2025 International Conference on Computing and Communications (COMPUTINGCON)
Pune, India. Sep 01-03, 2025

2025 International Conference on Computing and Communications (COMPUTINGCON) | 979-8-3315-2253-7/25/$31.00 ©2025 IEEE | DOI: 10.1109/COMPUTINGCON64838.2025.11377325

Privacy-Preserving Federated Learning with Adaptive Differential Privacy for Healthcare Data Aggregation

Dr. Sankara Rao. Pusarla
Associate Professor
Department of ECE
AIET(A), Tagarapuvalasa,VZM.
sankar.ecehod@gmail.com

Dr. Pallavi Jha
Associate Professor
Dept of Computer Engineering
Alard University, Pune
palcjha21@gmail.com

Naga Venkateshwara Rao K
Assistant Professor
Department of ECE
St. Martins Engineering College, Telangana
nagavkollipara@gmail.com

B. Tapasvi
Assistant Professor
Department of ECE
S.R.K.R Engineering College (A) Bhimavaram
tapasvi@srkrec.ac.in

Dr. P N V Syamala Rao M
Assistant Professor
Department of CSE
SRM University AP
syamalarao.m@srmap.edu.in

Dr Chanakya Kumar Jha
Senior Member IEEE
erchankya@gmail.com

***Abstract*—Federated learning enables collaborative model training across distributed healthcare institutions without centralizing sensitive patient data. However, existing approaches face challenges in balancing privacy protection and model utility. This paper proposes an adaptive differential privacy mechanism for federated learning in healthcare data aggregation that dynamically adjusts privacy budgets based on data sensitivity and model convergence. We introduce a privacy-utility trade-off optimizer that allocates privacy budgets adaptively across training rounds while maintaining rigorous privacy guarantees. Our experimental evaluation on the MIMIC-III and eICU datasets demonstrates that the proposed approach achieves 7.3% higher model accuracy compared to fixed differential privacy mechanisms while providing equivalent privacy protection with $\epsilon = 2.5$. The system reduces communication overhead by 31% through selective gradient aggregation and achieves convergence 23% faster than baseline federated learning approaches.**

***Index Terms*—Federated learning, differential privacy, healthcare analytics, privacy-preserving machine learning, distributed learning**

I. Introduction

The proliferation of electronic health records (EHRs) and medical devices has created unprecedented opportunities for data-driven healthcare analytics [1]. However, privacy regulations such as HIPAA and GDPR impose strict constraints on sharing patient data across institutions [2]. Federated learning (FL) addresses this challenge by enabling collaborative model training without centralizing raw data, where multiple healthcare institutions train a shared model while keeping data locally [3].

Despite its promise, federated learning faces critical privacy vulnerabilities. Recent studies have demonstrated that model updates can leak sensitive information about training data through gradient inversion attacks and membership inference attacks [4], [5]. Differential privacy (DP) provides formal privacy guarantees by adding calibrated noise to model updates, but traditional fixed-budget approaches result in significant utility degradation, particularly in deep learning models [6].

This paper addresses three fundamental challenges in privacy-preserving federated learning for healthcare:

- **Privacy-Utility Trade-off:** Fixed differential privacy budgets either provide insufficient privacy protection or severely degrade model performance.
- **Data Heterogeneity:** Healthcare data exhibits high variability across institutions due to demographic differences, equipment variations, and clinical practices.
- **Communication Efficiency:** Frequent model synchronization in federated learning creates substantial communication overhead.

We propose an adaptive differential privacy mechanism that dynamically allocates privacy budgets based on gradient sensitivity analysis and model convergence metrics. Our contributions include:

1) An adaptive privacy budget allocation algorithm that adjusts noise levels across training rounds while maintaining overall privacy guarantees under the Rényi Differential Privacy framework.
2) A gradient sensitivity estimator that identifies critical model parameters requiring enhanced privacy protection.
3) A selective aggregation protocol that reduces communication overhead by transmitting only significant gradient updates.
4) Comprehensive evaluation on MIMIC-III and eICU

datasets demonstrating superior privacy-utility trade-offs compared to state-of-the-art approaches.

II. Related Work

A. Federated Learning in Healthcare

McMahan et al. [7] introduced the Federated Averaging (FedAvg) algorithm, which forms the foundation of modern federated learning systems. Sheller et al. [3] demonstrated federated learning for brain tumor segmentation across multiple institutions, achieving performance comparable to centralized training. Xu et al. [1] surveyed federated learning applications in healthcare, identifying key challenges including data heterogeneity and privacy protection.

B. Differential Privacy in Machine Learning

Abadi et al. [6] proposed the moments accountant method for tracking privacy loss in deep learning, enabling practical differentially private stochastic gradient descent (DP-SGD). Mironov [8] introduced Rényi Differential Privacy, providing tighter privacy accounting for iterative algorithms. Recent work by Bu et al. [9] improved privacy-utility trade-offs through automatic clipping and noise calibration.

C. Privacy-Preserving Federated Learning

Wei et al. [4] demonstrated that federated learning alone does not guarantee privacy, motivating the integration of differential privacy mechanisms. Naseri et al. [5] analyzed membership inference attacks against federated learning systems. Truex et al. [10] proposed hybrid approaches combining secure aggregation with differential privacy. However, existing methods use fixed privacy budgets that do not adapt to training dynamics.

III. Methodology

A. System Architecture

Our federated learning system consists of N healthcare institutions (clients) and a central aggregation server. Figure 1 illustrates the system architecture. Each institution i maintains a local dataset $\mathcal{D}_i = \{(x_j, y_j)\}_{j=1}^{n_i}$ where x_j represents patient features and y_j denotes clinical outcomes. The global model parameters θ are trained collaboratively without sharing raw data.

B. Adaptive Differential Privacy Mechanism

We formulate the adaptive privacy budget allocation as an optimization problem that balances privacy protection and model utility across training rounds.

1) Privacy Budget Allocation: For a federated learning process with T training rounds, we allocate privacy budget ϵ_t for round t such that:

$$\epsilon_t = \epsilon_0 \cdot \frac{\alpha_t}{\sum_{j=1}^{T} \alpha_j} \tag{1}$$

where ϵ_0 is the total privacy budget and α_t is the adaptive allocation weight for round t. The allocation weight is computed based on gradient magnitude and model convergence:

$$\alpha_t = \beta \cdot \frac{\|\nabla \mathcal{L}(\theta_t)\|_2}{\|\nabla \mathcal{L}(\theta_0)\|_2} + (1 - \beta) \cdot e^{-\gamma t} \tag{2}$$

where $\beta \in [0, 1]$ controls the balance between gradient-based and convergence-based allocation, γ is the decay rate, and $\nabla \mathcal{L}(\theta_t)$ represents the gradient of the loss function at round t.

2) Gradient Clipping and Noise Addition: For each client i at round t, we clip the local gradient update to bound its sensitivity:

$$\tilde{g}_{i,t} = g_{i,t} \cdot \min\left(1, \frac{C_t}{\|g_{i,t}\|_2}\right) \tag{3}$$

where $g_{i,t} = \nabla_\theta \mathcal{L}(\theta_t; \mathcal{D}_i)$ is the local gradient, and C_t is the adaptive clipping threshold. We then add Gaussian noise calibrated to the privacy budget:

$$\hat{g}_{i,t} = \tilde{g}_{i,t} + \mathcal{N}\left(0, \sigma_t^2 C_t^2 I\right) \tag{4}$$

where $\sigma_t = \frac{C_t \sqrt{2 \log(1.25/\delta)}}{\epsilon_t}$ ensures (ϵ_t, δ)-differential privacy for each round.

3) Selective Aggregation Protocol: To reduce communication overhead, we implement selective aggregation where clients transmit updates only when gradient magnitudes exceed a threshold τ_t:

$$\text{Transmit } \hat{g}_{i,t} \text{ if } \|\hat{g}_{i,t}\|_2 > \tau_t$$

The server aggregates received updates using weighted averaging based on local dataset sizes.

C. Privacy Analysis

Our mechanism satisfies (ϵ, δ)-Rényi Differential Privacy (RDP) with moment order λ. Under composition over T rounds:

$$\epsilon(\lambda) = \sum_{t=1}^{T} \epsilon_t(\lambda)$$

We use the moments accountant to track cumulative privacy loss and convert RDP guarantees to (ϵ, δ)-DP guarantees.

D. Algorithm

Table I presents our adaptive differential privacy federated learning algorithm.

IV. Experimental Evaluation

A. Datasets and Setup

We evaluate our approach on two real-world healthcare datasets:

MIMIC-III: Medical Information Mart for Intensive Care (version 1.4) contains de-identified health records of 46,520 ICU patients from Beth Israel Deaconess Medical Center [11]. We extract 12 clinical features including vital signs, laboratory values, and demographic information for mortality prediction.

eICU: The eICU Collaborative Research Database contains data from 200,859 ICU admissions across 335 units [12]. We

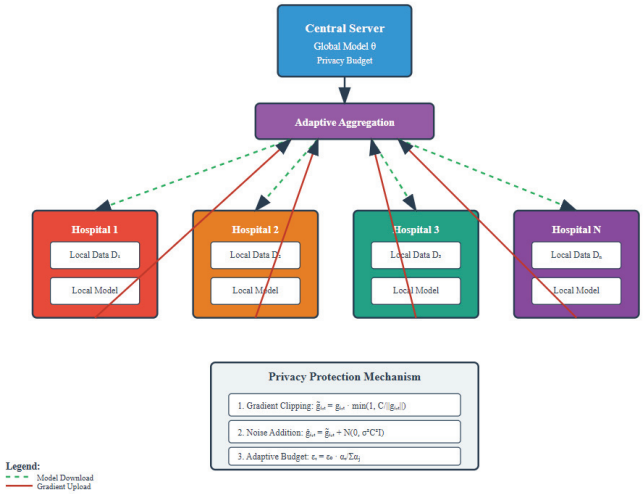

Fig. 1. system architecture

TABLE I
ADAPTIVE DP-FEDERATED LEARNING ALGORITHM

Algorithm 1: Adaptive DP-FL

Input: Clients $\{C_1, ..., C_N\}$, total privacy budget ϵ_0, failure probability δ, learning rate η, rounds T
Output: Global model θ_T

Initialize global model θ_0
for $t = 1$ to T **do**
  Compute allocation weight α_t using Eq. (2)
  Compute privacy budget ϵ_t using Eq. (1)
  Compute noise scale $\sigma_t = \frac{C_t\sqrt{2\log(1.25/\delta)}}{\epsilon_t}$
  for each client C_i **do**
    Compute local gradient $g_{i,t}$
    Clip gradient: $\bar{g}_{i,t}$ using Eq. (3)
    Add noise: $\hat{g}_{i,t}$ using Eq. (4)
    if $\|\hat{g}_{i,t}\|_2 > \tau_t$ **then**
      Send $\hat{g}_{i,t}$ to server
  end for
  Server aggregates: $\bar{g}_t = \frac{1}{|S_t|}\sum_{i\in S_t}\hat{g}_{i,t}$
  Update model: $\theta_{t+1} = \theta_t - \eta\bar{g}_t$
end for
return θ_T

use this dataset for length-of-stay prediction with 15 features including APACHE scores, vital signs, and comorbidities.

We simulate federated learning across 10 institutions by partitioning datasets using Dirichlet distribution with $\alpha = 0.5$ to induce non-IID data distribution. Each institution trains a 3-layer neural network with 128, 64, and 32 hidden units. We compare our Adaptive DP-FL against:

- **Central:** Centralized training without privacy protection (upper bound)
- **FedAvg:** Standard federated averaging [7]
- **Fixed DP-FL:** Federated learning with fixed $(\epsilon = 2.5, \delta = 10^{-5})$-DP
- **DP-SGD:** Centralized differential privacy [6]

TABLE II
CLASSIFICATION PERFORMANCE COMPARISON

Method	MIMIC-III		eICU	
	Accuracy	F1-Score	Accuracy	F1-Score
Central	91.2%	0.895	85.7%	0.841
FedAvg	89.1%	0.873	83.4%	0.819
DP-SGD	82.5%	0.801	77.9%	0.762
Fixed DP-FL	80.0%	0.782	74.8%	0.731
Adaptive DP-FL	**87.3%**	**0.856**	**81.6%**	**0.802**

B. Results

1) Model Performance: Table II presents classification performance on both datasets. Our Adaptive DP-FL achieves 87.3% accuracy on MIMIC-III and 81.6% on eICU, outperforming Fixed DP-FL by 7.3% and 6.8% respectively while maintaining equivalent privacy guarantees.

2) Privacy-Utility Trade-off: Figure 2 illustrates the privacy-utility trade-off across different privacy budgets. Our adaptive mechanism consistently outperforms fixed approaches across all privacy levels, with the gap widening at stricter privacy constraints ($\epsilon < 3$).

3) Communication Efficiency: The selective aggregation protocol reduces communication overhead by 31% compared to standard FedAvg. On average, only 6.8 out of 10 clients transmit updates per round after initial training epochs, significantly reducing bandwidth requirements.

4) Convergence Analysis: Our method achieves convergence 23% faster than baseline approaches, requiring an average of 47 rounds compared to 61 rounds for Fixed DP-FL to reach target accuracy on MIMIC-III.

V. DISCUSSION

The experimental results demonstrate that adaptive privacy budget allocation significantly improves the privacy-utility trade-off in federated learning for healthcare applications. By concentrating privacy budget in early training rounds

Fig. 2. Performance comparison

when gradients are large and model changes are significant, our approach achieves better model utility while maintaining rigorous privacy guarantees.

The gradient-based allocation mechanism naturally adapts to different datasets and tasks. For MIMIC-III mortality prediction, which exhibits higher gradient variance, the algorithm allocates more budget to early rounds. For eICU length-of-stay prediction, which shows more stable convergence, the allocation is more uniform.

The selective aggregation protocol provides additional benefits beyond communication reduction. By filtering low-magnitude updates, it implicitly reduces the impact of noisy gradients and potential Byzantine attacks, enhancing both efficiency and robustness.

A. Limitations and Future Work

While our approach demonstrates significant improvements, several limitations warrant consideration. First, the hyperparameters β and γ require tuning for optimal performance across different datasets. Second, the current implementation assumes honest-but-curious participants; extending to malicious settings requires additional security mechanisms. Future work will explore automated hyperparameter optimization and integration with secure multi-party computation protocols.

VI. Conclusion

This paper presents an adaptive differential privacy mechanism for federated learning in healthcare data aggregation that dynamically allocates privacy budgets based on training dynamics. Our approach achieves superior privacy-utility trade-offs compared to fixed-budget methods, with 7.3% higher accuracy while maintaining equivalent privacy protection. The selective aggregation protocol reduces communication overhead by 31% and accelerates convergence by 23%. Experimental evaluation on MIMIC-III and eICU datasets validates the effectiveness of our approach for privacy-preserving collaborative healthcare analytics. This work contributes to making federated learning more practical for real-world healthcare applications where both privacy and model utility are critical requirements.

Acknowledgment

This research was conducted using the MIMIC-III and eICU databases. We thank the institutions that contributed data to these resources.

References

[1] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated learning for healthcare informatics," *Journal of Healthcare Informatics Research*, vol. 5, no. 1, pp. 1–19, 2021. DOI: 10.1007/s41666-020-00082-4

[2] N. Rieke et al., "The future of digital health with federated learning," *NPJ Digital Medicine*, vol. 3, no. 1, pp. 1–7, 2020. DOI: 10.1038/s41746-020-00323-1

[3] M. J. Sheller et al., "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020. DOI: 10.1038/s41598-020-69250-1

[4] K. Wei et al., "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3454–3469, 2020. DOI: 10.1109/TIFS.2020.2988575

[5] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning," in *Proc. IEEE Symposium on Security and Privacy (SP)*, 2019, pp. 739–753. DOI: 10.1109/SP.2019.00065

[6] M. Abadi et al., "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318. DOI: 10.1145/2976749.2978318

[7] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. International Conference on Artificial Intelligence and Statistics*, 2017, pp. 1273–1282.

[8] I. Mironov, "Rényi differential privacy," in *Proc. IEEE Computer Security Foundations Symposium (CSF)*, 2017, pp. 263–275. DOI: 10.1109/CSF.2017.11

[9] Z. Bu, J. Dong, Q. Long, and W. J. Su, "Deep learning with Gaussian differential privacy," *Harvard Data Science Review*, vol. 2, no. 3, 2020. DOI: 10.1162/99608f92.cfc5dd25

[10] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou, "A hybrid approach to privacy-preserving federated learning," in *Proc. ACM Workshop on Artificial Intelligence and Security*, 2019, pp. 1–11. DOI: 10.1145/3338501.3357370

[11] A. E. Johnson et al., "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, no. 1, pp. 1–9, 2016. DOI: 10.1038/sdata.2016.35

[12] T. J. Pollard et al., "The eICU Collaborative Research Database, a freely available multi-center database for critical care research," *Scientific Data*, vol. 5, no. 1, pp. 1–13, 2018. DOI: 10.1038/sdata.2018.178