SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

FedCAF: Federated Confidence-Adaptive Fusion Framework for Dual-Modal CT and MRI Brain Tumor Classification

Malathi Janapati[1], Shaheda Akthar[2]

[1]Department of Computer Science and Engineering, Acharya Nagarjuna University, Guntur, India. E-mail: janapatimalathi@gmail.com, ORCID: 0000-0001-7096-0435
[2]Department of Computer Science and Engineering, Acharya Nagarjuna University, Guntur, India, and also with the Department of Computer Science, Government College for Women (A), Guntur, India. E-mail: shaheda@gcwguntur.ac.in,
ORCID: 0009-0000-8479-4060

Abstract
Automated brain-tumor classification from medical pictures is complex issue because CT and MRI provide various and complimentary representations of brain abnormalities, and also medical imaging data are widely spread across institutions due to privacy and data-sharing constraints. Most of the existing works were on single modality categorization, traditional multi-modal fusion or federated learning separately. Here, we propose a dual-modal CT and MRI brain-tumor classification via a federated confidence-adaptive system. The proposed FedCAF framework adopts a modality-aware dual-backbone architecture using EfficientNetV2-S for CT image feature extraction and Swin-Small for MRI image feature extraction. The extracted representations are projected onto a shared feature space and processed by employing Dynamic Hierarchical Confidence Estimation (D-HCES) to assess the reliability of features. The generated confidence information is used to direct the Confidence-Adaptive Fusion Module (CAFM) to adaptively weight the features. Then the revised representation is processed by the Dual-Gated Confidence-Guided Self-Attention Module (DGCGSAM), and is classified in binary classification. The full model is trained utilizing Federated Averaging (FedAvg) on several distributed clients, where the model parameters are transmitted between clients and the central server, rather than raw medical images. The suggested FedCAF framework achieved an overall accuracy of 96.74%, with a precision of 99.87%,recall of 94.24%,and F1-score of 96.97% on the final test set. The federated training over four clients showed a successful collaborative optimization while the medical images remained local. The results show that the proposed framework helps in reliable classification of brain tumors using CT and MRI images.

Keywords: Brain tumor classification, CT–MRI imaging; Federated learning, Modality-aware feature extraction, Confidence-adaptive fusion, Dynamic confidence estimation, Self-attention, EfficientNetV2-S, Swin Transformer, FedAvg.

1. Introduction

Brain tumors are a challenging topic in medical image processing due to their variable appearance between patients, imaging procedures, anatomical locations and disease states.Correct identification of brain tumors in medical pictures is vital for diagnosis and further treatment planning. Medical imaging allows us to obtain non-invasive information on intracranial abnormalities. However, tumors can vary greatly in their appearance depending on size, location, texture, contrast and imaging procedure. Magnetic resonance imaging (MRI) is one of the widely used imaging modalities since it may give detailed soft-tissue information. Thus, a significant part of recent computer-aided brain-tumor investigations has focused on categorization based on MRI employing convolutional neural networks (CNNs), transfer learning and transformer architectures (Reddy et al., 2024; Özkaraca et al., 2023; Pasvantis and Protopapadakis, 2024).

While MRI offers extensive structural information, computed tomography (CT) images remains medically important because of its accessibility, quick acquisition and complementing imaging properties.

Therefore, reliance on MRI alone may restrict a learning system's ability to utilize information from other imaging modalities. Recently, multimodal research has started to study the simultaneous application of CT and MRI for brain-tumor analysis. The MBTC-Net framework incorporates a deep neural network with a multi-head attention mechanism to fuse CT and MRI data, revealing the promise of multimodal representation learning for brain-tumor classification (Kar and Singh, 2025). This trend is particularly crucial because CT and MRI reflect complementing information, not equal representation of the underlying path physiology.

Another challenge is that CT and MRI images are processed using the same feature-extraction architectures. The two modalities have very distinct picture properties, intensity distributions, and structural representations. So a common feature extractor might not be able to capture the modality-specific properties well. This drives the use of modality-aware feature extraction, where various deep structures are used for distinct imaging modalities. In the suggested framework, EfficientNetV2-S is used for CT pictures and Swin-Small for MRI images. Such a design enables the feature representation to be learnt according to the features of each modality instead of forcing both modalities to be fit into a single representation-learning process.

Another main issue is the availability and privacy of medical imaging data. Traditional centralized deep learning systems need collection of patient photos from several institutions to a common place. This can pose practical and privacy challenges when medical imaging datasets are spread among hospitals and research centers. Federated learning (FL) offers a solution in which several clients train local models and share model parameters, rather than raw patient data. Recently, application of FL to brain-tumor analysis has attracted significant attention. Albalawi et al. suggested an integrated federated-learning and transfer-learning approach for brain-tumor classification using MRI images, showing the feasibility of collaborative learning without explicitly centralizing the imaging data (Albalawi et al., 2024).

Later works study more in depth federated brain-tumor classification using several deep-learning architectures. For example, federated CNN based algorithms have been studied for MRI based brain-cancer classification and other works have studied federated learning under non-IID data distributions, that represent the heterogeneous nature of data available across multiple clients (Awotunde et al., 2025; Muntaqim and Smrity, 2025). Not only on MRI, federated learning has also been applied to CT-based brain-tumor detection, indicating that privacy-preserving learning (Al-Saleh et al., 2025). The results emphasize the value of distributed learning for medical image interpretation. Nevertheless, current federated brain-tumor frameworks still target a single imaging modality.

Multimodal learning and federated learning thus signify two interdependent paths. Multimodal models can leverage data from several imaging modalities, while federated learning facilitates cooperative model training without the necessity of raw data transfer. So, merging these two prerequisites presents more difficulties. CT and MRI images can vary significantly among patients, and the reliability of each imaging technique may not be uniform for all cases. A static feature-fusion process can therefore attribute equal significance to features regardless of their dependability. Recent studies on multimodal medical-image fusion have demonstrated the efficacy of attention and feature-interaction processes in amalgamating complementing data from several modalities (Song et al., 2024).The dependability of acquired representations is especially crucial in a federated setting because to the potential variations in client datasets regarding size, class distribution, imaging attributes, and modality composition. A methodology that excels with one client may consequently yield less dependable outcomes with another.

Based on the above limitations, this research introduces FedCAF, a Federated Confidence-Adaptive Framework for the classification of brain tumors using CT and MRI. The architecture utilizes distinct modality-specific feature extractors, utilizing EfficientNetV2-S for CT and Swin-Small for MRI, followed by, Modality-Aware Feature Extraction, Dynamic Hierarchical Confidence Estimation, Confidence-Adaptive Fusion Module, Dual-Gated Confidence Guided Self Attention, Brain Tumor Classification.The resultant depiction is employed for ultimate classification, whilst cooperative training among clients is conducted utilizing the Federated Averaging (FedAvg) methodology.

The main contributions of the proposed FedCAF framework are a modality-aware dual-backbone architecture with EfficientNetV2-S for CT images and Swin-Small for MRI images, and a Dynamic Hierarchical Confidence Estimation (D-HCES) mechanism to evaluate the reliability of features. The estimated confidence is exploited by the Confidence-Adaptive Fusion Module (CAFM) to emphasize the informative representations and reduce the negative impacts of less reliable ones. Beside, we propose a Dual-Gated Confidence-Guided Self-Attention Module (DGCGSAM) to enhance the self-attention based confidence-guided feature refining. Finally, these components are combined into a federated learning framework with FedAvg, which enables collaborative brain-tumor classification among remote clients without directly sharing the underlying medical pictures.

2. Related Work

Deep Learning-Based Brain Tumor Classification:

Deep learning has been a popular method for automated brain-tumor research because CNNs can learn hierarchical picture representations directly from training data, decreasing the need for handmade features. Customized CNNs, transfer learning, attention processes and transformer structures have been examined in recent studies in MRI for enhanced classification accuracy. Albalawi et al. (2024) developed a customized multilayer CNN for MRI-based brain-tumor classification and demonstrated the usefulness of task-specific convolutional representations. These approaches are neutral to the dense feature engineering as in the classic machine learning approaches and are based on an autonomous feature learning. However, the study is limited to MRI, and in a centralized learning setup, thus cannot solve privacy-preserving collaboration among universities. Hence, their work gives a solid fundament for MRI classification, although the questions of multimodal learning and distributed optimization are still open.

Preetha et al. (2024) investigated a fine-tuned EfficientNet-B4 architecture for automated brain-tumor detection from MRI images. The study applied transfer learning to utilize pretrained visual representation and found EfficientNet to be successful for medical-image categorization. The key advantage of this method is that it is reasonably efficient in extracting features in comparison to conventional very deep networks. However, the model solely employs the MRI pictures and does not explicitly include the variations between the imaging modalities and the reliability of the individual feature representations. This constraint is crucial for the present work in that FedCAF utilizes EfficientNetV2-S as the CT feature-extraction branch and not one architecture across CT and MRI indiscriminately.

Ozdemir and Dogan (2024) presented MTAP model for brain-tumor classification employing a CNN-based method with ADASYN for class balance, network pruning and Avg-TopK pooling for feature extraction. Their approach tries to solve two practical problems concurrently, i.e. class imbalance and redundant model parameters. The authors demonstrated that the proposed architecture may obtain high classification performance with less redundant network components. The advantage of this design is its emphasis on issues of computing efficiency and class distribution . But this approach is still a single-modality centralized framework and does not involve federated learning or confidence-guided feature refining.

Rajput et al. (2024) investigated transfer learning for brain-tumor classification from MRI images.Transfer learning offers a promising technique to transfer visual representations pre-trained on large-scale datasets to the medical domain where the number of annotated samples may be rather low. Typically, such approaches lower the training requirement and can offer strong classification results. The disadvantage, however, is that pretrained representations are not expressly designed to consider modality-specific features of CT and MRI. Therefore, although transfer learning is effective for brain-tumor classification, it does not, in its own right, provide a mechanism for modality-aware feature extraction or privacy-preserving distributed training.

Balamurugan et al. (2024) tried to solve the feature-selection challenge by using channel-wise attention on a deep-learning framework for brain-tumor classification. They studied attention mechanisms to attend to informative feature channels, and showed the possibility of attention-based feature refinement for MRI classification. Its primary advantage is that the network may learn the relative importance of feature channels rather than treat all channels equally. However, the attention mechanism is not explicitly conditioned on a confidence estimate and the work does not incorporate federated learning or CT–MRI modality distinctions.

Transformer and Attention-Based Brain Tumor Analysis:

Transformer-based architectures have recently been introduced into brain-tumor classification to improve the modeling of long-range relationships in medical images. Pacal (2024) proposed a Swin Transformer-based approach incorporating residual multilayer perceptron components for MRI-based brain-tumor diagnosis. The hierarchical Swin architecture provides multi-scale representations while restricting attention to local windows, reducing the computational burden associated with global self-attention. The study demonstrates the suitability of transformer-based representation learning for brain-tumor analysis. Still, it remains an MRI-only centralized approach and does not incorporate confidence-aware feature refinement or federated optimization.

Reddy et al. (2024) evaluated Vision Transformer models fine-tuned for the multi-class brain-tumor classification using MRI imaging. Their study involved glioma, meningioma, pituitary tumors and no-tumor categories, and showed the feasibility of transformer based representations for MRI classification. The main advantage is that self-attention can model relationships that extend beyond the narrow receptive fields often associated with convolutional networks. However, the system is built for MRI only and does not explore diverse CT and MRI representations. Neither does it have a confidence mechanism in the attention process.

Aamir et al. (2025) proposed an automatic MRI classification framework incorporating image enhancement, morphological processing, multi-scale feature extraction, attention and ensemble classification. Their approach has a single-backbone classifier since it incorporates multiple processing stages before the final prediction. The work provides good classification performance and highlights the advantage of combining multiscale representations with attention and ensemble learning. But the entire workflow is still centralized and MRI-specific. The privacy restrictions of distributed medical datasets and the use of distinct architectures for CT and MRI are not addressed in the work.

Mao et al. (2025) suggested a Dilated SE-DenseNet architecture for MRI brain-tumor classification, which is an extension of DenseNet-121 with dilated convolutions and Squeeze-and-Excitation attention. The dilated convolutions can enlarge the effective receptive field while SE blocks can give the channel-level feature recalibration. This combination is beneficial since it tries to gather more contextual information while keeping channel selective feature learning. But, the technique is still a single-modality centralized classifier and cannot directly quantify the confidence of the prediction

Lv et al. (2025) suggested FoTNet for the multi-class diagnosis of malignant brain tumors from MRI. The architecture uses a frequency-based channel-attention mechanism and targeted loss to tackle class imbalance, especially for the significantly underrepresented primary central nervous system lymphoma class. The study used a multi-center MRI dataset and public data.The proposed model showed better performance than the comparison deep-learning algorithms. One key advantage is the utilization of multi-center data, which provides more variety than a dataset from a single source. Still, the work is MRI-specific and does not address CT – MRI representation learning or federated model aggregation.

Zarenia et al. (2025) proposed a multi-class MRI classification and segmentation paradigm based on deformable attention and saliency mapping. The strategy combines attention and simultaneous categorization and segmentation, enabling the model to learn both diagnostic and geographic information at the same time. This is beneficial, as tumor location might offer further structural information for categorization. The framework, however, is still centralized and MRI-only.

Panigrahi et al. (2025) introduced a hybrid transfer-learning and self-attention paradigm for MRI-based brain-tumor categorization. The FedCAF is especially interesting for the use of combined feature extraction and self-attention, since the present model additionally mixes the convolutional and transformer-based representations. The advantage of their approach is that it merges feature learning and attention based refining on pretrained features. However, it does not consider distributed learning or modality specific CT/MRI processing.

Multimodal CT–MRI Brain Tumor Analysis:

Usha et al. (2024) established a multimodal brain-tumor classification framework utilizing a Convolution tumnet architecture that combined MRI and CT information. They examined the application of multimodal image fusion for classification and segmentation in their research and showed that information fusion from imaging can be beneficial. The main advantage is that CT and MRI provide various kinds of anatomical information, leading to a multimodal depiction that includes information that may not be available from either modality alone. This fusion technique is based on multimodal image data and requires the relevant slices for the fusion process not patient-paired. This is different from the current dataset situation where the CT and MRI collections are not patient-paired.

Kar and Singh (2025) suggested MBTC-Net for multimodal brain-tumor classification utilizing CT and MRI data. They used a deep neural architecture with a multi-head attention mechanism for their framework and validated their framework on several open-access multimodal brain-tumor datasets. The present work is most linked to this work, which explicitly mixes CT and MRI and improves multimodal representation learning via attention. Thus, MBTC-Net has the benefit of directly taking both complimentary CT and MRI information into account.

Wang et al. (2025) proposed a multimodal supervised contrastive framework with image fusion, named MMSupcon, for brain tumor identification. The method mixes multimodal images and uses a supervised contrastive objective. Instead of concatenating features, fused samples are employed to drive the learning of cross-modal consistency and class discrimination. The work reports an enhanced accuracy and generalization on a real world brain tumor dataset and other public datasets. The fundamental feature of this method is the use of contrastive learning to preserve important information from multimodal data while preserving class separation. But , the method does not consider federated learning and the multi-modal formulation.

Recent works have also explored multimodal picture fusion beyond traditional CNN-based methods. A study published in Frontiers in Computational Neuroscience in 2026 on CT–MRI integration discussed performance disparities between separate modalities and integrated representations Almadhor et al. (2026) . This work is important in that it emphasizes the necessity of examining modality-specific and multimodal settings, rather than assume that adding another modality will automatically increase performance. But, this area of research often does not consider federated optimization, nor an explicit confidence mechanism.

Federated learning has become a significant paradigm for medical-image analysis, since healthcare datasets are often spread across institutions and cannot be easily centralized. Albalawi et al. (2024) proposed a federated brain-tumor classification framework based on transfer learning using a modified VGG16 architecture. They trained their model on MRI data from several sources like Figshare, SARTAJ and Br35H and attained 98% overall accuracy and stated high precision across the tumor types. The main value of this research is that it shows the possibility of integrating transfer learning and federated learning for brain-tumor classification, while keeping the basic data spread. However, the method is solely based on MRI, and does not require modality-specific feature extraction of CT/MRI and confidence-guided feature adaptation.

Joseph et al. (2025) used federated learning augmented with CNNs for brain cancer classification from MRI data was performed. This method obviates the drawbacks of centralized medical-image learning by dispersing the training process across clients. Its key advantage is a combination of CNN representation learning and a privacy-preserving optimization methodology. But, the analysis remains limited to MRI and conventional federated classification. It does not explicitly model the reliability of the extracted representations and use the confidence to drive the subsequent feature refining.

Al-Saleh et al. (2025) extended federated learning to CT-based brain-tumor detection. Their privacy-preserving framework is particularly relevant to the present study because it demonstrates that CT-based brain-tumor analysis can be performed within a federated setting. The work addresses the privacy constraints associated with centralized CT-image collection and demonstrates the applicability of distributed learning to CT data. However, CT is considered independently rather than jointly with MRI, and the framework does not introduce modality-specific dual-backbone learning or confidence-adaptive feature refinement. Thus, it provides evidence for the feasibility of federated CT learning but leaves the CT–MRI integration problem unresolved.

Paul et al. (2026) proposed a federated MobileNetV2 framework combined with ensemble meta-learning for privacy-preserving brain-tumor classification in 2026. The study explicitly considers FedAvg and FedProx and examines convergence within a federated MRI classification setting. Its lightweight MobileNetV2 design is advantageous for reducing computational and communication requirements. However, the framework remains MRI-centered and does not address dual-modal CT/MRI feature learning or confidence-guided feature refinement.

Melo et al. (2026) have studied test-time augmentation within the framework of federated brain-tumor MRI classification. Their findings show moderate improvements with only preprocessing and statistically significant improvements with a combination of preprocessing and test-time augmentation. This paper shows that the federated performance depends on the robustness of the inference technique, as well as the training algorithm. However, the study does not modify the fundamental feature learning architecture and does not explore modality-aware CT/MRI learning.

Recent studies also show that MRI architectures with an increasing degree of specialization can achieve great performance. Lu et al. proposed a deep learning framework for classification and segmentation based on non-contrast multichannel MRI, and tested it on internal and external datasets (Al-Saleh et al., 2025). In their framework, the classification accuracy and Dice score for the tumor segmentation are 98.3% and 0.937, respectively (Lu et al., 2025). The clinical relevance of non-contrast MRI lies in the scenarios where contrast-enhanced imaging might not be appropriate. However, the approach is still based on MRI and does not solve the federated CT–MRI learning

Ilani et al. (2025) explored T1-weighted MRI classification using U-Net and transfer-learned CNN architectures, including InceptionV3, EfficientNetB4, and VGG19. Their study reported 98.56% accuracy in the U-Net based technique and 96.01% accuracy in the cross-dataset validation. Inclusion of an external cohort is a strength since it offers evidence outside one test distribution. However, the method is still centralized and MRI-specific. The major focus is on segmentation assisted classification rather than confidence-guided federated multimodal representation learning.

As per the knowledge the above comparison studies indicates that there has been a significant progress in areas of automated brain-tumor classification through CNNs, transfer learning, transformer-based architectures, attention mechanisms, multimodal learning, and federated learning. As per the knowledge it is also observed that all these approaches have generally addressed the aspects independently ie: CNN and transformer-based approaches have been mostly adopted for MRI classification, whereas multimodal studies have investigated CT-MRI fusion using either traditional fusion or representation-learning approaches. In contrast, federated investigations have generally focused on privacy-preserving learning with a single imaging modality. As a result, there is a essential requirement for a unified approach that can balance the reliability of the learned representations in a distributed learning environment with the ability to accommodate modality-specific characteristics. The proposed FedCAF framework addresses this gap by combining Modality-Aware CT and MRI feature extraction (MAFE) with Dynamic Hierarchical Confidence Estimation (D-HCE), Confidence-Adaptive Fusion (CAFM), and Dual-Gated Confidence-Guided Self-Attention (DGCGSAM), followed by federated optimization using FedAvg for classification. In contrast to multimodal approaches that involve patient pairs, FedCAF doesn't require a one-to-one correspondence between CT and MRI images; rather, each image is analyzed in accordance with its modality prior to confidence-guided feature refinement.

3. Proposed System

The proposed architecture FedCAF (Federated Confidence-Adaptive Fusion) frame work uses federative learning approach to perform accurate brain tumor classification using dual data model of CT and MRI images by preserving privacy of individual client institutions. The overall architecture in shown in Figure1 where in each communication round the framework undergoes the following phases which includes dual data preprocessing, Modality-Aware Feature Extraction(MAFE), Dynamic Hierarchical Confidence Estimation, Confidence-Adaptive Fusion Module, Dual-Gated Confidence Guided Self Attention, Brain Tumor Classification.

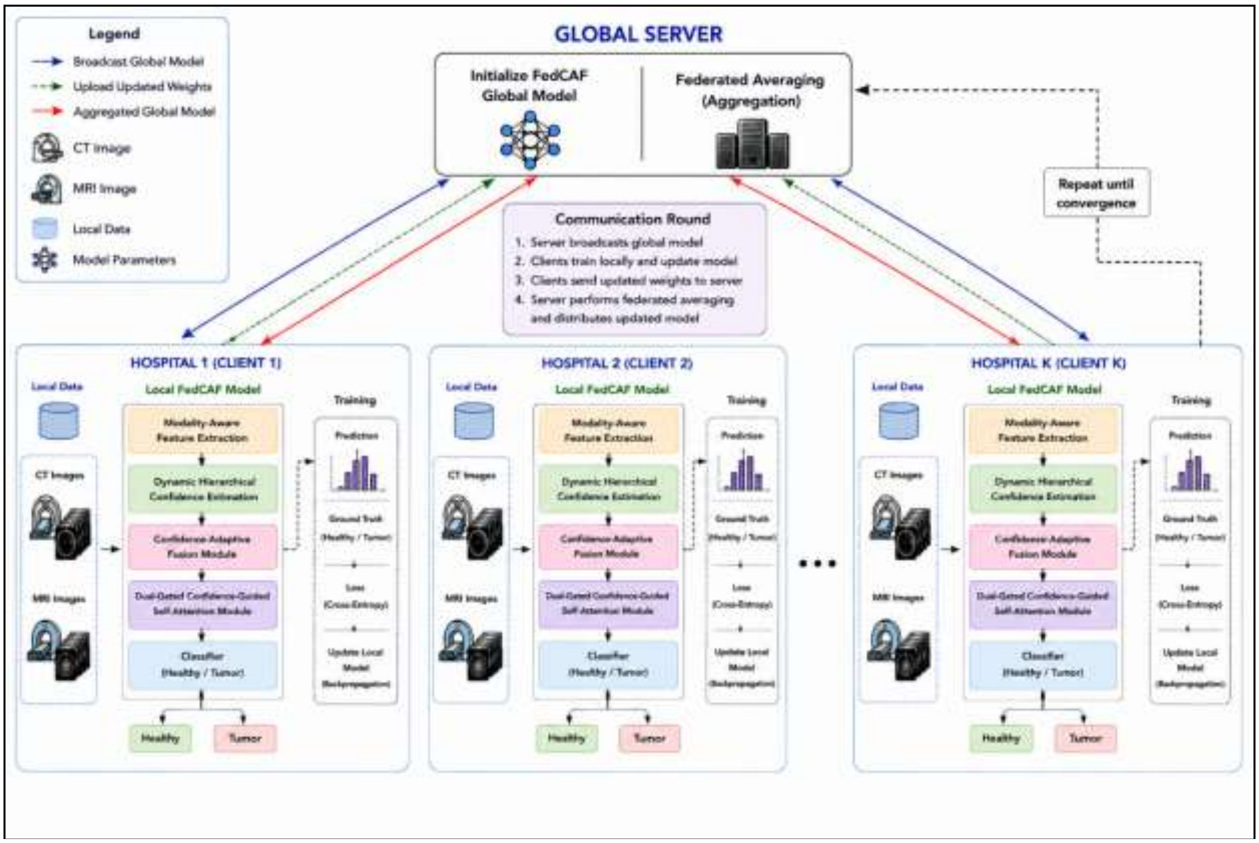


Figure 1. Overall architecture of the proposed FedCAF framework for federated brain tumor classification using dual-modal CT and MRI images.

The workflow of FedCAF framework is summarized in the below Algorithm1. Initially a global FedCAF model is initialized at the central server and training data are distributed to multiple federated clients. At each client preprocessing of raw CT and MRI data is done which undergoes image resizing, intensity normalization, and contrast enhancement and data augmentation on training data. Later at each communication round, every client independently performs modality aware feature extraction(MAFE) which enable to learn discriminative representations that preserve the anatomical characteristics of each imaging modality and helps in generating a unified feature embedding suitable for multimodal learning. The extracted features are then passed to D-HCE module which estimates the confidence associated with each feature .The estimated confidence from D-HCE is incorporated into CAFM which dynamically adjusts the contribution of features based on estimated confidence.ie: Features with higher confidence receive greater emphasis during the fusion process, whereas uncertain or less informative features are automatically suppressed. The obtained features are passed to DGCGSAM for further attention refinement. This module proposes two complementary gating mechanisms to modulate the feature interaction according to the confidence information before the self-attention is applied. The resulting attention mechanism allows the fused representation to be adaptively refined, while conserving the complimentary information gathered from both CT and MRI modalities. The final refined feature representation is then passed to a lightweight classification network. The classifier at each client generates local healthy/tumor predictions The classifier predicts at each client if the input image is of category Healthy or Tumor computes the loss and updates the parameters through back propagation. After local training, the updated local model parameters are sent to a central server where FedAvg aggregates the parameters to form an updated global model. This decentralized learning approach improves privacy preservation and enables collaborative model training on distributed medical datasets.

Algorithm 1: Over workflow of FedCAF
Input: D = {D1,D2,….Dk} are the distributed CT and MRI images. R is the no. of communication rounds. E is the no. of epochs. η is the learning rate.
Output: W^* is the optimized FedCAF model.
Step 1: Initialize global FedCAF model W^o
Step 2:Partition the CT and NRI images among P federated clients
Step 3:for each communication round t=1 to R do the following
Step 4: Broadcast global model W^t to all the clients
Step 5: for each client p=1 to P do the following
Step 6: Load local CT and MRI images
Step 7: Preprocess all the CT and MRI images.
Step 8: Perform Modality Aware Feature Extraction
Step 9: Estimate the confidence score
Step10: Perform Confidence Adaptive Feature Refinement
Step11: Perform Dual-gated Confidence-Guided Self-attention learning.
Step 12: Predict class probabilities and compute cross entropy loss
Step 13: Update and upload the local model parameters to FedCAF
Step 14: end for
Step 15: Aggregrate all the client models using Federated Averaging
Step 16: end for
Step 17:Return optimized global model W^*

3.1 Dual-Modal Dataset preprocessing:

To improve the quality and consistency of the publicly available dual-modal brain imaging dataset comprising CT and MRI images preprocessing is done which includes as follows.

Resize: Here, all images are resized to 224 × 224 pixels for studying the entire dataset in fixed dimension.

Normalization: All the pixel intensities are normalized using min-max normalization.

Contrast Enhancement: Contrast limited adaptive histogram equalization (CLAHE) is applied to improve local contrast and enhance visibility.

Data Augmentation: To improve model generalization, data augmentation is done which includes random horizontal flipping, vertical flipping , random rotate90,random brightness-contrast adjustment on the data. The preprocessed images were subsequently forwarded to the modality-aware feature extraction module for feature learning.

The preprocessing image is represented as

$$I' \leftarrow T(I) \quad (1)$$

where I denotes the original Ct or MRI brain image, $T(.)$ represents the preprocessing pipeline and I' is the transformed image used for subsequent feature extraction.

3.2 Modality-Aware Feature Extraction (MAFE):

The preprocessed images is from previous phase is forwarded to MAFE , here the discriminative representations are learned from CT and MRI images while retaining their modality-specific properties. As the CT scans mainly focus on bone features, calcifications and tissue density differences while MRI is superior in soft tissue contrast and tumor shape. Instead of directly integrating images from different modalities, the proposed framework handles two independent feature extraction branches separately according to their modality.

For a input image I' and its corresponding modality is labeled as $I' = \{I_{CT}, I_{MRI}\}$.The CT images are processed using EfficientNetV2-S as backbone and MRI images are processed using Swin Transformer small.

$$F = \varphi_{CT}(I_{CT}) \text{ and } F = \varphi_{MRI}(I_{MRI}) \quad (2)$$

From the CT images the features like local anatomical structures, tissue density variations and boundary information is extracted by using EfficientnetV2 as backbone. From MRI Images the Swin Transformer is used to learn soft tissue representation and long range contextual relationships. By using these independent feature extraction networks the proposed framework efficiently learns modality specific features reducing inter modularity interference.

The extracted features are thereafter given to common embedding space consisting of fully connected layer ,Relu activation and dropout regularization for modality specific feature projection process which is expressed as shown below,

$$F_{fp} = \delta(W_{fp}F + b_{fp}) \quad (3)$$

Where

W_{fp} and $b_{fp}$are learnable projection parameters.

δ is the Relu activation function,

F_{fp} denotes the projected feature representation.

The heterogeneous backbone features are transformed into unified 512-dimention feature embedding($F_{fp} \in R^{512}$).The extracted feature embeddings are forwarded to the Dynamic Hierarchical Confidence Estimation module for confidence estimation.

3.3 Dynamic Hierarchical Confidence Estimation (D-HCE):

The modality aware feature vector generated by MAFE is given as input to this module. The proposed D-HCE module dynamically estimates a confidence score for each extracted feature embedding. It estimates the discriminative quality of modality-specific features, and generates adaptive confidence values to guide the feature fusion stage. Therefore, features with high confidence will be more importance in multimodal fusion, and uncertain features will be less importance. Initially the extracted feature representation is converted into shared latent representation. This shared feature representation is expressed as

$$F_s = \delta(W_s F_{fp} + b_s) \quad (4)$$

W_s and $b_s$denote the learnable parameters of shared encoder,

δ represents the non linear transformation

$F_s \in R^{512}$ is the shared confidence representation.

These shared features are processed by prediction confidence branch, feature reliability branch and attention consistency branch where each branch independently produces the confidence score within a range [0,1].The three confidence scores are represented as $P_c, F_r, A_c \in [0,1]$.The three confidence estimates are adaptively combined using learnable fusion weights to generate a final unified confidence score and is represented as

$$C = \propto P_c + \beta\, F_r + \gamma A_c \quad (5)$$

Where $\propto, \beta\ and\ \gamma$ are learnable fusion coefficients satisfying the condition $\propto + \beta + \gamma = 1$.Finally the fused confidence score is fused. The confidence-enhanced features are forwarded to Confidence-adaptive Fusion Module for adaptive multimodal feature integration.

3.4 Confidence-adaptive Fusion Module (CAFM):

This module receives the modality aware feature representation F and the corresponding confidence score C generated by the D-HCES module. CAFM uses information to refine the extracted features before attention learning. Initially the feature representation is processed through channel attention network for estimating the importance of every feature channel.

The channel attention weights are computed as

$$W_{ch} = \sigma\,(W_2\delta(W_1F + b_1) + b_2) \quad (6)$$

Where

W_{ch} is the generated channel vector

$W_1$and W_2 denote learnable weights matrices,

δ represents the ReLU activation function,

σ denotes the sigmoid activation function.

The confidence score estimated by D-HCES is projected into feature space using learnable confidence projection layer,

$$W_c = \sigma(W_p C + b_p) \quad (7)$$

Where
W_p denotes the confidence projection matrix.
b_p represents the bias vector.

The extracted feature representation is refined by combining the original feature vector ,the channel attention weights and the projected confidence weights through element-wise multiplication.

$$F_e = F \odot W_{ch} \odot W_c \quad (8)$$

Where
$\odot$ denotes element wise multiplication.
During this process the channel attention mechanism determines the relative importance of individual feature channels , the projected confidence weights regulate the contribution of each feature according to the estimated reliability.
The residual learning strategy is employed to the network for preserving structural information extracted while incorporating confidence-guided feature enhancement.

$$F_r = F_e + F \quad (9)$$

The final F_r is processed through fully connected refinement layer

$$F_{CAFM} = \delta(W_r F_r + b_r) \quad (10)$$

This confidence-adaptive refined feature is forwarded to Dual-Gated Confidence-Guided Self-Attention module for global contextual learning.

3.5 Dual-Gated Confidence Guided Self Attention:
The feature representation from previous module are sent to Dual-Gated Confidence Guided Self Attention module (DGCGSA).The obtained features are first partitioned into feature tokens and they are projected through learnable tokenize for attention learning.

$$T = \mathrm{Token}(F_{CAFM}) \quad (11)$$

Where,
F_{CAFM} is the input feature representation.
T is the tokenized feature matrix.

From the D-HCES module estimated confidence score is injected into the token space of every feature token.

$$T_c = T \odot C_p \quad (12)$$

Where,
C_p is the projected confidence embedding and $\odot$ is the element wise multiplication. This operation enables feature reliability during contextual learning.
Later the tokens are processed by transformer encoder for learning the global contextual relationships which are further refine using two independent gating working process. The refined token process can be represented as

$$T_g = transformer(T_c \odot G_a \odot G_b) \quad (13)$$

Here, G_a and G_b are two confidence guided gates which selectively preserve information feature responses while the transformer captures the global contextual dependencies.
Finally through residual learning the refined feature tokens are reconstructed into original feature dimention.

$$F^* = Relu(Flatten(T_g) + F_{CAFM}) \quad (14)$$

Where F^* denotes the final refined feature representation and F_{CAFM} is original input feature vector. The final embedding is send to classification network for prediction.

3.6 Brain Tumor Classification:
The final refined feature representation from previous phase is taken as input to classification network. This network contains two fully connected layers separated by Rectified linear Unit(ReLU) and Dropout layer.The hidden fully connected layer learns higher-level nonlinear relationships in the refined feature representation, and the Dropout operation randomly deactivates neurons during training to reduce overfitting and improve the generalization capability of the model. The output layer contains two neurons corresponding to the Healthy and Tumor classes.
The classification logits are computed as

$$V = W_l F^* + b_l \quad (15)$$

where W_l and b_l denote the learnable weight matrix and bias vector of the classifier, respectively. V is the output classification logits vector.
The posterior probability of each class is obtained using the Softmax activation function

$$P(y_i) = \frac{\exp(V_i)}{\sum_{j=1}^{2} \exp(V_j)} \quad (16)$$

Where $P(y_i)$ denotes the predicted probability of class i.
During network optimization, the classifier parameters are learned using the Cross-Entropy loss function

$$loss_{CE} = -\frac{1}{N} \sum_{i=1}^{N} y_i \log(P(y_i)) \quad (17)$$

Where ,

- N denotes the total number of training samples,
- y_i represents the ground-truth class label, and
- $P(y_i)$ denotes the predicted class probability.

The network parameters are optimized through back-propagation in the FedCAF network to minimize Cross-Entropy loss, such that the classifier learns discriminative decision boundaries between healthy and tumor brain images. At inference time, the class with the highest Softmax probability is taken as the final prediction. The final classifier generated prediction is used for local model optimization at each federated client. After the local training is completed the updated model parameters are transferred to central server ,but the predicted labels and patient images preserved in local institution for maintaining data privacy.

3.7 Federated Learning:

Here in each communication round, the central server first initializes the global FedCAF model, and then sends the current global model parameters to all participating clients. Each client separately trains the receiving model on its local CT and MRI datasets for a preset number of local epochs. In the entire FedCAF pipeline, the local optimization of each client is composed of modality-aware feature extraction, confidence estimation, confidence-adaptive feature fusion, dual-gated confidence-guided self-attention, and final brain tumor classification.

After the local training is done, each client only sends the updated model parameters to the central server. Since only model weights are communicated, the original CT and MRI scans remain at the different universities, minimizing the requirement for centralized data gathering and facilitating privacy-preserving collaborative learning.

Let $W_k^{(t)}$ denote the model parameters obtained by the k[th] client after local training during communication round t. The central server updates the global model using the Federated Averaging (FedAvg) algorithm as

$$W^{(t+1)} = \frac{1}{K} \sum_{k=1}^{K} W_k^{(t)}$$

Where ,

K denotes the total number of participating clients,

$\sum_{k=1}^{K} W_k^{(t)}$ is the total number of training samples across all clients,and

W^{t+1} represents the updated global model after aggregation.

The FedAvg algorithm accomplishes aggregation in a weighted manner, where clients with larger training sets contribute more to the updated global model proportionally. The weighted average allows the global model to benefit from the combined expertise of all participating institutions and balances the contribution of each client according to the size of its dataset. The modified global model is then sent back to all the clients and the federated optimization process is repeated for further communication cycles until convergence. The final optimized global value is used tumor classification on unseen CT and MRI images. The proposed FedCAF framework learns a generalized classification model from distributed CT and MRI datasets through this iterative client–server collaboration without exchanging sensitive medical pictures.

4. Experiments and Results

This section presents the experimental setup and results obtained for the proposed FedCAF model.

4.1 Dataset:

To evaluate the performance of proposed model a dual modal brain imaging dataset that is publicly available in kaggle is used. This dataset consists of a total of 9618 images which are categorized into healthy and tumor classes. In CT images we have total of 4618 which are in turn classified as 2300 healthy and 2318 tumor. Similarly for MRI images we have total of 5000 which are classified as 2000 healthy and 3000 tumor.To maintain class balance the data is split into 6732 images i.e. 70% for training, 1443 images i.e. 15% for validation and 1443 images 15% for testing. Figure 2 shows some samples of the dataset which contain CT healthy, CT tumor, MRI healthy and MRI tumor images.

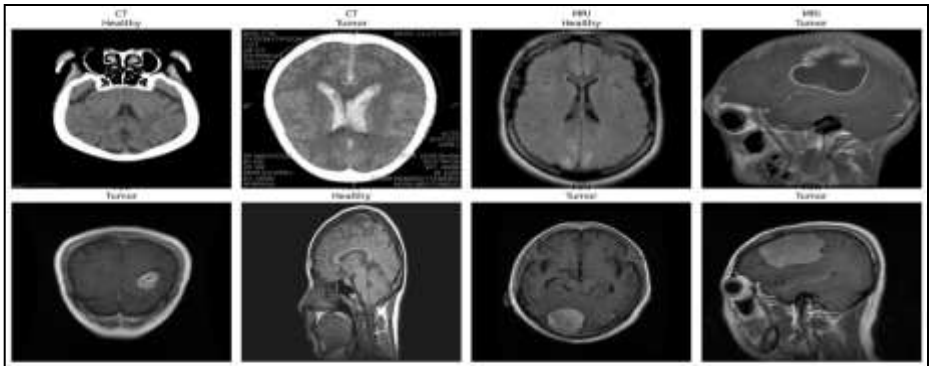


Figure 2: Samples of the dataset.

4.2 Experimental setup:

The implementing the configuration details for the proposed model are detailed below Table 2.

Table 2: Implementation and configuration details.

Parameter	Value / Description
Programming Language	Python
Deep Learning Framework	PyTorch
Hardware	NVIDIA GPU (RTX 2080) with CUDA Support.
Image Processing Library	Albumentations,OpenCV
CT Feature Extractor	EfficientNetV2-S
MRI Feature Extractor	Swin Transformer Small
Optimizer	Adam
Learning Rate	$1 * 10^{-4}$
Batch Size	16
Number of Clients	4
Communication Rounds	20
Federated Aggregation	FedAvg
Token Dimension	32
Number of tokens	16
Transformer Encoder layers	2
Multi-Head Attention	8 Heads
Dropout rate	0.30
Output Classes	Healthy / Tumor

4.3 Validations of proposed modules:

Initially in preprocessing all CT and MRI pictures were scaled to 224*224 pixels before training, providing a common input size to the deep neural networks. Contrast Limited Adaptive Histogram Equalization (CLAHE) was used to boost the local picture contrast and highlight significant anatomical structures. To improve the robustness and generalization ability of the proposed framework, we used the online data augmentation approaches, including random horizontal flipping, vertical flipping, random 90° rotation, and brightness-contrast adjustment during training. Figure 3 shows the samples of preprocessed CT and MRI images.

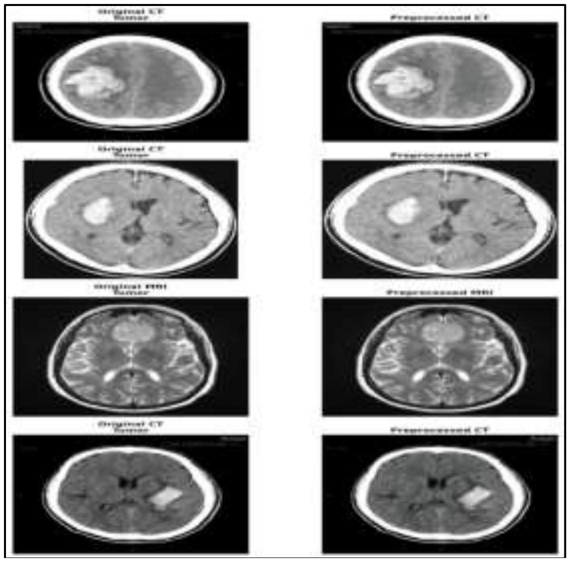


Figure 2: Samples of the dataset after preprocessing.

Further to verify the effectiveness of the proposed FedCAF framework, the intermediate outputs of the three main modules D-HCE module, CAFM and DGCGSAM are analyzed below.The proposed DHCE module learns to estimate the reliability of extracted feature representations by jointly learning three complementing confidence measures which include prediction confidence, feature reliability and attention consistency. The confidence ratings are integrated in an adaptive manner to produce a unified confidence value that drives the ensuing feature refining process. From the below Figure 3 it is observed that the three confidence branches provide varied confidence replies according to the properties of input samples. The adaptive fusion mechanism combines these complementary confidence estimates to create a stable final confidence score. This behavior shows that the suggested module can effectively discriminate the highly reliable feature representation from the uncertain one, leading the subsequent fusion module to attach more weight to informative features and mitigate the impact of unreliable feature replies.

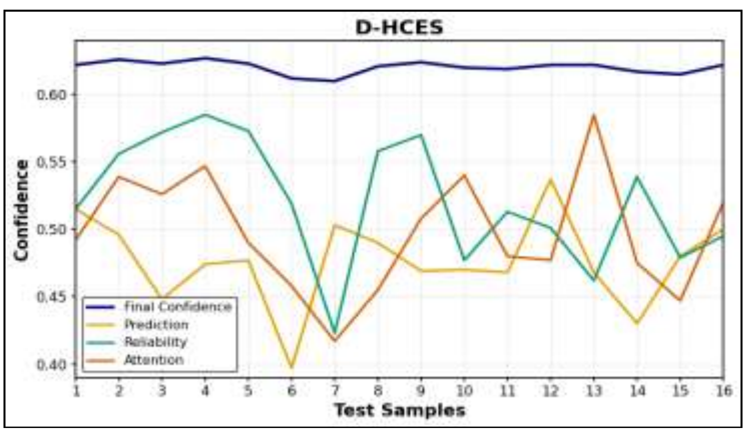


Figure 3: Confidence analysis of D-HCES module

For the effectiveness evaluation of the proposed CAFM, we compare the feature representations before and after the confidence-guided feature refining. The below Figure 4 shows the enchanced feature representation has a more compact and stable feature distribution compared to the original feature representation. This demonstrates that the suggested CAFM has effectively suppressed less informative feature responses while keeping highly discriminative features. This leads to a more powerful input for the succeeding attention module.

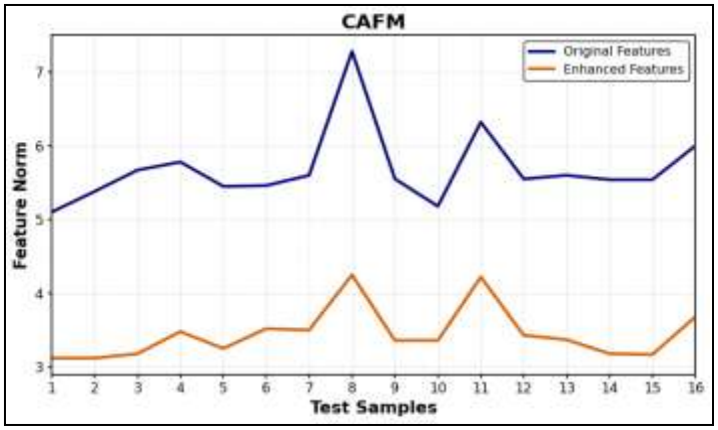


Figure 4: Feature analysis of CAFM

DGCGSAM then further refines the confidence-enhanced feature representations using a transformer-based self-attention mechanism and dual-gated feature selection. The below Figure 5 shows distribution of the modified feature representation after DGCGSAM processing It can be seen that only a subset of the feature dimensions has the high activation responses while the other dimensions keep the relatively lower activation levels. The selective activation pattern suggests that the proposed dual-gated attention mechanism is able to efficiently screen out redundant information and highlight the discriminative features for brain tumor categorization.

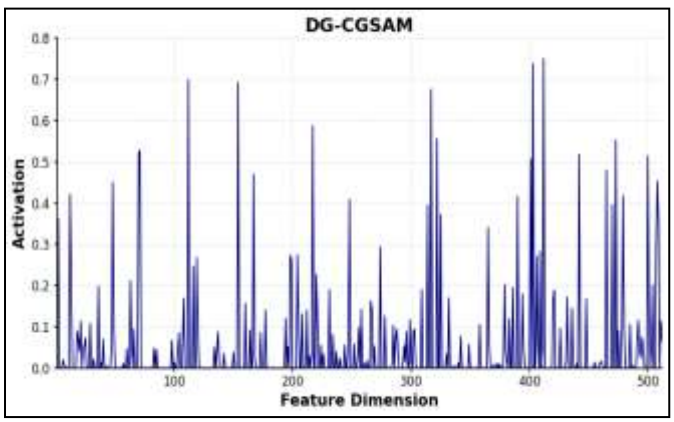


Figure 4: Feature enhancement analysis of DG-CGSAM

4.4 Federated Learning Performance Evaluation:

The convergence behavior of the proposed FedCAF framework is analyzed using different evaluation metrics.In each communication round, every client independently trained the local FedCAF using its private CT and MRI images. Instead of sending raw medical images, only the updated local model parameters were broadcast to the central server. The central server then used the Federated Averaging (FedAvg) method to combine the received values to generate an improved global model. This iterative collaborative optimization was conducted for 20 communication rounds, enabling the global model to gradually learn generalized feature representations while respecting patient data privacy.

The convergence behavior of the proposed FedCAF framework is analyzed using different evaluation metrics. Figure 5 shows the variation of training loss during federated optimization. The training loss decreases gradually with the communication rounds, which means the local client models gradually minimize the classification error throughout optimization.

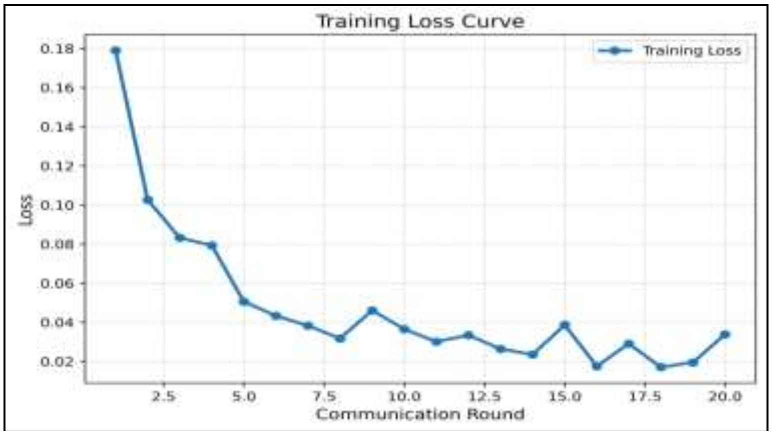


Figure 5: Training loss of FedCAF framework.

The below Figure 6 shows the validation loss of the framework. The validation loss falls gradually without any oscillation which shows that the aggregated global model generalizes well across the unseen test data.

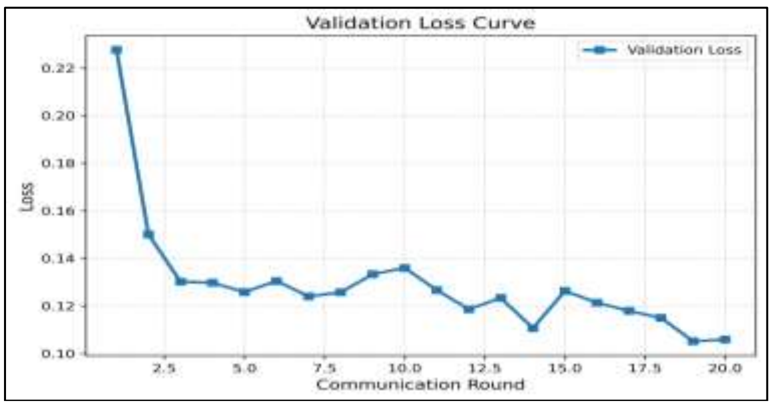


Figure 6: Validation loss of FedCAF framework

The classification accuracy during the federated learning process is shown below in Figure 7. The global model shows a steady increase in classification accuracy with respect to the communication rounds, demonstrating that the knowledge learned from different healthcare organizations has been well combined through collaborative learning.

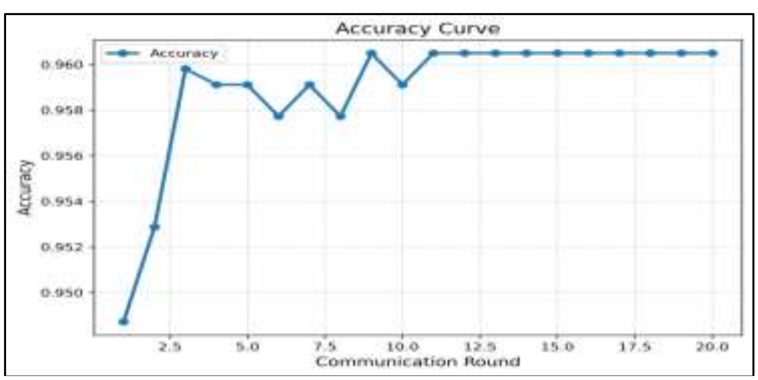
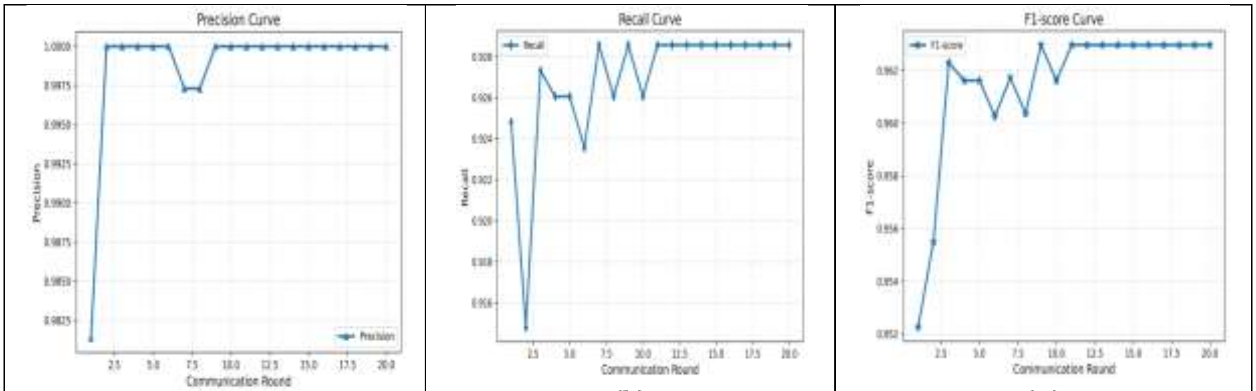


Figure 7: Classification Accuracy of FedCAF framework

Figure 8(a)-(c) depicts the precision, recall, and F1-score obtained after each communication round. The stable convergence of these performance indicators confirms the proposed federated learning technique can successfully reduce both false-positive and false-negative predictions with balanced classification performance.



(a) (b) (c)

Figure 8(a)-(c) shows precision, recall and F1-score curves during federated learning.

4.5 Classification Performance:

After the federated learning process is finished, the final global FedCAF model is assess on the independent test dataset to evaluate the brain tumor classification performance. The classification capabilities of the suggested framework was tested by Accuracy, Precision, Recall, F1-score, Area Under the Receiver Operating Characteristic Curve (AUC), Receiver Operating Characteristic (ROC) analysis and Confusion Matrix.

Table 3 shows the quantitative results of the proposed FedCAF framework on the test dataset.

Metric	Value
Accuracy	96.74 %
Precision	99.86 %
Recall	94.23 %
F1 Score	96.96 %
AUC	99.66 %

This framework achieves accuracy of 96.74% whish shows its ability of exhibiting Healthy and Tumor classification. Precision 99.86% says the model gives very few false positive predictions. Recall 94.23% how efficiently it can identify tumor classes.F1 score gives balance between Precision and Recall. AUC 99.66% shows the framework capability for brain tumor classification.

Figure 9 shows the ROC curve which is close to upper left of the graph that talks about high true positive rate and low false positive rate.

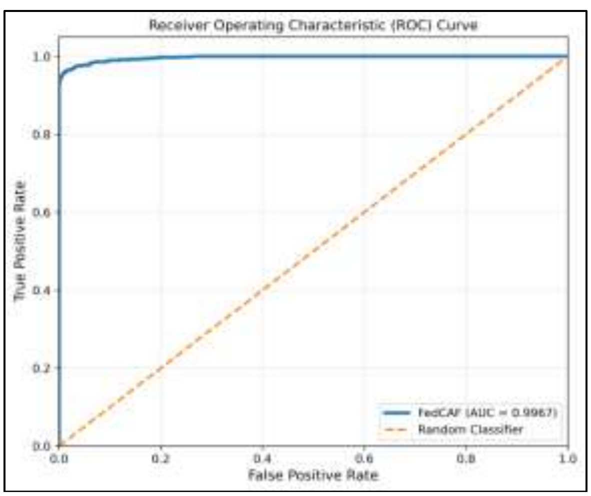


Figure 9: ROC curve of proposed framework.

Figure 10(a)-(b) shows the confusion matrix and normalized confusion matrix for the proposed framework. These results show that the framework is highly reliable in correct classification and has very low errors.

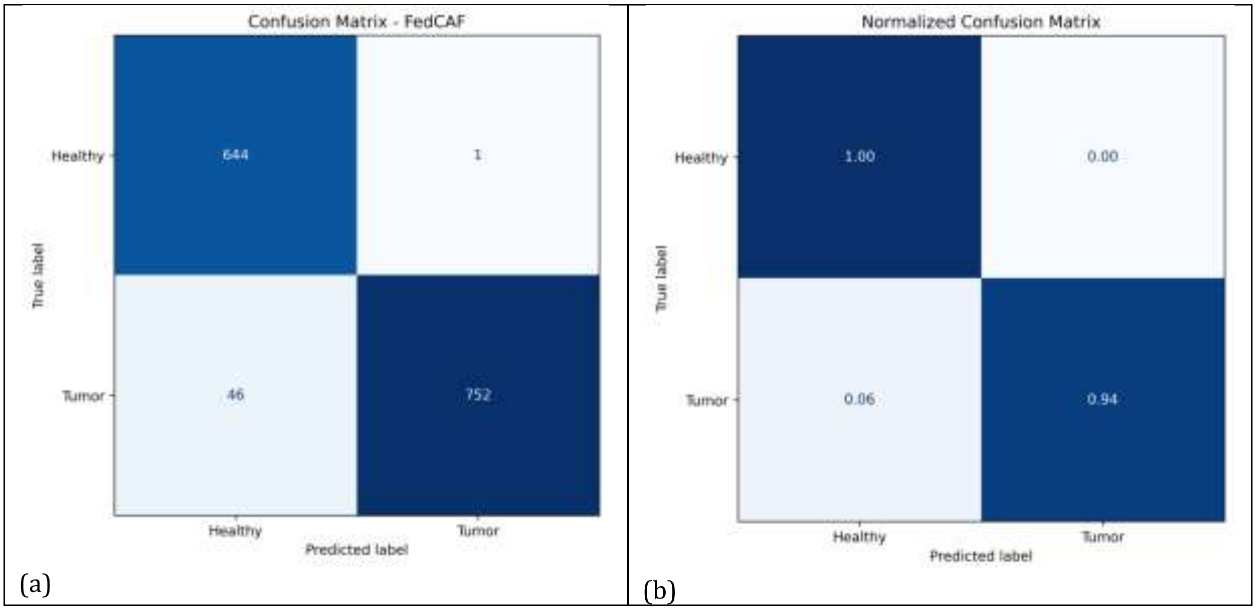


Figure 10(a)-(b) confusion matrix and normalized confusion matrix for framework

The qualitative predictive analysis of the proposed framework is shown in Figure 11. It displays the image type, ground truth(GT), perdition and confidence score.

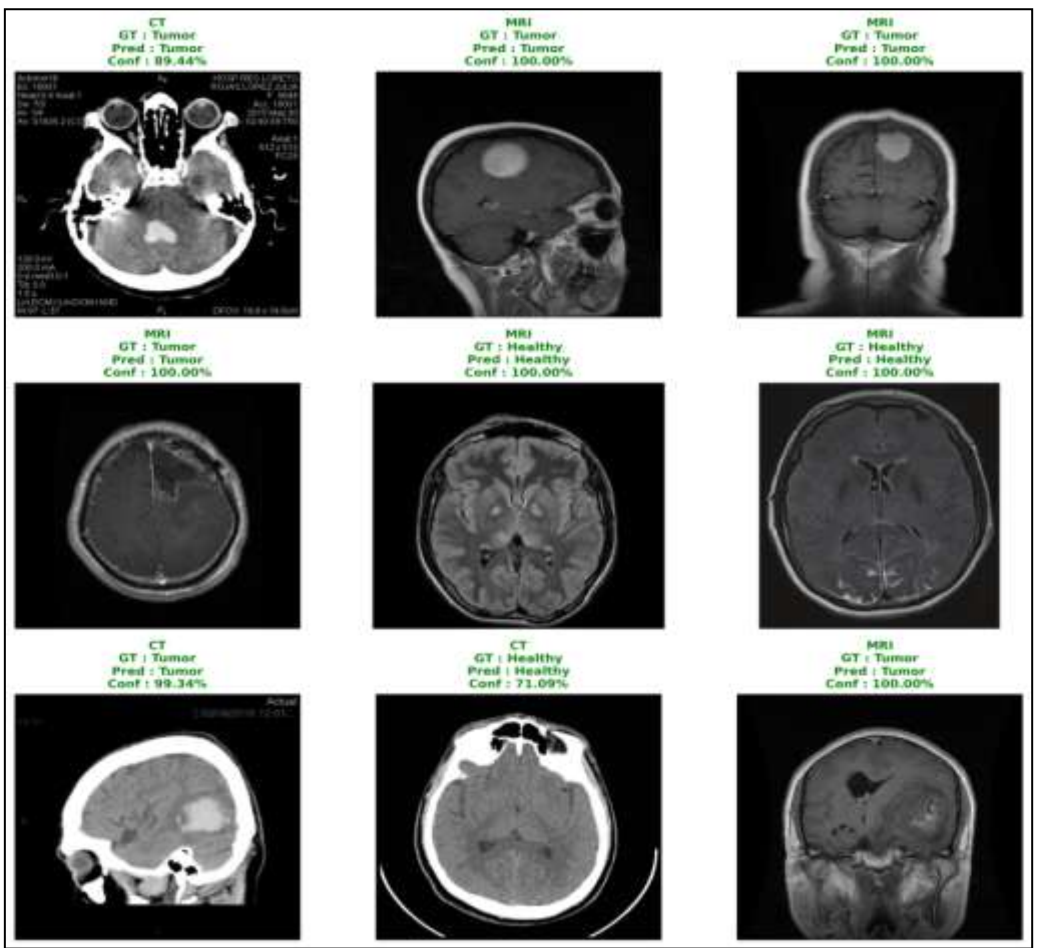


Figure 11: Prediction generated by FedCAF

Conclusion

The proposed FedCAF, a Federated Confidence-Adaptive Fusion Framework for brain-tumor classification using dual-modal CT and MRI images was designed to handle the issues of heterogeneous medical imaging data and privacy-preserving distributed learning. Here CT and MRI images were processed by modality-specific feature extraction networks: EfficientNetV2-S for CT and Swin-Small for MRI. Then, the extracted feature representations are projected onto a common feature space and are processed with Dynamic Hierarchical Confidence Estimation (D-HCES) which can offer confidence information about the derived representations. This information is then utilized in Confidence-Adaptive Fusion Module (CAFM) to adaptively weigh the feature representation. The Dual-Gated Confidence-Guided Self-Attention Module (DGCGSAM) further refines the confidence-weighted features before classification. The whole system is combined with Federated Averaging (FedAvg) to allow numerous clients to collectively train a global model without explicitly sharing their local medical data. The proposed system provides a unified approach to modality-aware representation learning, confidence-guided feature refinement, self-attention and federated optimization in a single pipeline for brain-tumor classification. The suggested framework does not require a one-to-one connection between the two modalities because it is intended for independently structured CT and MRI collections, unlike methods that rely on patient-level paired CT and MRI observations. The proposed federated training technique also facilitates collaborative learning among remote clients, while preserving the underlying medical images in their own local settings. The suggested architecture proposes a confidence-aware and modality-specific approach for distributed brain-tumor classification, and lays a foundation for further exploration of multimodal federated learning in medical imaging.

Acknowledgement

None

Data availability statements

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no potential conflict of interest.

References

1. Aamir, M., Rahman, Z., Bhatti, U. A., Abro, W. A., Bhutto, J. A., et al. (2025). An automated deep learning framework for brain tumor classification using MRI imagery. *Scientific Reports, 15*, 17593. https://doi.org/10.1038/s41598-025-02209-2
2. Albalawi, E., et al. (2024). Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor. *BMC Medical Imaging, 24*, 110. https://doi.org/10.1186/s12880-024-01261-0
3. Albalawi, E., Mahesh, M. T. R., Thakur, A., Kumar, V. V., Gupta, M., Bhatia Khan, S., et al. (2024). Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor. *BMC Medical Imaging, 24*, 110. https://doi.org/10.1186/s12880-024-01261-0
4. Albalawi, E., Thakur, A., Dorai, D. R., Khan, S. B., Mahesh, M. T. R., Almusharraf, A., et al. (2024). Enhancing brain tumor classification in MRI scans with a multi-layer customized convolutional neural network approach. *Frontiers in Computational Neuroscience, 18*, 1418546. https://doi.org/10.3389/fncom.2024.1418546
5. Almadhor, M., et al. (2026). CT–MRI integration for brain-tumor analysis. *Frontiers in Computational Neuroscience*, 1798561. https://doi.org/10.3389/fncom.2026.1798561
6. Al-Saleh, A., Tejani, G. G., Mishra, S., Sharma, S. K., & Mousavirad, S. J. (2025). A federated learning-based privacy-preserving image processing framework for brain tumor detection from CT scans. *Scientific Reports, 15*, 23578. https://doi.org/10.1038/s41598-025-07807-8
7. Al-Saleh, A., Tejani, G. G., Mishra, S., Sharma, S. K., & Mousavirad, S. J. (2025). A federated learning-based privacy-preserving image processing framework for brain tumor detection from CT scans. *Scientific Reports, 15*, 23578. https://doi.org/10.1038/s41598-025-07807-8

8. Awotunde, J. B., et al. (2025). Federated learning augmented with convolutional neural networks for brain cancer classification. *Procedia Computer Science, 258*, 2617–2626. https://doi.org/10.1016/j.procs.2025.04.523
9. Balamurugan, A. G., Srinivasan, S., Preethi, D., Monica, P., Mathivanan, S. K., & Shah, M. A. (2024). Robust brain tumor classification by fusion of deep learning and channel-wise attention mode approach. *BMC Medical Imaging, 24*, 147. https://doi.org/10.1186/s12880-024-01323-3
10. *Federated learning augmented with convolutional neural networks for brain cancer classification*. (2025). *Procedia Computer Science, 258*, 2617–2626. https://doi.org/10.1016/j.procs.2025.04.523
11. Ilani, M. A., Shi, D., & Banad, Y. M. (2025). T1-weighted MRI-based brain tumor classification using hybrid deep learning models. *Scientific Reports, 15*, 7010. https://doi.org/10.1038/s41598-025-92020-w
12. Kar, S., & Singh, P. K. (2025). MBTC-Net: Multimodal brain tumor classification from CT and MRI scans using deep neural network with multi-head attention mechanism. *Medicine in Novel Technology and Devices, 27*, 100382. https://doi.org/10.1016/j.medntd.2025.100382
13. Kar, S., & Singh, P. K. (2025). MBTC-Net: Multimodal brain tumor classification from CT and MRI scans using deep neural network with multi-head attention mechanism. *Medicine in Novel Technology and Devices, 27*, 100382. https://doi.org/10.1016/j.medntd.2025.100382
14. Lu, N.-H., Huang, Y.-H., Liu, K.-Y., et al. (2025). Deep learning-driven brain tumor classification and segmentation using non-contrast MRI. *Scientific Reports, 15*, 27831. https://doi.org/10.1038/s41598-025-13591-2
15. Lv, J., Wu, L., Hong, C., Wang, H., Wu, Z., Chen, H., & Liu, Z. (2025). Multi-class brain malignant tumor diagnosis in magnetic resonance imaging using convolutional neural networks. *Brain Research Bulletin, 225*, 111329. https://doi.org/10.1016/j.brainresbull.2025.111329
16. Mao, Y., Kim, J., Podina, L., & Kohandel, M. (2025). Dilated SE-DenseNet for brain tumor MRI classification. *Scientific Reports, 15*, 3596. https://doi.org/10.1038/s41598-025-86752-y
17. Melo, T. L. de D., Moreira, R., Moreira, L. F. R., & Backes, A. R. (2026). Exploiting test-time augmentation in federated learning for brain tumor MRI classification. *arXiv*.
18. Muntaqim, M. D. Z., & Smrity, T. A. (2025). Federated learning framework for brain tumor detection using MRI images in non-IID data distributions.
19. Ozdemir, C., & Dogan, Y. (2024). Advancing brain tumor classification through MTAP model: An innovative approach in medical diagnostics. *Medical & Biological Engineering & Computing, 62*, 2165–2176. https://doi.org/10.1007/s11517-024-03064-5
20. Özkaraca, M., et al. (2023). Multiple brain tumor classification with dense CNN architecture using brain MRI images. *Life, 13*(2), 349. https://doi.org/10.3390/life13020349
21. Pacal, I. (2024). A novel Swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in MRI images. *International Journal of Machine Learning and Cybernetics, 15*, 3579–3597. https://doi.org/10.1007/s13042-024-02110-w
22. Panigrahi, S., Adhikary, D. R. D., & Pattanayak, B. K. (2025). Hybrid transfer learning and self-attention framework for robust MRI-based brain tumor classification. *Scientific Reports, 15*, 21343. https://doi.org/10.1038/s41598-025-09311-5
23. Pasvantis, K., & Protopapadakis, E. (2024). Enhancing deep learning model explainability in brain tumor datasets using post-heuristic approaches. *Journal of Imaging, 10*(9), 232. https://doi.org/10.3390/jimaging10090232
24. Paul, M., Tiwari, A., Barmashe, B., Rai, K., et al. (2026). Federated MobileNetV2 with ensemble meta-learning for privacy-preserving brain tumor classification. *Scientific Reports*. https://doi.org/10.1038/s41598-026-55847-5
25. Preetha, R., Priyadarsini, M. J. P., & Nisha, J. S. (2024). Automated brain tumor detection from magnetic resonance images using fine-tuned EfficientNet-B4 convolutional neural network. *IEEE Access, 12*, 112181–112195. https://doi.org/10.1109/ACCESS.2024.3442979
26. Rajput, I. S., Gupta, A., Jain, V., & Tyagi, S. (2024). A transfer learning-based brain tumor classification using magnetic resonance images. *Multimedia Tools and Applications, 83*, 20487–20506. https://doi.org/10.1007/s11042-023-16143-w
27. Reddy, C. K. K., Reddy, P. A., Janapati, H., Assiri, B., Shuaib, M., Alam, S., & Sheneamer, A. (2024). A fine-tuned vision transformer based enhanced multi-class brain tumor classification using MRI scan imagery. *Frontiers in Oncology, 14*, 1400341. https://doi.org/10.3389/fonc.2024.1400341

28. Reddy, C. K. K., Reddy, P. A., Janapati, H., Assiri, B., Shuaib, M., Alam, S., & Sheneamer, A. (2024). A fine-tuned vision transformer based enhanced multi-class brain tumor classification using MRI scan imagery. *Frontiers in Oncology, 14*, 1400341. https://doi.org/10.3389/fonc.2024.1400341
29. Song, W., et al. (2024). CT and MRI image fusion via multimodal feature interaction network. *Network Modeling Analysis in Health Informatics and Bioinformatics, 13*(1), 13. https://doi.org/10.1007/s13721-024-00449
30. Usha, M. P., Kannan, G., & Ramamoorthy, M. (2024). Multimodal brain tumor classification using convolutional Tumnet architecture. *Behavioural Neurology, 2024*, 4678554. https://doi.org/10.1155/2024/4678554
31. Wang, H., Zhang, J., Wu, S., Wei, H., Chen, X., Ou, Y., & Sun, X. (2025). MMSupcon: An image fusion-based multi-modal supervised contrastive method for brain tumor diagnosis. *Artificial Intelligence in Medicine, 170*, 103253. https://doi.org/10.1016/j.artmed.2025.103253
32. Zarenia, E., Far, A. A., & Rezaee, K. (2025). Automated multi-class MRI brain tumor classification and segmentation using deformable attention and saliency mapping. *Scientific Reports, 15*, 8114. https://doi.org/10.1038/s41598-025-92776-1