Towards Privacy-Preserving Medical Imaging: Federated Learning with Differential Privacy and Secure Aggregation Using a Modified ResNet Architecture

Mohamad Haj Fares
Department of Computer Engineering
Istanbul University-Cerrahpasa
Istanbul, Türkiye
ai.mohres@gmail.com

Ahmed Mohamed Saad Emam Saad
School of Computing
Queen's University
Kingston, ON K7L 3N6, Canada
a.saad@queensu.ca

Abstract

With increasing concerns over privacy in healthcare, especially for sensitive medical data, this research introduces a federated learning framework that combines local differential privacy and secure aggregation using Secure Multi-Party Computation for medical image classification. Further, we propose DPResNet, a modified ResNet architecture optimized for differential privacy. Leveraging the BloodMNIST benchmark dataset, we simulate a realistic data-sharing environment across different hospitals, addressing the distinct privacy challenges posed by federated healthcare data. Experimental results indicate that our privacy-preserving federated model achieves accuracy levels close to non-private models, surpassing traditional approaches while maintaining strict data confidentiality. By enhancing the privacy, efficiency, and reliability of healthcare data management, our approach offers substantial benefits to patients, healthcare providers, and the broader healthcare ecosystem.

1 Introduction

Deep learning for medical imaging faces significant challenges in leveraging advancements from other fields of computer vision due to privacy concerns and limited data access [2, 24]. Privacy-enhancing technologies, such as Federated Learning (FL) and Differential Privacy (DP), provide viable solutions by enabling insights from sensitive data while safeguarding individual privacy [2, 19]. FL enables decentralized model training directly on user devices or within hospital systems, removing the need for central data aggregation and thereby reducing the risk of data breaches [15, 12]. In this setup, data remains stored locally, while a central server coordinates training by exchanging model updates, ensuring data confidentiality throughout the process.

DP further strengthens privacy by introducing controlled noise to data, which masks specific information while preserving model accuracy [7, 6, 5, 9]. This method includes gradient clipping to limit individual data contributions [3], followed by the addition of noise proportional to data sensitivity, making it challenging for adversaries to reconstruct individual data points, even if they gain access to the trained model. To enhance security in FL further, Secure Multi-Party Computation (SMPC) [29, 20] is employed for Secure Aggregation (SecAgg), ensuring that individual model updates remain private from both the central server and any unauthorized parties during transmission.

In this paper, we introduce a privacy-preserving FL framework that combines the Federated Averaging (FedAvg) algorithm [18] with local DP, using fixed gradient clipping, alongside SMPC-based SecAgg.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

Additionally, we propose DPResNet, a modified ResNet-9 architecture optimized for differential privacy within federated settings, further enhancing model performance under privacy constraints. Applied to the BloodMNIST datasets, this approach simulates realistic data-sharing environments in medical imaging. Our results show that this framework achieves high accuracy while adhering to strict privacy standards, enabling secure and effective deployment of Deep Neural Networks (DNNs) in healthcare applications.

Contributions Our main contributions are as follows:

- We develop a privacy-preserving FL framework for medical imaging, integrating the FedAvg algorithm with local DP, including gradient clipping, and SMPC-based SecAgg to protect data confidentiality during model training and aggregation.
- We propose DPResNet, a modified ResNet architecture tailored for differential privacy in federated settings to enhance compatibility with privacy-preserving protocols.
- We apply this framework to the BloodMNIST dataset, creating a realistic simulation of multi-institutional data-sharing scenarios across decentralized healthcare environments.
- We evaluate the performance of our framework, demonstrating that it achieves competitive accuracy levels while maintaining strong privacy guarantees, close to those of non-private models.

2 Related Work

DP and FL are foundational for privacy-preserving medical imaging. DP-Stochastic Gradient Descent (DP-SGD) [1] introduced gradient clipping and noise addition to protect data privacy, with subsequent adaptations improving scalability and refining privacy-accuracy trade-offs [30, 22]. Combined with FL, DP enables near-centralized performance without raw data aggregation in multi-institutional setups [24, 10]. However, model update transmission remains vulnerable to data leakage, which SMPC [29, 20] addresses by preserving privacy during aggregation, even when DP alone may fall short [14, 18].

Personalized FL frameworks tackle non-IID data heterogeneity by customizing model weights per client [31, 17]. Approaches like FedProx [13] use regularization to handle data variability, informing cross-device FL approaches [27]. Domain-specific adaptations of DP for medical fields, including histopathology and brain imaging, often employ Gaussian noise but lack sensitivity clipping, affecting privacy guarantees [16, 26]. Methods such as transformer-based pre-training and domain adaptation improve FL performance with non-IID data distributions [28, 14].

FL optimization for resource-constrained edge devices involves methods like FedSup [32], incorporating neural architecture search and Bayesian Convolutional Neural Networks (BCNN) for uncertainty, and asynchronous aggregation to reduce communication costs [11, 15].

Despite these advancements, privacy vulnerabilities remain, with adversarial attacks, as demonstrated by MediLeak [25], and Generative Adversarial Networks (GANs) capable of reconstructing private data, underscoring the need for robust defenses [8, 21]. Although promising, FL with DP and SMPC faces challenges like privacy-accuracy trade-offs and a lack of standardized datasets, complicating reproducibility.

While FL with DP and SMPC-based aggregation shows promise, challenges like the privacy-accuracy trade-off and the scarcity of standardized datasets hinder reproducibility and benchmarking. Our study addresses these gaps by applying FL with DP and SMPC to the MedMNIST dataset, exploring privacy and accuracy trade-offs in privacy-preserving medical imaging.

3 Methodology

3.1 Secure Federated Learning Framework

Our framework combines FL with DP and Secure SMPC for privacy-preserving medical image classification. The FL objective is to train a model $f_\theta : \mathcal{X} \to \mathcal{Y}$ with parameters θ using distributed

datasets $\mathcal{D} = \{\mathcal{D}_i\}_{i=1}^N$ across N hospitals or devices, where data $\mathcal{D}_i$ remains local. The federated optimization problem is defined as:

$$\min_{\theta} \sum_{i=1}^{N} p_i \cdot \mathcal{L}(\mathcal{D}_i; \theta), \tag{1}$$

where p_i is the weight of client i and $\mathcal{L}(\mathcal{D}_i; \theta)$ is the loss on the local data $\mathcal{D}_i$. Each client trains locally and transmits updates $\Delta\theta_i$ to the server for SecAgg.

Privacy Mechanisms To ensure privacy, clients apply gradient clipping to limit the sensitivity of updates. The clipped gradient is:

$$g_i^{\text{clipped}} = \frac{g_i}{\max\left(1, \frac{\|g_i\|}{C}\right)}, \tag{2}$$

where C is the clipping norm. After clipping, Gaussian noise calibrated to an (ϵ, δ)-DP guarantee is added:

$$\Delta\theta_i^{\text{DP}} = g_i^{\text{clipped}} + \mathcal{N}(0, \sigma^2 I), \tag{3}$$

where σ is the noise scale. These mechanisms ensure that individual contributions remain obfuscated in the transmitted updates.

Secure Aggregation To protect model updates during aggregation, SMPC aggregates the updates securely without exposing individual contributions. The aggregated model update is:

$$\theta_{t+1} = \text{SecAgg}\left(\{\Delta\theta_i^{\text{DP}}\}_{i=1}^N\right) = \sum_{i=1}^{N} p_i \cdot \Delta\theta_i^{\text{DP}}, \tag{4}$$

where $\text{SecAgg}(\cdot)$ denotes the Secure Aggregation operation. This protocol prevents an honest-but-curious server from accessing individual updates while enabling global model updates.

Modified ResNet Architecture Our model, DPResNet, adapts ResNet-9 by replacing Batch-Normalization with GroupNormalization (32 groups per layer) and removing max-pooling layers. GroupNormalization computes statistics over grouped channels, making it compatible with DP requirements while maintaining simplicity and effectiveness.

Overall Objective The overall objective integrates FL, DP, SMPC-based SecAgg, and the DPRes-Net architecture. This can be expressed as:

$$\min_{\theta'} \sum_{i=1}^{N} p_i \cdot \mathbb{E}\left[\mathcal{L}(\mathcal{D}_i; \theta')\right] \quad \text{subject to } (\epsilon, \delta)\text{-DP constraints.} \tag{5}$$

Equations (1) through (5) encapsulate the integration of the proposed secure FL framework.

3.2 Workflow and Experimental Setup

To illustrate the workflow of our privacy-preserving framework, Figure 1 presents the interaction between hospitals during the training process. Each hospital i trains a local model f_{θ_i} using a differentially private approach, ensuring that sensitive data $\mathcal{D}_i$ remains protected.

Locally trained updates $\Delta\theta_i^{\text{DP}}$ (Eq. 3) are aggregated securely via the SecAgg+ protocol [4], ensuring confidentiality of individual contributions, even with client dropouts. The aggregated model is computed as shown in Eq. 4 and redistributed for subsequent rounds. This iterative process continues until convergence, producing a global model ready for inference while preserving privacy.

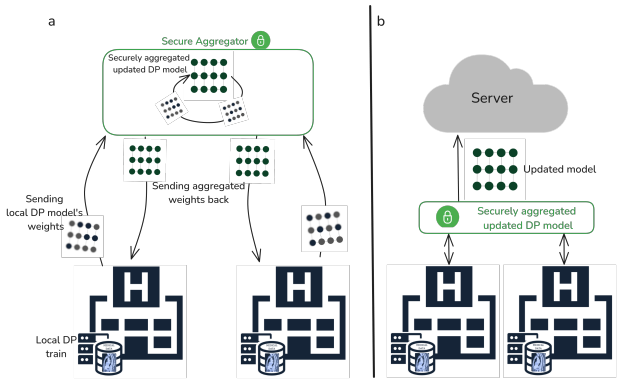

Figure 1: Illustration of the FL workflow. (a) Models are trained locally with differential privacy and aggregated securely using SecAgg+. (b) The final aggregated model (green) is updated after multiple rounds.

The framework is evaluated on the BloodMNIST dataset, partitioned non-IID across $N \in \{5, 10\}$ clients to simulate real-world data heterogeneity. Each client receives between $0.45 \times \overline{n}$ and $0.55 \times \overline{n}$ samples, where $\overline{n}$ is the average sample count per client, mimicking imbalanced federated settings.

Federated training employs the FedAvg algorithm for aggregation, with each client performing $E = 3$ local epochs per round. Privacy parameters are $(\epsilon = 6.0, \delta = 1.9 \times 10^{-4})$, and gradients are clipped with a norm $C = 7$ (Eq. 2) to balance accuracy and privacy trade-offs. The SecAgg+ protocol uses a reconstruction threshold of four shares to ensure robustness against client dropouts. Together, these components form a secure and efficient framework for FL in medical imaging.

4 Experiments and Results

We evaluated our privacy-preserving FL framework on the BloodMNIST dataset under three configurations: without DP or SecAgg (DP-/SecAgg-), with only SecAgg enabled (DP-/SecAgg+), and with both techniques enabled (DP+/SecAgg+). Table 1 compares our results to PriMIA [10], the pioneering work in this domain, and FEDMIC [23], the current state-of-the-art.

Our experiments involved training FL models for 50 global rounds with non-IID data distributions across 10 and 20 clients to simulate real-world heterogeneity. In each configuration, local updates were secured using a fixed clipping norm ($C = 7$) and privacy budget parameters ($\epsilon = 6.0, \delta = 1.9 \times 10^{-4}$). The SecAgg+ protocol ensured secure model aggregation, protecting individual client updates during training.

Table 1: Accuracy of different configurations on the BloodMNIST dataset for varying client sizes. Results from PriMIA and FEDMIC are included for comparison.

Dataset	Approach	Client Size	DP-/SecAgg-	DP-/SecAgg+	DP+/SecAgg+
BloodMNIST	Ours	10	98.76	98.11	97.78
		20	97.77	97.01	96.89
	PriMIA [10]	10	90.00	89.00	85.00
	FEDMIC [23]	20	–	–	96.33

Conclusion In this work, we present a privacy-preserving federated learning framework that integrates differential privacy and secure aggregation to address privacy challenges in medical imaging. By leveraging a modified ResNet architecture tailored for DP and simulating realistic non-IID data distributions, our approach achieves superior accuracy compared to PriMIA, the pioneering work in this domain, and exceeds FEDMIC, the current state-of-the-art. Our results demonstrate the efficacy of the proposed framework, offering a robust solution for privacy-preserving medical image classification.

References

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, oct 2016. doi: 10.1145/2976749.2978318. URL https://doi.org/10.1145%2F2976749.2978318.

[2] Mohammed Adnan, Shivam Kalra, Jesse C Cresswell, Graham W Taylor, and Hamid R Tizhoosh. Federated learning and differential privacy for medical image analysis. *Scientific reports*, 12(1): 1953, 2022.

[3] Galen Andrew, Om Thakkar, Brendan McMahan, and Swaroop Ramaswamy. Differentially private learning with adaptive clipping. *Advances in Neural Information Processing Systems*, 34:17455–17466, 2021.

[4] James Henry Bell, Kallista A Bonawitz, Adrià Gascón, Tancrède Lepoint, and Mariana Raykova. Secure single-server aggregation with (poly) logarithmic overhead. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, pages 1253–1269, 2020.

[5] Christopher A Choquette-Choo, H Brendan McMahan, Keith Rush, and Abhradeep Thakurta. Multi-epoch matrix factorization mechanisms for private machine learning. *arXiv preprint arXiv:2211.06530*, 2022.

[6] Christopher A Choquette-Choo, Arun Ganesh, Ryan McKenna, H Brendan McMahan, John Rush, Abhradeep Guha Thakurta, and Zheng Xu. (amplified) banded matrix factorization: A unified approach to private training. *Advances in Neural Information Processing Systems*, 36, 2024.

[7] Cynthia Dwork. Differential privacy: A survey of results. In *International conference on theory and applications of models of computation*, pages 1–19. Springer, 2008.

[8] Briland Hitaj, Giuseppe Ateniese, and Fernando Perez-Cruz. Deep models under the gan: information leakage from collaborative deep learning. In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security*, pages 603–618, 2017.

[9] Peter Kairouz, Brendan McMahan, Shuang Song, Om Thakkar, Abhradeep Thakurta, and Zheng Xu. Practical and private (deep) learning without sampling or shuffling. In *International Conference on Machine Learning*, pages 5213–5225. PMLR, 2021.

[10] Georgios Kaissis, Axel Ziller, Jonathan Passerat-Palmbach, Thomas Ryffel, Dmitrii Usynin, Andrew Trask, Ivan Lima, James Mancuso, Florian Jungmann, Michael Steinborn, Ali Saleh, Marcus Makowski, Daniel Rueckert, and Rickmer Braren. End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nature Machine Intelligence*, 3(6): 473–484, 2021. doi: 10.1038/s42256-021-00337-8. URL https://doi.org/10.1038/s42256-021-00337-8.

[11] Taehyeon Kim and Se-Young Yun. Supernet training for federated image classification under system heterogeneity, 2022. URL https://arxiv.org/abs/2206.01366.

[12] Anusha Lalitha, Shubhanshu Shekhar, Tara Javidi, and Farinaz Koushanfar. Fully decentralized federated learning. In *Third workshop on bayesian deep learning (NeurIPS)*, volume 2, 2018.

[13] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks, 2018.

[14] Wenqi Li, Fausto Milletarì, Daguang Xu, Nicola Rieke, Jonny Hancox, Wentao Zhu, Maximilian Baust, Yan Cheng, Sébastien Ourselin, M. Jorge Cardoso, and Andrew Feng. Privacy-preserving federated brain tumour segmentation. In Heung-Il Suk, Mingxia Liu, Pingkun Yan, and Chunfeng Lian, editors, *Machine Learning in Medical Imaging*, pages 133–141, Cham, 2019. Springer International Publishing. ISBN 978-3-030-32692-0. URL https://doi.org/10.1007/978-3-030-32692-0_16.

[15] Wei Yang Bryan Lim, Nguyen Cong Luong, Dinh Thai Hoang, Yutao Jiao, Ying-Chang Liang, Qiang Yang, Dusit Niyato, and Chunyan Miao. Federated learning in mobile edge networks: A comprehensive survey. *IEEE communications surveys & tutorials*, 22(3):2031–2063, 2020.

[16] Ming Y Lu, Richard J Chen, Dehan Kong, Jana Lipkova, Rajendra Singh, Drew FK Williamson, Tiffany Y Chen, and Faisal Mahmood. Federated learning for computational pathology on gigapixel whole slide images. *Medical image analysis*, 76:102298, 2022. doi: 10.1016/j.media.2021.102298. URL https://doi.org/10.1016/j.media.2021.102298.

[17] Jun Luo and Shandong Wu. Adapt to adaptation: Learning personalization for cross-silo federated learning, 2021. URL https://arxiv.org/abs/2110.08394.

[18] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[19] H Brendan McMahan, Galen Andrew, Ulfar Erlingsson, Steve Chien, Ilya Mironov, Nicolas Papernot, and Peter Kairouz. A general approach to adding differential privacy to iterative training procedures. *arXiv preprint arXiv:1812.06210*, 2018.

[20] Payman Mohassel and Yupeng Zhang. Secureml: A system for scalable privacy-preserving machine learning. In *2017 IEEE symposium on security and privacy (SP)*, pages 19–38. IEEE, 2017.

[21] Milad Nasr, Reza Shokri, and Amir Houmansadr. 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 (SP)*, pages 739–753. IEEE, 2019.

[22] Nicolas Papernot, Shuang Song, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Úlfar Erlingsson. Scalable private learning with pate. *arXiv preprint arXiv:1802.08908*, 2018.

[23] Sufen Ren, Yule Hu, Shengchao Chen, and Guanjun Wang. Federated distillation for medical image classification: Towards trustworthy computer-aided diagnosis. *arXiv preprint arXiv:2407.02261*, 2024.

[24] Micah J Sheller, Brandon Edwards, G Anthony Reina, Jason Martin, Sarthak Pati, Aikaterini Kotrotsou, Mikhail Milchenko, Weilin Xu, Daniel Marcus, Rivka R Colen, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific reports*, 10(1):12598, 2020.

[25] Shanghao Shi, Md Shahedul Haque, Abhijeet Parida, Marius George Linguraru, Y. Thomas Hou, Syed Muhammad Anwar, and Wenjing Lou. Harvesting private medical images in federated learning systems with crafted models, 2024. URL https://arxiv.org/abs/2407.09972.

[26] Sara Silva, Andre Altmann, Boris Gutman, and Marco Lorenzi. Fed-BioMed: A General Open-Source Frontend Framework for Federated Learning in Healthcare. In *Lecture Notes in Computer Science*, pages 201–210. 2020. doi: 10.1007/978-3-030-60548-3_20. URL https://doi.org/10.1007/978-3-030-60548-3_20.

[27] Alysa Ziying Tan, Han Yu, Lizhen Cui, and Qiang Yang. Towards personalized federated learning. *IEEE transactions on neural networks and learning systems*, 34(12):9587–9603, 2022.

[28] Rui Yan, Liangqiong Qu, Qingyue Wei, Shih-Cheng Huang, Liyue Shen, Daniel L. Rubin, Lei Xing, and Yuyin Zhou. Label-efficient self-supervised federated learning for tackling data heterogeneity in medical imaging. *IEEE Transactions on Medical Imaging*, 42(7):1932–1943, July 2023. ISSN 1558-254X. doi: 10.1109/tmi.2022.3233574. URL http://dx.doi.org/10.1109/TMI.2022.3233574.

[29] Andrew C Yao. Protocols for secure computations. In *23rd annual symposium on foundations of computer science (sfcs 1982)*, pages 160–164. IEEE, 1982.

[30] Da Yu, Huishuai Zhang, Wei Chen, Jian Yin, and Tie-Yan Liu. Large scale private learning via low-rank reparametrization. In *International Conference on Machine Learning*, pages 12208–12218. PMLR, 2021.

[31] Michael Zhang, Karan Sapra, Sanja Fidler, Serena Yeung, and Jose M. Alvarez. Personalized federated learning with first order model optimization, 2020. URL https://arxiv.org/abs/2012.08565.

[32] Chen Zhao, Zhipeng Gao, Qian Wang, Kaile Xiao, Zijia Mo, and M Jamal Deen. Fedsup: A communication-efficient federated learning fatigue driving behaviors supervision approach. *Future Generation Computer Systems*, 138:52–60, 2023.