Federated Learning for Brain Tumor Diagnosis: Methods, Challenges and Future Prospects

Yuhan Ma

Computer Science and Technology, Anhui University, Hefei, China

Abstract. With the widespread application of Artificial Intelligence (AI) and Machine Learning (ML) in medical field, early diagnosis of brain tumors has become increasingly significant. However, traditional methods face challenges such as data privacy, model interpretability, and data heterogeneity. This paper presents a detailed literature review of the application of Federated Learning (FL) in brain tumor classification, focusing on the necessity of privacy-preserving ML using Magnetic Resonance Imaging (MRI) technology. This paper analyzes various Convolutional Neural Network (CNN) models, including VGG16, ResNet50, DenseNet121, and EfficientNet, exploring their integration within the FL framework to enhance diagnostic accuracy while preserving patient data privacy. Through the discussion, core issues in this field emerge, including model interpretability, non-independent and identically distributed (Non-IID) data distributions, and computational challenges in FL architecture. Although these factors limit the widespread application of FL in medical settings, this paper also proposes potential solutions, such as improving algorithm interpretability through interpretable tools, and utilizing transfer learning and domain adaptation methods to enhance model effectiveness across different datasets. Techniques like knowledge distillation, model quantization, and pruning are proposed to enhance computational efficiency and minimize communication costs. Future research should focus on the application of these methods to enhance model reliability and efficiency.

1 Introduction

Tumors, abnormal masses of tissue, are broadly categorized as benign (non-cancerous) and malignant (cancerous) neoplasms, with over 200 different types affecting human health (Prabukumar, Agilandeeswari, & Ganesan, 2019). According to the World Health Organization (WHO), cancer, another name for malignant tumors, is the second leading cause of death globally, accounting for 1 in 6 deaths in 2018 (World Health Organization, 2024). Among these, brain tumors represent a significant challenge, as they account for approximately one-quarter of all cancer-related deaths in children, according to the National Brain Tumor Foundation (NBTF) (El-Dahshan, Mohsen, Revett, & Salem, 2014). Furthermore, the NBTF has reported that research from developed nations indicates a

Corresponding author: e02114162@stu.ahu.edu.cn

significant rise in the incidence and mortality rates of brain tumors, with an estimated increase of up to 300% over the past three decades (El-Dahshan, Mohsen, Revett, & Salem, 2014).

For decades, Artificial Intelligence (AI) and Machine Learning (ML) have been utilized in many aspects of human daily lives, including financial, political, surveillance and healthcare systems. These technologies enable the analysis of massive amounts of data generated from various sources, allowing for the identification of patterns and prediction of events, which is crucial in fields like healthcare (Shahriar, Allana, Hazratifard, & Dara, 2023).

With the growing impact of tumors, especially brain tumors, the need for early and accurate diagnosis has become increasingly significant. Magnetic Resonance Imaging (MRI), a commonly used non-invasive imaging technique, produces three-dimensional images to assist neurologists in identifying brain anomalies (Soomro et al., 2023). However, the manual interpretation of MRI scans is time-consuming, making it critical to implement advanced computational methods to monitor the important factors and predict events before they may occur (Bagave, Westberg, Dobbe, Janssen, & Ding, 2022). In recent years, machine learning and deep learning methods have been applied to assist in tumor detection, significantly improving the accuracy and efficiency of diagnoses (Soomro et al., 2023). It is becoming increasingly general to implement advanced brain tumor diagnosing approaches globally to improve the accuracy of brain tumor detection and significant to increase the survival rate of an individual through early diagnosis and immediate treatment of brain tumor (Hemanth, Janardhan, & Sujihelen, 2019).

However, despite the significant benefits artificial intelligence, machine learning, and deep learning bring to brain tumor diagnosis, these technologies are also simultaneously confronted with several challenges, including the common issues of integrating medical data from across the world that people frequently face in normal AI field and the accompanying privacy concerns (Kalpana, Chowdary, Sravya, Reddy, Pravallika, & Gnanasri, 2023). To address these issues, Federated Learning (FL) is introduced to brain tumor diagnosis to fix privacy problems and to leverage different types of ML and deep learning algorithms (Kalpana, Chowdary, Sravya, Reddy, Pravallika, & Gnanasri, 2023; Albalawi et al., 2024). The emergence of FL can well fix the privacy concerns while just training medical data locally without sending the raw data to centralized server through the Internet and enable the integration of medical imaging resources from various geographical regions. However, FL also faces its unique challenges brought up from data heterogeneity of various sources and computational heterogeneity of decentralized servers, which are two common issues when it comes to FL. People cannot deny the significance of implementing FL in brain tumor diagnosis using MRI images, which may be a powerful tool in advancing brain tumor diagnosis. Therefore, it is vital and essential to conduct a comprehensive review of FL for brain tumor diagnosis in recent years, focusing on its methods, challenges and future directions.

The remainder of the paper is organized as follows. First, this study will recapitulate the organization, solutions as well as results of several recent applications of FL for brain tumor diagnosis in Section 2. Then, in Section 3, this paper will discuss the challenges and future prospects in the field of brain tumor diagnosis in general. Finally, Section 4 summarizes the paper, and presents conclusions drawn from the contents discussed here.

2 Method

2.1 Introduction of ML and Federated Learning Workflow

The workflow of ML typically involves data collection, data preprocessing, researching the model that will be best for the type of data, training and testing the model, and evaluation. The workflow of ML in detection of brain tumor using MRI also follows this routine. First of all, scientists gather data that are MRI images of brains from every patient with brain tumor or volunteers with healthy brains. Then, scientists convert these images into a labeled and numerical format suitable for computer processing and deal with certain issues, such as data missing or noisy data and split the data into training data and testing data, following certain rules. The most significant goal is to find and train the most optimal model with pre-processed data. In this stage, scientists would employ different models and algorithms and test each model and algorithms until they find an optimal model which yields the best results and the highest classification accuracy (Hemanth, 2016; Mittal, 2018).

FL takes advantage of a central server and offers an innovative and decentralized approach to train these MRI data locally among institutions without sharing the sensitive and private data. These locally trained models then share their own updates with the central server, which aggregates the updates to form an optimal global model. That means implementing ML model separately in each institution, trying to secure an optimal brain tumor detecting model as well as protecting data privacy.

2.2 Federated Learning Based on Key CNN Models for Brain Tumor Classification and Diagnosis

2.2.1 VGG16

The VGG16 is a prototypical convolutional neural network architecture developed by Visual Graphics Group from Oxford in 2014, comprising a total of 16 layers, including 13 convolutional layers and 3 fully connected layers. Despite its relatively straightforward design, it has demonstrated remarkable efficacy in image classification and gained renown for its unparalleled performance in the ImageNet competition (Simonyan & Zisserman, 2015).

Eid Albalawi and his group members applied VGG16 to classify brain tumors using MRI images within a federated learning framework. The dataset contained MRI images from figshare, SARTAJ, and Br35H. Images of these dataset were classified into four classes: glioma, meningioma, no tumor and pituitary and these images underwent several preprocessing stages including augmentation, normalization and resizing to target pixel dimensions. To better leverage the model's existing feature extraction capabilities, transfer learning strategy which involved utilization of pre-trained weights from the ImageNet dataset was conducted, complemented by fine-tuning the top layers of the model to align with specific brain tumor classification task. At the same time, additional dropout layers were employed to prevent overfitting, which means randomly drops some neurons during training to improve the model's performance.

Federated learning was implemented to protect data privacy throughout the process. This method involves client selection which is randomly selecting a subset of clients (approximately 50% of the total clients) for training at each iteration or round. The approach combined local training and model aggregation, ensuring accurate and privacy-preserving classification of brain tumors (Albalawi et al., 2024).

2.2.2 ResNet50

ResNet50, a deep convolutional neural network model with 50 layers, was initially proposed by Kaiming He and colleagues in 2015 as part of the broader Residual Network (ResNet) family. This innovation introduced the conception of residual blocks to effectively address the vanishing gradient problem commonly encountered in very deep neural network (He, Zhang, Ren, & Sun, 2015).

Khanh Le Dinh Viet and his team integrated the classic federated average (FedAvg), a type of algorithm in federated learning, into ResNet50 and employed this model to a dataset comprising 3064 MRI T1-weithted brain tumor images. These MRI images were classified into 3 classes: meningioma, glioma and pituitary tumor. The results of this experiment demonstrated exceptional performance of the model (Le Dinh Viet, Le Ha, Quoc, & Hoang, 2023).

2.2.3 DenseNet121

DenseNet121 is a convolutional neural network characterized by dense connectivity. Dense connectivity means not only the output of each layer was propagated to next layer, but also it can be propagated to every subsequent layers, thereby promoting efficient information flow throughout the network. DenseNet121 contains 121 layers, making it suitable for complex tasks such as image classification and medical image analysis (Huang, Liu, van der Maaten, & Weinberger, 2018).

In Vadlamudi Kalpana's study, a dataset from kaggle, consisting of 255 MRI images of brain, underwent preprocessing, including resizing to a uniform 128×128-pixel dimension, conversion to grayscale, and division into 75% training data and 25% testing data. The classic federated learning approach enables the formation of global models without the need to centralize data, preserving patient privacy. The relatively small number of parameters in DenseNet121 resulted in reduced communication load between local clients and the central server during each round or iteration, which is particularly advantageous in resource-constrained environments or with limited bandwidth. Additionally, transfer learning was implemented to enhance feature extraction capabilities. The study also utilized the Adam optimizer with a learning rate of 0.001 for training, leveraging its adaptive learning rate to expedite model aggregation (Kalpana, Chowdary, Sravya, Avinash, Pravallika, & Gnanasri, 2023).

2.2.4 EfficientNet-B0

EfficientNet-B0, a convolutional neural network architecture developed by researchers at Google in 2019, introduces a novel approach to model scaling through systematic adjustments to depth, width, and resolution using fixed scaling coefficients. This innovation enhances computational efficiency while maintaining high accuracy. Serving as the baseline model in the EfficientNet family, EfficientNet-B0 aims to achieve superior performance with minimal computational resources (Tan & Le, 2020).

This research employed 3,260 MRI brain tumor images classified into glioma, meningioma, pituitary tumor, and no tumor categories. Common preprocessing including enhancing and normalizing were conducted, ensuring consistency across clients' local datasets. Each client trains the EfficientNet-B0 model locally and due to its characteristic of having fewer parameters, communication between the local clients and central server is minimized, lowing bandwidth requirements. The AdamW optimizer (Loshchilov & Hutter, 2019), which was built upon Adam optimizer, was used to enhance both model's generalization ability and convergence rate. Due to its efficient scaling and adaptability, EfficientNet-B0 performs better in these heterogeneous environments compared to models like ResNet (Zhou, Wang, & Zhou, 2023).

2.2.5 InceptionV3

InceptionV3 is a deep convolutional neural network architecture that was introduced as an enhancement over previous Inception models. One of its key advancements lies in the utilization of multi-scale convolutions, allowing the model to extract features at various scales simultaneously. By utilizing inception modules, InceptionV3 adeptly captures both fine and coarse details within an image, thereby enhancing its efficacy in tasks such as classification and segmentation. The inception modules are designed to conduct convolutions at multiple filter sizes (e.g., 1x1, 3x3, 5x5), which enables efficient feature extraction while maintaining low computational costs relatively (Szegedy et al., 2015).

InceptionV3 is effective in classifying multi-modal MRI brain tumor data due to its multi-scale feature extraction. Its minimal computational demands render it well-suited for deployment within federated learning envronments, where fast inference and efficient training are imperative for the facilitation of distributed, privacy-preserving medical applications (Islam, Reza, Kaosar, & Parvez, 2023).

2.2.6 ConvNeXt

ConvNeXt is a modern convolutional neural network (ConvNet) designed to integrate the latest deep learning advancements and optimize the performance of traditional convolutional networks. ConvNeXt enhances convolution operations and utilizes modern architecture improvements (such as normalization and activation function optimization) to achieve better performance in image classification and segmentation tasks (Liu et al., 2022).

In the federated learning framework, ConvNeXt performs exceptionally well in handling brain tumor classification tasks. Due to its powerful feature extraction capabilities and optimized computational efficiency, ConvNeXt excels when processing multi-modal MRI images distributed across multiple clients. According to the research by Khanh Le Dinh Viet et al., ConvNeXt achieved high classification accuracy after weight aggregation using the FedAvg algorithm. Particularly, ConvNeXt reached 98.69% accuracy on IID data and also showed good stability and convergence on Non-IID data. Thus, ConvNeXt is well-suited for federated learning environments, striking a balance between high performance and low computational overhead, providing an effective solution for multi-modal medical image classification tasks (Le Dinh Viet, Le Ha, Quoc, & Hoang, 2023).

3 Discussion

3.1 Limitations and Challenges

3.1.1 Interpretability

Nowadays, in the medical field, various deep learning models are designed for brain tumor classification based on images of MRI and these models have been reaching excellent predictive performance. Notwithstanding, these models usually act as black boxes and lack of interpretability, making the decision-making progress hard to comprehend. For example, some models mentioned earlier have at least 16 layers or very complicated architectures, making them somewhat hard to comprehend. Like ResNets (He, Zhang, Ren, & Sun, 2015), a classic deep learning model, some have over 200 layers and surpass human performance in visual cognition.

In the field of medicine, the need for model interpretation is particularly urgent. Doctors and patients should know why the model makes a certain diagnostic strategy to ensure the scientificity and rationality of therapeutic schedule. Although FL protects data privacy through training locally, it restricts the accessibility of data meanwhile, leading to the further decrease of transparency. Because the weight aggregation process of the model itself is opaque, it is challenging for researchers and doctors to track how the model handles different data, and this “black box” problem is particularly prominent in distributed environments.

Therefore, how to improve interpretability of models under the FL framework is one of the significant challenges. For example, tools such LIME and SHAP can provide local interpretation of models, but how to effectively apply these tools in a distributed environment remains an unsolved issue.

3.1.2 Applicability

Although FL can effectively protect patient privacy, it faces issues of applicability across different medical institutions. Data from different institutions may exhibit a non-independent and identically distributed (Non-IID) phenomenon, where each institution has different data distribution, leading to inconsistent local training outcomes and limited generalization capabilities. According to research on the effects of great data heterogeneity, it shows that heterogeneity across different local clients lead to a significant decline of FL model, including a 9.2% decline in accuracy and 2.32 times longer training time, which makes the global model less effective (Yang et al., 2021).

Due to factors that data from different institutions may come from different patient groups or MRI images obtained by different devices may have different resolution ratio, contrast ratio and noise level, it can lead to inconsistent outcomes and performances on different clients, eventually affecting the aggregation capabilities of global model. In addition, heterogeneity of MRI data also poses potential challenges. MRI devices vary in different medical institutions, meaning that a model that performs well on a certain group or model may behave not so well in another group or model.

Therefore, how to handle Non-IID data appropriately in classification of brain tumor based on FL and ML and improve model’s applicability under various environments are becoming one of the key challenges in FL.

3.1.3 Computational Efficiency and Communication Costs

FL reduces reliance of medical institution on central server by training models just on local clients, but it also poses challenges, like computational efficiency and communication cost. For resource-constrained environments (such as small hospitals or equipment), the cost of training and communicating models can be a bottleneck, especially if deep learning models, like ResNet or DenseNet, require significant computational resources. Reducing communication overhead, improving local training efficiency, and how to optimize models under limited computing resources are the technical challenges faced by federated learning.

3.2 Future Prospects

3.2.1 Computational Efficiency and Communication Costs

Interpretability is actually one of the most significant issues to be considered in this architecture. To improve interpretability on this model based on FL, some classic and prominent methods such as Local Interpretable Model-agnistic Explanations (LIME) and SHapley Additive exPlanations (SHAP) could be used for interpreting and explaining ML models locally.

LIME achieve its interpretative goal by perturbing the input data and observing how the predictions change, making it suitable for even any specific ML model. However, in medical field, due to some uncontrollable factors under FL environment, heterogeneity commonly exists across different medical institutions, which means the effectiveness of LIME could be somehow restrained because it could only provide reliable explanations in environments with less data heterogeneity. To improve applicability of LIME under FL environment, researchers can explore methods to integrate LIME's insights with the parameters of the FL model. For instance, FL would not only aggregate weights across different clients, but also collect interpretability index, such as some scores of vital characteristics and visual results. Through this method, the overall interpretability of the model will be improved, as each client contributes not only their model updates but also valuable interpretive insights that reflect their data characteristics. And also, as the interpretability of local clients improves, the interpretability of global model would improve.

Like LIME, SHAP could also be implemented with FL architecture. SHAP leverages the concept of Shapley value to quantify contributions of each characteristic to provide explanation for each prediction. In FL architecture, it can aggregate values of Shapley across various clients or institutions, making it more helpful for doctors to understand how local data changes influence global model decisions and improving medical practitioners' trust and acceptance for this model.

3.2.2 Transfer Learning and Domain Adaptation

To solve applicability issues discussed earlier, researchers can consider using technologies like transfer learning and domain adaptation. Transfer learning allows a pre-trained model, which has learned from one dataset, to be fine-tuned and applied to another dataset. This capability is particularly significant in the medical field, where models can rapidly adapt to new data without undergoing expensive and time-consuming re-training. Domain adaptation can improve models' performance across different institutions, especially when these institutions use different data collection protocols or patient demographics. Under the circumstance of FL, data across various institutions commonly remain localized and non-IID, and it would improve the generalization ability of the aggregated model. Firstly, researchers can utilize domain adaptation techniques to align the data distributions from different clients before model aggregation. By ensuring that the models are trained on similarly distributed data, it can reduce the negative influence of data heterogeneity on model performance. Once data distributions are aligned, researchers can continue to conduct federated averaging of model weights, making sure the resulting model is both accurate and applicable in real-world clinical background.

3.2.3 Solutions for Improving Computational Efficiency and Reducing Communication Costs

To address the challenges of computational efficiency and communication costs in FL, several solutions can be implemented. Knowledge distillation is one method that reduces the model size by training a smaller student model to mimic a larger teacher model, ensuring performance while lowering computational demands. Methods like model quantization can decrease the precision of the parameters, thus reducing both the demands for storage and computational load, reducing costs of communication. Pruning techniques, which are a method that can eliminate redundant parameters, can make the model perform better in resource-constrained settings. By utilizing these approaches, FL can enhance computational efficiency and minimize the communication costs, eventually facilitating its application in medical settings with limited resources.

4 Conclusions

This paper presents a detailed literature review of FL for brain tumor classification with MRI. This paper illustrated the importance of privacy-preserving ML using FL for distributed medical institutions and demonstrated how it can be done, as well as the hurdles associated with data heterogeneity, interpretability and model generalization. This paper performs the study on different CNNs including VGG16, ResNet50, DenseNet121 and EfficientNet along with reviewing their integration with FL frameworks to improve diagnostic accuracy whilst preserving patient data privacy.

From the discussion, central issues in the area were outlined: understanding of models, non-IID data distributions, and computational challenges behind FL environments. This is a limitation to the wider usage of FL in medical settings. However, this study outlined potential mitigations too, such as algorithmic legibility via interpretability/inspection tools like LIME and SHAP, and transfer learning and domain adaptation methods to make models more effective across different datasets. In the future, work should focus on applying these methods, while combing with more sophisticated technologies to improve both reliance and efficiency.

References

1. Albalawi, E., Mahesh, T. R., Thakur, A., Kumar, V. V., Gupta, M., Khan, S. B., & Almusharraf, A. 2024. Correction to: Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor. BMC Medical Imaging, 24(1), 161.
2. Bagave, P., Westberg, M., Dobbe, R., Janssen, M., & Ding, A. Y. 2022. Accountable AI for healthcare IoT systems. In 2022 IEEE 4th International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA) (pp. 20-28). IEEE.
3. El-Dahshan, E. A., Mohsen, H. M., Revett, K., & Salem, A. B. M. 2014. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. Expert Systems with Applications, 41(11), 5526-5545.
4. He, K., Zhang, X., Ren, S., & Sun, J. 2015. Deep residual learning for image recognition. arXiv. https://arxiv.org/abs/1512.03385
5. Hemanth, G., Janardhan, M., & Sujihelen, L. 2019. Design and implementing brain tumor detection using machine learning approach. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1289-1294). IEEE.
6. Hemanth, J. 2016. Biomedical Signal and Image Processing in Patient Care. Springer.
7. Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. 2018. Densely connected convolutional networks. arXiv. https://arxiv.org/abs/1608.06993
8. Islam, M., Reza, M. T., Kaosar, M., & Parvez, M. Z. 2023. Effectiveness of federated learning and CNN ensemble architectures for identifying brain tumors using MRI images. Neural Processing Letters, 55(4), 3779-3809.
9. Kalpana, V., Chowdary, J., Sravya, T., Avinash, A., Pravallika, P., & Gnanasri, V. 2023. Implemented global model for brain tumor detection using federated learning. In Proceedings of the 2023 International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech) (pp. 1-5). IEEE.
10. Le Dinh Viet, K., Le Ha, K., Quoc, T. N., & Hoang, V. T. 2023. MRI brain tumor classification based on federated deep learning. In 2023 Zooming Innovation in Consumer Technologies Conference (ZINC) (pp. 131-135). IEEE.

11. Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. 2022. A ConvNet for the 2020s. arXiv. https://arxiv.org/abs/2201.03545
12. Loshchilov, I., & Hutter, F. 2019. Decoupled weight decay regularization. arXiv. https://arxiv.org/abs/1711.05101
13. Mittal, A. 2018. Workflow of a machine learning project. Towards Data Science. https://towardsdatascience.com/workflow-of-a-machine-learning-project-ec1dba419b94
14. Prabukumar, M., Agilandeeswari, L., & Ganesan, K. 2019. An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier. Journal of Ambient Intelligence and Humanized Computing, 10(1), 267-293.
15. Shahriar, S., Allana, S., Hazratifard, S.M., & Dara, R. 2023. A Survey of Privacy Risks and Mitigation Strategies in the Artificial Intelligence Life Cycle. IEEE Access, 11, 61829-61854.
16. Simonyan, K., & Zisserman, A. 2015. Very deep convolutional networks for large-scale image recognition. arXiv. https://arxiv.org/abs/1409.1556
17. Soomro, T. A., Zheng, L., Afifi, A. J., Ali, A., Soomro, S., Yin, M., & Gao, J. 2023. Image Segmentation for MR Brain Tumor Detection Using Machine Learning: A Review. IEEE reviews in biomedical engineering, 16, 70–90.
18. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. 2015. Rethinking the Inception architecture for computer vision. arXiv. https://arxiv.org/abs/1512.00567
19. Tan, M., & Le, Q. V. 2020. EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv. https://arxiv.org/abs/1905.11946
20. World Health Organization. 2024. Cancer. World Health Organization. https://www.who.int/health-topics/cancer#tab=tab_1
21. Yang, C., Wang, Q., Xu, M., Chen, Z., Bian, K., Liu, Y., & Liu, X. 2021. Characterizing impacts of heterogeneity in federated learning upon large-scale smartphone data. In Proceedings of the Web Conference 2021 (pp. 935–946). Association for Computing Machinery.
22. Zhou, L., Wang, M., & Zhou, N. 2023. Distributed Federated Learning-Based Deep Learning Model for Privacy MRI Brain Tumor Detection. ArXiv, abs/2404.10026.