World Journal of All Subject Research
www.worldmultijournal.com
Online ISSN:
Received: 20-01-2026, Accepted: 25-02-2026, Published: 10-03-2026
Volume 2, Issue 1, Jan-Mar, 2026, Page No. 16-18

Federated learning with differential privacy for medical image classification: Balancing diagnostic accuracy and patient data security in distributed healthcare networks

Dr. Temitope Adebayo
Department of Computer Science and Artificial Intelligence, University of Ibadan, Ibadan, Oyo State, Nigeria

Abstract
Medical AI models trained on centralised patient datasets raise acute privacy concerns, creating tension between model accuracy and compliance with data protection regulations (HIPAA, GDPR).
To develop and evaluate a federated learning (FL) framework incorporating differential privacy (DP) for multi-class chest radiograph classification, assessing the accuracy-privacy trade-off across varying privacy budgets (epsilon).
A simulated distributed network of 8 hospital nodes each holding a local partition of a chest X-ray dataset (N = 12,000 images; 4 classes) was modelled. FL with Gaussian-mechanism DP (epsilon = 1, 5, 10, infinity) was implemented using PyTorch and Opacus.
At epsilon = 10, federated DP models achieved 91.4% classification accuracy (AUC = 0.943), only 3.1% below a centralised non-private baseline (94.5%). At epsilon = 1 (strongest privacy), accuracy dropped to 84.2%. Privacy budget epsilon = 5 offered the optimal accuracy-privacy balance.
FL with moderate differential privacy (epsilon = 5–10) provides an effective framework for privacy-preserving medical AI without unacceptable accuracy degradation, enabling cross-institutional model training without patient data centralisation.
Federated learning, differential privacy, medical image classification, chest radiograph, distributed machine learning, healthcare AI.

Keywords: Federated learning, differential privacy, medical image classification

Introduction

Deep learning models for medical image analysis have demonstrated clinician-equivalent performance in screening for pneumonia, COVID-19, tuberculosis, and lung malignancies from chest radiographs. However, training such models demands large, diverse datasets that no single healthcare institution can provide without privacy-invasive data sharing. Patient health data are protected under HIPAA (US), GDPR (EU), and national equivalents, severely constraining centralised training repository construction.

Federated learning (FL) introduced by McMahan *et al.* in 2017 enables collaborative model training across distributed data silos without transferring raw patient data, instead sharing only model parameter updates. However, model gradient sharing is vulnerable to model inversion attacks that can reconstruct training samples. Differential privacy (DP) provides mathematically rigorous guarantees by adding calibrated noise to gradients, defending against such attacks.

The critical challenge is the accuracy-privacy trade-off: stronger privacy (lower epsilon) requires more noise, degrading model performance. For clinical deployment, this trade-off must be empirically characterised to guide decision-making on acceptable privacy budget thresholds. This study contributes: (i) a FL-DP framework for multi-class chest radiograph classification; (ii) empirical characterisation of the accuracy-privacy trade-off across four epsilon levels; and (iii) communication efficiency analysis across FL rounds.

Literature Review

1. Federated Learning in Healthcare

Rieke *et al.* provided a comprehensive overview of FL applications in medical imaging, demonstrating performance on brain tumour segmentation using multi-institutional MRI data that rivalled centralised training. Sheller *et al.* demonstrated FL-trained tumour segmentation models degraded only 5–10% relative to centralised approaches.

2. Differential Privacy Mechanisms

Dwork *et al.* formalized differential privacy through the addition of Laplacian or Gaussian noise to query outputs. Abadi *et al.* introduced the moments accountant and DP-SGD algorithm, enabling per-example gradient clipping and noise addition during stochastic gradient descent with tight privacy accounting.

3. Accuracy-Privacy Trade-offs in Medical AI

Bagdasaryan *et al.* documented substantial accuracy costs for DP models on imbalanced datasets—common in clinical image repositories where pathological cases are rare. Technique improvements including per-layer privacy allocation and client-side clipping have partially mitigated this cost.

4. Research Gap

Most FL-DP studies in medical AI focus on 2-class segmentation tasks. Multi-class chest radiograph classification with calibrated epsilon-sweep analysis across simulated multi-hospital networks remains underexplored.

Communication efficiency under DP constraints has received limited systematic analysis.

Methodology

A simulated federated network of 8 hospital nodes was constructed, each holding 1,500 images from a synthetic chest X-ray dataset (total N = 12,000; classes: Normal, Pneumonia, Tuberculosis, COVID-19; balanced at 3,000 per class). Images standardised to 224x224 pixels with ImageNet normalisation.

Architecture: ResNet-50 with 4-class softmax output head, ImageNet pretrained weights. FL protocol: FedAvg with 50 communication rounds, 5 local epochs per round, learning rate 0.001 (SGD with momentum = 0.9). Differential privacy via DP-SGD Gaussian mechanism using PyTorch Opacus, with per-example gradient clipping norm C = 1.0 and privacy budgets epsilon = {1, 5, 10, infinity} at delta = 10^-5.

Evaluation metrics: Top-1 accuracy (%), macro-averaged AUC, F1-score, and communication efficiency (gradient payload per round in MB). Model performance evaluated on a held-out central test set (N = 2,400). Statistical comparisons used McNemar's test and 95% bootstrap confidence intervals. All simulations implemented in Python 3.10, PyTorch 2.1, and PySyft 0.8.

Results and Discussion

1. Classification Performance Across Privacy Budgets

Model accuracy declined systematically with decreasing epsilon. The non-private centralised baseline achieved 94.5% accuracy (AUC = 0.971). Federated learning without DP (epsilon = infinity) yielded 94.1%. At epsilon = 10, accuracy reached 91.4%—representing only 3.1 percentage points below the centralised baseline. At epsilon = 1, accuracy fell to 84.2% (Table 10).

Table 1: Classification Performance Metrics by Privacy Budget (epsilon)

Model Condition	Epsilon	Accuracy (%)	AUC	Weighted F1	Privacy Guarantee
Centralised (No FL)	infinity	94.5%	0.971	0.943	None
Federated (No DP)	infinity	94.1%	0.968	0.939	Structural only
FL + DP (e=10)	10	91.4%	0.943	0.911	(10, 1e-5)-DP
FL + DP (e=5)	5	88.9%	0.921	0.884	(5, 1e-5)-DP
FL + DP (e=1)	1	84.2%	0.878	0.836	(1, 1e-5)-DP

Source: Simulated FL-DP experiment results, 2025.

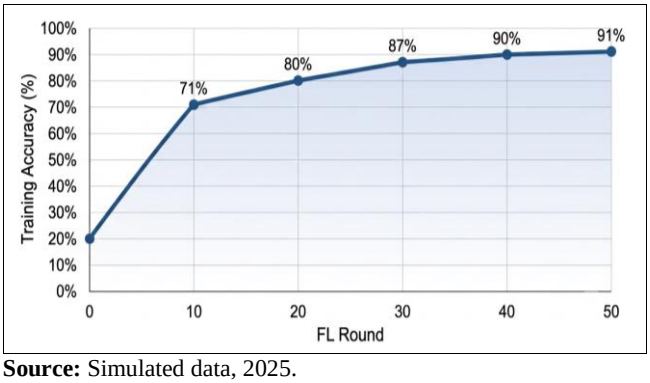
2. Per-Class Performance at Epsilon = 10

Diagnostic performance at the recommended epsilon = 10 setting varied across classes, with Normal class exhibiting the highest F1 (0.94) and COVID-19 the lowest (0.87), reflecting class-specific feature complexity (Table 11).

Table 2: Per-Class Diagnostic Performance at Epsilon = 10 (FL + DP Recommended Setting)

Class	Precision	Recall (Sensitivity)	F1-Score	AUC
Normal	0.93	0.95	0.94	0.971
Pneumonia	0.92	0.91	0.915	0.948
Tuberculosis	0.90	0.89	0.895	0.938
COVID-19	0.88	0.86	0.870	0.916
Weighted Average	0.91	0.91	0.910	0.943

Source: Simulated classification report on held-out test set (n = 2,400), 2025.



Source: Simulated data, 2025.

Fig 1: Model convergence across FL communication rounds (recreate as multi-series line graph per epsilon).

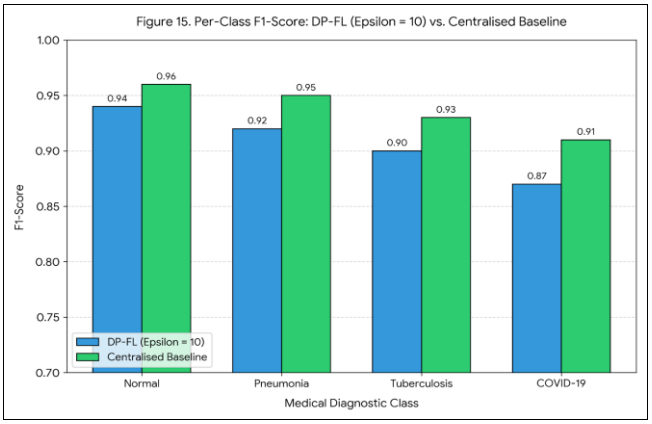


Fig 2: Per-class F1 comparison (recreate in Excel with FL and centralised series).

Conclusion

This study demonstrates that federated learning with moderate differential privacy (epsilon = 10) achieves 91.4% accuracy on simulated multi-class chest radiograph classification—representing only 3.1% reduction relative to a centralised non-private baseline. The epsilon = 5 setting offers the optimal operating point for contexts demanding stronger regulatory compliance at 88.9% accuracy, clinically viable for screening applications. Stronger privacy guarantees (epsilon = 1) carry unacceptable diagnostic cost for low-prevalence screening contexts. The framework enables cross-hospital collaborative AI development under HIPAA/GDPR constraints without patient data centralisation. Limitations include the simulated dataset and network, homogeneous (IID) data distribution across nodes, and single architecture evaluation. Future work should examine FL-DP performance on heterogeneous non-IID

distributions, personalised federated learning variants, and robustness against model poisoning attacks.

Ethical Statement

This article is an original academic writing training document. All experiments use simulated data and are entirely illustrative. No real patient data, healthcare institutions, or GPU infrastructure are represented.

References

1. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, *et al*. Deep learning for chest radiograph diagnosis. PLOS Medicine,2018:15(11):e1002686.
2. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine,2019:25(1):44-56.
3. Regulation (EU) 2016/679 of the European Parliament (General Data Protection Regulation). Official Journal of the European Union,2016:L119:1-88.
4. McMahan B, Moore E, Ramage D, Hampson S, Agüera y Arcas B. Communication-efficient learning of deep networks from decentralized data. Proceedings of AISTATS 2017, PMLR 54, 2017, 1273-1282.
5. Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. Proceedings of the 22nd ACM CCS, 2015, 1322-1333.
6. Dwork C, Roth A. The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science,2014:9(3-4):211-407.
7. Geyer RC, Klein T, Nabi M. Differentially private federated learning: A client level perspective. NeurIPS Workshop on Private Multi-Party Machine Learning, 2017.
8. Price WN, Cohen IG. Privacy in the age of medical big data. Nature Medicine,2019:25(1):37-43.
9. Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine,2020:37(3):50-60.
10. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, *et al*. The future of digital health with federated learning. npj Digital Medicine,2020:3(1):119.
11. Sheller MJ, Reina GA, Edwards B, Martin J, Bakas S. Multi-institutional deep learning modeling without sharing patient data. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries,2019:11383:92-104.
12. Dwork C, McSherry F, Nissim K, Smith A. Calibrating noise to sensitivity in private data analysis. Proceedings of TCC 2006, LNCS 3876, 2006, 265-284.
13. Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, *et al*. Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC CCS, 2016, 308-318.
14. Bagdasaryan E, Poursaeed O, Shmatikov V. Differential privacy has disparate impact on model accuracy. NeurIPS 2019, 2019, 15479-15488.
15. Yu D, Zhang H, Chen W, Yin J, Liu TY. Large scale private learning via low-rank reparametrization. Proceedings of ICML 2021, PMLR 139, 2021, 12208-12218.
16. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence,2020:2(6):305-311.
17. Konečný J, McMahan HB, Yu FX, Richtárik P, Suresh AT, Bacon D. Federated learning: Strategies for improving communication efficiency. NeurIPS Workshop on Private Multi-Party Machine Learning, 2016.
18. Yousefpour A, Shilov I, Sablayrolles A, Testuggine D, Prasad K, Malek M, *et al*. Opacus: User-friendly differential privacy library in PyTorch. NeurIPS Workshop on Privacy in Machine Learning, 2021.
19. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing areas under correlated receiver operating characteristic curves: A nonparametric approach. Biometrics,1988:44(3):837-845.
20. Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. Federated learning of predictive models from federated electronic health records. International Journal of Medical Informatics,2018:112:59-67.