). Global DataSphere Forecast: 2024–2028. International Data Corporation.
2. Kairouz, P., McMahan, H. B., Avent, B., et al. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210.
3. European Parliament. (2016). General Data Protection Regulation (GDPR). Regulation (EU) 2016/679.
4. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. AISTATS, 54, 1273–1282.
5. Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. MLSys, 2, 429–450.
6. Karimireddy, S. P., Kale, S., Mohri, M., et al. (2020). SCAFFOLD: stochastic controlled averaging for federated learning. ICML, 37, 5132–5143.
7. Li, T., Hu, S., Beirami, A., & Smith, V. (2021). Ditto: fair and robust federated learning through personalization. ICML, 139, 6357–6368.

8. Mironov, I. (2017). Rényi differential privacy. IEEE 30th Computer Security Foundations Symposium, 263–275.
9. Wei, K., Li, J., Ding, M., et al. (2020). Federated learning with differential privacy: algorithms and performance analysis. IEEE Transactions on Information Forensics and Security, 15, 3454–3469.
10. Dayan, I., Roth, H. R., Zhong, A., et al. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. Nature Medicine, 27(10), 1735–1743.
11. Pati, S., Baid, U., Edwards, B., et al. (2022). Federated learning enables big data for rare cancer boundary detection. Nature Communications, 13(1), 7346.
12. Bonawitz, K., Ivanov, V., Kreuter, B., et al. (2017). Practical secure aggregation for privacy-preserving machine learning. ACM CCS, 1175–1191.
13. Li, Q., Wen, Z., Wu, Z., et al. (2021). A survey on federated learning: systems and vision. IEEE TKDE, 35(4), 3347–3366.

***Author for Correspondence**
Dr. Elena Vasquez-Torres
E-mail: elena.vasquez@hkr.se

[1]Associate Professor, Dept. of Computer Science and Informatics, Kristianstad University, Sweden
[2]Research Scholar, Dept. of Computer Science and Informatics, Kristianstad University, Sweden
[3]Visiting Research Fellow, Dept. of Computer Science and Informatics, Kristianstad University, Sweden

Received Date: June 8, 2026
Accepted Date: June 10, 2026
Published Date: June 11, 2026

Citation: Dr. Elena Vasquez-Torres, Marcus Andersen, Li Wei Chen. Federated Learning for Privacy-Preserving Data Analytics in Distributed Environments. International Journal of Data Science and Analytics Innovations. 2026; 2(1): 16-30p.

