



Efficient Federated Aggregation Technique for Brain Tumor Classification Using Transfer Learning Approaches

Ashwini Jewalikar^{1,2*} , Rais Abdul Hamid Khan¹ , Deepak T. Mane³ 

¹ Department of CSE, Sandip University, Nasik 422001, India

² Pune Institute of Computer Technology, SPPU, Pune 411043, India

³ Department of Computer Engineering, VIT, Pune 411037, India

Corresponding Author Email: ashwinijewalikar@gmail.com

Copyright: ©2025 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300415>

ABSTRACT

Received: 6 February 2025

Revised: 18 April 2025

Accepted: 24 April 2025

Available online: 30 April 2025

Keywords:

federated learning, aggregation techniques, transfer learning, brain tumor classification, MRI scan images

Federated learning (FL) is a decentralized machine learning technique enabling machine learning models to be trained on local datasets, preserving data privacy. Aggregation techniques play a vital role in FL since updates from multiple clients located at remote places are shared and combined to build a global model without directly accessing raw data. Most widely used averaging Algorithms such as FedAvg and FedAdam, are often difficult to tune and exhibit unfavorable convergence behavior when the clients participating in averaging vary. We propose a novel approach to allow a selected number of clients to participate in aggregation and hence speed up the convergence by increasing the performance of the machine learning model used to classify the brain tumor types, meningioma, glioma, no tumor, and pituitary utilizing a combination of Figshare, SARTAJ and Br35H dataset. We adopted an innovative approach which modifies the Xception model architecture optimized for Brain tumor classification. We compared the performance of Xception, VGG19 and DenseNet201 pretrained classification models, where Xception model outperformed with a measured accuracy of 99.4. The proposed algorithm PC_FedAvg (Priority-based Client selection Federated Learning) is compared with existing FedAvg, FedAdam and FedProx, in terms of the number of communication rounds, along with the accuracy of the classification model and results show that PC_FedAvg has demonstrated improved performance than the selected approaches in terms of model accuracy, precision, and recall. PC_FedAvg gives the best results for the Xception model with a Precision of 90.52, Recall of 91.8, and accuracy of 91.6%.

1. INTRODUCTION

Federated learning (FL) is a decentralized machine learning paradigm that enables machine learning models to be trained collaboratively across multiple nodes while preserving data privacy. This approach allows the organizations to share and contribute to training the ML models without sharing the data [1]. The aggregation methodology adopted in the collaborative approach plays a vital role in the model's performance and reliability. Initially, a simple averaging method was adopted for aggregation. As per the studies and survey, there is a limited contribution in addressing communication efficiency, scalability issue and selecting the appropriate clients who can take part in FL [2, 3].

Significant progress has been made in classifying brain tumors from MRI scans using a variety of machine learning and deep learning techniques. These techniques help doctors plan treatments and improve the accuracy of diagnoses. Gliomas, meningiomas, and pituitary tumors are among the various forms of brain tumors that can be successfully classified by combining convolutional neural networks (CNNs) with other machine learning models [4, 5].

We are trying to adopt a FL approach for brain tumor

classification. Since early detection plays a vital role in brain tumor treatment [5, 6]. Hence, a novel approach PC_FedAvg (Priority Based Federated Averaging) is adopted to select a genuine and limited number of clients to take part in FL, which reduces communication rounds between clients and servers and increases the performance of the global model.

CNN is one of the appropriate machine learning models to accurately detect the tumors in the brain for the early start of the treatment [6]. Due to privacy concerns of medical health data and HIPAA guidelines, usually medical institutes are very rigid in sharing the data. Hence federated machine learning approach is suitable for medical health applications. Various aggregation techniques have been proposed such as parameter-based, output-based. In parameter-based aggregation weight parameters are objects for consideration. Few researchers worked on a precision-based FL approach [7-11]. We considered a parameter-based approach for our study. Some studies focused on aggregation with fine tuning and designed new aggregation algorithms based on the application requirements [12-15]. Emerging FL systems now incorporate secure global aggregation techniques designed for heterogeneous device networks, addressing critical security challenges in multi-device environments [16-18].

Medical image classification problems require large amounts of data on which models can be trained, because of the restrictions in sharing data of the patients, transfer learning is the most suitable approach for data-hungry applications. FL aggregation methodologies have been advanced to improve brain tumor classification accuracy while maintaining data privacy across distributed systems [19-21]. We adopted the Federated Transfer learning approach and used Xception, DenseNet201, and VGG19 models to classify brain tumor types.

We propose PC_FedAvg (Priority Based Client Federated Averaging), a novel aggregation algorithm which selects a limited number of clients based on their local model performance results, to take part in federated aggregation.

We adopted Cross-Silo federated Transfer learning approach for brain tumor classification and used the combination of Figshare, SARTAJ and Br35H benchmark datasets to classify the brain tumors into four types: glioma, meningioma, no tumor and pituitary. We analyzed the performance of our proposed PC_FedAvg with widely used aggregation algorithms FedAvg, FedProx and FedAdam. Results shows that PC_FedAvg is communication efficient as it reduces the communication rounds between client and server. We adopted Federated Transfer Learning approach by choosing Pretrained models Xception and VGG19 and DenseNet201. Initially, we compared the performance of VGG19, DenseNet201 and Xception models for brain tumor classification in centralized machine learning approach. Classification results show that the Xception model outperforms VGG19 and DenseNet with Precision 98.4, Recall 98.4, F1 score 98 and accuracy 99.4%, hence the Exception model is adopted for a federated approach for classification.

1.1 FL in healthcare

FL is essentially machine learning in a decentralized form. Conventional machine learning methods often include training models with data that is collected from several edge devices, such as laptops, and smartphones, and then transferred to a central server. This centralized data storage is the learning platform, where machine learning algorithms, such as neural networks, train themselves on aggregated data before making predictions on fresh data [1]. Improved data privacy, lower bandwidth requirements than transmitting raw data, and the capacity to use insights from various data sources without directly accessing the raw data are just a few benefits of this decentralized machine-learning technique [2, 3]. The FL approach includes 4 steps as shown in the general architecture in Figure 1. Client nodes or edge devices has their own dataset available with them and have a global model initially sent by the global server to all nodes.

Step1: Individual clients send their local parameters(weights) obtained by training the model on local dataset to the Global server.

Step2: The global server updates the global model by incorporating the global weight obtained by taking the average of all weights received (FedAvg).

Step3: Global server sends the global model to all the clients.

Step4: Clients update the local mode.

FL can be classified Based on devices involved and based on the data or models used as shown in Figure 2.

I Based on the devices involved

i) Cross-Silo: Suitable when large number of devices (Especially Homogeneous).

E.g., when different medical institutes taking part in FL process, referred to as cross-silo federation.

ii) Cross-Device: Limited number of devices are involved (Especially Heterogeneous).

E.g., when different wearable devices from different patients are collaborating for model training, referred to as cross-device.

II Based on the data or models used

i) Horizontal: Same features and same sample space.

ii) Vertical: Different features in same sample space.

iii) Federated Transfer Learning: Use of pre-trained machine learning models.

There are many application areas where FL is beneficial such as smart security systems, transportation systems health care, etc. some of the applications are explained below:

Healthcare: Hospitals can collaboratively train models for disease prediction without sharing patient data [1, 2].

Smart Transportation: FL can enhance Smart Transportation systems by aggregating insights from multiple devices while maintaining data confidentiality [1, 2].

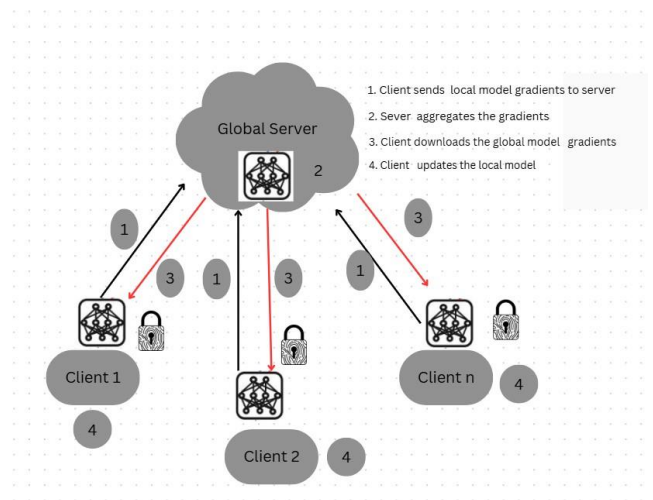


Figure. 1. FL general architecture

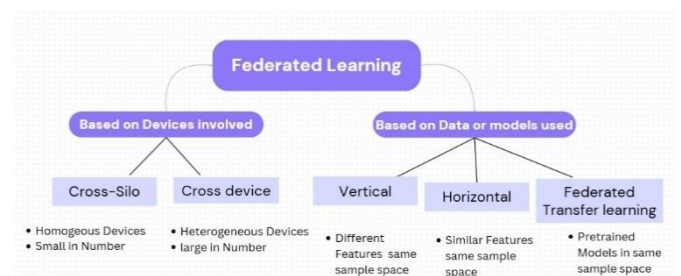


Figure. 2. FL classification types

1.2 Aggregation techniques used in FL and its limitations

As shown in Figure 1, the central server performs model aggregation on the weights and builds the global model. Aggregation Techniques can be classified into two types [7, 8].

A) Synchronous aggregation: All clients or nodes may not take part in the aggregation process. One can set the limit such as certain number of epochs or time duration to take part in aggregation. Delayed updates are not considered by the server.

B) Asynchronous aggregation: Usually all Clients or

nodes take part in the aggregation process, and delay in the model updates can be considered by the central server [7].

Traditional FedAvg algorithm which is widely used adopts a synchronous approach for communication rounds. The complete aggregation process can be represented with help of steps shown below.

Different Aggregation techniques or algorithms used in FL most widely used are discussed here:

FedAvg: FedAvg is the most commonly used weighted averaging technique based on the available data on the client nodes. It is the generalization of FedSGD (Federated Stochastic Gradient decent). More than one local gradient decent update is performed in FedAvg [1, 2, 7, 8]

Initialize the Global Model: Let the global model at round t be represented as: w^t where w^t is the parameter vector of the global model. Start with an initial global model w^t on the central server.

Local Training: Each client k receives the global model w^t , performs local training on the dataset and performs local training on its dataset D_k . where $k \in \{1, 2, \dots, K\}$. Each client minimizes a locally regularized objective function, The local optimization (e.g., SGD) updates the model parameters:

$$W_k^{t+1} = w^t - \eta \nabla L_k(w^t) \quad (1)$$

where, η is the learning rate and $\nabla L_k(w^t)$ is the gradient of the loss function L_k with respect to w^t using the local data D_k .

Model Aggregation: After local training, instead of each client sending its updated weights w_k^{t+1} back to the central server in FedAvg, a selected number of clients will update the weights to the server. At each round, FedAvg randomly selects the clients to participate in the aggregation process. Weights are assigned to the clients based on the quantity of data volume available to them. It adopts the synchronous approach. The limitation of FedAvg is it drops the clients if a certain number of epochs are not performed by them [7, 8, 11, 12, 14].

FedProx: FedProx addresses the heterogeneity of devices and performs well when devices with different computational powers are involved the FL. It allows partial work instead of eliminating clients like FedAvg. In FedProx the proximal term is added to the local objective function of clients to make the local updates stabilize. Each client minimizes a locally regularized objective function, adding a proximal term to reduce the discrepancy between its local model and the global model:

$$\min [F_k(w_k) + \frac{\mu}{2} \|w_k - w^t\|^2] \quad (2)$$

where, $F_k(w_k)$ is the local objective function for client k (e.g., empirical risk or loss function on D_k).

This proximal term $\frac{\mu}{2} \|w_k - w^t\|^2$ acts as a penalty that limits how much each client's local model can deviate from the global model.

Clients perform this optimization for a set number of epochs or until convergence.

FedAdam: Adam optimizer is adopted for FL by FedAdam. Parameters are updated using both first and second moment of gradients. The formula for updating is:

$$W_{t+1} = W_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} m_t \quad (3)$$

where, m_t is mean of gradients, v_t is variance of the gradients, ϵ is small constant added and η is the learning rate [14].

FedAdam address the challenges of data heterogeneity.

FedYogi: This algorithm is based on Yogi optimizer and it is gradient based optimization which address the limitation of FedAdam (excessively large variance) [14, 15]:

$$w_{t+1}^k = w_t - \frac{\eta}{\sqrt{v_t^k} + \epsilon} m_t^k \quad (4)$$

FedNova: FedNova addresses the non-IID distribution of data in the federated set up and enhances the model aggregation phase. This algorithm introduces a technique to normalize the updates from each client on its local iteration number, before updating the global model.

One of the concerns of these algorithms is limited focus on the number of clients taking part in the aggregation process. It is observed that when the number of clients taking part in the aggregation increases, the performance of few (FedAvg) decreases and some require more communication rounds for convergence [1, 14].

Below are the limitations of aggregation techniques used in FL:

Participating Nodes Availability: All clients may not be available during each training round; hence aggregation results may be biased based on available clients

Data Heterogeneity: Clients usually having non-IID Data distribution with them and traditional algorithms, e.g., FedAvg performs poorly in such cases [1, 2].

Fairness Issues: Clients with smaller datasets may have less influence on the global model and hence global model may be biased.

Communication Overhead: Frequent communication between client and server consumes network bandwidth and time for global model convergence.

Research Significance:

- To identify and emphasize the latest research trends in aggregation techniques used by federated machine learning.
- Design communication efficient aggregation technique for brain tumor classification in 4 types and analyze the performance of the proposed technique with existing approaches.
- Improve the performance of the global model used to classify brain tumors.

We are trying to balance the limitations of FedAvg which struggles with non-IID data and FedProx, which struggles with slow convergence by selecting a limited number of clients based on local model performance. To address the performance limitation of the machine learning model due to appropriate client selection for aggregation, we proposed a novel approach for aggregation called PC_Fed average which selects the limited number of clients based on their overall performance during local training. Detail algorithm is explained in Section 4. Table 1 gives the details of the aggregation algorithm, its limitations and its applications.

We provide extensive evaluations of the proposed PC_FedAvg algorithm using benchmark image classification datasets, demonstrating its robustness to unbalanced data distributions; and also compared the proposed method to Federated Averaging on empirical experiments, and with fewer communication rounds we obtain comparable accuracy.

Table 1. Survey on transfer learning approach for brain tumor classification

Reference	Year	Methodology	Pretrained Models Used	Performance Parameter	Model with the Highest Score
Albalawi et al. [19]	2024	Federated transfer learning	Modified VGG16	Accuracy	VGG 68%
Amarnath et al. [20]	2024	Transfer learning	Xception, EfficientNetV2-S, ResNet152V2, ResNet50, VGG16	F1 score	Xception 0.9817
Malakouti et al. [21]	2024	Transfer learning	Modified GoogleNet	Accuracy, F1 score	GoogleNet 99.3, 99
Chattopadhyay and Maitra [22]	2022	CNN (Brats dataset)	-----	Accuracy	99.74
Rahman and Islam [23]	2023	CNN (Figshare dataset)	-----	Accuracy	97.60
Khan et al [24]	2022	Transfer learning	VGG-16 combined with 23 Layer CNN (Figshare dataset)	Accuracy	97.8
Khaliki and Başarslan [25]	2024	Transfer learning	InceptionV3, EfficientNet, VGG19, VGG16	Accuracy, F1 score, AUC, Recall, Precision	VGG16 98%, 97%, 99%, 98%, 98%

2. LITERATURE SURVEY AND METHODOLOGY

In general research Literature survey is performed by referring to Web of Science and Scopus database. The methodology is shown in Figure 3. Firstly, we have gone through the publications in this emerging domain in last few years and performed the search for research articles, proceedings and review articles with help of the keywords “Federated Learning+Healthcare applications”, “Federate Learning+deep learning” and “Federated Learning+Aggregation techniques”.

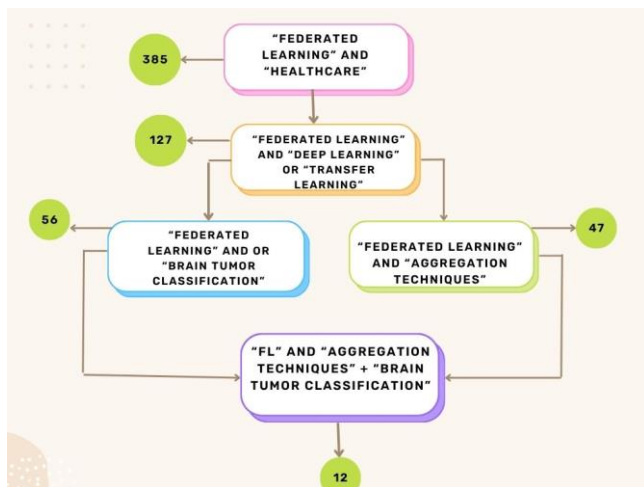
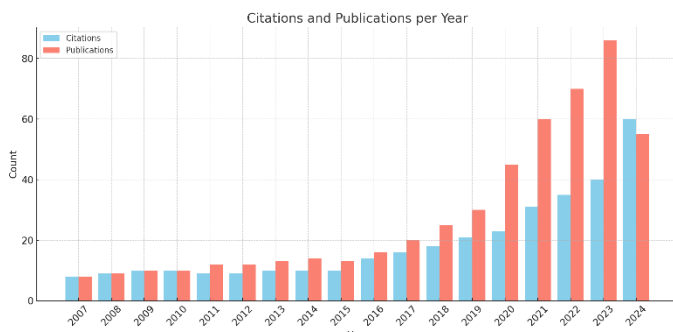
**Figure 3.** Methodology of literature review**Figure 4.** Research publications and citations

Figure 4 shows the recent research trends (year wise) which highlights the popularity of the trend. The graph shown publications Vs Citations and the count is highest in 2024.

Literature survey is performed by considering the following two aspects.

2.1 Brain tumor classification using MRI images

Deep learning models were found to be more effective in the classification of brain tumors. Authors performed extensive studies in brain tumor detection using deep learning algorithms and techniques using different datasets along with Brain tumor MRI dataset [18]. Brain tumors are on the 10th position for the cause of mortality. We performed a literature survey with a centralized machine learning approach and a federated machine learning approach. The present research focused on centralized machine-learning approaches. Traditional machine-learning approaches are used to perform image analysis and classification to detect brain tumors. Albalawi et al. [19] adopted a federated approach to classify the brain tumors in 4 classes, glioma, meningioma, and no tumor and pituitary on the same dataset Figshare. Traditional FedAvg is used for aggregation. The machine learning model used is CNN. Amarnath et al. [20] performed experiments using transfer learning on five pre-trained deep learning models-ResNet50, Xception, EfficientNetV2-S, ResNet152V2, and VGG16 models on this dataset resulting Xception model achieving the highest performance with a test F1 score of 0.9817, followed by EfficientNetV2-S with a test F1 score of 0.9629. Malakouti et al. [21] chose modified GoogleNet to classify the people in to two classes sick and healthy people leading to accuracy 99.3 and F1 score 99. Table 1 mentions the details of work done in brain tumor detection.

2.2 Aggregation techniques in FL survey

As mentioned, the researchers worked on various aggregation approaches by considering the limitations. details of state of artwork in mentioned in Table 2.

Observations on Literature Survey: Limited focus is given by researchers to select appropriate clients to take part in the aggregation process. Multi classification and federated transfer learning approach is rarely applied to classify the brain tumors using Brain tumor MRI dataset in a federated set up.

Table 2. Survey on aggregation techniques

Reference	Summary	Aggregation Algorithms	Limitations	Dataset	Model	Medical Application
Aledhari et al. [1]	Survey on technologies and protocols, frameworks in FL	FedAvg, FedProx, FedAdam	Limited discussion on client selection	NA	NA	Yes
Qi et al. [7]	Comprehensive survey on model aggregation techniques and limitations in many applications including smart healthcare such that security and statistical heterogeneity are mentioned	FedAvg, FedProx, FedNova, Scaffold, MOON, Zeno, FedAvg are discussed	No specific client selection techniques are discussed	NA	NA	Yes
Lee et al. [10]	FedAvg and FedSGD are compared for client parallelism and communication efficiency, FedAvg performs better than FedSGD	FedAvg, FedSGD	No specific client selection techniques are discussed	MNIST, CIFAR	CNN, LSTM	No
Reyes et al. [11]	Proposed precision-weighted FL algorithm, a novel variance-based averaging scheme to aggregate model weights across clients	FedAvg, FedSGD	Precision weighted client's selection, excludes the genuine clients	MNIST, Fashion-MNIST, CIFAR	DNN	No
Tao Sun et al. [12]	Decentralized averaging is proposed to reduce the communication rounds	DFedAvgM, FedAvg	Random selection of clients reduces the performance	MNIST, CIFAR	CNN	No
Collins et al. [13]	D-GD Fedavg with one local update is analysed with FedAvg	D-GD and FedAvg	Random	CIFAR	DNN	No
Mansour et al. [14]	FedAvg is fine tuned to get the better accuracy this paper explores and investigates several FL aggregation strategies and algorithms	FedAvg, Fine-tuned FedAvg	Random selection of clients	MNIST, Fashion-MNIST	No	
Moshawrab et al. [15]	Gives technical comparison on existing aggregation rules	NIL	No specific discussion on client selection strategies	NA	NA	NA
Nanayakkara et al. [16]		FedAvg, FedProx, FedYogi, FedAvgm	General discussion on client selection strategies	NA	NA	NA

3. PROPOSED PC_FEDAVG AND ARCHITECTURE OF SYSTEM

By considering the limitations found in the literature survey, we identified that aggregation of local models trained on divergent data can result in poor global model performance across the entire population traditional aggregation algorithms treats all the clients equally while some clients may have better quality or more representative data.

We designed PC_FedAvg (Priority-based Federated Averaging) by selecting a limited number of clients prioritizing those who are likely to benefit in each round of communication between client and server. Which helps in better representation and faster convergence and reduced gradient conflicts.

Some common priority Matrices can be Data Diversity, Model contribution or local accuracy and resource availability or reliability have chosen local accuracy as priority mastics and chosen the 60% clients with highest local model accuracy.

Prioritizing the client selection in FL aids in the following manner non-IID data makes local models go in opposite directions when optimizing. By choosing more aligned or complementary customers, the system eliminates conflicting updates and converges better.

We used pre-trained deep learning models, Xception, DenseNet201, and VGG19 to classify the brain tumors in to four types glioma, meningioma, and no tumor and pituitary. Cross-silo Federated approach is used and hence number of clients taking part in FL is limited to max 5.

3.1 Mathematical representation of PC_FedAvg algorithm

Let

K : Total number of clients.

D_k : Local dataset for client k .

w^t : Global model weights at communication round t .

w_k^t : Client k 's local model weights at round t .

The detailed mathematical representation of PC_FedAvg is explained in Table 3.

3.1.1 Details of representation of the selected clients set in PC_FedAvg

Let: K represent set of all available clients $K=\{1,2,3,4,5,...,K\}$, S_t K represent the subset of clients selected for participation in round t , $S_t=\{k1,k2,k3,...,Kn\}$.

The server aggregates the weights using a weighted average based on each client's data size

FedAvg Aggregation:

$$w^{t+1} = \sum_{k=1}^K \frac{n_k}{n} \omega_k^{t+1} \quad (5)$$

$$w^{t+1} = \sum_{k \in S^t}^K \frac{n_k}{\sum_{j \in S^t} n_j} \omega_k^{t+1} \quad (6)$$

3.1.2 PC_FedAvg aggregation incorporating selected clients

In FedAvg, if only the updates from the selected clients S^t used to update the global model. The aggregation formula then becomes:

where, S^t is set of top 60 % of available clients with highest accuracy in that communication round.

Table 3. Mathematical representation of PC_FedAvg algorithm

Steps	Description
Initialize the Global Model	Let the global model at round t be represented as: w^t , where w^t is the parameter vector of the global model. Start with an initial global model w^t on the central server.
Local Training	Each client k receives the global model w^t , performs local training on dataset and performs local training on its dataset D_k , where $k \in \{1, 2, \dots, K\}$. Each client minimizes a locally regularized objective function. The local optimization (e.g., SGD) updates the model parameters: $w_k^{t+1} = w^t - \eta \nabla L_k(w^t)$, where, η is the learning rate and $\nabla L_k(w^t)$ is the gradient of the loss function L_k with respect to w^t using the local data D_k .
Model Aggregation PC_FedAvg Aggregation Incorporating Selected Clients	After local training, instead of each client sends its updated weights w_k^{t+1} back to the central server in FedAvg Selected number of clients will update the weights to server. In FedAvg, if only the updates from the selected clients S^t used to update the global model. The aggregation formula then becomes: $w^{t+1} = \sum_{k \in S^t}^K \frac{n_k}{\sum_{j \in S^t} n_j} \omega_k^{t+1}$. Here, only the clients in S^t contribute to the global model update.
Global Model Update	The weights are normalized over the selected clients to ensure proper averaging. This ensures that even with partial client participation, the global model updates remain consistent and balanced. The central server updates the global model w^{t+1} using the aggregated weights. The updated global model w^{t+1} is redistributed to the clients for further rounds of training, and the process is repeated over multiple rounds.

4. EXPERIMENTAL SETUP AND RESULTS

We adopted the Federated Transfer Learning approach with a modified aggregation technique for brain tumor classification (dataset brain tumor MRI) We used well known Xception, VGG19, and Densenet201 models as shown in Figures 5 and 6, and modified the Xception model architecture by adding dense and pooling layers to get the improved results for classification of brain tumor types in to four classes.

Initially, the modified global pre-trained Xception model available with the server is trained on the dataset available with the server and then shared with the client nodes. The experiment is explained in detail below.

4.1 Dataset description and preprocessing

Brain tumor MRI dataset is used for experimentation, which is combination of Figshare, SARTAJ dataset and Br35H. It contains 6695 images of human brain MRI images classified in to glioma, meningioma, no tumor and pituitary. Figure 7 represents the detailed class wide distribution of images in the dataset. Training data contains 4400 images, and validation and, testing data contain total 1225 images and 1070 images respectively with 4 classes dataset distribution is shown in Table 4.

For better results data preprocessing is performed on dataset. The scanning of directories for images starts, from where the region of interest is cropped with a cropping technique on the relevant area. In addition to this, each image then gets resized to a particular dimension so that all the images are of the same size. Then, the pixel values are normalized to the range $[0, 1]$ for improved model performance. Images are to be aligned in specific required dimensions, so a standard size of images is required. So here, 240×240 is assumed as the standard image size, which is resized to $224 \times 224 \times 3$. Augmentation further

increases the data availability and improves the future extraction process. After standard augmentation as shown in Figure 8, the number of images in dataset becomes 14954.

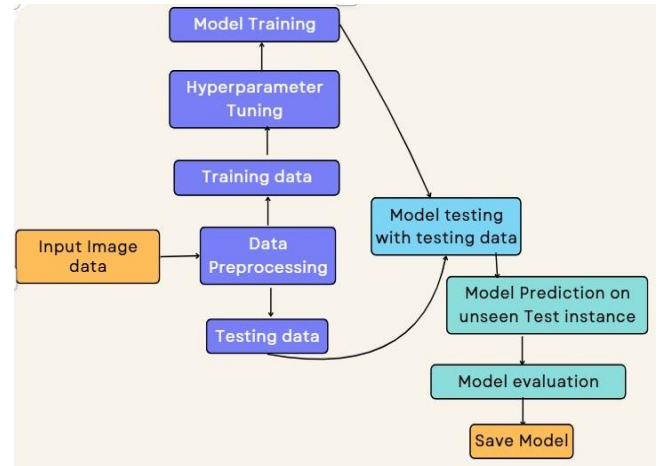


Figure 5. Flowchart federated model training and sharing

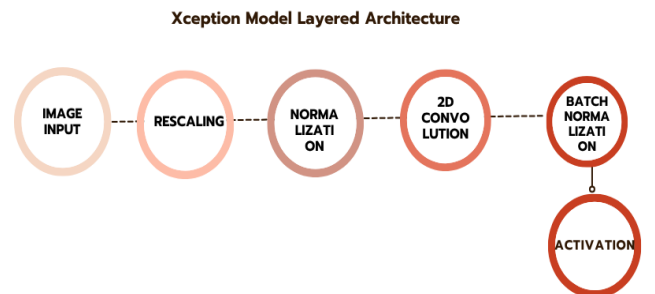


Figure 6. Modified Xception transfer learning model

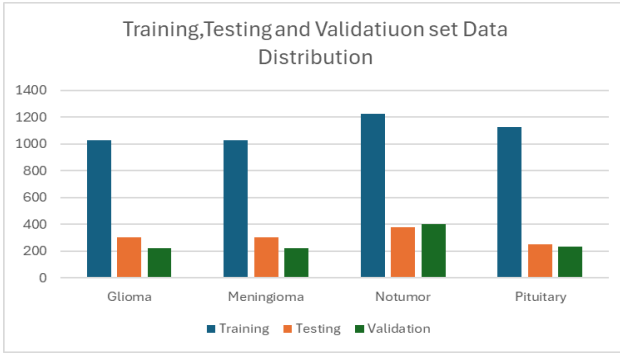


Figure 7. Visualization of the images across four classes

Table 4. Training and testing image distribution for each class

Set	Glioma	Meningioma	No Tumor	Pituitary
Training	1025	1025	1225	1125
Testing	300	300	375	250
Validation	220	220	400	230

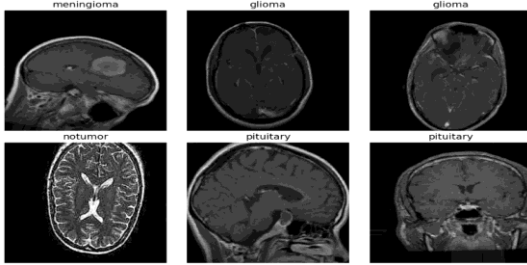


Figure 8. Sample images from the brain tumor MRI dataset after augmentation

4.2 Xception model architecture

Figures 6 and 9 show the architecture of modified Xception, along with the parameter details and hyperparameters set in Tables 4 and 5, respectively.

Xception model (Extreme Inception) is a deep Convolutional Neural Network abbreviated as CNN, which enables improvement of computational efficiency, offering the high feature extraction capabilities from an Image. The standard convolutions are replaced here with depth-wise separable convolutions which reduce the number of parameters required and increase efficiency.

The task of tumor classification using Brain Tumor MRI dataset. The architecture consists of three important stages:

a. Entry flow: Primarily reduces spatial dimensions while capturing low-level features using convolutions, depth wise separable convolutions, and max pooling layers.

b. Middle flow: It refines the extracted features via 8 repeated blocks of depth wise convolutions.

c. Exit flow: Extracts the discriminative features by using additional convolutions, followed by global average pooling layer, a fully connected layer of 256 neurons and a softmax classifier for the final predictions with 4 classes as output.

Table 5. Xception model parameter details

Model Parameters	Size
Total Parameters	21,124,268 (80.58MB)
Trainable Parameters	21,069,740 (80.37MB)
Non-Trainable parameters	54,528 (213.00KB)

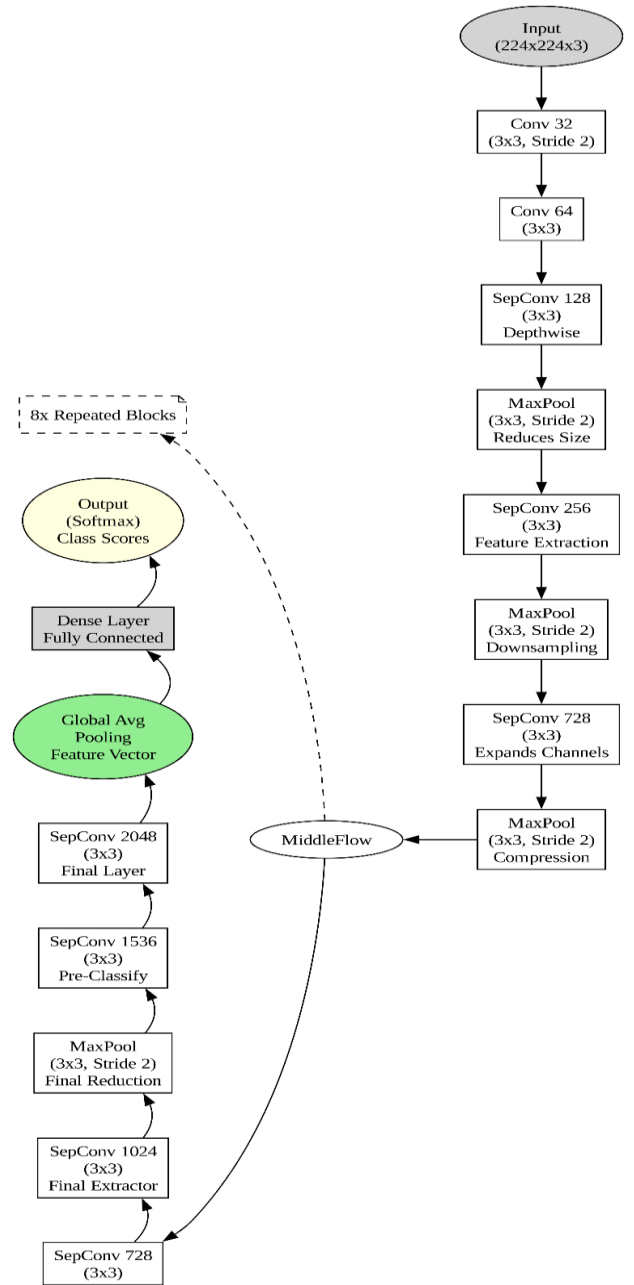


Figure 9. Xception model architecture

4.3 Experimental results and analysis

The modified Xception model along with widely used VGG19 and DenseNet201 is used for experimentation on the same dataset. The dataset is divided in to 80% and 20% for Training and testing respectively. Number of clients selected were 5 and data distribution among the clients is IID. Adam optimizer function is used along with softmax as an output activation function. Learning rate for the experiment was kept 0.001. Tables 5 and 6 show the hyperparameter settings for the Xception model for experimentation. Experimental results shows that the modified Xception model outperforms VGG19 and DenseNet201 with 99.4% of accuracy where as VGG19 gives 98.6% and DenseNet measured accuracy is 96.6%. For details refer to Table 7. Class-wise performance is also shown in Tables 8-10.

This experiment is further extended to adopt the federated approach and based on earlier results Xception is chosen as a base model for decentralized learning.

Table 6. Hyperparameters of Transfer Learning model

Hyper Parameter	Value
Input image shape	224×224×3
Batch Size	32
Output Activation function	Softmax
optimizer	Adam
Epochs	5
Learning Rate	0.001
Criteria	Cross Entropy loss

Table 7. Performance of Xception, VGG19 and DenseNet201 models

Model	Training Accuracy	Training Loss	Testing Accuracy	Testing Loss
Xception	0.994	0.15	0.988	0.21
VGG19	0.986	0.204	0.965	0.33
DenseNet201	0.966	0.321	0.957	0.41

Table 8. Class wise performance of the Xception model

Tumor types	Precision	Recall	F1 Score
Glioma	0.991	0.966	0.978
Meningioma	0.958	0.982	0.97
No tumor	1	0.987	0.994
Pituitary	0.983	0.997	0.99
Accuracy			0.984

Table 9. Class wise performance of the VGG19 model

Tumor types	Precision	Recall	F1 Score
Glioma	0.985	0.954	0.969
Meningioma	0.933	0.953	0.943
No tumor	0.985	0.99	0.987
Pituitary	0.973	0.979	0.976
Accuracy			0.969

Table 10. Class wise performance of DenseNet201

Tumor types	Precision	Recall	F1 Score
Glioma	0.984	0.913	0.947
Meningioma	0.901	0.932	0.917
No tumor	0.987	0.99	0.988
Pituitary	0.948	0.983	0.965
Accuracy			0.956

4.3.1 Performance parameters

Evaluation metrics used for brain tumor classification to gain comprehensive insights are Accuracy, F1 score, Precision and Recall which serve as a benchmark for model performance measures.

Accuracy: It is the proportion of total images Vs Correctly classified images. In term of confusion matrix

$$Accuracy = \frac{(TP + TN)}{TP + TN + FP + FN}$$

Precision: It is the measure of calculating the correctness of positive predictions

$$Precision = \frac{TP}{TP + FP}$$

Recall: Recall is measure of models ability to correctly identify all instances of particular class among all instances

belonging to that class.

$$Recall = \frac{TP}{TP + FN}$$

F1 score: F1 score is a harmonic mean of precision and recall. It balances both positive and negative values. It is crucial metric for brain tumor classification where each type of tumor identification is vital task.

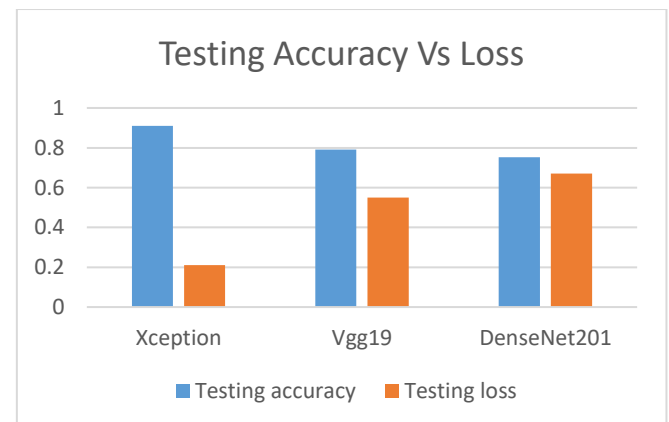
4.3.2 Performance evaluation of PC_FedAvg using Xception model

We concluded the experiment with a Brain tumor MRI dataset for analyzing the performance of the newly designed PC_FedAvg explained in Section 4.

By focusing on clients with different patient populations, imaging modalities, or types of pathology, the model learns to generalize over anatomy variations, disease appearance, and scanner variation. This minimizes model bias towards any one data source. If you sample the clients at random, the model could overfit to strong patterns (e.g., more from one hospital). Prioritization prevents minority cases or unusual disease types from other institutions from being overwhelmed. A better balanced, more generalizable model that will work well in all populations. Figure 10 represents the testing accuracy Vs testing loss graph when PC_Fedavg is used for aggregation along with Xception model to classify the tumors with 91.6 % accuracy in 3 rounds as described in Table 11. Figures 11 and 12 show the class wise classification performance of PC_FedAvg with Xception model resulting better performance compared to VGG19 and DenseNet201.

We compared the performance of PC_FedAvg with FedAvg, FedProx, and FedAdam when experimenting with the Xception model. Results are represented in Figure13 and 14 which magnifies the performance of PC_FedAvg with 91.6% accuracy. The number of clients taking part in aggregation is chosen to be 5. Results show that Accuracy of Xception model is 91.6% along with a smaller number of communication rounds detailed results are shown in the Table 11.

Since PC_FedAvg has sorted 60% of the total clients based on the local model accuracy, the convergence speed has increased, and hence total number of rounds reduced to 3 compared with FedAvg and FedAdam who waits for all the clients to complete the process. As local model accuracy increased, global model accuracy also increases in the aggregation.

**Figure 10.** Performance evaluation of Xception VGG19 and DenseNet201 on BrainMRI dataset

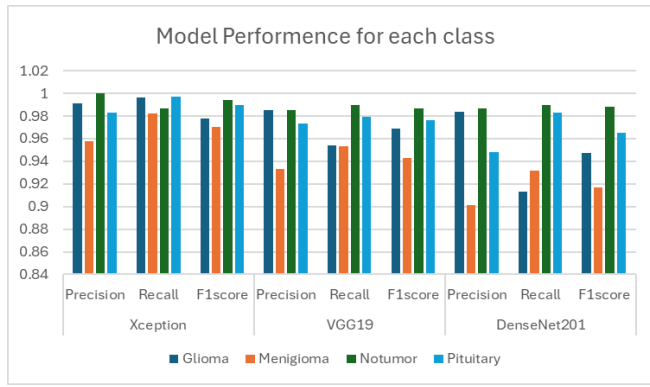


Figure 11. Xception, VGG19 and DenseNet201 model performance for meningioma, glioma, no tumor and pituitary

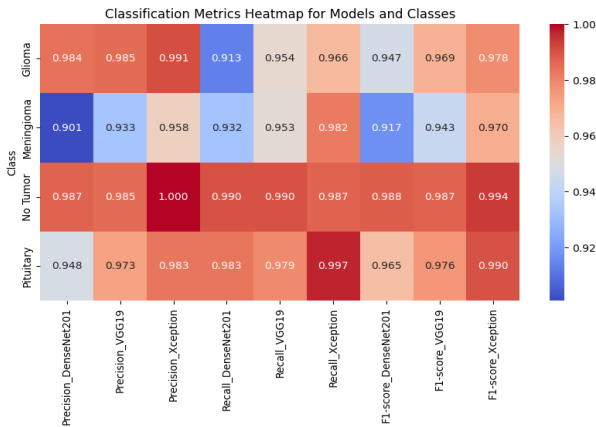


Figure 12. Model comparison results

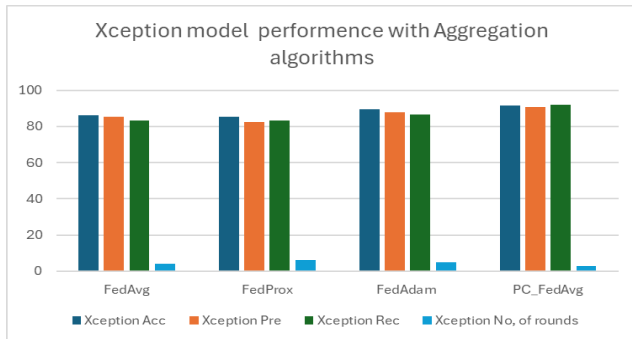


Figure 13. Xception model performance model with aggregation

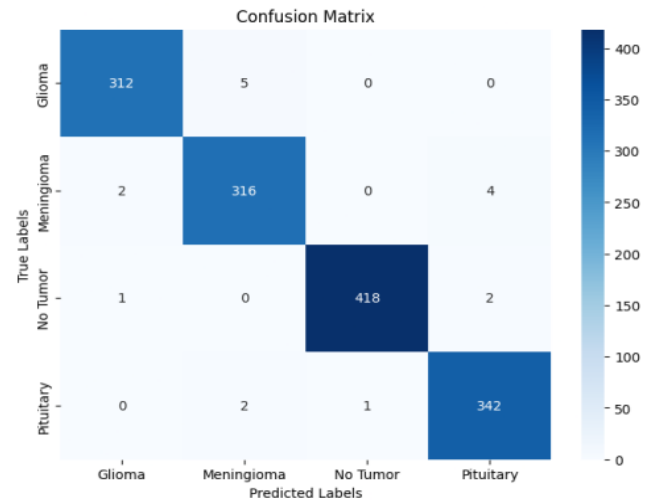


Figure 14. Confusion matrix

Table 11. Xception model performance with FedAvg, FedProx and FedAdam

Aggregation Algorithm	Acc	Pre	Rec	No. of Rounds
FedAvg	86.2	85.18	83.2	4
FedProx	85.2	82.4	83.2	6
FedAdam	89.3	87.8	86.5	5
PC_FedAvg	91.6	90.52	91.8	3

4.4 Comparison of the proposed approach with existing studies

Table 12 shows the details of state of art work using transfer learning models for brain tumor classification very few researchers adopted the privacy-preserving set up for brain tumor classification. FL frameworks employing convolutional neural networks and FedAvg aggregation achieve 85.55% classification accuracy on the BrainMRI dataset for brain tumor detection [4, 19]. We adopted to federated setup with modified PC_FedAvg leading to 91.6% accuracy and less communication rounds compared with FedAvg and FedProx due to limited number of clients participating in the aggregation process based on the local performance and hence faster global model convergence. PC_FedAvg generalizes the performance of a machine learning model with a uniform client selection approach to all image classification applications.

Table 12. Comparison of results with existing state-of-the-art work using transfer learning

Reference	ML Technique/Algorithm	Dataset	Accuracy	Privacy of Data
Khaliki and Başarslan [25]	3 Layer CNN and VGG16	Brai MRI	98%	No
Sangui et al. [26]	Unet with image segmentation	BRATS2020	99%	No
Pravitasari et al. [27]	Unet+VGG16	BrainMRI	96.1%	No
Rasool et al. [28]	CNN	Brain MRI	NA (Strait of art survey)	No
Shyamala and Mahaboob Basha [29]	CNN with texture analysis	BRATS	95.21	No
Mitra et al. [30]	VGG16	Brain MRI	97.2%	No
Al-Asfoor et al. [31]	DenseNet	BRats, Figshare, SARTAJ	96.2%	No
Prakash et al. [32]	DenseNet121	BrainMRI	97.39	No
Ay et al. [33]	CNN	Brain MRI	85.55%	Yes
Proposed PC_FedAvg	Xception	Figshare, Brats and SARTAJ	91.6% (with less communication rounds)	Yes

5. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

A novel Federated approach for brain tumor classification in to 4 types glioma, meningioma, no tumor and pituitary by using modified Xception transfer learning model had successfully developed and evaluated. Experimental results show the remarkable accuracy of Xception model, 99.6% compared to VGG19 and DenseNet201, who pertains 98.6 and 96.6% accuracy. Xception model is then adopted for proposed PC_FedAvg aggregation technique in FL setup for brain tumor classification. Results show that PC_FedAvg outperforms with 91.6% accuracy with 3 communication rounds. These results showcase the notable improvement in existing results for brain tumor classification. Due to adoption of federated approach the privacy of the data is preserved, and the radiologist and oncologist may get benefit as they can carry diagnosis quickly.

Because of the limited clients involved in the aggregation step according to the local performance and thus quicker global model convergence. PC_FedAvg is able to generalize the performance of the machine learning model with uniform client choose strategy to any image classification task this approach can be extended further for various medical imaging applications such as X-Ray scan diagnosis, cervical cancer detection etc. We believe that this research can contribute in designing the vertical federated architecture for medical applications.

REFERENCES

- [1] Aledhari, M., Razzak, R., Parizi, R.M., Saeed, F. (2020). Fedmodels, inception a survey on enabling technologies, protocols, and applications. *IEEE Access*, 8: 140699-140725.
<https://doi.org/10.1109/ACCESS.2020.3013541>
- [2] Ding, J., Tramel, E., Sahu, A.K., Wu, S., Avestimehr, S., Zhang, T. (2022). Federated learning challenges and opportunities: An outlook. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, Singapore, pp. 8752-8756.
<https://doi.org/10.1109/ICASSP43922.2022.9746925>
- [3] Nguyen, D.C., Pham, Q.V., Pathirana, P.N., Ding, M., Seneviratne, A., Lin, Z., Dobre, O., Hwang, W.J. (2022). Federated learning for smart healthcare: A survey. *ACM Computing Surveys*, 55(3): 60.
<https://doi.org/10.1145/3501296>
- [4] Mahesh, G., Yogesh, K.M. (2024). Brain tumor detection and classification using MRI images. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*. *International Journal for Science Technology and Engineering*, Undefined.
<https://doi.org/10.22214/ijraset.2024.64719>
- [5] Onuri, E.E., Adeyemi, J., Umeaka, K.C. (2024). MRI Based brain tumour classification using convolutional neural networks: A systematic review and meta-analysis. *British Journal of Computer, Networking and Information Technology*, 7(4): 27-46.
<https://doi.org/10.52589>
- [6] Mohanty, N., Sarmadi, M. (2024). Brain tumor MRI classification and identification using an image classification model via convolutional neural networks. *MEDRxiv*.
<https://doi.org/10.1101/2024.09.13.23299832>
- [7] Qi, P., Chiaro, D., Guzzo, A., Ianni, M., Fortino, G., Piccialli, F. (2024). Model aggregation techniques in federated learning: A comprehensive survey. *Future Generation Computer Systems*, 150: 272-293.
<https://doi.org/10.1016/j.future.2023.09.008>
- [8] Shailesh, S., James, J. (2024). Types of federated learning and aggregation techniques. In *Federated Learning*. Apple Academic Press, pp. 23-45.
<https://doi.org/10.1201/9781003497196-2>
- [9] McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pp. 1273-1282.
- [10] Lee, S., Sahu, A.K., He, C., Avestimehr, S. (2023). Partial model averaging in federated learning: Performance guarantees and benefits. *Neurocomputing*, 556: 126647.
<https://doi.org/10.1016/j.neucom.2023.126647>
- [11] Reyes, J., Di Jorio, L., Low-Kam, C., Kersten-Oertel, M. (2021). Precision-weighted federated learning. *arXiv Preprint arXiv: 2107.09627*.
<https://doi.org/10.48550/arXiv.2107.09627>
- [12] Sun, T., Li, D., Wang, B. (2022). Decentralized federated averaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4289-4301.
<https://doi.org/10.1109/tpami.2022.3196503>
- [13] Collins, L., Hassani, H., Mokhtari, A., Shakkottai, S. (2022). Fedavg with fine tuning: Local updates lead to representation learning. *Advances in Neural Information Processing Systems*, 35: 10572-10586.
<https://arxiv.org/abs/2205.13692>
- [14] Mansour, A.B., Carenini, G., Duplessis, A., Naccache, D. (2022). Federated learning aggregation: New robust algorithms with guarantees. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, Nassau, Bahamas, pp. 721-726.
<https://doi.org/10.1109/icmla55696.2022.00120>
- [15] Moshawrab, M., Adda, M., Bouzouane, A., Ibrahim, H., Raad, A. (2023). Reviewing federated learning aggregation algorithms; strategies, contributions, limitations and future perspectives. *Electronics*, 12(10): 2287. <https://doi.org/10.3390/electronics12102287>
- [16] Nanayakkara, S.I., Pokhrel, S.R., Li, G. (2024). Understanding global aggregation and optimization of federated learning. *Future Generation Computer Systems*, 159: 114-133.
<https://doi.org/10.1016/j.future.2024.05.009>
- [17] Pillutla, K., Kakade, S.M., Harchaoui, Z. (2022). Robust aggregation for federated learning. *IEEE Transactions on Signal Processing*, 70: 1142-1154.
<https://doi.org/10.1109/TSP.2022.3153135>
- [18] Rasool, N., Bhat, J.I. (2024). Brain tumour detection using machine and deep learning: A systematic review. *Multimedia Tools and Applications*, 84: 11551-11604.
<https://doi.org/10.1007/s11042-024-19333-2>
- [19] Albalawi, E., TR, M., Thakur, A., Kumar, V.V., Gupta, M., Khan, S.B., Almusharraf, A. (2024). Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor. *BMC Medical Imaging*, 24(1): 110.
<https://doi.org/10.1186/s12880-024-01261-0>

- [20] Amarnath, A., Al Bataineh, A., Hansen, J.A. (2024). Transfer-learning approach for enhanced brain tumor classification in MRI imaging. *BioMedInformatics*, 4(3): 1745-1756. <https://doi.org/10.3390/biomedinformatics4030095>
- [21] Malakouti, S.M., Menhaj, M.B., Suratgar, A.A. (2024). Machine learning and transfer learning techniques for accurate brain tumor classification. *Clinical eHealth*, 7: 106-119. <https://doi.org/10.1016/j.ceh.2024.08.001>
- [22] Chattopadhyay, A., Maitra, M. (2022). MRI-based brain tumour image detection using CNN CNN-based deep learning method. *Neuroscience Informatics*, 2(4): 100060. <https://doi.org/10.1016/j.neuri.2022.100060>
- [23] Rahman, T., Islam, M.S. (2023). MRI brain tumor detection and classification using parallel deep convolutional neural networks. *Measurement: Sensors*, 26: 100694. <https://doi.org/10.1016/j.measen.2023.100694>
- [24] Khan, M.S.I., Rahman, A., Debnath, T., Karim, M.R., Nasir, M.K., Band, S.S., Mosavi, A., Dehzangi, I. (2022). Accurate brain tumor detection using deep convolutional neural network. *Computational and Structural Biotechnology Journal*, 20: 4733-4745. <https://doi.org/10.1016/j.csbj.2022.08.039>
- [25] Khaliki, M.Z., Başarslan, M.S. (2024). Brain tumor detection from images and comparison with transfer learning methods and 3-Layer CNN. *Scientific Reports*, 14(1): 2664. <https://doi.org/10.1038/s41598-024-52823-9>
- [26] Sangui, S., Iqbal, T., Chandra, P.C., Ghosh, S.K., Ghosh, A. (2023). 3D MRI segmentation using U-Net architecture for the detection of brain tumor. *Procedia Computer Science*, 218: 542-553. <https://doi.org/10.1016/j.procs.2023.01.036>
- [27] Pravitasari, A.A., Iriawan, N., Almuhyar, M., Azmi, T., Irhamah, I., Fithriasari, K., Purnami, S.W., Ferriastuti, W. (2020). UNet-VGG16 with transfer learning for MRI-based brain tumor segmentation. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 18(3): 1310-1318. <http://doi.org/10.12928/telkomnika.v18i3.14753>
- [28] Rasool, M., Noorwali, A., Ghandorh, H., Ismail, N.A., Yafooz, W.M. (2024). Brain tumor classification using deep learning: A state-of-the-art review. *Engineering, Technology & Applied Science Research*, 14(5): 16586-16594. <http://doi.org/10.48084/etasr.8298>
- [29] Shyamala, N., Mahaboob Basha, S. (2024). Brain tumor dissection and categorization using texture characteristics and deep learning techniques. In *2024 5th International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, pp. 1609-1613. <https://doi.org/10.1109/ICOSEC61587.2024.10722066>
- [30] Mitra, A., Sridar, K., Rathna, S., Chowdhury, R., Kumar, P. (2024). Optimizing brain tumor MRI classification using modified Vgg16 model. In *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)*, Hassan, India, pp. 1-7. <https://doi.org/10.1109/IACIS61494.2024.10721808>
- [31] Al-Asfoor, M., Abed, M.H., Maher, K. (2024). Brain tumor classification based on federated learning. In *2024 10th International Conference on Optimization and Applications (ICOA)*, Almeria, Spain, pp. 1-4. <https://doi.org/10.1109/ICOA62581.2024.10754056>
- [32] Prakash, R.M., Kumari, R.S.S., Valarmathi, K., Ramalakshmi, K. (2023). Classification of brain tumours from MR images with an enhanced deep learning approach using densely connected convolutional network. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(2): 266-277. <https://doi.org/10.1080/21681163.2022.2068161>
- [33] Ay, Ş., Ekinici, E., Garip, Z. (2024). A brain tumour classification on the magnetic resonance images using convolutional neural network based privacy-preserving federated learning. *International Journal of Imaging Systems and Technology*, 34(1): e23018. <https://doi.org/10.1002/ima.23018>