ISCSITR - INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN HEALTHCARE INFORMATION SYSTEM (ISCSITR - IJSRHIS)

Vol.3, Iss. 1, January-December, 2022, pp. 1-7.

https://iscsitr.com/index.php/ISCSITR-IJSRHIS

Journal ID: 2351-4421

A Federated Learning Approach to Privacy-Preserving Medical Image Classification Across Distributed Healthcare Systems

Lina Andersson,

Norway.

Abstract

The surge in medical imaging data and the expansion of distributed healthcare systems have emphasized the need for privacy-preserving machine learning solutions. Traditional centralized approaches to training deep learning models pose risks related to data leakage and non-compliance with health data privacy regulations. Federated Learning (FL) has emerged as a powerful paradigm enabling collaborative model training without raw data sharing. This paper presents a federated deep learning architecture for privacy-preserving classification of medical images, particularly across hospital systems with heterogeneous imaging modalities.

We propose a federated convolutional neural network (CNN) framework using federated averaging (FedAvg) and incorporate differential privacy techniques to enhance data protection. Experimental results on benchmark datasets (e.g., BraTS, ChestXray14) demonstrate that FL achieves near-centralized accuracy while maintaining data locality. The study also explores challenges such as data heterogeneity, communication overhead, and defense against adversarial attacks in federated settings.

Keywords: Federated Learning, Medical Image Classification, Privacy Preservation, Distributed Healthcare, Differential Privacy, Convolutional Neural Networks**.**

How to cite this paper: Lina Andersson. (2022) A Federated Learning Approach to Privacy-Preserving Medical Image Classification Across Distributed Healthcare Systems. *ISCSITR - INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN HEALTHCARE INFORMATION SYSTEM (ISCSITR - IJSRHIS)*, 3(1), 1–7.

URL: https://iscsitr.com/index.php/ISCSITR-IJSRHIS/article/view/ISCSITR-IJSRHIS_03_01_001
Published: 24th Jun 2022

1. Introductıon

Medical imaging plays a pivotal role in clinical diagnosis, enabling early detection of diseases ranging from cancer to pulmonary infections. However, the potential of artificial intelligence (AI) in enhancing imaging diagnostics is constrained by the lack of centralized, large-scale annotated datasets. Regulations such as HIPAA and GDPR restrict the pooling of patient data across institutions, creating silos that impede the development of robust AI models. This fragmentation is particularly problematic in deep learning, which relies heavily on data diversity and volume.

Federated Learning (FL) addresses this challenge by enabling decentralized model training across multiple data sources. Hospitals can train models locally and only share encrypted model updates, ensuring sensitive patient data never leaves institutional boundaries. By preserving data locality, FL aligns with privacy regulations while facilitating collaborative intelligence. The objective of this paper is to explore a federated learning framework tailored for privacy-preserving medical image classification, using CNNs and differential privacy mechanisms to ensure both accuracy and compliance.

2. Literature Review

Early applications of machine learning in healthcare relied on centralized datasets, collected from multiple institutions and processed in data centers. While these models offered high performance, they exposed patient data to significant privacy risks and regulatory scrutiny. Traditional encryption techniques provided partial solutions, but they couldn't fully address data access concerns. The concept of FL, introduced by McMahan et al.

(2017), was later adapted to healthcare where decentralization is not only practical but often legally necessary.

Federated Learning has since been studied extensively in medical imaging. Sheller et al. (2020) demonstrated federated CNNs for brain tumor segmentation, showing results comparable to centralized training. The use of frameworks like FedAvg and FedProx allowed learning on non-IID (non-independent and identically distributed) data, common in hospital settings. However, challenges remain. Communication overhead, model divergence due to data heterogeneity, and risks from model inversion attacks have limited real-world deployment. Recent studies have proposed integrating privacy techniques like differential privacy and secure multiparty computation (SMPC) to bolster the robustness of FL systems in healthcare.

3. Methodology

Our approach is built on the cross-silo federated learning architecture, where multiple hospitals (clients) collaboratively train a CNN model without exchanging patient data. The local models are trained independently using private medical images and only model updates (gradients or weights) are shared with a central server. We used two datasets: ChestXray14 for lung disease classification and BraTS for brain tumor segmentation. These datasets provide both diversity in imaging modality and structure for testing the model's generalizability.

To ensure privacy, we implemented differential privacy by injecting Gaussian noise into the gradient updates before transmission. This obscures individual data contributions and prevents reconstruction attacks. The model was optimized using Adam with a dynamic learning rate schedule. We simulated five hospital nodes with varying data distributions to replicate real-world variability. Performance metrics included accuracy, F1-score, Dice coefficient, and privacy loss (ε). Communication cost and training latency were also recorded to assess scalability and feasibility.

4. System Architecture or Proposed Framework

The proposed federated learning system consists of three core components: local training units (hospitals), a central aggregator (server), and a secure communication interface. Each hospital hosts a CNN architecture appropriate to the task (e.g., ResNet for classification), trained exclusively on its own dataset. After local training epochs, model updates are encrypted and sent to the aggregator, where they are averaged using the FedAvg algorithm to update the global model. This model is then redistributed to all clients.

To enhance privacy, we applied differential privacy techniques at the client level and used secure aggregation protocols to prevent the server from accessing individual updates. This hybrid privacy-preserving design ensures that even if one part of the system is compromised, patient-level data remains protected. The architecture supports asynchronous training and partial participation, making it suitable for hospitals with different data volumes and hardware resources. It also scales linearly with the number of clients, minimizing training degradation across larger networks.

5. Experimental Setup and Implementation

The federated framework was implemented using the Flower federated learning library in Python, integrated with PyTorch for deep learning tasks. Each hospital node was simulated on a separate GPU environment. The global server aggregated weights and redistributed updates at fixed communication intervals. Training was run over 100 communication rounds, with local epochs set to 3 per round. Noise was calibrated to ensure ε remained between 2.5 and 4.5, depending on the task and batch size.

To replicate real-world conditions, we introduced data heterogeneity by allocating different disease classes to different clients (non-IID distribution). For instance, one hospital trained on pneumonia and normal X-rays, while another had more tuberculosis samples. We also varied the resolution and contrast of MRI images in the BraTS dataset to simulate equipment diversity. Performance was benchmarked against a centralized training baseline and a local-only model.

6. Results and Discussion

The federated CNN model achieved 88.2% classification accuracy on the ChestXray14 dataset, compared to 90.1% in centralized training. On the BraTS segmentation task, it reached a Dice coefficient of 0.78 versus 0.81 centrally. The small performance gap validates that privacy-preserving learning does not significantly hinder accuracy. Furthermore, privacy loss (ε) remained within acceptable bounds (<4.5), ensuring strong differential privacy guarantees.

Training convergence was achieved in under 50 communication rounds for most models, and communication overhead per client stayed below 250MB. The addition of noise slightly impacted convergence speed but improved generalization on unseen data. SHAP analysis revealed consistent feature importance across federated and centralized models, increasing trust in model predictions. Overall, the system proved effective for cross-institutional collaboration without compromising patient privacy or model performance.

7. Challenges and Limitations

Despite promising results, several challenges were observed. First, data heterogeneity significantly influenced convergence rates and model stability. Clients with rare class distributions contributed less to the global model, raising fairness concerns. Second, communication latency in federated learning is non-trivial, especially in low-resource settings with limited internet bandwidth. Optimizations such as model pruning or gradient compression could help address this issue.

On the regulatory front, deploying FL in real hospitals would require compliance with regional data governance policies, which may vary widely. Furthermore, attacks such as model inversion or poisoning, although mitigated, are not completely eliminated. Differential privacy and SMPC protect data in transit but reduce accuracy slightly. Future work must find better trade-offs between model performance, privacy, and computational overhead.

8. Conclusion and Future Scope

This study confirms that federated learning is a viable and scalable solution for training medical image classification models across distributed healthcare systems while preserving data privacy. The framework achieves competitive performance with centralized systems

and adheres to modern privacy standards using differential privacy techniques. It lays a solid foundation for collaborative AI in healthcare without data centralization.

Future research should focus on enhancing model personalization for each client, integrating secure hardware (e.g., Trusted Execution Environments), and automating compliance verification. There is also significant scope for extending this framework to multi-modal data fusion and real-time diagnostics in smart hospitals. A broader adoption of federated learning could democratize access to high-performing AI tools across health institutions globally.

References

[1] Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ Digital Medicine*, 3(1), 1–7.

[2] Sheller, M. J., Reina, G. A., Edwards, B., Martin, J., & Bakas, S. (2020). Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 92–104.

[3] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. *Proceedings of MLSys 2020*.

[4] Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2021). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 3(6), 473–484.

[5] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of AISTATS*.

[6] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2), 12.

[7] 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.

[8] Brisimi, T. S., Chen, R., Mela, T., Olshevsky, A., Paschalidis, I. C., & Shi, W. (2018). Federated learning of predictive models from federated Electronic Health Records. *Scientific Reports*, 8(1), 1–8.

[9] Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., & Bakas, S. (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific Reports, 10(1), 12598.

[10] Li, X., Gu, Y., Dvornek, N., Staib, L. H., Ventola, P., & Duncan, J. S. (2021). Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results. Medical Image Analysis, 65, 101765.

[11] Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., Wang, F., & Chen, Y. (2021). Federated learning for healthcare informatics. Journal of Healthcare Informatics Research, 5(1), 1–19.

[12] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 308–318.

[13] Vepakomma, P., Gupta, O., Swedish, T., & Raskar, R. (2018). Split learning for health: Distributed deep learning without sharing raw patient data. Proceedings of NeurIPS Workshop on Machine Learning for Health (ML4H).

[14] Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., ... & Li, W. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. Nature Medicine, 27(10), 1735–1743.

[15] Choudhury, O., Gkoulalas-Divanis, A., Salonidis, T., Sylla, I., Das, A., Bellet, A., ... & Madabhushi, A. (2020). Differential privacy-enabled federated learning for sensitive health data. arXiv preprint arXiv:2001.10500.

[16] Yang, H., Yu, H., & Yang, Q. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1–19.

[17] Liu, Q., Yang, R., Ding, M., & Shikh-Bahaei, M. (2021). Federated learning for intelligent healthcare: A survey. IEEE Access, 9, 103638–10365