Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 2, April 2026, pp: 672-688 e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

FedBrain-3DMRI: Federated Self-Supervised Learning for 3D Brain Tumor Segmentation using SCAFFOLD Algorithm

Neeshu Kumari and Chintan Thacker

Computer Science and Engineering, Parul University, Gujarat, India

Corresponding author: Neeshu Kumari (e-mail: 212300470010@paruluniversity.ac.in), **Author(s) Email**: Chintan Thacker (e-mail: chintan.thacker19435@paruluniversity.ac.in).

Abstract **Brain tumor segmentation is the most important way to separate tumor areas from healthy brain tissue in medical imaging. This is necessary for making an accurate diagnosis and planning treatment. But building strong deep learning models is often hard because there isn't much labeled medical data available, and strict privacy rules stop data from being shared in one place. Federated Learning (FL) helps keep patient data private by keeping it local, but its performance often drops when data from different hospitals have big differences in quality, imaging protocols, and distribution. Our research seeks to create a privacy-preserving federated learning framework that adeptly manages significant data heterogeneity while ensuring high segmentation accuracy across various institutions. We propose a new two-stage FL framework that allows multiple institutions to work together while keeping their privacy and effectively dealing with complicated non-IID data distributions. To start, we use a Federated Masked Autoencoder (MAE) for self-supervised pre-training. This lets the model learn strong anatomical features from unlabeled MRI scans. Second, the model is carefully fine-tuned using an Attention ResUNet3D architecture to get very accurate tumor segmentation. We use the SCAFFOLD optimization algorithm to keep training stable across all clients, even when the scanner varies from site to site, thereby directly addressing client drift. We also use strategic foreground-biased sampling and Test-Time Augmentation (TTA) techniques to greatly improve segmentation accuracy in difficult, uneven tumor sub-regions. We ran extensive experiments on the BraTS 2024 dataset in simulated federated settings with 10, 50, and 100 different clients. The Dice coefficients we got were 0.826, 0.824, and 0.818, which demonstrate strong performance. In the end, these strong results show that the suggested method works well on a larger scale and can be used in a clinical setting.**

Keywords **Brain Tumor Segmentation, Federated Learning, Masked Autoencoder, SCAFFOLD, BraTS 2024.**

I. Introduction

Brain tumors arise from the uncontrolled growth of brain cells, mainly due to disruptions in normal cellular functions and genetic mutations [1]. A brain tumor not detected on time may affect the patient's survival. Magnetic Resonance Imaging (MRI) is the most popular imaging modality since doctors mostly utilize it to look at tumors. Manual Interpretation of the MRI scans is a reliable approach, despite requiring time and expertise [2]. For automatic tumor detection, deep learning models have proven to be outstanding benchmarks. A large dataset is needed to segment and classify medical images. [3] Due to privacy laws such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR), it is challenging to collect data because it contains patient personal details, resulting in confined data silos within the hospital and subsequent data silos [4]. The privacy concerns are addressed by the federated learning setup, which maintains the dataset locally and distributes the weights and biases of the data to train the deep learning model, as demonstrated in Fig. 1 (on the next page). At every communication round, each client's medical institute and healthcare center sends the updated model weight to the server for aggregation. Federated learning has this potential advantage, but it still suffers from non-independent and Identically Distributed data, the reason being that it collects data from various sources. It leads to client drift, where the local model departs from its main objective, resulting in degraded model performance [5, 6]. In order to handle these issues, we proposed a novel two-stage federated learning framework for Brain tumor segmentation. The primary contributions of this work are as follows:

1. Novel Two-Stage Federated Pipeline: We introduce a privacy-preserving framework that couples Federated Self-Supervised Pre-training

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

(via 3D MAE) with supervised fine-tuning. By leveraging unlabeled data to learn robust volumetric features, this strategy reduces reliance on large, annotated datasets.

2. Drift-Resilient Optimization via SCAFFOLD: We adapt the SCAFFOLD optimization algorithm for 3D medical segmentation. This method explicitly corrects for client drift caused by non-IID data distributions, ensuring stable convergence across statistically diverse medical centers.
3. Attention-Enhanced 3D Architecture: We propose a specialized Attention ResUNet3D backbone. By integrating residual blocks with attention gates, the network effectively suppresses background noise and enhances feature extraction for small, irregular targets, specifically improving segmentation metrics for the Enhancing Tumor (ET) and Tumor Core (TC) regions.
4. Comprehensive Training and Inference Strategy: To tackle the severe class imbalance inherent in volumetric MRI, we implement a foreground-biased patch sampling technique. Furthermore, we employ Test-Time Augmentation (TTA) during inference to maximize prediction stability. We validate the scalability of our approach on the BraTS 2024 dataset, demonstrating robust performance across federated networks of 10, 50, and 100 clients.

The rest of the paper is organized as follows. Section 2 includes related work. Section 3 is dedicated to Methodology. Section 4 provides Experimental Setup. Section 5 discusses the results achieved by our proposed work. Section 6 describes a detailed discussion and Section 7 concludes the study.

II. Related Work

This section highlights key advancements in brain tumor segmentation using deep learning and federated learning, covering both basic techniques and modern deep learning approaches.

A. Deep Learning Approaches for Tumor Segmentation

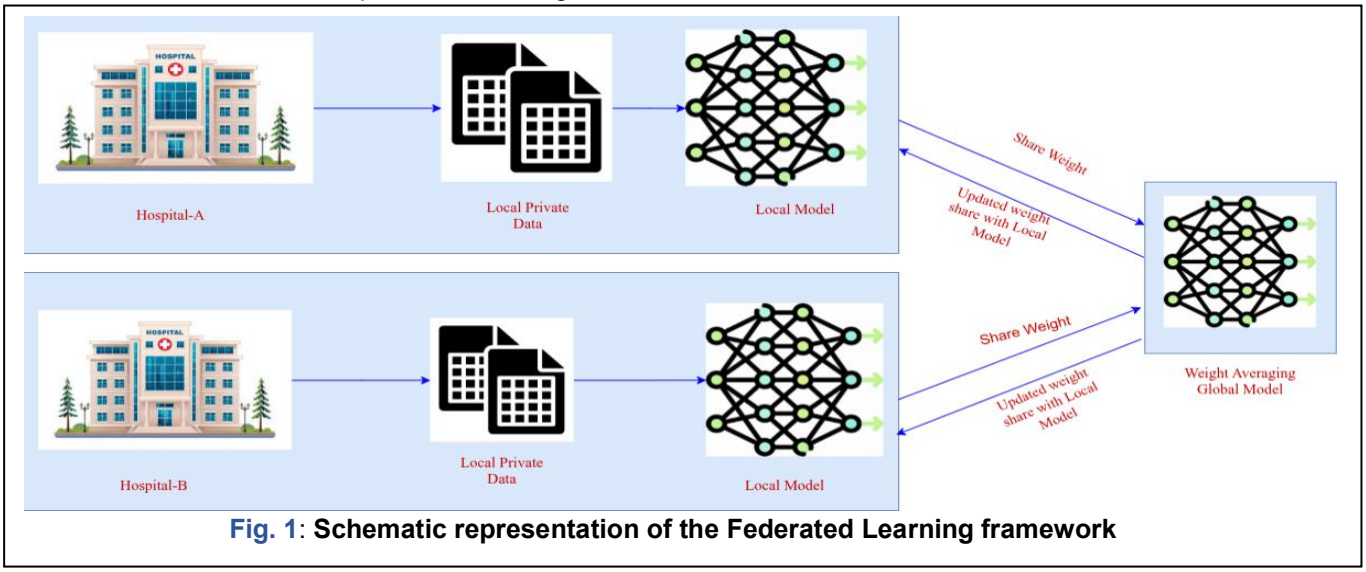


Fig. 1: **Schematic representation of the Federated Learning framework**

The automation of feature extraction has completely transformed medical image analysis through deep learning techniques. Hybrid and lightweight models have been prioritized in recent literature to avoid computational restrictions and data availability issues. S. Alphonse et.al [7] performed segmentation task using the attention enhanced multi scale network However, the model relies on fully supervised learning and may not generalize well in federated or data-scarce environments and R. Ahsan et.al [8] implemented the encoder decoder based architecture that focused on the spatial and contextual information and as the base architecture for the segmentation they have used the convolution neural networks but the approach does not address data heterogeneity across institutions, limiting its applicability in federated settings. The YOLOv5 model has been implemented followed by a 2D U-Net for segmentation. A. Anaya-Isaza et al. [9], they deal with the constraint of limited data by incorporating the Principal Component Analysis (PCA) and the ResNET50, which is utilized as a base model but relies on handcrafted dimensionality reduction only. T. Shelatkar et al. [10], YOLOv5 models in this study are used to perform the segmentation process on the dataset BraTS 2021. YOLO loss function is used to calculate the segmentation error. YOLO-based architectures may struggle to accurately delineate fine tumor boundaries, as they are primarily designed for object detection.

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

In line with this direction, M. F. Almufareh et al. [11] compare the evaluation of YOLO based models, incorporating a mask alignment scheme to improve the localization quality. Furthermore, the Gaussian Convolutional Neural Network (GCNN) has been shown to be an effective preprocessing strategy for noise reduction before feature extraction, but the process increases the overall complexity.

M. Rizwan et al. [12] Preprocessing strategies to reduce noise have been proven in this study. However, the method primarily focuses on preprocessing and lacks advanced feature extraction for complex tumor structures, while A. S. Musallam et al. [13] proposed a streamlined multi-stage Deep CNN preprocessing protocol to expedite convergence. Hybrid and ensemble-based methods have been explored by several studies to enhance the accuracy of boundary segmentation, but the model requires extensive data to perform well.

S. Li et al. [14] proposed a Region-of-Interest (ROI) guided framework that merges 2D localization with 3D volumetric segmentation, enabling the depiction of the boundary more precisely, but the process heavily depends on the extraction of ROI; if any error occurs during the extraction of the ROI, it can negatively impact segmentation performance. In the same line, U. Bhimavarapu et al. [15] The Fuzzy C-Means clustering method is used with Extreme Learning Machines (ELM) to optimize texture-based features. However, an unsupervised clustering approach may fail to accurately capture complex tumor boundaries. Recently, Anantharajan et al. [16] presented the Ensemble Deep Neural Support Vector Machine (EDN-SVM) approach, which includes adaptive contrast enhancement to upgrade the precision of tissue classification, which increases the computational cost and complexity. J. Walsh et al. [17], they suggested the lightweight U-Net design for tumor segmentation from the MRI image. 3D MRI images do not proceed immediately; instead, MRI scans are segmented into three orthogonal 2D images: axial, coronal, and sagittal planes. The representation of this brain in multiple views simplifies the segmentation task and reduces complexity. The author uses encoder-decoder architecture to concentrate on contextual and geographical information. The model was tested on the BITS dataset. However, converting 3D data into 2D slices may lead to loss of volumetric information.

Y. Jiang et al. [18] present SwinBTS, a transformer-based framework for 3D segmentation of multimodal MRI brain tumors. The proposed architecture follows an encoder-decoder structure. In both cases, the structure 3D Swin transformer is used to gather information about long-range dependencies and the context, and a down and up sampling convolution layer is utilized to keep the local spatial feature. To extract more detailed features from the data, we implemented a bottleneck-layer transformer model. The assessment of model efficacy on the BRATS datasets demonstrates that SwinBTS exceeds the performance of the majority of prominent 3D CNN-based models, but the transformer model requires high computational resources. In [19], L. ZongRen et al. developed DenseTrans, which utilizes the Swin transformer within the UNet++ framework to facilitate the modeling of the global context in multimodal brain tumor segmentation. The UNet++ convolutional layers in this method draw out local features, and the Swin transformer blocks are added to the high-resolution layers to pull out long-range dependencies. A depth-controlled transformer and convolutions are employed to make the network lighter and minimize the cost of self-attention techniques. The BraTS 2021 dataset was used to test the model.The model is complex, which increases the training time.

Z. Zhu et al. [20] introduced a multimodal brain tumor segmentation model that clearly integrates deep features and edge data to enhance the accuracy of the boundary. The proposed approach includes three key modules: a segmentation framework that uses Swin Transformer to extract high-level contextual features, an edge detection module that uses convolutional neural networks with an edge spatial attention block to improve boundary information, and a feature fusion module that uses graph convolution to combine semantic and edge features. The approach combines semantic context and edge cues modeling between contrasting MRI modalities to enhance the localization of tumors and the delineation of the boundaries that delimit tumors more precisely. The role of the experiment conducted on the BraTS benchmarks proved to have better performance than a number of the existing work techniques. Nonetheless, the multi-module design adds complexity to the architecture and computation cost, which are likely to impact the training efficiency and deployment. Due to combining the Transformer, CNN, and graph convolution modules, the model is computationally complex.

C. Yan et al. [21] developed an improved U-Net architecture known as SEResU-Net, which makes use of channel-wise attention and residual learning to distinguish brain tumors that consist of more than a single kind of cell. The issue of degradation is compensated for by the residual blocks to allow the extraction of more features during the down-sampling process. The squeeze-and-excitation (SE) blocks introduced in the upsampling path target the feature channels that are valuable and ignore the non-valuable feature channels. This improves the precision of the segmentation, particularly in the case of a very small tumor. The model was tested on the BraTS 2018 and 2019 datasets, and it performed significantly better than the U-Net baseline. However, this study depends on a large annotated dataset.

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

B. Federated Approaches for Medical Image Analysis

Federated learning (FL) is a standard framework widely used for privacy-preserving collaboration in the healthcare domain. M. E. Yahiaoui et al. [22], each client trained the 3D Unet model on a local dataset. To protect the data from unauthorized users, they have utilized privacy-preserving techniques. This approach does not explicitly address data heterogeneity across clients, which may affect model convergence. A. Giri et al. [23] proposed a hybrid approach that combines the U-Net and ResNet architectures, which merge the deep feature with the detailed segmentation. The hybrid architecture requiring strong hardware support.

FL deployment deals with heterogeneous data at the client site. M. Islam et al. [24], to mitigate this problem, a regularized similarity aggregation (RegSimAgg) protocol is used to select the client for each round. They have used the selection process in order to enhance the communication efficiency, but it introduces the selection overhead. Q. Dai et al. [25], the authors address both data security and multimodal heterogeneity by using a federated learning framework. The model design is complex, it requires synchronization among clients. F. Wagner et al. [26] proposed Fed-Unibrain, to deal with missing modality during training random modality dropout is applied. Random modality dropout may lead to loss of critical information and reduce the accuracy of the segmentation.

M. Grama et al. [27], Adaptive Federated Averaging (AFA) is proposed that provides privacy and robustness. At the server-side robust aggregation is used to aggregate the clients update. This approach may suffer from instability when dealing with highly non-IID data. Likewise, T. S. Armstrong et al. [2] give the empirical correlation that FL architectures maintain a high degree of computational efficiency and communication viability to cater to a large number of clients as the network scales up. The study is symptom-based and subjective, relying on patient-reported outcomes, which may introduce bias.

Recently, federated learning has mainly focused on enhancing the accuracy of the diagnostic process through the use of neural networks. Mastoi et al. [28] explored explainable AI, which integrated GRAD-CAM, a visualization technique within a federated GoogleNet model, to facilitate four-class classification. This approach focuses only on classification rather than segmentation, limiting its clinical applicability. M. Islam et al. [29] address the dual objectives of segmentation performance and data privacy, employing a pre-trained ensemble of convolutional networks that includes DenseNet-121 and VGG-19, but it may suffer from data heterogeneity. S. Sharma et al. [30] incorporated MobileNetV2 for IID and non-IID data within a federated learning framework to detect tumors. The model is primarily designed for classification and does not address segmentation.

The study [7] introduced Federated Averaging (RL-FedAvg), reinforcement learning in a federated learning setup to aggregate the updates, combined with a Dual attention multiscale dense U-Net architecture. D. H. Mahlool et al. [31] proposed a distributed diagnostics model of the brain for focusing on federated learning with the help of convolutional neural networks (CNNs). In it, local CNN models are trained locally in various clinical locations, and only model parameters are transmitted to a central server, which performs aggregation, thus protecting patient privacy. This technique was tested on two benchmark datasets, BT-small-2c and BT-large-3c, respectively, and the classification yields 82% and 96%, respectively. The research shows that FL is viable for privacy-sensitive brain tumor classification in decentralized healthcare settings. The work is however more on classification and does not attempt at segmenting tumors pixel by pixel but instead depends on 2D CNNs and therefore is not able to capture the fine tumor boundaries at a spatial level. The model relies on a 2D CNN that limits its ability to capture detailed information.

Y. M. Elbachir et al. [32] suggested a distributed diagnostics model of brain tumors focusing on federated learning with the help of convolution neural networks (CNNs). In it, local CNN models are trained at various clinical locations, thus protecting patient privacy and modeling only locally. Parameters are transmitted to a central server, which performs aggregation, thus protecting patient privacy and yielding classification accuracies of patient privacy. This technique was tested on two benchmark datasets, BT-small-2c and BT-large-3c. However, the work focuses more on classification, thus protecting patient privacy, and depends on 2D CNNs instead of attempting to segment tumors pixel by pixel; therefore, it cannot capture the fine tumor boundaries at the spatial level. However, the work focuses more on classification, thus protecting patient privacy, and depends on 2D CNNs instead of attempting to segment tumors pixel by pixel; therefore, it cannot capture the fine tumor boundaries at a spatial level. The research shows that FL is viable in terms of privacy-sensitive brain tumor classification in decentralized healthcare settings.

E. Albalawi et al. [33] came up with both federated learning and transfer learning to automatically classify brain tumors using MRI data. The proposed model utilizes a VGG16 architecture with adjusted and optimized weights to enhance feature representation, employing federated learning to ensure decentralized and privacy-conscious training with different clients. It first performs an image-level classification approach, rendering it unsuitable for fine-grained tumor

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

localization or segmentation to facilitate clinical treatment planning.

A. Al-Saleh et al. [34] provided a privacy-preserving system for detecting brain tumors based on federated learning. This is a mix of ResNet-50 and capsule networks model in a deep learning model. Anisotropic diffusion filtering and morphological preprocessing are used to improve features and reduce noise. The hybrid Gorilla Badger Optimization Algorithm (HGBOA) is used for feature selection and to find the best parameter. Blockchain technology is used to make sure the model is safe and secure. However, the complex multiple-component design increases computational overhead and implementation difficulty.

A decentralized federated learning structure of a healthcare network has been suggested by B. C. Tedeschini et al. [35] and has been confirmed on brain tumor segmentation using a modified U-Net structure. The studies conducted in geographically distributed medical nodes have shown that federated learning (FL) may achieve performance comparable to centralized learning in the segmentation job while preserving data locality. However, even if the framework takes tumor segmentation a step further by using real-world FL deployment, it mostly uses CNN/U-Net backbones and doesn't use global context modeling or interpretability techniques directly. G. Luo et al. [36] synthetically evaluate the effects of the heterogeneity of data distribution on federated learning framework for segmentation on both CT and MRI images, with multi-institutional brain tumor MRI (FeTS). Training the nnU-Net models on centralized and federated data and measuring inter-site distribution differences using measures of statistical distance, the study showed a strong negative correlation between distributional differences and the performance of the federated models. This paper presents some very important pieces of evidence that non-IID data at various institutions can significantly deteriorate FL performance, which is why domain adaptation, solid aggregation strategies, and harmonization methods should be considered in federated tumor segmentation pipes. The study highlights the problem but does not propose a robust solution to handle non-IID data.

V. Kukreja et al. [37] proposed a segmentation-classification synergy framework, which uses dual U-Net segmentation and federated learning of hybrid CNN-Random Forest (CNN-RF) classifiers to analyze MRI-based brain tumors. The approach uses image de-noising (medians, Gaussian filters, Wiener filters), segmentation by means of a dual U-Net architecture, and a federated structure (training classifiers) to maintain privacy among clients. The federated aggregation offered a better global classification compared to the local models that operated in privacy-limited conditions, which is a benefit of collaborative learning. Although such a framework successfully incorporates segmentation and classification into FL, it prioritizes image-level classification and does not use more sophisticated models of global context (e.g., transformers) to segment fine-grained volumes.

Although deep learning and federated learning have achieved significant progress in brain tumor segmentation, there are several limitations that remain. The majority of the methods rely on fully supervised learning, which requires large annotated datasets that are difficult to collect in medical domains due to privacy concerns. Most methods used in federated environments do not scale to work well with non-IID data among clients, and thus cause client drift and performance loss. Additionally, limited work has investigated the use of self-supervised learning techniques such as Masked Autoencoders (MAE) within federated frameworks. Likewise, 3D medical image segmentation is not extensively implemented with optimization methods such as SCAFFOLD, which aims to reduce client drift. Our proposed framework provides the solution by merging self-supervised Masked Autoencoder (MAE) pre-training with client drift-aware SCAFFOLD optimization within the 3D Attention ResUNet architecture.

III. Methodology

This paper presents a novel two phase federated learning approach for the segmentation task of the brain tumor from the MRI input image. The primary objective is not only to focus on segmentation but also to improve the model's robustness and generalization in non-IID federated environments while reducing dependence on labeled data.

In phase 1, the architecture consists of self-supervised federated pre-training via a 3D Masked Autoencoder (MAE) that learns the rich anatomical features of the brain. During the MAE pre-training phase, unlabeled MRI data are utilized by dismissing the segmentation label. The model has the ability to utilize the structural patterns, spatial dependencies, as well as the contextual information of the brain tissues by incorporating a significant part of the input volume and repairing the missing parts. Learned features are highly valued in brain tumor segmentation, after the pre-training phase, is able to distinguish between normal and abnormal regions very effectively. MAE is used in unlabeled data, which decreases the reliance on manually annotated data. The encoder trained during this stage is then transferred to the AttentionResUNet3D segmentation model and it gives strong initialization, enhancing faster convergence, more precise segmentation, and robustness in non-IID federated settings. In phase 2, the SCAFFOLD algorithm is utilized for fine-tuning on a deep 3D Attention-ResUNet that mitigates the client drift

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

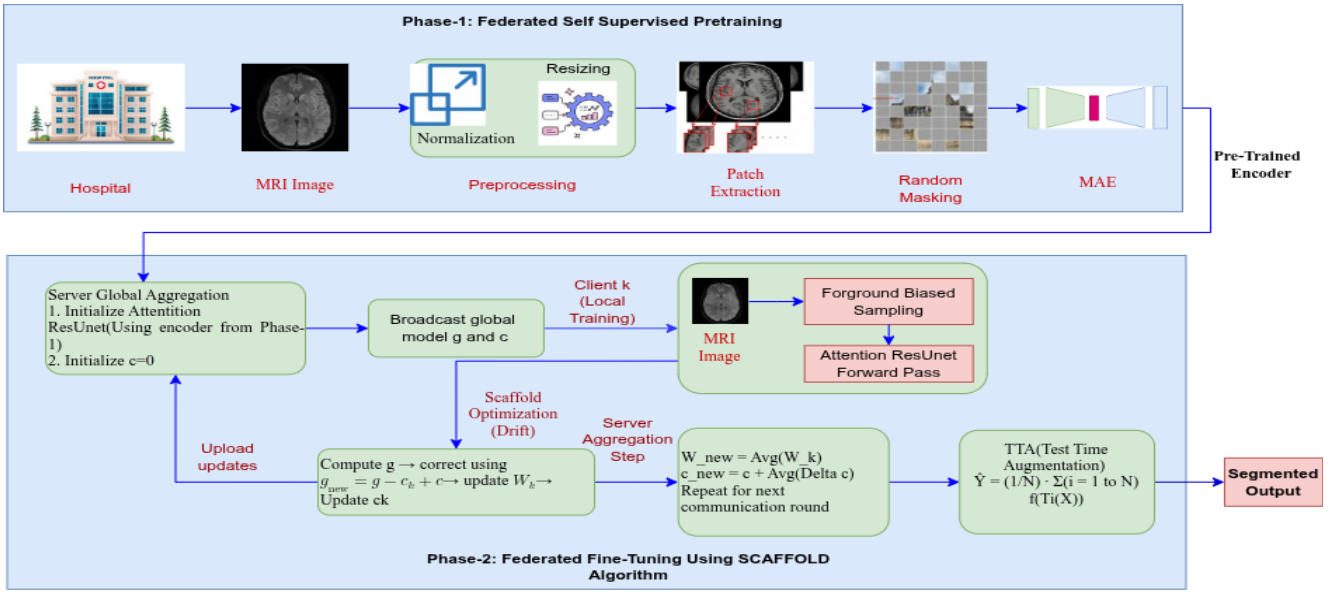


Fig. 2: The proposed two-phase federated learning framework.

problem that arises due to the non-IID data distribution. In contrast to the traditional methods of aggregation, like FedAvg, which might be prone to divergence because of the heterogeneous data, SCAFFOLD adds control variates to fix local updates. The client under this scheme has a local control variate, and the server has a global control variate. During the local training, adaptive changes of the gradient are made between the global and local control variates in such a way that the transformations applied to the client do not stray out of the global objective of the optimization. This error correction The method is useful in enhancing the variation that is a result of data heterogeneity. In order to provide a realistic simulation of a non-IID federated system, the BraTS-2024 dataset was split into K clients (e.g., 10, 50, or 100). The clients were given different subsets of data that have unequal sample sizes, tumor distributions, and different groups of patients in a manner that replicates the real-world scenario where data are different across hospitals.

A. Federated Learning Environment

FL facilitates a decentralized model that is trained across K distinct clients, where each client holds a private MRI datasetD_K. Unlike the centralized paradigms, FL eliminates the need of data sharing, thereby satisfying strict privacy, regulatory, and bandwidth limitations. The global Optimization objective is formulated as a weighted empirical risk minimization as illustrated in Eq. (1) [24].

$$F(\theta) = \sum_{k=1}^{k} \sum_{j} \frac{|D_j|}{|D_k|} F_K(\theta) \quad (1)$$

Here, θ refers to the global model parameters, K is the total number of clients, and Dk refers to the dataset of the k th client. The value of $\frac{|D_j|}{|D_k|}$ shows the contribution of each client depending on its data size and $F_K(\theta))$ denotes the loss function specific to client k and Each client influence on the global model is proportional to the size of the dataset. A criticality of this scheme is the optimization of the divergence that is caused by the non-IID data distribution among the clients. To handle this problem, we incorporate the SCAFFOLD algorithm, as detailed in Section 3.

B. Phase 1: Federated Self-Supervised MAE Pre-training

The first phase incorporates self-supervised learning, extracting semantic anatomical features by using Masked Autoencoders (MAE) without depending on manual annotations. This process is depicted in Fig. 2. Mostly, tissue boundaries captured are uncertain and modality-dependent with reference to 3D MRI images; therefore, the knowledge of such robust representations is essential. For a given input volume x, using mask M, the MAE framework proportionally hides patches, thereby enabling the encoder to infer the missing spatial dependencies. From the Eq.2 [44] anatomical coherence and shape consistency are

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

preserved; furthermore, the loss of foster is also significantly reduced.

$$L_{MAE} = \| M \odot (x - \hat{x}) \|_{2^2} \quad (2)$$

Federated Averaging (FedAvg) is performed for model aggregation at this stage. In this equation L_{MAE} represent the auto encoder loss. Here x is the original MRI input and $\hat{x}$ shows the reconstructed image produced by the model. The term $(x - \hat{x})$ calculate the difference between the original and reconstructed Image. M denotes the masking matrix that selects the hidden regions, and $\odot$ represents element-wise multiplication. This initialization step is vital for stabilizing the fine-tuning phase by enriching anatomical priors. In this equation Eq. (3) [24], where θ^{t+1} represents the updated global model after round *t*. The term θ_{k^t} denotes the local model parameters from client *k* at round *t*. $\sum_{J \epsilon S_T} |D_J|$ represent the total data size of all selected clients in that round. The summation indicates that the global model is updated by taking a **weighted average of local models**, where each client's contribution is proportional to its data size.

$$\theta^{t+1} = \sum_{k \epsilon s} \frac{|D_K|}{\sum_{J \epsilon S_T} |D_J|} \theta_{k^t} \quad (3)$$

C. Phase 2: Federated Fine-Tuning Using SCAFFOLD

Traditional aggregation methods, such as FedAvg, are used to aggregate the results from different clients. Due to the source from which the data was collected, the data at each site are varied, and the aggregation often suffers from client drift. To reduce this issue, we incorporate the SCAFFOLD algorithm, which utilizes control variates to suppress variance. The algorithmic workflow for Phase 2 is illustrated in Fig. 2. In this scheme, each client k maintains a control variate that captures the discrepancy between local and global variates. The updates is defined as in Eq.4 [45].

$$\theta_k^{t+1} = \theta^t - \eta [g_k(\theta^t) - c_k + c] \quad (4)$$

In this equation, θ_k^{t+1} represents the updated local model for client *k*, and θ^t is the global model at round *t*. η is the learning rate, and $g_k(\theta^t)$ is the gradient computed at client *k*. The terms c_k **and** c are the local and global control variates, which help correct client drift and stabilize training. This scheme aligns the local update with the global control variate direction. This mechanism optimizes the gradient noise that happens due to label skew. It assures stabilized convergence in large-scale federated networks (K ≥ 100).

D. Attention ResUNet3D Architecture

The 3D Attention ResUNet is a backbone of the network for the segmentation task, well designed to capture the complex and heterogeneous lesion morphologies of tumor.

1. Attention Mechanism

The attention gates are deployed to enhance the feature selectivity that bypasses the encoder features x using a gating signal g. The mathematical formula for calculating the attention coefficient map is as in Eq.(5) [21].

$$\alpha = \sigma(W_x x + W_g g + b) \quad (5)$$

where α represents the attention coefficient, x is the input feature map from the encoder and g is the gating signal from deeper layers. W_x and W_g are learnable weight parameters, while b is the bias term. The function σ is the sigmoid activation, which highlights important features and suppresses irrelevant ones.

The resulting feature map is calculated as in Eq. (6) [21].

$$\hat{x} = \alpha \cdot x \quad (6)$$

In this equation, $\hat{x}$ represents the refined output feature map, α is the attention coefficient, and x is the input feature map. The operation $\alpha \cdot x$ performs element-wise multiplication, which enhances important features and suppresses irrelevant regions.

2. Residual Learning

In the network architecture, residual blocks are integrated within the encoder of the ResUnet.

E. Foreground-Biased Patch Sampling

To address class imbalance situations, we utilize a biased patch sampling method that mainly focuses on selecting the patches intentionally emphasized in tumorous regions.

The probability Ps(p) of choosing a patch p is derived by Eq. (7) [18].

$$Ps(p) = \begin{cases} \beta, & if\ p\ contains\ tumor \\ 1 - \beta, & if\ p\ contains\ no\ tumor \end{cases} \quad (7)$$

β set to 0.8, frequently check the model for abnormal lesion regions.

F. Supervised Segmentation Loss

A composite loss function is used for model optimization and computed using Eq. (8) [20].In this equation, L represents the total loss used to train the model. L_{Dice} measures the overlap between predicted and ground truth segmentation, while L_{Dice} (Binary Cross-Entropy) measures pixel-wise classification error. The terms λ_1 and λ_2 are weighting factors that control the contribution of each loss component.

$$L = \lambda_1 L_{Dice} + \lambda_2 L_{Dice} \quad (8)$$

By evaluating the overlap between the predicted mask and the ground truth mask, dice l oss evaluates the quality of the model in Eq. (9) [18]. In this equation, L_{Dice} represents the Dice loss used to measure segmentation accuracy. y_i denotes the ground truth value of pixel *i*, and $\hat{y}_i$ is the predicted value. The numerator represents the overlap between prediction and ground truth, while the denominator represents the

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

Algorithm 1: Federated Two-Stage Tumor Segmentation

(1) Initialize global model parameters θ and distribute them to all K clients.
(2) **Stage 1: Federated MAE Pre-training**
(2.1) Each client masks **75%** of the 3D MRI patches and computes the MAE loss (Eq. (1)).
(2.2) Local encoder parameters θ_k are updated independently on each client.
(2.3) The server aggregates client models using FedAvg (Eq. (3)).
(3) Transfer the pre-trained encoder to the AttentionResUNet3D segmentation model.
(4) **Stage 2: Federated Fine-Tuning with SCAFFOLD**
(4.1) Each client performs foreground-biased patch sampling (Eq. (7)).
(4.2) Each client computes the corrected **SCAFFOLD update** (Eq. (4))
(4.3) Optimize the segmentation loss: $L = \lambda_1 L_{Dice} + \lambda_2 L_{Dice}$(Eq. (8))
(4.4) Server aggregates updated model parameters and controls variates from all clients.
(5) **Inference: Test-Time Augmentation**
(5.1) Apply N spatial transformations Ti(X) to the input volume.
(5.2) Compute the final prediction: $\hat{Y}$(Eq. (13))
(6) Return the final segmentation output $\hat{Y}$.

total pixels in both. The loss is subtracted from 1 so that minimizing it improves the overlap between predicted and actual tumor regions.

$$L_{\text{Dice}} = 1 - \frac{2\sum_i y_i \hat{y}_i}{\sum_i y_i + \sum_i \hat{y}_i} \quad (9)$$

Cross-Entropy Loss: L_{CE} (cross-entropy loss) measures the difference between the ground truth y_i and predicted probability $\hat{y}_i$ It penalizes incorrect predictions, helping the model improve pixel-wise accuracy as illustrated as in Eq. (10) [18].

$$L_{CE} = -\sum_i y_i \log(\hat{y}_i) \quad (10)$$

G. Stabilization Mechanisms

1. Gradient Clipping

We apply norm-based gradient clipping to deal with the heterogeneous data in a federated learning setup and compute it using Eq. (11) [12].

This equation ensures that the gradient norm $\|\nabla_\theta\|_2$ does not exceed a threshold τ. It is used for gradient clipping, which helps prevent unstable updates and improves training stability.

$$\|\nabla_\theta\|_2 \leq \tau \quad (11)$$

2. Cosine Annealing Learning Rate

Cosine annealing scheduler is applied to modulate the learning rate ηt. This approach facilitates smooth convergence by gradually reducing the learning rate from a maximum η_{max} to a minimumη_{min} and illustrated as in Eq. (12) [11].

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})(1 + cos(\frac{t}{T}\pi)) \quad (12)$$

H. Test-Time Augmentation (TTA)

To account for predictive uncertainty during inference, we utilize Test-Time Augmentation (TTA). The final prediction is derived by averaging the outputs of N augmented views of the input, and it is illustrated as in Eq. [13] [30]. In this equation, $\hat{Y}$ is the final prediction obtained by averaging outputs over N augmented versions of the input X. Where N = 7 includes rotations and flips along cardinal axes. $(T_i(X)$represents different transformations (like flip/rotation), f and is the model. This improves prediction robustness using test-time augmentation.

$$\hat{Y} = \frac{1}{N}\sum_{i=1}^{N} f(T_i(X)) \quad (13)$$

The Process of implementing stage 1 and stage 2 is presented in Algorithm 1.

IV. Experimental Setup

A. Dataset Description

The proposed method was evaluated using the BraTS 2024 challenge dataset [38], which is widely used for benchmarking glioblastoma segmentation algorithms. The dataset consists of multi-institutional, pre-operative MRI scans and includes four commonly acquired modalities per subject: T1-weighted (T1), contrast-enhanced T1-weighted (T1c), T2-weighted (T2), and Fluid Attenuated Inversion Recovery (FLAIR) [39]. All images were processed using standardized pipelines to maintain spatial alignment, including registration to a common anatomical reference, skull removal, and resampling to an isotropic voxel spacing of 1×1×1. Ground truth annotations describe: Enhancing (ET): This region captures the active, hyperintense tissues. Core (TC): A region formed by merging the necrotic core with the enhancing. Whole (WT): The most inclusive region, combining the peritumoral oedema with the active core and enhancing structures.

B. Implementation Details

This section explains how the proposed model was trained and tested. It includes information about the data preprocessing and how the dataset was distributed across the client for training and testing in a federated learning environment, the network architecture, the hyperparameters selected, and the methods used to evaluate the model's performance. Lastly, Comparative Analysis.

Our proposed model was executed on the PyTorch ecosystem (version 2.1) on a high-performance

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

Table 1: Comparison of Segmentation Performance Metrics with other Methods.

Method	Dice WT	Dice TC	Dice ET	Average Dice	Sensitivity	Specificity	Precision
Wang et al. [41]	0.88	0.81	0.76	0.817	0.807	0.80	0.85
Satushe, et al. [42]	0.87	0.87	0.49	0.743	0.743	0.98	0.99
Kamnitsas, et al. [43]	0.87	0.77	0.783	0.808	0.816	0.82	0.87
Proposed Method	0.862	0.811	**0.804**	**0.826**	**0.882**	**0.988**	**0.999**

workstation utilizing a single NVIDIA A100 GPU (80GB VRAM), and the dataset was partitioned across K=4 clients.

1. Preprocessing and Augmentation

In this study, we have used Z-Score normalization to achieve the input data standardization for each MRI image. Volumetric Patching is used to reduce the impact of irrelevant details in the background of the MRI image. Volumes were cropped to the brain region. We have used a foreground- biased sampling scheme to handle the class imbalance problem.

2. Network Architecture

We have used 3D Attention ResUNet as the backbone for segmentation, which has three modules: encoder path, attention mechanism, and decoder path. The encoder consists of a series of residual convolutional blocks that expand with the feature depth of 32, 64, 128, and 256 at each layer. The size of the convolution is 3x3x3, followed by batch-relevant information and ignoring irrelevant information. Lastly, a decoder is used to regenerate the features with the help of a transpose convolution layer.

3. Training Protocols

The training task is divided into two phases to optimize feature representation. In phase 1, federated self-supervised pre-training (MAE) is incorporated to learn the anatomical structure of the brain. In phase 2, federated fine-tuning (SCAFFOLD), the encoder of the attention ResUNet3D was initialized with pre-trained weights obtained from phase 1. The SCAFFOLD algorithm was used to tackle non-IID data.

4. Hyperparameter Configuration

A total of 250 communications rounds were conducted for the
training purpose. The client performed one local epoch at each round. AdamW algorithm is used to optimize the model.

$$L_{Total} = L_{Dice} + L_{BCE} \quad (14)$$

Adjusting the learning rate by adopting a cosine annealing strategy. It is illustrated as in Eq. (14) [18] [20].

5. Evaluation Metrics

We have used Dice Similarity Coefficient, Sensitivity, Specificity, and Precision to evaluate model performance, and computed them using Eq. (15) & (16) [40].

In such equations, TP (True Positive) is used to depict correctly predicted tumor pixels, TN (True Negative) is used to show correctly predicted non-tumor pixels, FP (False Positive) is used to present non-tumor pixels falsely predicted to be tumor, and FN (False Negative) depicts tumor pixels that the model missed. The Dice Similarity Coefficient (DSC) is a measure of overlap between prediction and ground truth. Sensitivity is a measure of the ability to find tumor regions correctly, Specificity is a measure of the ability to avoid false positives, and Precision is a measure of how well the positive prediction can be relied upon.

$$DSC = \frac{2TP}{2TP + FP + FN} \qquad Sensitivity = \frac{TP}{TP + FN} \quad (15)$$

$$SpecificitY = \frac{TN}{TN + FP} \qquad Precision = \frac{TP}{TP + FP} \quad (16)$$

6. Comparative Analysis Strategy.

The summary of the quantitative results is represented in Table 1 and is discussed exhaustively in the Results section. Our proposed approach has the advantage of doing the segmentation precisely over the baseline ACNN model [41]. The model achieves a Dice coefficient of 0.804, sensitivity of 0.988, and specificity of 0.999, thus showing better performance as compared to the comparative methods such as [42] and [43].

V. Results

In our experiment, we analyzed the performance of our model based on different quantitative metrics. We tested our model on 10 clients and then increased the number of clients by 50 and then by 100. The final result is evaluated from the final stage 2. We have also performed validation for each communication round and applied test-time augmentation (TTA) for the final

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

inference. The aggregation predictions from seven transformations like flip and rotation, were used to improve the segmentation results.

A. Comparison with State-of-the-Art Benchmark

To evaluate the effectiveness of the proposed framework, a comparative analysis was conducted against existing state-of-the-art methods using standard segmentation metrics, including Dice coefficient, sensitivity, specificity, and precision. These metrics provide a comprehensive understanding of the model's ability to accurately segment tumor regions while minimizing false predictions. Table 1 presents the comparison of the proposed method with Wang et al. [41], Satushe et al. [42], and Kamnitsas et al. [43] in terms of Dice coefficient, sensitivity, specificity, and precision. The proposed approach has Dice scores of 0.862 (WT), 0.811 (TC), and 0.804 (ET), which leads to the greatest average Dice coefficient of 0.826 of all methods. Comparatively, Wang et al. [41] give an average of 0.817, Kamnitsas et al. [43] have 0.808, and Satushe et al. [42] have the lowest of 0.743. This means that the overall segmentation accuracy of the proposed model is better. In the case of the Whole Tumor (WT) region, the suggested approach has a Dice score of 0.862, which is similar to Wang et al. [41] (0.88) and slightly worse than other algorithms, yet competitive. In the case of the Tumor Core (TC) region, the proposed model attains 0.811 and is higher than Kamnitsas et al. [43] (0.77) and similar to Wang et al. [41] (0.81), thus demonstrating consistent segmentation of core tumor structures. In the Enhancing Tumor (ET) region, the proposed method reaches 0.804, which is much higher than Satushe et al. [42] (0.49) and a little higher than Wang et al. [41] (0.76) and Kamnitsas et al. [43] (0.783). This enhancement is critical since ET is the most difficult area since it is small and is bounded by irregularities. The proposed method has a sensitivity of 0.882, which is better than wang et al. [41], Satushe et al. [42] (0.743), and Kamnitsas et al. [43] (0.82), meaning it is more effective in detecting tumors. Specificity of 0.988 is also greater than that of Wang et al. [41] (0.80) and Kamnitsas et al. [43] (0.82), indicating improved ability to reduce false positives. Similar to Satushe et al. [42] (0.98), indicating a high capability of reducing false positives. The proposed method is the most precise, with 0.999, which is greater than the precision of Wang et al. [41] (0.85), Satushe et al. [42] (0.99), and Kamnitsas et al. [43] (0.87), meaning that the predictions are highly reliable.

B. Quantitative Performance Analysis

Table 2 presents the segmentation performance of the proposed model under the federated learning Configurations 10, 50, and 100 clients, providing a detailed analysis of scalability and robustness under increasing data heterogeneity. In the 10-client setting,

Table 2. Segmentation performance for 10, 50 and 100 clients in FL environment.

Metric	WT	TC	ET	Mean
Segmentation performance for 10 clients.				
Dice Coefficient	0.862	0.811	0.804	0.826
Sensitivity	0.952	0.847	0.848	0.882
Specificity	0.987	0.998	0.998	0.995
Precision	0.788	0.787	0.767	0.781
Segmentation performance for 50 clients.				
Dice Coefficient	0.8909	0.8041	0.7783	0.8244
Sensitivity	0.8796	0.8172	0.8075	0.8348
Specificity	0.9954	0.9984	0.9982	0.9974
Precision	0.9028	0.7956	0.7588	0.8191
Segmentation performance for 100 clients.				
Dice Coefficient	0.8797	0.8017	0.7735	0.8183
Sensitivity	0.8974	0.8261	0.8241	0.8492
Specificity	0.9934	0.9984	0.9981	0.9966
Precision	0.8635	0.7846	0.7370	0.7950

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

the model shows stable performance with Dice coefficients of 0.862 (WT), 0.811 (TC), and 0.804 (ET), indicating its ability to effectively learn tumor representations across different regions.The high WT score shows the global tumor structures, whereas the large TC and ET scores reveal the ability of the model to learn complex sub-regions. The model shows strong sensitivity, with values of **0**.952 (WT), 0.847 (TC), and 0.848 (ET)**,** giving an overall average of **0.882**, which indicates its ability to effectively detect tumor regions. In addition, it achieves a high specificity, with values of 0.987 (WT), 0.998 (TC), and 0.998 (ET), mean of 0.8348, indicating a slight reduction in detection capability due to increased data heterogeneity.

The specificity improves further to 0.9954 (WT), 0.9984 (TC), and 0.9982 (ET), with a mean of 0.9974, confirming that the model becomes even more effective at avoiding false positives. The precision increases significantly to 0.9028 (WT), 0.7956 (TC), and 0.7588 (ET), with a mean of 0.8191, indicating improved prediction confidence, especially for WT. The performance is stable because the SCAFFOLD optimization algorithm is resistant to client drift, as it

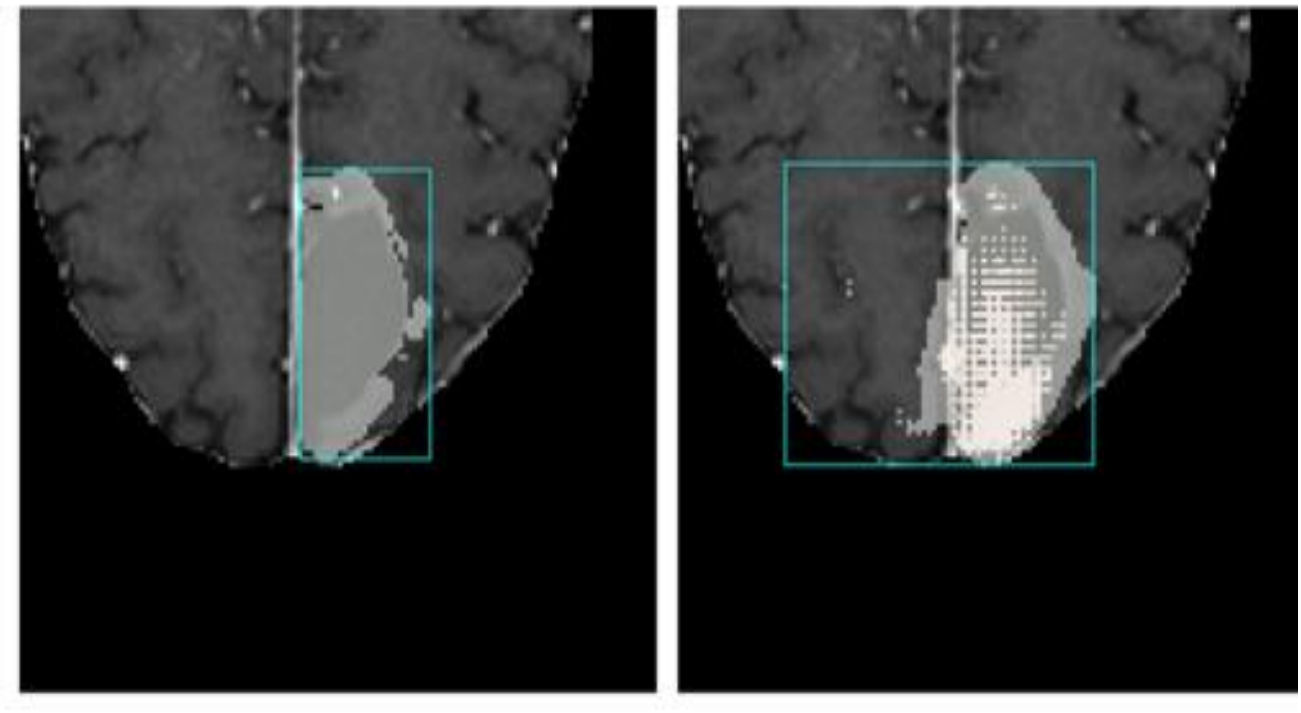


Fig. 3: Qualitative comparison: Ground Truth Mask (Left) vs. Predicted Segmentation (Right). The model accurately delineates the tumor boundaries with high fidelity.

and the mean specificity of **0.995** indicates that the model effectively reduces false positives in all tumor areas. This sensitivity and specificity imply that the model is accurate as well as reliable in clinical segmentation activities. Foreground-biased sampling is a crucial component of this performance in that it focuses on tumor areas during training, thus enhancing the segmentation quality, particularly of smaller and irregular areas like ET. Scaling to the 50-client environment, the model still has a mean Dice coefficient of 0.8244 with region-wise Dice values of 0.8909 (WT), 0.8041 (TC), and 0.7783 (ET), which shows a small change of about 0.19% relative to the 10-client setup. Such low degradation means that the model is efficient in the case of higher data heterogeneity among clients. The sensitivity decreases to 0.8796 (WT), 0.8172 (TC), and 0.8075 (ET), with a

Table3: **Impact of number of clients on Mean Dice Coefficient.**

Clients	Mean Dice Coefficient
10	0.826
50	0.8244
100	0.8183

uses control variates to correct local updates. The fact that the performance of the model is similar in WT, TC, and ET regions is yet another confirmation that the model has not lost its segmentation ability within both large and small tumor substructures.

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

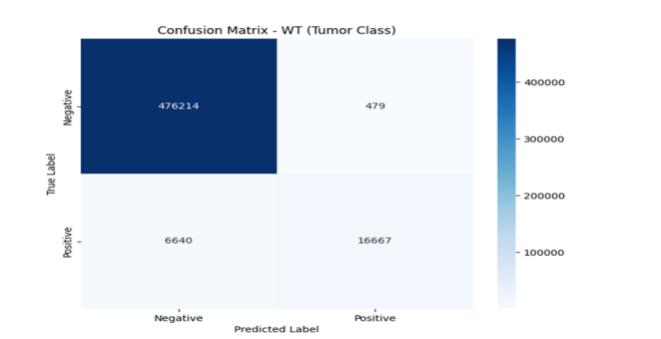


Fig. 4: Confusion Matrix for the Tumor Core (TC). High diagonal values indicate robust classification accuracy

Table 3, in this table, we consolidate all client results across all federated learning frameworks. Although data distribution has become more varied, the model does not lose its generalization across various domains of clients, and it is highly robust. It means that the suggested framework is effective to scale even in the most distributed environment. At the more demanding **100-client** scenario, the model has 0.8796 (WT), 0.8172 (TC), and 0.8075 (ET) with the mean Dice coefficient of 0.8183 that represent, a total reduction of about 0.93% in comparison with the 10-client scenario Even when the number of clients increases to 10 the model achieves a mean Dice coefficient of 0.826, representing the baseline performance with relatively lower data heterogeneity. When the number of clients increases to 50, the mean Dice coefficient slightly decreases to 0.8244. The Dice coefficient proves that the proposed framework is well-suited for a federated environment. This small decrease suggests that the model is relatively stable in performance when it is trained on more heterogeneous and distributed data. In the 100-client configuration, the Dice coefficient further decreases to 0.8183 as compared to the 50- client setup. The overall reduction between 10 and 100 clients is only 0.93, which is very low. The minimal performance decrease shows that the model is robust to increasing data heterogeneity and can generalize effectively across multiple clients.

C. Qualitative Analysis

As shown in Fig. 3, we present a visual comparison between the ground truth and the actual predicted segmentation of the brain tumor region. The left side of the image shows the ground-truth mask, with the tumor manually annotated by the expert, and the region highlighted within the bounding box. This serves as the standard for assessing our model performance. The right side of the image shows the predicted segmentation output evaluated by the proposed model. The model successfully detected the tumor region within the same area. The predicted output closely matches the ground truth, indicating that our model accurately captures the tumor location and overall structure. The bounding box on both sides of the image represents the region of interest (ROI), which helps to clearly assess the alignment between the ground truth and the prediction.

Then, we further calculate the confusion matrix for the Tumor core (TC), and it indicates a dense diagonal, high voxel level accuracy, and very little confusion is there between classes, as shown in Fig. 4.(on the next page). The Dice curve stabilizes after a few rounds of communication, signifying convergence of the model. This action proves that the model has acquired ideal feature representations of the tumor. The Dice Coefficient steadily increases, confirming that the MAE-based setup helps our model converge faster in the initial training rounds.

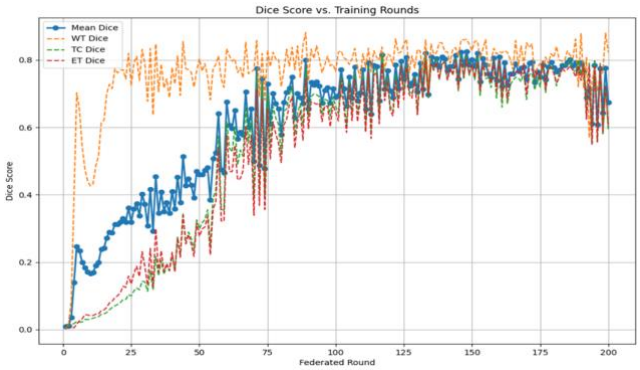


Fig. 5: Training Evolution: Dice Coefficients over communication rounds for the 50-client federated setup.

Finally, our model demonstrates training stability, as shown in Fig. 5 . The gradual nature of the curve, without significant variations, evidences consistent training in the presence of a non-IID data distribution, which is one of the main problems of federated learning.

VI. Discussion

The experimental results demonstrate that the proposed FedBrain-3DMRI framework achieves robust and scalable performance for brain tumor segmentation under challenging federated learning environments. The following discussion provides a detailed interpretation of the results, comparison with existing methods, and analysis of limitations and clinical implications.

A. Effectiveness of the Two-Stage Federated Framework

The two-stage pipeline (self-supervised Masked Autoencoder (MAE) pre-training with federated fine-tuning) proposed is a successful solution to the problem of limited annotated medical data. When using

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

10 clients, the model has a Mean Dice coefficient of 0.826 (Table 3), meaning that there is high overlap between the predicted and ground truth tumor regions. Such high results may be explained by the fact that the MAE-based pre-training is trained on the unlabeled MRI data, which learns strong anatomical representations. This methodology is especially effective in the Enhancing Tumor (ET) area, with the model scoring a Dice of 0.804, which outperforms several fully supervised baselines that struggle with the irregular boundaries of ET regions.

B. Impact of SCAFFOLD on Federated Optimization

The use of the SCAFFOLD algorithm can greatly enhance the stability of training in non-IID federated settings. As shown in Table 3, increasing the number of clients from 10 to 100 results in only a significant drop in the mean Dice score from 0.826 to 0.8183, corresponding to a reduction of approximately 0.93%. This low degradation shows that SCAFFOLD is very effective in reducing client drift because it allows local updates to be synchronized with the global target. This stability is essential in a real-life healthcare system, where the heterogeneity of data in different institutions is unavoidable.

C. Comparison with State-of-the-Art Methods

The proposed framework outperforms several existing methods in terms of segmentation accuracy. As shown in Table 1, the model achieves an average Dice score of 0.826, which is higher than Wang et al. [41] (0.817), Kamnitsas et al. [43] (0.808), and Satushe et al. [42] (0.743). The proposed approach also shows high performance in the ET region (0.804), indicating its ability to effectively delineate complex tumor subregions. This enhancement proves the benefit of combining self-supervised learning and federated optimization.

D. Analysis of Class-wise Performance

The segmentation performance across different tumor sub-regions, as presented in Table 2, provides important insights into the behavior of the proposed framework. The Dice scores of 0.90 for Whole Tumor (WT), 0.84 for Tumor Core (TC), and 0.804 for Enhancing Tumor (ET) indicate that the model responds differently to regions with varying structural complexity. The higher performance in the WT region (0.90) suggests that the model effectively captures larger and more continuous tumor regions, where spatial patterns are easier to learn. In contrast, the comparatively lower scores for TC (0.84) and ET (0.804) highlight the inherent challenges associated with segmenting smaller, heterogeneous, and irregular tumor sub-regions. Despite these challenges, the performance in the ET region (0.804) remains strong, indicating that the model is capable of identifying clinically significant active tumor regions. Similarly, the TC score (0.84) demonstrates that the model maintains reliable detection of core tumor structures, which are critical for assessing tumor progression. These variations in performance reflect the contribution of the model design, where attention mechanisms enhance the localization of relevant features, and foreground-biased sampling improves the representation of difficult tumor regions during training. In addition, MAE-based pre-training supports improved feature learning, enabling better delineation of complex tumor structures.

E. Metric-Level Interpretation and Structural Contributions

The model has a sensitivity of 0.882, specificity of 0.995, and a precision of 0.999 (Table 1), which means that the model has a strong detection and a high degree of reliability. The large sensitivity indicates that the model is effective at detecting tumor regions, whereas the large specificity indicates that the model is effective at reducing false positives. The accuracy figure also substantiates that predicted tumor regions have great accuracy. The Attention ResUNet3D architecture supports these findings by allowing attention mechanisms to increase tumor-relevant features and residual connections to learn deep features stably.

E. Training Strategies and Clinical Implications

Foreground-biased sampling yields better segmentation results, as it targets tumor areas in training, especially small and imbalanced areas, like ET. Test-Time Augmentation (TTA) is an additional method of improving the robustness of the prediction by averaging several augmented predictions. Clinically, the proposed framework is appropriate to be implemented in multi-institutional healthcare systems because of its capacity to attain a high level of segmentation without distributing raw patient information, which is needed to adhere to privacy laws including HIPAA and GDPR.

F. Limitations

Despite the strong performance of the proposed framework, several limitations should be acknowledged. The first is that the model presupposes that all four MRI modalities (T1, T1c, T2, and FLAIR) are available at each client, which might not always be possible in a real-world clinical setting when partial data is frequently encountered. Second, federated learning and the use of 3D architectures generate large computational and communication overhead. Multi-client multi-volume training can be expensive in terms of memory and processing resources, and is thus potentially limited by resource-constrained healthcare environments.

VII. Conclusion

In this study, we proposed FedBrain-3DMRI, a two-stage federated learning framework for 3D brain tumor segmentation aimed at addressing challenges related to data privacy, limited annotations, and non-IID data

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

distributions in multi-institutional environments. The experimental results demonstrate strong performance, achieving a mean Dice coefficient of 0.826 in the 10-client setting, with slight reductions to 0.8244 and 0.8183 for 50 and 100 clients, respectively, indicating high scalability. Additionally, the model achieves a sensitivity of 0.882, a specificity of 0.988, and a precision of 0.999, confirming reliable tumor detection with reduced false positives. Furthermore, integrating self-supervised MAE pre-training enhances feature representations while reducing reliance on labeled data, and using the SCAFFOLD algorithm effectively mitigates client drift, ensuring stable performance across heterogeneous datasets. Techniques such as foreground-biased sampling and test-time augmentation further improve segmentation accuracy and prediction stability. However, the proposed framework assumes complete multimodal MRI availability and involves higher computational complexity due to 3D modeling and federated training. In future work, efforts will focus on handling missing modalities, improving computational efficiency through lightweight architectures, and incorporating explainability and uncertainty estimation to enhance clinical applicability.

Acknowledgment

The authors would like to express sincere gratitude to the supervisor for invaluable guidance and unwavering support throughout the course of this research.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

No datasets were generated or analyzed during the current study.

Author Contribution

First author contribution: Conceptualization, Methodology, Experiments, Coding Result analysis, Write & review original draft. The second author contributed to supervision and investigation.

Declarations

Ethical Approval

Not Applicable

Consent for Publication Participants.

Consent for publication was given by all participants

Competing Interests

The authors declare no competing interests.

References

[1] M. L. Bondy, M. E. Scheurer, B. Malmer, J. S. Barnholtz-Sloan, F. G. Davis, D. Il'Yasova, *et al.*, "Brain tumor epidemiology: Consensus from the Brain Tumor Epidemiology Consortium," *Cancer*, vol. 113, no. S7, pp. 1953–1968,2008. https://doi.org/10.1002/cncr.23741

[2] T. S. Armstrong, E. Vera-Bolanos, A. A. Acquaye, M. R. Gilbert, H. Ladha, and T. Mendoza, "The symptom burden of primary brain tumors: Evidence for a core set of tumor- and treatment-related symptoms," *Neuro-Oncology*, vol. 18, no. 2, pp. 252–260, 2015. https://doi.org/10.1093/neuonc/nov166

[3] F. Ullah, M. Nadeem, M. Abrar, F. Amin, A. Salam, and S. Khan, "Enhancing brain tumor segmentation accuracy through scalable federated learning with advanced data privacy and security measures," *Mathematics*, vol. 11, Art. no. 4189, 2023. https://doi.org/10.3390/math11184189

[4] A. N. Onaizah, Y. Xia, and K. Hussain, "FL-SiCNN: An improved brain tumor diagnosis using Siamese convolutional neural network in a peer-to-peer federated learning approach," *Alexandria Engineering Journal*, vol. 114, pp. 1–11, 2025. https://doi.org/10.1016/j.aej.2024.11.063

[5] D. Li, D. Han, T. H. Weng, Z. Zheng, H. Li, H. Liu, *et al.*, "Blockchain for federated learning toward secure distributed machine learning systems: A systemic survey," *Soft Computing*, vol. 26, no. 9, pp. 4423–4440, 2022. https://doi.org/10.1007/s00521-021-06491-5

[6] T. K. Dang, X. Lan, J. Weng, and M. Feng, "Federated learning for electronic health records," *ACM Trans. Intell. Syst. Technol.*, vol. 13, pp. 1–17, 2022. https://doi.org/10.1145/3485730

[7] S. Alphonse, F. Mathew, K. Dhanush, and V. Dinesh, "Federated learning with integrated attention multiscale model for brain tumor segmentation," *Scientific Reports*, vol. 15, Art. no.11889,2025. https://doi.org/10.1038/s41598-025-96416-6

[8] R. Ahsan, I. Shahzadi, F. Najeeb, and H. Omer, "Brain tumor detection and segmentation using deep learning," *Magn. Reson. Mater. Phys. Biol. Med.*, vol. 38, pp. 13–22, 2025. https://doi.org/10.1007/s10334-024-01203-5

[9] A. Anaya-Isaza and L. Mera-Jiménez, "Data augmentation and transfer learning for brain tumor detection in magnetic resonance imaging," *IEEE Access*, vol. 10, pp. 23217–

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

23233, 2022.
https://doi.org/10.1109/ACCESS.2022.3154061

[10] T. Shelatkar, Urvashi, M. Shorfuzzaman, A. Alsufyani, and K. Lakshmanna, "Diagnosis of brain tumor using lightweight deep learning model with fine-tuning approach," *Comput. Math. Methods Med.*, vol. 2022, Art. no. 2858845, 2022.
https://doi.org/10.1155/2022/2858845

[11] M. F. Almufareh, M. Imran, A. Khan, M. Humayun, and M. Asim, "Automated brain tumor segmentation and classification in MRI using YOLO-based deep learning," *IEEE Access*, vol. 12, pp. 16189–16207, 2024. https://doi.org/10.1109/ACCESS.2024.3359418

[12] M. Rizwan, A. Shabbir, A. R. Javed, M. Shabbir, T. Baker, and D. A. J. Obe, "Brain tumor and glioma grade classification using Gaussian convolutional neural network," *IEEE Access*, vol. 10, pp. 29731–29740, 2022.
https://doi.org/10.1109/ACCESS.2022.3153108

[13] A. S. Musallam, A. S. Sherif, and M. K. Hussein, "A new convolutional neural network architecture for automatic detection of brain tumors in magnetic resonance imaging images," *IEEE Access*, vol. 10, pp. 2775–2782,2022. https://doi.org/10.1109/ACCESS.2022.3140289

[14] S. Li, J. Liu, and Z. Song, "Brain tumor segmentation based on region of interest-aided localization and segmentation U-Net," *Int. J. Mach. Learn. Cybern.*, vol. 13, pp. 2435–2445, 2022.
https://doi.org/10.1007/s13042-022-01536-4

[15] U. Bhimavarapu, N. Chintalapudi, and G. Battineni, "Brain tumor detection and categorization with segmentation using an improved unsupervised clustering approach and machine learning classifier," *Bioengineering*, vol. 11, no. 3, Art. no. 266, 2024. https://doi.org/10.3390/bioengineering11030266

[16] S. Anantharajan, S. Gunasekaran, T. Subramanian, *et al.*, "MRI brain tumor detection using deep learning and machine learning approaches," *Measurement: Sensors*, vol. 31, Art. no. 101026, 2024. https://doi.org/10.1016/j.measen.2024.101026

[17] J. Walsh, A. Othmani, M. Jain, and S. Dev, "Using U-Net network for efficient brain tumor segmentation in MRI images," *Healthcare Analytics*, vol. 2, Art. no. 100098, 2022. https://doi.org/10.1016/j.health.2022.100098

[18] Y. Jiang, Y. Zhang, X. Lin, J. Dong, T. Cheng, and J. Liang, "SwinBTS: A method for 3D multimodal brain tumor segmentation using swin transformer," *Brain Sciences*, vol. 12, no. 6, Art. no. 797, 2022. https://doi.org/10.3390/brainsci12060797

[19] L. ZongRen, W. Silamu, W. Yuzhen, and W. Zhe, "DenseTrans: Multimodal brain tumor segmentation using swin transformer," *IEEE Access*, vol. 11, pp. 42895–42908, 2023. https://doi.org/10.1109/ACCESS.2023.3272055

[20] Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, and Y. Liu, "Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI," *Information Fusion*, vol. 91, pp. 376–387, 2023. https://doi.org/10.1016/j.inffus.2022.10.022

[21] C. Yan, J. Ding, H. Zhang, K. Tong, B. Hua, and S. Shi, "SEResU-Net for multimodal brain tumor segmentation," *IEEE Access*, vol. 10, pp. 117033–117044, 2022. https://doi.org/10.1109/ACCESS.2022.3214309

[22] M. E. Yahiaoui, M. Derdour, R. Abdulghafor, S. Turaev, M. Gasmi, A. Bennour, A. Aborujilah, and M. A. Sarem, "Federated learning with privacy-preserving techniques for multi-institutional three-dimensional brain tumor segmentation," *Diagnostics*, vol. 14, Art. no. 2891, 2024. https://doi.org/10.3390/diagnostics14242891

[23] A. Giri, P. Thapa, J. S. Banu, S. Poudyal, B. Rijal, and S. Karki, "Harnessing ResUHybridNet with federated learning: A new paradigm in brain segmentation," *Revue d'Intelligence Artificielle*, vol. 38, pp. 765–772, 2024. https://doi.org/10.18280/ria.380303

[24] M. Islam, M. T. Reza, M. Kaosar, and M. Z. Parvez, "Effectiveness of federated learning and CNN ensemble architectures for identifying brain tumors using MRI images," *Neural Processing Letters*, vol. 55, no. 4, pp. 3779–3809, 2023. https://doi.org/10.1007/s11063-022-11014-1

[25] Q. Dai, D. Wei, H. Liu, J. Sun, L. Wang, and Y. Zheng, "Federated modality-specific encoders and multimodal anchors for personalized brain tumor segmentation," in *Proc. AAAI Conf. Artif. Intell.*, pp. 1445–1453, 2024. https://doi.org/10.1609/aaai.v38i2.27909

[26] F. Wagner, W. Xu, P. Saha, Z. Liang, D. Whitehouse, D. Menon, V. Newcombe, N. Voets, J. A. Noble, and K. Kamnitsas, "Feasibility of federated learning from client databases with different brain diseases and MRI modalities," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, pp. 357–367, 2025.

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

http://doi.org/10.1109/WACV61041.2025.00045

[27] M. Grama, M. Musat, L. Muñoz-González, J. Passerat-Palmbach, D. Rueckert, and A. Alansary, "Robust aggregation for adaptive privacy-preserving federated learning in healthcare," *arXiv preprint*, 2020. https://doi.org/10.48550/arXiv.2009.08294

[28] Q. U. A. Mastoi, S. Latif, S. Brohi, J. Ahmad, A. Alqhatani, M. S. Alshehri, *et al.*, "Explainable AI in medical imaging: An interpretable and collaborative federated learning model for brain tumor classification," *Frontiers in Oncology*, vol. 15, Art. no. 1535478, 2025. https://doi.org/10.3389/fonc.2025.1535478

[29] M. Islam, M. T. Reza, M. Kaosar, and M. Z. Parvez, "Effectiveness of federated learning and CNN ensemble architectures for identifying brain tumors using MRI images," *Neural Processing Letters*, vol. 55, no. 4, pp. 3779–3809, 2023. https://doi.org/10.1007/s11063-022-11014-1

[30] S. Sharma, K. Guleria, A. Dogra, D. Gupta, S. Juneja, S. Kumari, and A. Nauman, "A privacy-preserved horizontal federated learning approach for malignant glioma detection using distributed data silos," *PLOS ONE*, vol. 20, Art. no. e0316543, 2025. https://doi.org/10.1371/journal.pone.0316543

[31] D. H. Mahlool and M. H. Abed, "Distributed brain tumor diagnosis using a federated learning environment," *Bull. Electr. Eng. Informatics*, vol. 11, no. 6, pp. 3313–3320, 2022. https://doi.org/10.11591/eei.v11i6.4131

[32] Y. M. Elbachir, D. Makhlouf, G. Mohamed, M. M. Bouhamed, and K. Abdellah, "Federated learning for multi-institutional 3D brain tumor segmentation," in *Proc. Int. Conf. Pattern Anal. Intell. Syst. (PAIS)*, pp. 1–8, Apr. 2024. https://doi.org/10.1109/PAIS62114.2024.10541292

[33] E. Albalawi, M. T. R., A. Thakur, V. V. Kumar, M. Gupta, S. B. Khan, and A. Almusharraf, "Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor," *BMC Med. Imaging*, vol. 24, no. 1, Art. no. 110, 2024. https://doi.org/10.1186/s12880-024-01261-0

[34] A. Al-Saleh, G. G. Tejani, S. Mishra, S. K. Sharma, and S. J. Mousavirad, "A federated learning-based privacy-preserving image processing framework for brain tumor detection from CT scans," *Scientific Reports*, vol. 15, no. 1, Art. no. 23578, 2025. https://doi.org/10.1038/s41598-025-07807-8

[35] B. C. Tedeschini, S. Savazzi, R. Stoklasa, L. Barbieri, I. Stathopoulos, M. Nicoli, and L. Serio, "Decentralized federated learning for healthcare networks: A case study on tumor segmentation," *IEEE Access*, vol. 10, pp. 8693–8708, 2022. https://doi.org/10.1109/ACCESS.2022.3141913

[36] G. Luo, T. Liu, J. Lu, X. Chen, L. Yu, J. Wu, et al., "Influence of data distribution on federated learning performance in tumor segmentation," Radiology: Artificial Intelligence, vol. 5, no. 3, Art. no. e220082, 2023. https://doi.org/10.1148/ryai.220082

[37] V. Kukreja, A. Dogra, R. K. Kaushal, S. Mehta, S. Vats, and B. Goyal, "Segmentation synergy with a dual U-Net and federated learning with CNN-RF models for enhanced brain tumor analysis," Current Medical Imaging, vol. 20, no. 1, 2024. https://doi.org/10.2174/0115734056312765240905104112

[38] S. Bakas, et al., "The 2024 Brain Tumor Segmentation (BraTS) challenge," Synapse, 2024. https://doi.org/10.48550/arXiv.2405.18368

[39] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. H. Ha, M. Rozycki, et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BraTS challenge," arXiv preprint, 2018. https://doi.org/10.48550/arXiv.1811.02629

[40] A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, no. 1, p. 29, 2015. https://doi.org/10.1186/s12880-015-0068-x

[41] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in Proc. MICCAI BrainLesion Workshop, pp. 178–190, 2017. https://doi.org/10.48550/arXiv.1709.00382

[42] V. Satushe, V. Vyas, S. Metkar, and D. P. Singh, "Advanced CNN architecture for brain tumor segmentation and classification using the BraTS-GoAT 2024 dataset," Current Medical Imaging, early access, 2025. https://doi.org/10.2174/0115734056344235241217155930

[43] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," Medical Image Analysis, vol. 36, pp. 61–78, 2017. https://doi.org/10.1016/j.media.2016.10.004

Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596

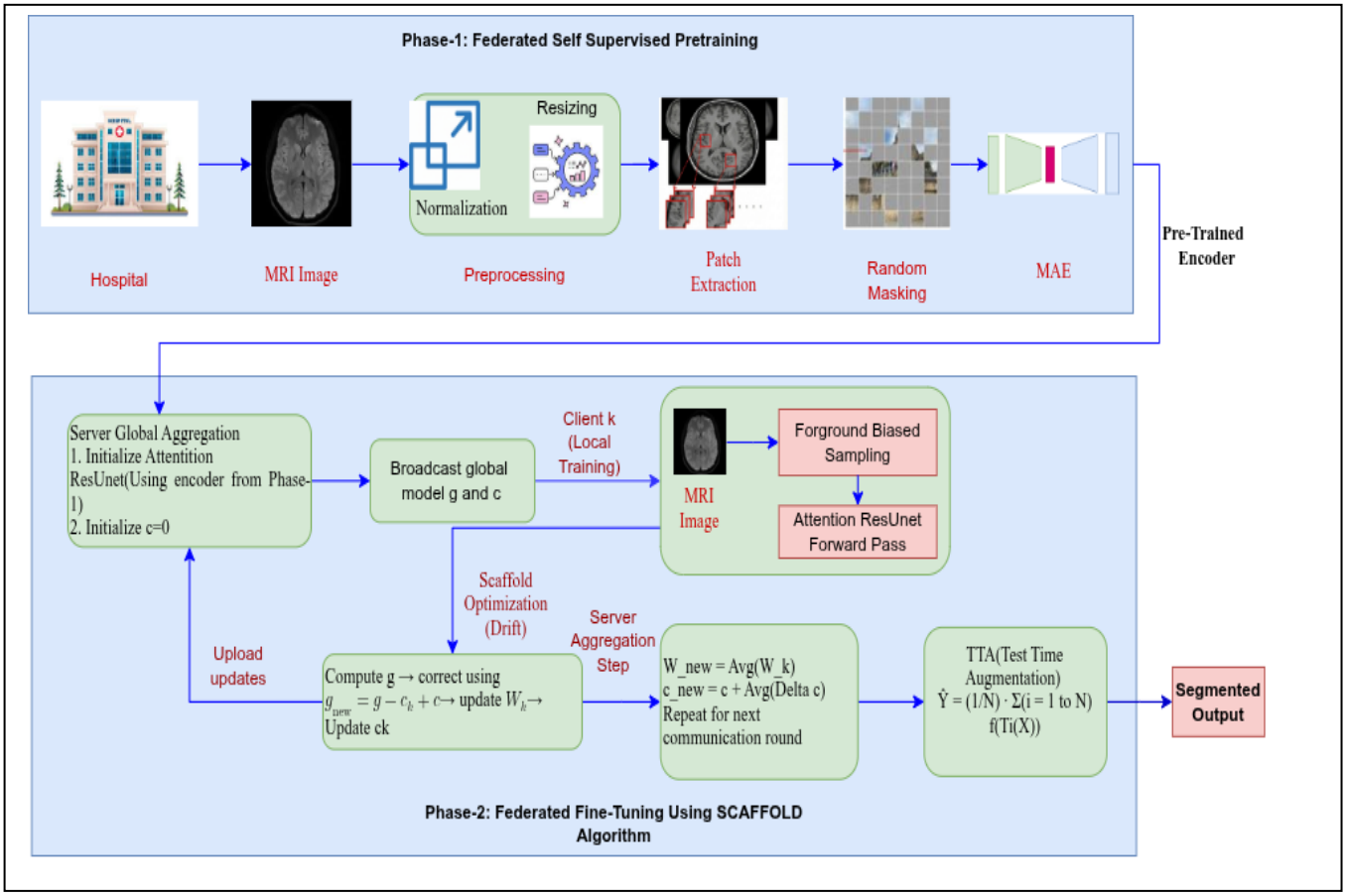
[44] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 16000–16009. https://doi.org/10.48550/arXiv.2111.06377

[45] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. U. Stich, and A. T. Suresh, "SCAFFOLD: Stochastic controlled averaging for federated learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2020, pp. 5132–5143. https://doi.org/10.48550/arXiv.1910.06378

Author Biography

Neeshu Kumari completed her Bachelor of Engineering in Computer Science and Engineering from Uttar Pradesh Technical University (UPTU) in 2010, and received her M.Tech degree in Computer Science and Engineering from GLA University in 2013. She is currently pursuing a Ph.D. in Computer Science and Engineering from Parul University. She has over 8 years of experience in academia. Currently, she serves as an Assistant Professor in the Department of Computer Science and Engineering at GSFC University, and her research interests include machine learning, deep learning, computer vision, and wireless networking. She had guided many major and minor projects of undergraduate students.
She is eager to develop digital study materials to provide her students with easy access during their studies.

Dr. Chintan Thacker received his Ph.D. in Artificial Intelligence and Computer Vision from Gujarat Technological University in 2021. He previously served as the Head of the Department of Computer Science and Engineering at HJD Institute of Technical Education and Research, Kera, India. He has over 13 years of experience in academia and more than 1 year of experience in the industry. Currently, he is an Associate Professor in the Department of Computer Science and Engineering at Parul Institute of Engineering and Technology, Parul University, Vadodara, Gujarat. He has mentored several Ph.D. students and regularly conducts workshops and leads projects in the field of computer vision.



Manuscript received 13 February 2026; Revised 10 April 2026; Accepted 12 April 2026; Available online 16 April 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i2.1596