

Privacy-Preserving CNN Frameworks for Brain Tumor Diagnosis Using Federated Learning

Jiahao Ren *

Maseeh College of Engineering & Computer Science, Portland State University, Portland OR,
United States

* Corresponding Author Email: clarkere@pdx.edu

Abstract. Diagnosis of brain tumors is complicated, considering issues of data privacy, imaging from multiple sources, and the demand for reliable automated tools. In this context, the Federated Learning (FL) has emerged as a privacy-preserving paradigm that enables collaborative training of Convolutional Neural Networks (CNNs) across institutions without sharing raw medical data. In this paper, this paper reviews recent works on FL in CNNs for the classification and segmentation of brain tumors. Indeed, there is much to be said about advances related to FedAvg-based ResNet-18, weight-sharing optimization frameworks, and personalized distillation approaches. Other strategies, such as the use of multi-encoders and privacy-preserving mechanisms, deserve consideration. Moving forward, related to data heterogeneity, communication costs, privacy risks, interpretability, and scalability, the discussed challenges present corresponding future directions. The analysis indicates that integrating elements of compression, robust privacy, explainability, and large-scale validation into one element will be relevant to pursuing FL research into clinical practice.

Keywords: Federated Learning, Brain Tumor, Convolutional Neural Networks.

1. Introduction

Brain tumors have been a major health issue worldwide due to their high mortality rates and complex management. According to the World Health Organization (WHO), cancer remains the second leading cause of death globally, with different types of brain tumors having a considerable percentage of neurological tumors and are important causes of cancer-related mortality and morbidity [1, 2]. Indeed, better treatment results are obtained after early and accurate diagnosis, leading to appropriate, timely, and life-saving interventions like surgery, chemotherapy, or radiotherapy. Unfortunately, most of the traditional diagnostic methodologies rely on the use of Magnetic Resonance (MRI) scans for manual interpretation, which consumes a lot of time and could lead to human errors that directly influence the delay in treatment and reduce survival ratios [1].

Recent advancements in Artificial Intelligence (AI), and Deep Learning (DL) methods such as Convolutional Neural Networks (CNNs), have shown remarkable success in medical image analysis, including detecting, segmenting, and classifying brain tumors [3]. In this work, a CNN model extracts complex patterns in MRI scans to classify tumor types, such as glioma, meningioma, and pituitary tumors. However, the traditional centralized DL approach needs a massive, annotated dataset in one location. This constitutes a real problem regarding data privacy, ownership, and security in the medical field since patients' imaging data are under extremely severe legal and ethical protective regulations, such as Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR) [4].

Considering these drawbacks, Federated Learning (FL) could serve as a transforming principle. FL differs from centralized learning in that it allows the collaboration of multiple institutions to develop a global model without transferring any sensitive patient data outside the institution. Each client trains its model on local MRI data and only shares model parameters or gradients with a central server for aggregation, thereby ensuring privacy preservation and compliance with regulatory frameworks [2, 4]. This approach not only mitigates privacy concerns but also enhances model robustness by leveraging diverse, heterogeneous datasets across different institutions.

Despite its promise, the application of FL in brain tumor diagnosis faces several challenges. Non-independent and identically distributed (non-IID) data, computational heterogeneity, and communication overhead can hinder convergence and degrade performance [2]. Moreover, the "black-box" nature of deep models raises issues on transparency and interpretability, especially for clinical practices where trust in AI-driven processes is vital [5]. To address these issues, recent research has proposed explainable AI (XAI) methods integrated within FL frameworks, aiming to provide visual interpretations of model decisions through techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) and saliency maps [5]. Other contributions have focused on segmentation accuracy by adding architectural innovations, such as multi-encoder CNNs and advanced aggregation strategies, reducing bias and improving generalization across distributed datasets [6, 7].

2. Method

2.1. Federated Learning (FL)–Based CNN

Federated Learning (FL) has turned into an indispensable paradigm in privacy-preserving artificial intelligence for medical imaging. Since it allows decentralized training across hospitals and research centers without having to share raw patient data, Federated Learning (FL) solves the single most important issue regarding sensitive data centralization while still enabling multi-institutional collaboration [1, 4, 7]. This work then looks at CNN-based federated models for brain tumor classification and segmentation, with a strong focus on optimization strategies such as model aggregation, weight compression, and personalized learning.

2.1.1. FedAvg–Based ResNet-18

McMahan et al. were the first to introduce Federated Averaging (FedAvg), the most widely used aggregation strategy in FL. In brain tumor imaging, CNNs based on FedAvg, like ResNet-18, yielded relatively good results, given the trade-off between depth and computational efficiency [4, 7]. In this respect, Wen et al. [4] build upon this baseline by incorporating Virtual Adversarial Training (VAT) into a 3D U-Net to mitigate data scarcity and data imbalance effects within federated settings. Thus, their Federated Heterogeneous Graph (FedHG) algorithm improves vanilla FedAvg by 2.2% in terms of Dice Similarity Coefficient (DSC) score. Likewise, Zhang et al. [8] described a federated transfer learning model in which ResNet-18 based backbones are locally trained and gradients aggregated at the server level. The novelty of these frameworks has been their capability to balance generalization and personalization through refining global weights while preserving institution-specific traits.

2.1.2. Weight Sharing Optimization (FedWSOcomp)

A constant bottleneck in FL is communication overhead, as frequent exchanges of high-dimensional weight updates between clients and servers significantly increase energy and bandwidth costs. Therefore, Onaizah et al. [6] proposed a federated architecture with weight sharing optimization called FedWSOcomp. In their work, they applied top-z sparsification, clustering-based quantization, and Huffman-inspired encoding techniques that achieved compression rates of up to $1085\times$ with no apparent loss of model accuracy. The FedWSOcomp framework demonstrates a 99% reduction in bandwidth and energy for client usage, which is another significant innovation toward scalability in FL. This is really sound compression for large CNNs, like ResNet-18 and U-Net, where the parameter size reaches hundreds of megabytes. Compressing communication costs allows FedWSOcomp to attain a proper margin for implementing FL in resource-constrained real-world settings for healthcare networks.

2.1.3. W Personalized Federated Distillation (FedSPD)

Traditional Federated Learning (FL) does not work well with medical data that is non-independent and identically distributed (non-IID), which usually occurs when there are imaging devices and protocols in various hospitals. Wu et al. [9] put forward FedSPD, a personalized FL framework using

similarity-preserving knowledge distillation. Unlike the direct weight averaging approach, in which global knowledge would not have been transferred, knowledge transfer occurs in FedSPD by aligning feature representations through cosine similarity. Besides, it manifested up to 10.55% improvement in classification accuracy against other state-of-the-art personalized FL algorithms, with a 67% reduction in computational overhead. Each client can personalize the model within FedSPD while still being able to take advantage of global collaboration, which is highly significant in the context of heterogeneous datasets of brain tumor MRIs.

2.2. Multi-Encoder and Transfer Learning Approaches

Beyond single-stream CNNs, multi-encoder strategies have been developed to capture the complex heterogeneity of brain tumors. Soni et al. [2] proposed a multi-encoder U-Net, in which each MRI modality is processed independently-T1, T2, T1ce, FLAIR, and the features are fused at the bottleneck layer. This late fusion approach achieved higher Dice coefficients than conventional early fusion methods. Similarly, Mastoi et al. [3] integrated XAI with federated CNNs and made clinicians visualize model decisions via Grad-CAM, which increased trust and improved interpretability.

Transfer learning in federated settings includes Zhang et al. [8]. In this regard, the authors presented a federated transfer learning framework wherein locally pre-trained models were shared across institutions without raw data exposure. Its generalization and robustness, addressing privacy concerns, has been improved dating back to that time. By leveraging pre-trained ResNet architectures and fine-tuning them within FL cycles, this approach achieved state-of-the-art performance compared to purely centralized training.

2.3. Privacy, Security, and Scalability Enhancements

While FL preserves data locality, it remains vulnerable to privacy leakage through gradient inversion and model update reconstruction. Ullah et al. [5] proposed a scalable FL framework that integrates advanced privacy-preserving techniques such as differential privacy and secure aggregation. Their experiments showed that segmentation accuracy improved as the number of clients scaled from 50 to 100, with Dice coefficients reaching 0.89 while maintaining strict privacy guarantees.

Other frameworks emphasized robustness and scalability. Onaizah et al. [6] showed that weight compression can reduce the communication footprint, while Wu et al. [9] and Wen et al. [4] highlighted the role of model personalization in combating non-IID distributions. Collectively, these methods demonstrate that federated CNN models for brain tumor classification must balance accuracy, privacy, interpretability, and scalability.

3. Discussion

3.1. Challenge

However, the following main challenges remain for CNN-based Federated Learning (FL) to classify brain tumors:

Non-IID and Imbalanced Data: Due to variations in gradient intensities, resolution, available modalities, and even the distribution of tumors, each institution will obtain a unique MRI dataset. This kind of heterogeneity damages model convergence and hence generally reduces accuracy [4, 9]. A few personalized approaches have been proposed to FedSPD, the question of generalization remains unsolved.

Communication and Resource Costs: Models such as ResNet-18 and U-Net are large, with millions of parameters. In other words, full gradient exchange is virtually infeasible at resource-constrained hospitals in this application area. Most compression techniques have been developed so far, for example, FedWSOcomp, yet balancing compression and accuracy remains an open issue [6].

Privacy and Security Issues: The gradients revealed sensitive patient information to an adversary, although patient data were never transferred. For the moment, this requires the adoption of other

privacy-preserving methods, which are time-consuming and sometimes reduce accuracy, such as differential privacy [5] and homomorphic encryption with secure aggregation [9].

Interpretability and Trust: Clinicians need explainable AI to know the reasons behind model predictions. Most current FL frameworks hardly consider interpretability; their performance focus remains dominant. Incorporating XAI techniques is critical to clinical translation, as indicated by Mastoi et al. [3].

Scalability and Real-World Deployment: Most FL works are benchmarked against datasets. Hospitals, however, have messy, incomplete, and even constantly evolving data. Thus, recent works have focused on scaling up to hundreds of clients.

3.2. Future Prospects

Pursuing research along the following lines will promise to embrace the earlier-stated challenges:

Personalized Aggregation Strategies: Next-generation FL architectures will probably embed adaptive aggregation methods like FedSPD or meta-learning-based personalization into their architecture. Thus, global generalization would be balanced with local customization to ensure robust performance across institutions [4, 9].

Communication-Efficient FL: Further development of FedWSOcomp, model pruning, and adaptive communication scheduling may overcome the high resource and communication costs. Such methods may yield enormous bandwidth savings with almost no loss in accuracy, which makes FL far more interesting for resource-poor hospitals [6].

Stronger Privacy Mechanisms: At the same time, differential privacy [10], secure multiparty computation, and homomorphic encryption can be combined to solve privacy and security issues at the front line of care. Hence, future work in these mechanisms will have to minimize the trade-off between data protection and model accuracy while ensuring all healthcare regulations are complied with [5, 8].

Explainable FL Frameworks: Federated CNNs need to consider ad hoc explainable AI techniques, such as Grad-CAM, saliency maps, or attention-based visualization, to gain interpretability and clinical trust. Since liability primarily resides in interpretability during training, such systems may be known as reliable for real-life use [3].

Scalable Clinical Validation: Further trials need to be made, considering these aspects, in multicenter hospital networks with various devices, heterogeneous data, and many patients, to pursue the development of FI systems with assurance across the dimensions of robustness, fault tolerance, and regulatory compliance, as required for the application of laboratory findings in the clinic [1,5].

4. Conclusion

Federated Learning (FL) combined with CNNs is a highly promising approach for developing brain tumor diagnosis while maintaining patient privacy. In particular, the collaborative nature of FL enables institutions to benefit from shared model improvements without compromising sensitive MRI data, making it a powerful framework for large-scale medical AI deployment. The methods discussed have improved the model's performance through adaptive aggregation, weight compression, personalized distillation, and explainable AI. These techniques collectively demonstrate how both computational efficiency and model interpretability can be enhanced through thoughtful architectural and algorithmic design within federated settings. However, the issues of non-IID data, communication overhead, privacy risks, interpretability, and scalability in real-world applications remain. Such limitations underscore that current FL systems must adapt to diverse clinical environments where imaging devices, patient populations, and computational resources vary significantly across hospitals. Hence, further research into resource-efficient, trustworthy, and clinically validated frameworks across institutions is required. Developing solutions that are simultaneously accurate, communication-efficient, and compliant with regulatory standards will be essential for widespread adoption. Addressing such challenges may allow federated CNNs to become

one of the leading technologies in medical imaging collaboration, accelerating the transition of research into healthcare practice. If these advancements continue, FL has the potential to form the backbone of future intelligent diagnostic platforms, supporting more equitable and interoperable healthcare systems worldwide.

References

- [1] Naeem A, Anees T, Naqvi RA, Loh WK. A comprehensive analysis of recent deep and federated-learning-based methodologies for brain tumor diagnosis. *J Pers Med*. 2022; 12 (2): 275.
- [2] Ma Y. Federated learning for brain tumor diagnosis: Methods, challenges and future prospects. *ITM Web Conf*. 2025; 70: 03028.
- [3] Mastoi QUA, Gumaei A, Al-Turjman F, Aloqaily M. Explainable AI in medical imaging: An interpretable and collaborative federated learning model for brain tumor classification. *Front Oncol*. 2025; 15: 1535478.
- [4] Wen J, Li X, Ye X, Li X, Mao H. A highly generalized federated learning algorithm for brain tumor segmentation. *Sci Rep*. 2025; 15: 21053.
- [5] Soni V, Singh NK, Singh RK, Tomar DS. Multiencoder-based federated intelligent deep learning model for brain tumor segmentation. *Int J Imaging Syst Technol*. 2024; 34 (1): e22981.
- [6] Onaizah AN, Xia Y, Obaid AJ, Hussain K. Deep learning-based brain tumour architecture for weight sharing optimization in federated learning. *Expert Syst*. 2025; 42 (2): e13643.
- [7] Ullah F, Nadeem M, Abrar M, Amin F, Salam A, Khan S. Enhancing brain tumor segmentation accuracy through scalable federated learning with advanced data privacy and security measures. *Mathematics*. 2023; 11 (19): 4189.
- [8] Zhang W, Jin W, Rho S, Jiang F, Yang CF. A federated learning framework for brain tumor segmentation without sharing patient data. *Int J Imaging Syst Technol*. 2024; 34: e23147.
- [9] Wu B, Shi D, Aguilar J. Brain tumors classification in MRIs based on personalized federated distillation learning with similarity-preserving. *Int J Imaging Syst Technol*. 2025; 35: e70046.
- [10] Zhao Y, Chen J. A survey on differential privacy for unstructured data content. *ACM Comput Surv*. 2022; 54 (10s): 1 – 28.