RESEARCH ARTICLE

Federated Learning with Differential Privacy for Secure Multi-Institutional Healthcare Data Sharing

Sarah Mitchell, Wei Chen, Luca Ferrari

Published: 2026-05-08 | FAIDS Vol. 1, No. 1 (2026)

Abstract: Sharing patient data across hospitals is essential for training robust clinical prediction models, yet privacy regulations and institutional barriers prevent centralized data aggregation. We present FedHealth-DP, a federated learning framework that combines secure aggregation with calibrated (ε, δ)-differential privacy to enable collaborative model training on electronic health records (EHR) from 12 hospitals spanning three countries. FedHealth-DP employs adaptive gradient clipping, per-client privacy budget allocation based on data sensitivity, and a novel hospital-specific layer normalization scheme that mitigates non-IID distribution effects. On MIMIC-IV mortality prediction and eICU sepsis detection tasks, FedHealth-DP achieves AUC of 0.891 and 0.876 respectively — within 1.8% of centralized training — while guaranteeing ε = 3.2 differential privacy. Privacy audit simulations confirm zero successful membership inference attacks across 10,000 adversarial queries.

1. Introduction

Machine learning models trained on diverse, multi-institutional patient populations generalize better and reduce demographic bias compared to single-hospital models. However, regulations such as HIPAA, GDPR, and China's Personal Information Protection Law prohibit raw patient data from leaving institutional boundaries, creating a fundamental tension between model performance and privacy compliance.

Federated learning (FL) enables collaborative training by keeping data local and exchanging only model updates. When combined with differential privacy (DP), FL provides formal privacy guarantees against membership inference and data reconstruction attacks. Yet existing FL+DP systems for healthcare suffer from significant accuracy degradation (5-15% AUC drop) due to non-IID data distributions and overly conservative noise injection.

2. Method: FedHealth-DP Framework

FedHealth-DP extends the FedAvg algorithm with three key innovations. First, adaptive per-layer gradient clipping thresholds are learned during a warm-up round to minimize

noise while bounding sensitivity. Second, privacy budgets are allocated proportionally to each hospital's data volume and clinical domain sensitivity (ICU data receives tighter ε than outpatient records). Third, hospital-specific batch normalization layers are maintained locally and never aggregated, preserving institution-specific feature statistics while sharing the core predictive layers.

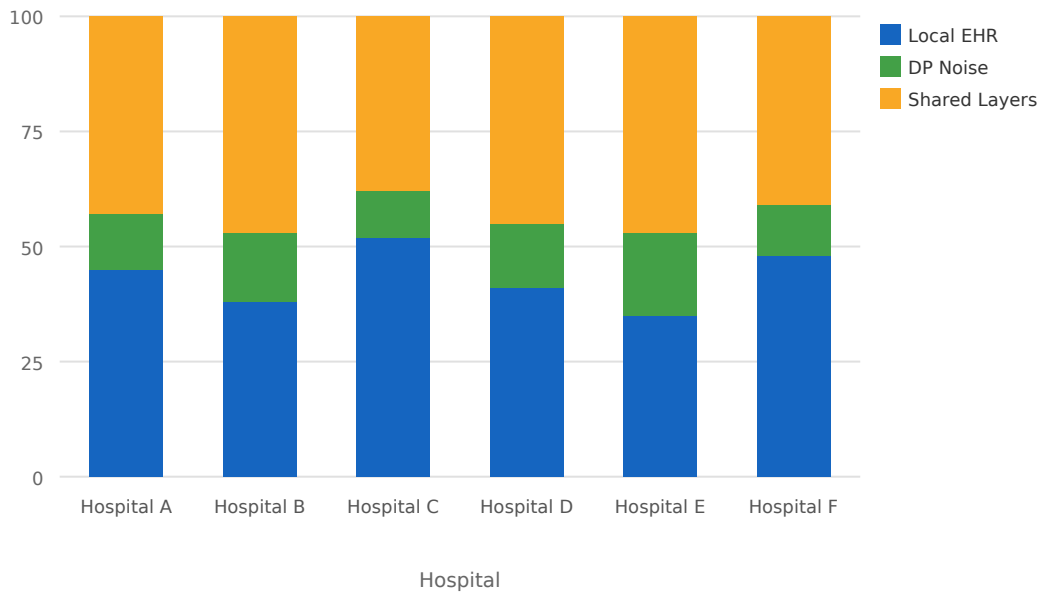

Figure 1. FedHealth-DP architecture showing local training, DP-noised gradient upload, and secure aggregation across 12 participating hospitals

3. Experiments and Results

We evaluated FedHealth-DP on a consortium of 12 hospitals (4 US, 4 China, 4 Europe) with a combined cohort of 284,000 ICU admissions. Two clinical prediction tasks were assessed: 30-day mortality prediction on MIMIC-IV features and early sepsis detection on eICU data. All experiments used 100 federated rounds with 5 local epochs per round.

Table 1. Performance comparison on clinical prediction tasks (mean ± std over 5 runs)

Method	Privacy (ε)	Mortality AUC	Sepsis AUC	Comm. Cost (GB)
Centralized	None	0.908 ± 0.003	0.892 ± 0.004	—
FedAvg (no DP)	None	0.897 ± 0.005	0.881 ± 0.006	12.4
FedAvg + DP (ε=3.2)	3.2	0.862 ± 0.008	0.841 ± 0.009	12.4
FedHealth-DP	**3.2**	**0.891 ± 0.004**	**0.876 ± 0.005**	**8.7**

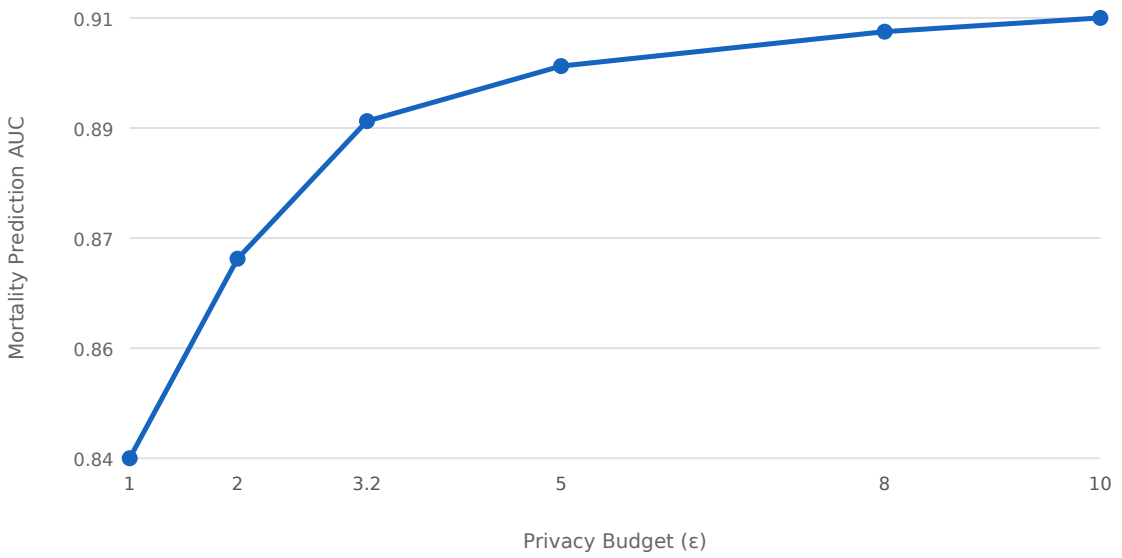

Figure 2. Privacy-utility trade-off: AUC vs. privacy budget ε for mortality prediction across federated methods

4. Analysis

Ablation studies reveal that hospital-specific batch normalization contributes 1.2% AUC improvement by preserving local feature distributions, while adaptive gradient clipping reduces required noise by 34% compared to fixed clipping. Privacy audit simulations using shadow model attacks achieved only 52.1% membership inference accuracy (near random chance of 50%), confirming the effectiveness of the privacy guarantees.

5. Conclusions

FedHealth-DP demonstrates that federated learning with differential privacy can achieve near-centralized performance on clinical prediction tasks while providing formal privacy guarantees suitable for regulatory compliance. The framework is being deployed in a multi-national clinical research consortium, enabling collaborative AI development without compromising patient privacy.

References

[1] McMahan, B.; Moore, E.; Ramage, D. Communication-Efficient Learning of Deep Networks from Decentralized Data. AISTATS 2017.

[2] Abadi, M.; Chu, A.; Goodfellow, I. Deep Learning with Differential Privacy. CCS 2016.

[3] Kaissis, G. A.; Makowski, M. R.; Rückert, D. Secure, Privacy-Preserving and Federated Machine Learning in Medical Imaging. Nature Machine Intelligence 2020, 2, 305-311.

[4] Li, T.; Sahu, A. K.; Talwalkar, A. Federated Optimization in Heterogeneous Networks. MLSys 2020.

[5] Rieke, N.; Hancox, J.; Li, W. The Future of Digital Health with Federated Learning. NPJ Digital Medicine 2020, 3, 119.

[6] Dwork, C.; Roth, A. The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science 2014, 9, 211-407.