Received: 12 August 2023 | Revised: 6 December 2023 | Accepted: 17 December 2023

DOI: 10.1002/ima.23018

RESEARCH ARTICLE

WILEY

# A brain tumour classification on the magnetic resonance images using convolutional neural network based privacy-preserving federated learning

Şevket Ay | Ekin Ekinci | Zeynep Garip

Computer Engineering Department, Faculty of Technology, Sakarya University of Applied Sciences, Serdivan, Turkey

**Correspondence**
Ekin Ekinci, Computer Engineering Department, Faculty of Technology, Sakarya University of Applied Sciences, Serdivan 54187, Turkey.
Email: ekinekinci@subu.edu.tr

**Abstract**

The healthcare industry has found it challenging to build a powerful global classification model due to the scarcity and diversity of medical data. The leading cause is privacy, which restricts data sharing among healthcare providers. Federated learning (FL) can contribute to developing classification models by protecting data privacy. This study has tested various federated techniques in a peer-to-peer setting to classify brain Magnetic Resonance Images (MRI). The authors propose various aggregation strategies for FL, including Federated Averaging (FedAvg), Quantum FL with FedAVG (QFedAvg) and Fault Tolerant FedAvg (Ft-FedAvg) and FedAvg with differential privacy (Dp-FedAvg). In each approach, a custom Convolutional Neural Network (CNN) model is applied to compute locally run nodes with different parts of the same brain MRI dataset for 10, 20 and 30 training and test rounds. A central server and CNN-based three federated clients are included in the FL-based brain tumour classification model to exchange data and combine the model weights on the server, which are sent from local devices to the server. The superiority of the performance of the proposed model is demonstrated by comparing it with traditional methods on various performance metrics. Experimental results show that in brain MRI dataset classification using FL approaches, FedAVg showed the best performance with 85.55% and 84.60% success for 10 and 20 rounds, respectively, while Ft-FedAvg showed the best performance with 85.80% success for 30 rounds for test set. Statistical results obtained from FL approaches showed that FedAvg and Ft-FedAvg have superior performance with regard to accuracy and robustness in comparison with the others.

**KEYWORDS**

classification, deep learning, federated learning, privacy-preserving

## 1 | INTRODUCTION

The lack of data privacy has always been an inherent problem in deep learning, especially in healthcare.[1,2] Deep learning methods include collecting data, sharing it with another party, processing it and making it ready for the model.[3] Due to patient privacy, collecting data from different institutions and performing deep learning models with these data emerges as a challenge. Overcoming these challenges has become an important research area where researchers work and develop new methods, and eventually, federated learning (FL) has emerged.[4]

FL is a learning method that enables algorithms to learn collaboratively by providing data privacy.[4] In other words, it is the collaborative training of a machine learning model across a server with local data obtained from many independent devices without exchanging data. During each training round of FL, each local client updates the local model with their data and then uploads the local model to the server to update the global model by aggregating the local model.[5] The training process of FL is shown in Figure 1. FL has arisen as a significant research topic in machine learning in recent years. Furthermore, it has been used in a variety of applications, including medical image classification,[6] disease classification,[7] medical image segmentation[8] and so on. Additional research has been conducted on the effects of FL on a range of other cancers, including breast cancer,[9] prostate cancer,[10] lung cancer,[11] larynx cancer,[12] rectal cancer,[13,14] skin cancer[15] and others.

Compared with classical methods in Table 1, FL has been used in a limited number of studies for brain tumour classification. Islam et al. studied data from 22 brain tumour patients from the UK dataset. They created an average Convolutional Neural Network (CNN) model from DenseNet121, VGG19 and InceptionV3 models, integrated them into the federated learning structure, and achieved an accuracy of 91.05%.[28] In another study, optimize weight sharing was proposed by rating the weight percentage of each client and using average weights. In order to assess the performance of the proposed model, it was examined how well support vector machine (SVM) and VGG16 performed in the FL environment together with the average weights of proposed CNN and VGG16. The experimental findings were 98% accuracy on BT_large-1c and 97.14% on BT-large-2c for rating weight percentage.[29] Viet et al. applied FedAVG with VGG16, ResNet50, ConvNext and MaxViT to the Figshare dataset, and ConvNeXt obtained 98.69% accuracy on independently and identically distributed (IID) data.[30] Bhati and Samed proposed a framework for evaluating data based on the blockchain in FL. They created a smart contract that assesses every local update by building a local model. It is only combined with the global model when the local model meets a predetermined accuracy criterion. The brain tumour dataset's lowest accuracy was 89.15%, and the highest was 93.34%.[31]

In this study, we propose FL approaches to solve the sharing of medical data with third institutions or individuals and to enable institutions to develop joint modelling in collaboration. These approaches are FedAvg, QFedAvg and Ft-FedAvg strategies. Additionally, the differential privacy method has been applied to the FedAvg strategy to see the effects of data privacy. The advantages of the FedAvg model include data privacy, reduced communication costs which can be beneficial for devices with limited bandwidth, better model generalization by preventing overfitting and the ability to be scalable to accommodate more devices.[32] No tolerance to data imbalance, slowly converging when the data samples among the end devices are unbalanced and identically and independently dispersed (non-IID) and straggler problems are considered disadvantages of FedAvg.[33] Several advantages of QFedAvg over FedAvg are providing better convergence and accuracy by reducing the impact of

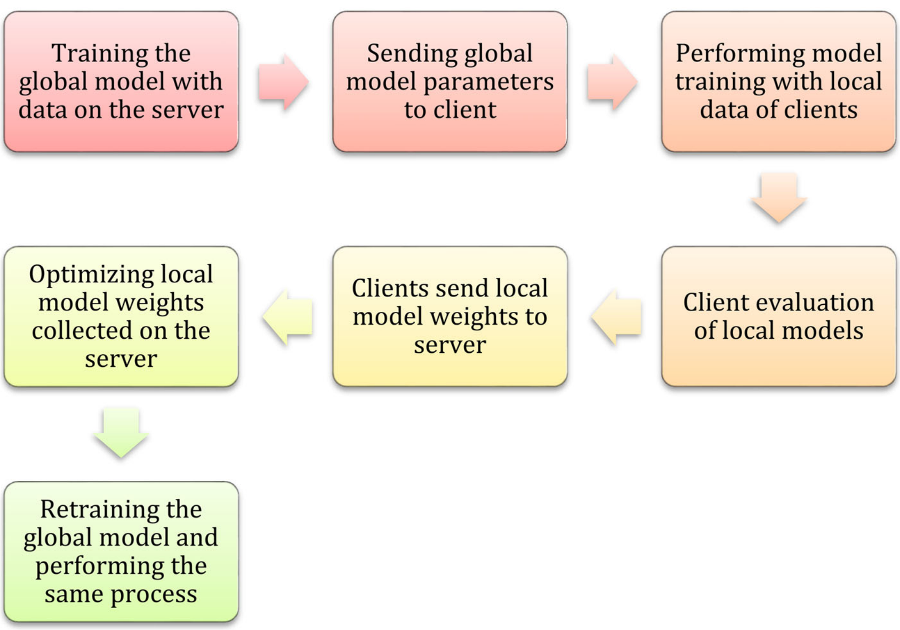



FIGURE 1 The training process of federated learning.

**TABLE 1** Comparison with existing literature those used Figshare, Sartaj and BR35H datasets.

| Ref. | Dataset | Algorithm | Results | Performance criteria (Accuracy %) |
|---|---|---|---|---|
| 16 | Figshare | PDCNN | A parallel deep convolutional neural network (PDCNN) was used for detecting and classifying brain cancers in MRI images. | 97.60 |
| 17 | Figshare Br35H | BMRI-Net + PFpM | They suggested PFpM, a novel parametric activation function for brain tumour classification. | Figshare: 99.57 Br35H: 99.00 |
| 18 | Figshare | Modified CNNBCN + ER | CNN with modified activation function or classification of brain tumours generated by the Erdos-Renyi (ER) algorithm, Watts-Strogatz (WS) algorithm and Barabasi-Albert (BA) algorithm was devised. | 95.49 |
| 19 | Figshare | ResNet18 + ShallowNet + SVM | The authors employed the fusion of deep and shallow data extracted from the extended tumour region. | 97.25 |
| 20 | Figshare | Border Collie Firefly Algorithm-based Generative Adversarial network + | For severity-level classification in brain tumours, a novel technique called BCFA-based GAN is developed with the Spark framework. | 97.515 |
| 21 | Figshare | Channel split dual attention + a backbone network | For efficient brain tumour classification that is end-to-end learnable, they proposed a channel split dual attention based network called CDANet. | 96.60 |
| 22 | Figshare | Bayesian Capsule Network | They demonstrated the possibility for creating the CapsNet architecture for the purpose of classifying the various types of brain tumours. | - |
| 23 | Br35H | PO+ PSO+ CNN | CNN with the political optimizer (PO) based particle swarm optimization (PSO) algorithm was used to detect brain tumours from MRI. | 97.09 |
| 24 | Br35H | DenseNet121 + ResNet101 + NasNet | They combined deep features from pre-trained deep convolutional neural networks with ML classifiers. | 98.67 |
| 25 | Br35H | 12-layered CNN | This study used CNN for classifying brain tumours. | 98.8 |
| 26 | Sartaj | Ensemble of deep CNN models | A deep CNN feature ensemble with two stages was presented for the exact and automatic classification of brain cancers. | 98.16 |
| 27 | Sartaj | Inceptionresnetv2 | They compared nine deep-learning models for classifying brain tumours through transfer learning. | 98.91 |

non-representative data, improving privacy by reducing the impact of clients with different data distributions and heterogeneous data and domain adaptation.[34] One of the limitations of QFedAvg is the difficulty in setting data privacy because it requires access to data distributions. In addition, computational cost, training time and resource consumption are among the limitations. The advantage of Ft-FedAvg over FedAvg is that it strengthens the robustness of the system by allowing clients to recover from communication errors and continue their participation in the training process.[4] Some limitations of Ft-FedAvg are that in cases where communication failures are frequent, the communication load of the system increases significantly. Additionally, if the data distribution changes during the training process or while the system is recovering from a failure, the assumption of stationary data distribution may be violated, which can lead to suboptimal performance or even failure of the Ft-FedAvg algorithm. Differential privacy has been used to eliminate the FedAvg privacy problem. Differential privacy is a rigid privacy-preserving method that forbids unauthorized individuals from obtaining personal data using publicly accessible data or service interfaces.[35]

To evaluate the effectiveness of these approaches, we conduct experiments on the brain tumour Magnetic Resonance Imaging (MRI) dataset combination of Figshare, Sartaj and BR35H.[36] FedAVG outperforms other FL methods. The results show that the FedAVG model trained and tested in 10 rounds, the last round is 41% more successful than the model in the first round with FedAVG, which is the most successful strategy. The contribution of the research can be summarized as follows:

- To improve the classification performance in terms of accuracy and F1-score, we suggest FedAvg, QFedAvg, Ft-FedAvg and Dp-FedAvg federated learning strategies be merged with the CNN model based brain tumor classification models.
- We evaluate the performance of brain tumour classification by FL approaches against CNN and CNN-based deep learning algorithms namely DenseNet and VGG19.

The novelty of this study can be summarized as a comparative study examining the effect of FL strategies for brain tumour classification. Comparative studies have been conducted with FedAvg, QFedAvg, Ft-FedAvg and Dp-FedAvg. There are no similar comparative analyses, according to our best knowledge.

# 2 | METHODOLOGY AND PROPOSED MODEL

## 2.1 | Convolutional Neural Network

CNN is a deep learning model commonly used in computer vision tasks such as image recognition, object detection and image classification.[37] CNN is specifically designed to analyze image data by taking advantage of spatial correlations and local patterns in the images.[38] The architecture of a CNN consists of several layers, including convolutional layers, pooling layers and fully connected layers.

This study integrates a custom model based on CNN with FL to classify brain MRIs. The three-layer CNN model for cross-device collaboration has been used in this model. The entire architecture and overall layout of a custom CNN is shown in Figure 2. Our learning model employs two-dimensional CNNs. As a result, the filters of the convolutional layer only have two dimensions. Additionally, for Max-pooling layers, the pool operations happen in two dimensions. The ReLU activation function is applied to each dense layer and convolutional layer. We also employ the Softmax activation and the Adam optimizer in the output layer.

## 2.2 | Federated learning approaches

FL enables predictions through collaborative models without taking medical data outside of the institution in which it is located. The deep learning process is carried out locally in each participating institution. The local model weights are then transmitted to the server in order to be used on the model on the device, which will be used as a server. The local model weights are then transmitted to the server in order to be used on the model of the device, which will be used as a server. There are certain challenges encountered during this whole decentralized

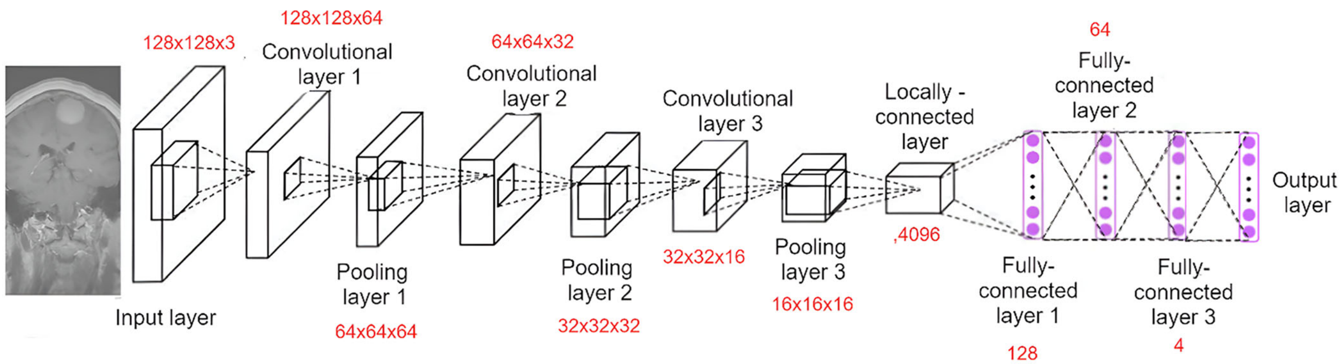



FIGURE 2 The architecture of the proposed model.

process. These challenges are statistical heterogeneity, data confidentiality and distributed optimization. The scope of this study lies in distributed optimization using various optimization methods and in data privacy utilizing differential privacy. Distributed optimization is training local models on local data while the central server is responsible for the global collection of updates. The updates collected on the central server must be optimized. As shown in Figure 3, there are FedAvg, QFedAvg and Ft- FedAvg strategies for optimization. Strategy preferences vary according to the distribution of data, the model's structure and the problem's kind. Data privacy is approached to observe the effect of differential privacy on the model. Differential privacy obfuscates each local model parameter by adding noise to the model weights before sending the model weights to the server for aggregation. This study implements Dp-FedAvg as one of the optimizers.

In this study, four different FL strategies are used FedAvg, QFedAvg, Ft-FedAvg and Dp-FedAvg. In the

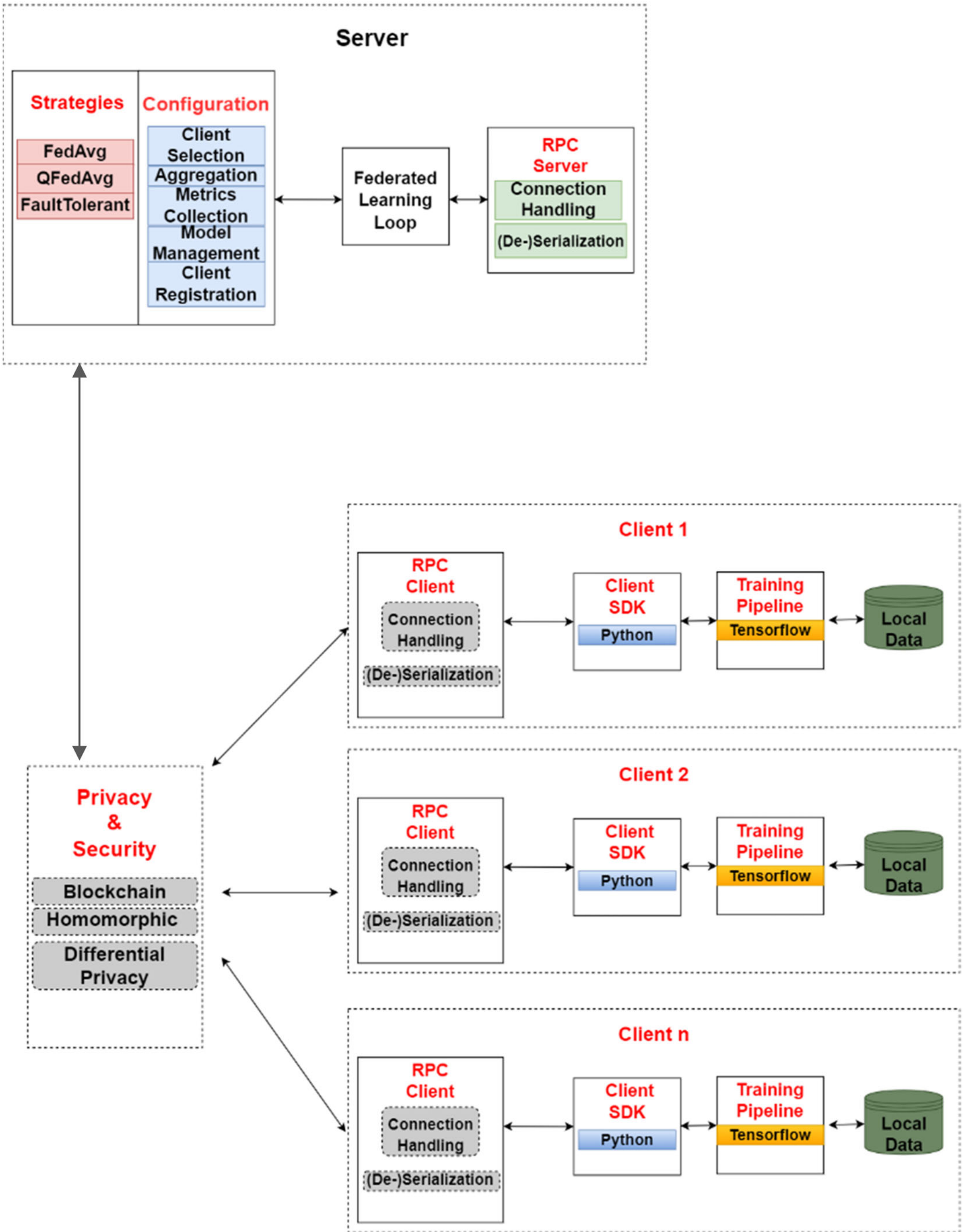



FIGURE 3 The general architecture of federated learning.

first strategy, FedAvg, the model weights from three different clients are optimized by averaging the model weights on the server. In the second strategy, the QFedAvg method, a q parameter is added to the standard FedAvg method. This parameter considers differences between clients in optimizing the collected model weights. The third strategy is the Ft-FedAvg strategy. This approach includes tolerating errors or interruptions in a client for any reason during the training process. In the last strategy, Dp-FedAvg, standard FedAvg model weights are used on the server, but unlike the first method, differential privacy and noise are added to a client's model weights before the model weight is sent to the server. The goal of the fourth strategy is to prevent model weights from being acquired in any possible attack, even though only the model weights are transmitted.

### 2.3 | Federated learning-based brain magnetic resonance imaging classification framework

The primary goal of this work is to improve the impact of custom CNN-based FL methods on classifying brain cancers while ensuring data confidentiality. The training methodology is to use FL approaches, namely FedAvg, QFedAvg, Ft-FedAvg and Dp-FedAvg for multi-institution datasets by collecting weighted model updates through a central aggregator server, as shown in Figure 4. Figure 4 illustrates the FL-based brain tumour classification model, which includes a central server and three federated clients to exchange information. FL approaches are used to combine the model weights on the server, which are sent from local devices to the server. Within the system, the global model on the server is first trained by using the data on the server. Then the server sends the model parameters to the active client devices. After receiving the model weights from the server, the client devices train their models with local data and perform their evaluations. As a next step, the client devices send the current model parameters to the server to combine the updated model parameters of other client devices. Then, the combined models, which are updated with various optimizers, are sent back from the server to all client devices. This process continues for a determined number of rounds. This study determines the number of rounds as 10.

## 3 | EXPERIMENTS AND RESULTS

We implement the CNN-based unified learning proposal with the help of the Python programming language and the Tensorflow and Flower libraries. In addition, other libraries, such as Numpy and Pandas, are also used in parts of the study, such as data pre-processing. We use a machine with 16 GB RAM, 12th Gen Intel(R) Core(TM) i5-12500H 2.50 GHz and NVIDIA GeForce RTX 3050 Ti GPU to train the CNN model within the Flower library. Three clients and one server are simulated on this machine.

### 3.1 | Dataset

This study uses the MRI dataset provided by the public Kaggle website to classify brain images.[36] This dataset was obtained by combining three datasets (Figshare, Sartaj, Br35H). The dataset contains 7023 greyscale brain MR images in four classes: 1600 Meningioma, 1600 Glioma, 1350 Pituitary tumour and 1850 no tumour, examples of which are shown in Figure 5. In addition, the distribution of classes is shown in Figure 6. The dataset is split before the separation between clients and servers into a training set that contains 5712 samples of the total dataset and a testing set that contains 1311 samples of the total dataset. The Kaggle brain tumour MRI dataset has dimensions of $512 \times 512$, and samples are reduced to $128 \times 128$ in a pre-processing step. The reason for resizing is to reduce the parameters of the images and, thus, the costs of model studies.

### 3.2 | Data sharing

This study divides the training and test datasets into four parts for collaborative model development while maintaining data privacy. These are three simulated devices and one simulated server as clients. The samples are shared as equally as possible between the clients and the server in FL. The distribution of training and test data by classes for each simulated device is listed in Table 2.

### 3.3 | Evaluation metrics

To evaluate the performance of the FL model, F1-score, precision and recall are used as performance metrics. After each round, the optimizers' performance and the models' progress are compared according to the F1-score metric. True Positive (TP) indicates that the model classifies tumour samples as the tumour, while True Negative (TN) indicates that the model classifies non-tumour samples as non-tumour. False Positive (FP) shows that the model incorrectly classifies non-tumour samples as tumour, and False Negative (FN) shows that the model incorrectly classifies tumour samples as non-tumour. All the metrics used for evaluation are given in Equations 1–3 regarding TP, FP, FN and TN.

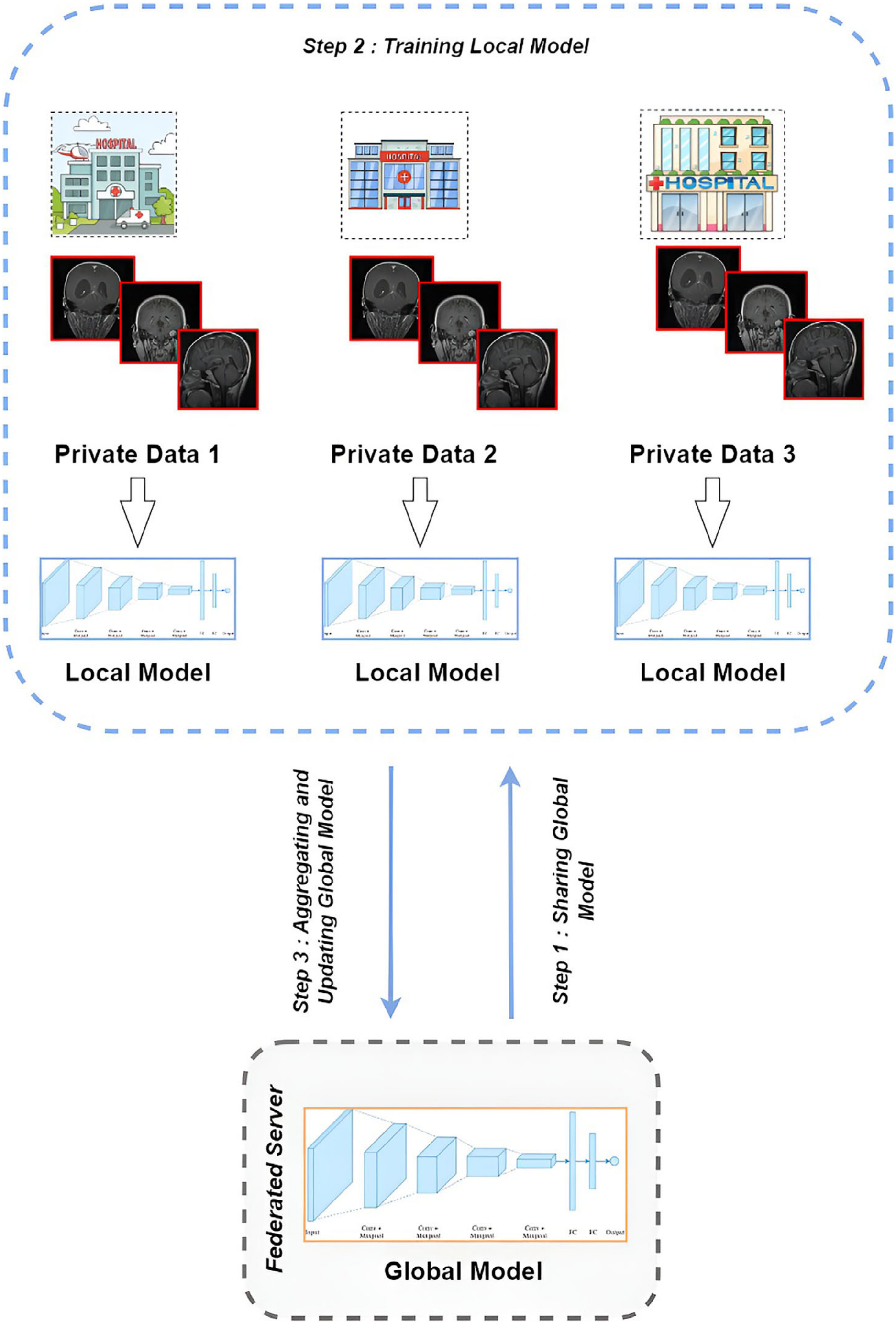



**FIGURE 4** The scenario of the proposed model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

In addition to evaluation metrics, a confusion matrix is used to better see the model's performance. The confusion matrix shows correct and incorrect predictions.

## 3.4 | Experimental results of FL methods

In this study, a hybrid approach has emerged by combining the FL and CNN model to accurately determine whether there is a tumour in the brain MRI and to which class it belongs. With this proposed approach, it is focused on how data privacy and data sharing are realized.

### 3.4.1 | Implementation details

Within the scope of the study, the three-layer CNN model is used for cross-device collaboration. Both server

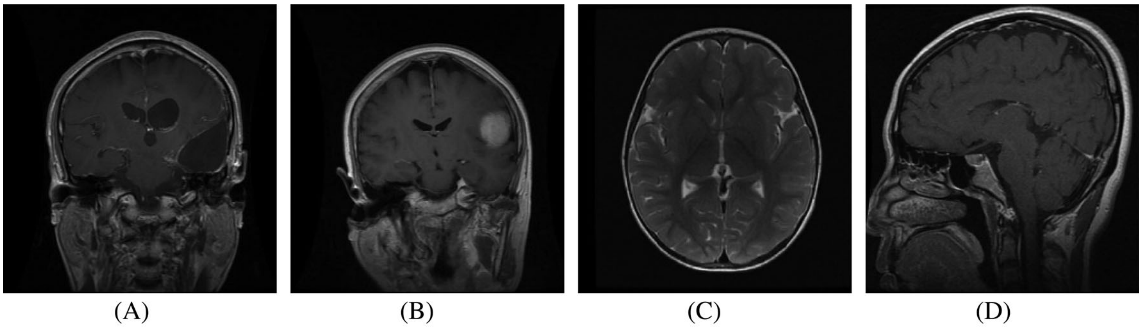



**FIGURE 5** Data samples from each class (A) Glioma, (B) Meningioma, (C) No Tumour, (D) Pituitary.

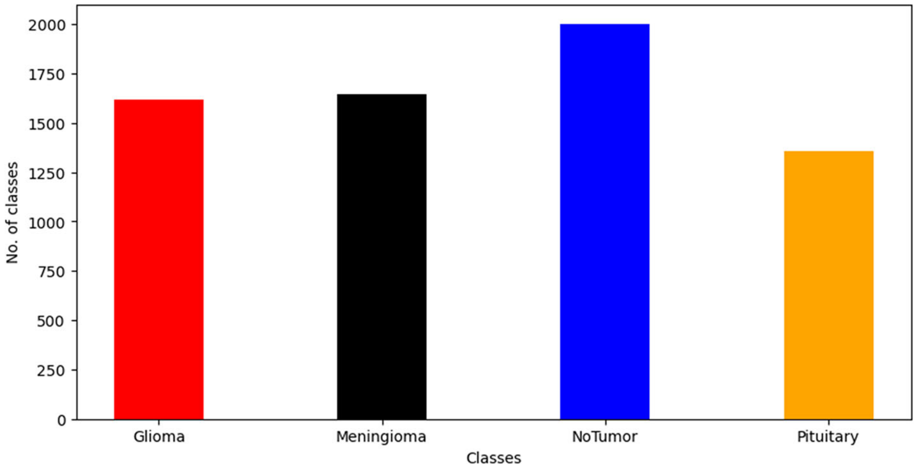



**FIGURE 6** Distribution of the classes.

**TABLE 2** Data distribution between simulated devices.

| Devices | Glioma | Meningioma | No tumour | Pituitary |
|---|---|---|---|---|
| Client 1—Train | 330 | 335 | 398 | 365 |
| Client 1—Test | 75 | 75 | 104 | 75 |
| Client 2—Train | 330 | 336 | 398 | 364 |
| Client 2—Test | 75 | 75 | 100 | 75 |
| Client 3—Train | 330 | 334 | 398 | 364 |
| Client 3—Test | 75 | 75 | 101 | 75 |
| Server—Train | 331 | 334 | 401 | 364 |
| Server—Test | 75 | 81 | 100 | 75 |

and client devices have the same model architecture. Using the CNN architecture, a local model for FL is constructed in this methodology. The model consists of three convolutional layers with 64, 32 and 16 filters, respectively, followed by a maximum pooling layer, each with a pool size (2.2). Drop layers of 0.2 are added after each maximum pooling layer to avoid overfitting. The output of the third maximum pooling layer is flattened and fed into two dense layers with 128 and 64 neurons, respectively, which have ReLU activation functions. A learning rate of 0.0001 is chosen for the CNN model. The loss function used is the categorical cross-entropy. The model is trained with five epochs and eight batch sizes.

### 3.4.2 | Federated learning experiments

This study compares recent FL approaches, FedAvg, QFedAvg, Ft-FedAvg and Dp-FedAvg. The aim is to compare the goodness of the models obtained by applying the FL strategies, which properly work over the complete set of MRI datasets. Experiments have been conducted with FedAvg, QFedAvg, Ft-FedAvg and Dp-FedAvg models

**TABLE 3** Model F1-score on the magnetic resonance imaging test dataset for 10 rounds.

| | F1-score | | | |
|---|---|---|---|---|
| **Round** | **FedAvg** | **QFedAvg** | **Ft-FedAvg** | **Dp-FedAvg** |
| 0 | 0.607 | 0.607 | 0.565 | 0.622 |
| 1 | 0.686 | 0.653 | 0.692 | 0.598 |
| 2 | 0.743 | 0.622 | 0.695 | 0.647 |
| 3 | 0.749 | 0.650 | 0.728 | 0.644 |
| 4 | 0.782 | 0.631 | 0.758 | 0.647 |
| 5 | 0.764 | 0.713 | 0.755 | 0.707 |
| 6 | 0.843 | 0.695 | 0.779 | 0.698 |
| 7 | 0.798 | 0.707 | 0.837 | 0.725 |
| 8 | 0.825 | 0.710 | 0.816 | 0.722 |
| 9 | 0.852 | 0.728 | 0.861 | 0.710 |
| 10 | 0.855 | 0.737 | 0.831 | 0.743 |

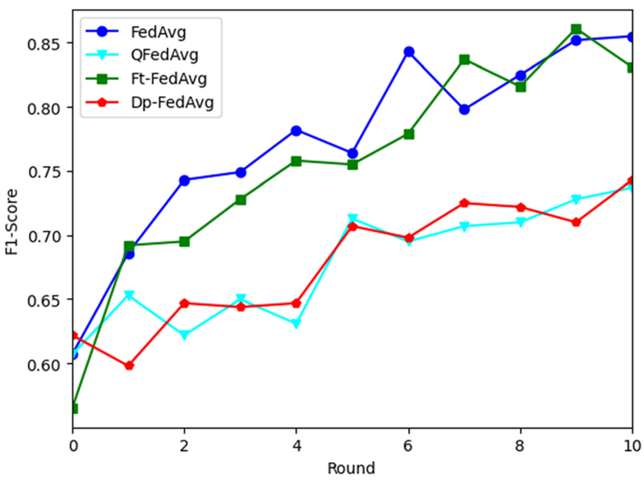



**FIGURE 7** F1-score of different federated learning approaches on the global model for test set.

trained and tested in 10 rounds. The experimental results are shown in Table 3. Figures 7 and 8 depict the F1-score and loss graph, and Figure 9 shows the confusion matrix for four-class in the brain MRI dataset.

The approaches are trained five epochs in each computation round. Ten rounds of communication are performed under the dataset experiments, as shown in Figure 8. In Table 3, the development of F1-score values according to the number of rounds of the global model on the server is shown.

When Table 3 is examined, global models starting from a low F1-score have improved over time for each strategy. Round 0 shows the evaluation that results from the global model training without any client-specific model weights. When the results of each round are examined, there is an overall improvement in the global models according to the local model weights from the three clients. However, there are cases where there is no positive change in the evolution of the global model between rounds. This is because any client needed to learn better in that round. As a result, when the first and last rounds are compared in the global model, it is observed that the F1-score performance has increased following the score. After 10 computation rounds, FedAvg appears to be the most successful strategy. This is an explainable result because the distribution of classes between clients is approximately even. Therefore, FedAvg has been successful because it is an averaging strategy. Practically, using Ft-FedAvg can be more valuable than FedAvg, so the preference between two strategies that close performances can be based on the problem and hardware features. When reaching the final round, FedAvg's F1-score is 85.50%, which improves 11.8%, 2.4% and 11.2% compared to the other approaches. Figure 8 shows the loss change curves. In the last training round, the model loss value trained by the FedAvg method is the largest, and the model loss of the other approaches on FedAvg differs.

Figure 9 shows the complexity matrices in the first round and the complexity matrices in the last round for the global model for each FL approach for the test set. When the first and last rounds for the server model were compared on the test set, it was seen that the local model, which made many incorrect predictions at the end of the first training round, reduced the incorrect predictions at the end of the last training round. When examining confusion matrices, FedAvg is the most successful strategy, but FaultTolerant may be preferred over FedAvg. After examining all complexity matrices, a standard global model was successfully developed by collecting model information on the server without sharing data between devices, and client devices developed their local models in this way while preserving data confidentiality.

### 3.4.3 | Comparison with state-of-the arts

We evaluate the performance of brain tumour classification by FL approaches against CNN and CNN-based deep learning algorithms namely DenseNet and VGG19 on the above metrics. Table 4 shows the performances of FL approaches. The FedAvg presents a high F1-score, 85.5%, compared with the others, and the strategy approach with the shortest training period was Q-FedAvg. This is expected because Q-FedAvg applies compression to the model weights, so the communication time is shorter than other strategies as the models are not used at full capacity. QFedAvg showed the lowest performance with

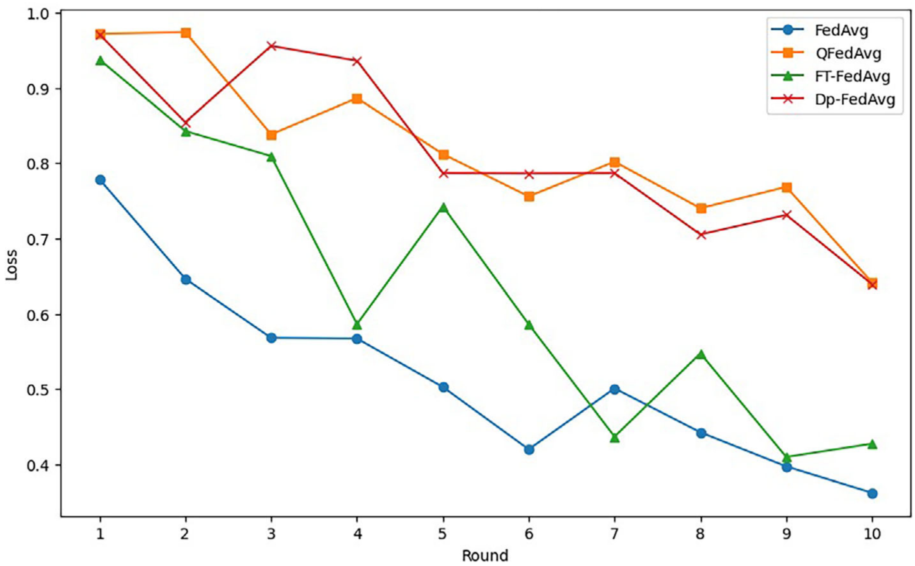



**FIGURE 8** Loss value of different federated learning approaches on the global model.

an F1-score of 73.7%. This is because QFedAvg is useful for heterogeneously distributed data rather than evenly distributed data. QFedAvg addresses this issue by using quantization to compress updates before they are sent to the server. For this reason, there may be relative losses in model performance as a result of quantization. FedAvg, on the other hand, gives its best performance in homogeneous data distributions. Therefore, in our experiment, it is expected that FedAvg will perform better than QFedAvg. On the other hand, federated learning methods' performance is typically not able to outperform that of non-federated learning methods. This demonstrates that FedAvg can improve performance and successfully address the privacy-preserving classification problem.

### 3.4.4 | Ablation analysis of our method

In order to evaluate the accuracy and effectiveness of the proposed FL models, an ablation study was carried out on different tasks.

*Effect of training and test rounds of FL*

We also examine the implications of varying the number of rounds in federated learning for our techniques' generalization performance. The test results for 10 rounds are compared to those for 20 and 30 rounds in Table 5. As can be seen from Table 5, the generalization performance of FL strategies will decline with the increase of the number except Ft-FedAvg for 30 rounds. Additionally, F1-scores of FL approaches on the global model for test set for 20 and 30 rounds are given in Figures 10 and 11, respectively.

*Analysis of p-values of statistical test*

Significance test is an important metric for statistically analyzing the performance of observations. In this section, we tested statistical significance for federated learning models for different rounds. The analysis of *p*-values of the Wilcoxon signed-rank test with a significant difference of 0.05 and obtained ranking values of Friedman's test for the designed strategies are shown in Table 6. When Table 6 is examined, it is seen that the *p*-values are less than 0.05. Therefore, it turns out that the results of these FL strategies are significant.

Examining the results of both tests, we can see that using the FedAvg strategy on brain tumour data is significantly better than the others for training 10 and 20 rounds. In training 30 rounds, the Ft-FedAvg strategy is more successful than other strategies.

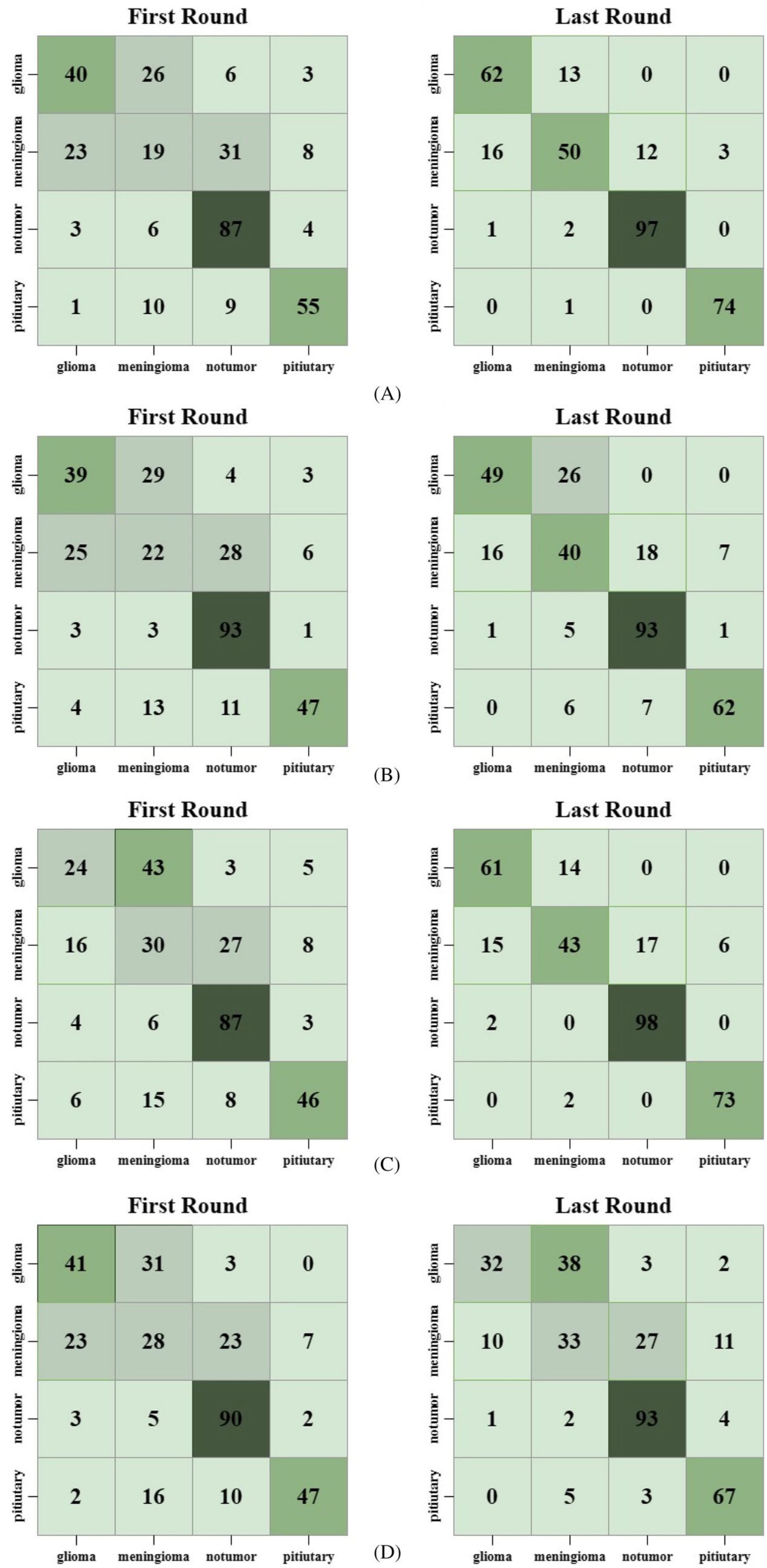
## 4 | DISCUSSION

### 4.1 | Privacy

To address the privacy issue, FL has emerged as a promising alternative work, enabling the development of machine learning models using data from multiple sources without the need to share data. Thanks to FL, meta information MRI images containing patient information such as the patient's name and age does not leave the organizations.

This paper offers a privacy-preserving paradigm for a collaborative MRI-based brain tumour classification framework across devices. The framework we proposed is designed for the multi-institutions training sample MRI-

FIGURE 9 Confusion matrices of global model of (A) FedAvg, (B) QFedAvg, (C) Ft-FedAvg, (D) Dp-FedAvg for test set.




based brain tumour classification task, which ensures data security by requiring data healthcare providers to share only the trained model parameters rather than their datasets. In addition, since the amount of local data is equal on each simulated device, FedAvg, QFedAvg and Ft-FedAvg strategies were used in the study. To see the

**TABLE 4** Performance of federated learning approaches on test set against others.

| Algorithms | F1 score (%) | Time (sn) |
|---|---|---|
| FedAvg | 0.855 | 4490.27 |
| QFedAvg | 0.737 | 3970.61 |
| Ft-FedAvg | 0.831 | 4982.88 |
| Dp-FedAvg | 0.743 | 5011.36 |
| CNN | 0.835 | - |
| DenseNet | 0.853 | - |
| VGG19 | 0.850 | - |

**TABLE 5** Performance of FedAvg, QFedAvg, Ft-FedAvg and Dp-FedAvg with Convolutional Neural Network as backbone under different numbers of rounds.

| | F1-score | | | |
|---|---|---|---|---|
| Round | FedAvg | QFedAvg | Ft-FedAvg | Dp-FedAvg |
| 10 | **0.855** | 0.737 | 0.831 | 0.743 |
| 20 | **0.846** | 0.719 | 0.819 | 0.718 |
| 30 | 0.852 | 0.736 | **0.858** | 0.728 |

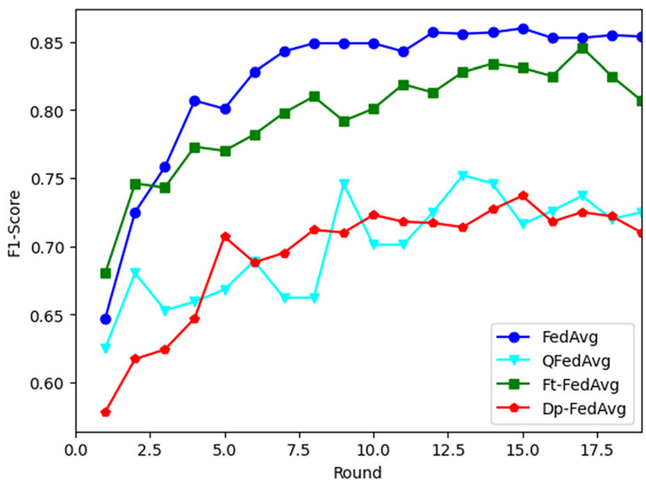



**FIGURE 10** F1-score of different federated learning approaches on the global model for test set for 20 rounds.

effects of data privacy, the differential privacy method was applied to the FedAvg strategy (Dp-FedAvg), which is the most successful strategy, to observe the difference between the results.

## 4.2 | Implications

Accurate classification of brain tumours is crucial for successful treatment planning, and the use of machine learning models has shown great potential in the healthcare field. However, privacy concerns around sharing medical

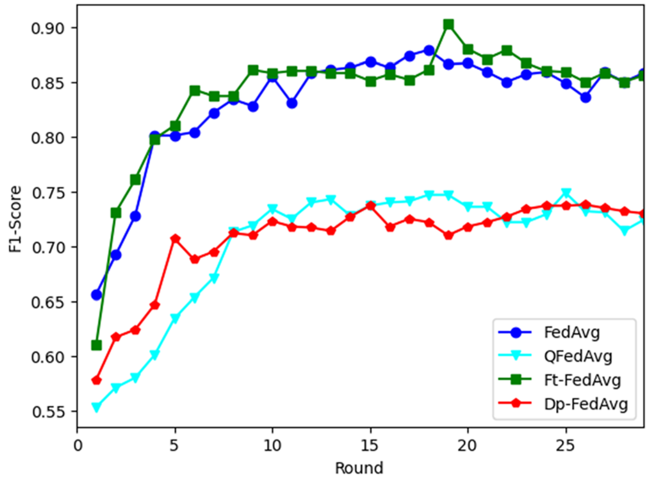



**FIGURE 11** F1-score of different federated learning approaches on the global model for test set for 30 rounds.

data have limited the application of traditional machine learning methods. There are many legal requirements for processing medical data in ML applications. Confidentiality is essential to comply with requirements and avoid legal consequences. In summary, medical data privacy is very important for ML. Healthcare providers and ML developers must take all necessary precautions to ensure the confidentiality of patient data to protect patient privacy, maintain ethical standards, prevent bias and discrimination and comply with legal requirements. Federated learning offers a promising solution to this problem by allowing the development of machine learning models that use data from multiple sources while preserving data privacy. Federated learning enables predictions to be made through collaborative models without taking medical data outside the institution where it is stored.

This study sought answers to the questions of whether a model can be developed by protecting the privacy of medical data with FL, whether models trained with relatively little data can perform like models trained with high data thanks to cooperation, how different strategies affect the models and whether adding differential privacy affects model performance.

As a result, the importance of research in this field is of great importance in terms of distributed training of learning models, as well as data security and privacy issues, in today's world where an increasing amount of data is available.

## 4.3 | Limitations and future works

This study can be improved. For example, in this study, data sharing between clients and server is distributed equally. In the next step, data sharing between clients

**TABLE 6** Analysis of the designed method for the MRI brain tumour classification model.

| | | Wilcoxon signed rank test | |
|---|---|---|---|
| Round | Algorithms | *p*-value | Ranking |
| 10 | FedAvg | - | **1.3** |
| | QFedAvg | 0.0020 | 3.6 |
| | Ft-FedAvg | 0.0635 | 1.7 |
| | Dp-FedAvg | 0.0020 | 3.4 |
| 20 | FedAvg | - | **1.10** |
| | QFedAvg | 8.8324e-05 | 3.35 |
| | Ft-FedAvg | 7.7877e-04 | 1.90 |
| | Dp-FedAvg | 8.8324e-05 | 3.65 |
| 30 | FedAvg | - | 1.65 |
| | QFedAvg | 1.7235e-06 | 3.48 |
| | Ft-FedAvg | 0.0490 | **1.35** |
| | Dp-FedAvg | 1.7279e-06 | 3.52 |

and servers and distribution between classes can be done in an unbalanced way. Thus, we can focus on solving the problem of statistical heterogeneity. In this case, the strategy approaches used will differ. Optimizers such as FedProx, created for statistical heterogeneity, may be preferred to solve this problem. In this study, there is one server and three clients. Therefore, central communication architecture was used. Real life problems do not always require a server or a server may not be available. In this case, decentralized communication architecture can be preferred. With decentralized communication architecture, all clients will develop a common model by sharing model weights among themselves, without a server. The difficulties of developing a common model with decentralized communication will also be discussed. For example, difficulties such as resolving the communication problem that will occur if a client is slower than other clients due to device heterogeneity, and finding solutions to ensure that the low model performance that may occur due to a client's local data does not affect other clients and the development of the common model should be addressed.

In the future, we will focus on the distribution of data according to the clients' statistical heterogeneity or develop models based on different strategies. Moreover, it is possible to carry out studies on vertical learning by simulating a non-IID situation.

## 5 | CONCLUSION

Data privacy is a significant concern in the analysis of medical data. The FL method used in this study allows more comprehensive use of data while maintaining data privacy. FL is a distributed learning method, eliminating the need to collect and store data in a central location. In this study, the FL approach has been proposed to perform the analysis of medical imaging data accessible to a broader use without the need for central data storage infrastructure. In this approach, we have designed and integrated a custom CNN model in an FL setting to classify the brain MRI dataset. A scenario is designed by dividing the brain MRI dataset into three local global models with data-sharing. Local CNN models were initiated using MRI of local devices from three different organizations simulated. The global CNN model on the server is fed with the weights of the local CNN models. Local model weights have been updated with the local model weight optimized on the server. FedAvg, QFedAvg, Ft-FedAvg and Dp-FedAvg are preferred as FL approaches. Experimental results show that in brain MRI dataset classification using FL approaches, FedAVg showed the best performance for 10 and 20 rounds, while Ft-FedAVg showed the best performance for 30 rounds. Additionally, on the basis of the statistical tests on brain tumours, we can observe that the FedAvg strategy is clearly superior to the others. A standard client-to-client model has been developed by keeping the local data on the clients by using all the suggested strategies. Compared to classical deep learning models, where confidentiality is not protected, and data is collected and trained on a server, confidentiality is preserved, and clients are provided with high model performance with less local data. The results show that thanks to FL, more data in an institution is needed to prevent model development. Additionally, the models have been developed with fewer epochs, thus reducing the time cost considerably. As FL continues to evolve and gain popularity, several research directions can be explored to improve its efficiency, scalability and security.

## ACKNOWLEDGEMENT

This study was not supported by any individual or institution.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Kaggle at https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset. These data were derived from the following resources available in the public domain: Kaggle, https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset.

## ORCID

*Ekin Ekinci* https://orcid.org/0000-0003-0658-592X

## REFERENCES

1. Wang B, Li H, Guo Y, Wang J. PPFLHE: a privacy-preserving federated learning scheme with homomorphic encryption for healthcare data. *Appl Soft Comput*. 2023;146:110677. doi:10.1016/j.asoc.2023.110677
2. Sun Y, Liu J, Yu K, Alazab M, Lin K. PMRSS: privacy-preserving medical record searching scheme for intelligent diagnosis in IoT healthcare. *IEEE Trans Industr Inform*. 2021;18(3):1981-1990. doi:10.1109/TII.2021.3070544
3. Nguyen DH, Pham Q, Pathirana PN, et al. Federated learning for smart healthcare: a survey. *ACM Comput Surv*. 2023;55(3):1-37. doi:10.1145/3501296
4. Zhang C, Xie Y, Bai H, Yu B, Li W, Gao Y. A survey on federated learning. *Knowl-Based Syst*. 2021;216:106775. doi:10.1016/j.knosys.2021.106775
5. Singh S, Rathore S, Alfarraj O, Tolba A, Yoon B. A framework for privacy-preservation of IoT healthcare data using federated learning and blockchain technology. *Futur Gener Comput Syst*. 2022;129:380-388. doi:10.1016/j.future.2021.11.028
6. Liu Z, Wu F, Wang Y, Yang M, Pan X. FedCL: federated contrastive learning for multi-center medical image classification. *Pattern Recogn*. 2023;109739:109739. doi:10.1016/j.patcog.2023.109739
7. Noman AA, Rahaman M, Pranto TH, Rahman RM. Blockchain for medical collaboration: a federated learning-based approach for multi-class respiratory disease classification. *Healthcare Anal*. 2023;3:100135. doi:10.1016/j.health.2023.100135
8. Liu J, Liang X, Yang R, et al. Federated learning-based vertebral body segmentation. *Eng Appl Artif Intell*. 2022;116:105451. doi:10.1016/j.engappai.2022.105451
9. Kumbhare S, Kathole AB, Shinde S. Federated learning aided breast cancer detection with intelligent heuristic-based deep learning framework. *Biomed Signal Process Control*. 2023;86:105080. doi:10.1016/j.bspc.2023.105080
10. Rajagopal A, Redekop E, Kemisetti A, et al. Federated learning with research prototypes: application to multi-center MRI-based detection of prostate cancer with diverse histopathology. *Acad Radiol*. 2023;30(4):644-657. doi:10.48550/arXiv.2206.05617
11. Heidari A, Javaheri D, Toumaj S, Navimipour NJ, Rezaei M, Unal M. A new lung cancer detection method based on the chest CT images using federated learning and blockchain systems. *Artif Intell Med*. 2023;141:102572. doi:10.1016/j.artmed.2023.102572
12. Hansen CR, Price G, Field M, et al. Larynx cancer survival model developed through open-source federated learning. *Radiother Oncol*. 2022;176:179-186. doi:10.1016/j.radonc.2022.09.023
13. Bermejo I, Mateus P, Savino M, et al. MO-0059 prediction of pathological response to chemo-radiotherapy in rectal cancer using federated learning. *Radiother Oncol*. 2023;182:S27.
14. Nergiz M. Federated learning-based colorectal cancer classification by convolutional neural networks and general visual representation learning. *Int J Imaging Syst Technol*. 2023;33(3):951-964. doi:10.1002/ima.22875
15. Yaqoob MM, Alsulami M, Khan MA, et al. Symmetry in privacy-based healthcare: a review of skin cancer detection and classification using federated learning. *Symmetry*. 2023;15(7):1369. doi:10.3390/sym15071369
16. Rahman T, Islam MS. MRI brain tumor detection and classification using parallel deep convolutional neural networks. *Meas: Sens*. 2023;26:100694. doi:10.1016/j.measen.2023.100694
17. Mondal A, Shrivastava VK. A novel parametric flatten-p mish activation function based deep CNN model for brain tumor classification. *Comput Biol Med*. 2022;150:106183. doi:10.1016/j.compbiomed.2022.106183
18. Huang Z, Du X, Chen L, et al. Convolutional neural network based on complex networks for brain tumor image classification with a modified activation function. *IEEE Access*. 2020;8:89281-89290. doi:10.1109/access.2020.2993618
19. Abirami S, Venkatesan GP. Deep learning and spark architecture based intelligent brain tumor MRI image severity classification. *Biomed Signal Process Control*. 2022;76:103644. doi:10.1016/j.bspc.2022.103644
20. Dutta TK, Nayak DR. CDANet: channel Split dual attention based CNN for brain tumor classification in Mr images. *In 2022 IEEE International Conference on Image Processing (ICIP)*. IEEE; 2022:4208-4212. doi:10.1109/ICIP46576.2022.9897799
21. Afshar P, Mohammadi A, Plataniotis KN. Bayescap: a bayesian approach to brain tumor classification using capsule networks. *IEEE Signal Process Lett*. 2020;27:2024-2028. doi:10.1109/LSP.2020.3034858
22. Öksüz C, Urhan O, Güllü MK. Brain tumor classification using the fused features extracted from expanded tumor region. *Biomed Signal Process Control*. 2022;72:103356. doi:10.1016/j.bspc.2021.103356
23. Bashkandi AH, Sadoughi K, Aflaki F, Alkhazaleh HA, Mohammadi H, Jimenez G. Combination of political optimizer, particle swarm optimizer, and convolutional neural network for brain tumor detection. *Biomed Signal Process Control*. 2023;81:104434. doi:10.1016/j.bspc.2022.104434
24. Kang J, Ullah Z, Gwak J. Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers. *Sensors*. 2021;21(6):2222. doi:10.3390/s21062222
25. Naseer A, Yasir T, Azhar A, Shakeel T, Zafar K. Computer-aided brain tumor diagnosis: performance evaluation of deep learner CNN using augmented brain MRI. *Int J Biomed Imaging*. 2021;2021:1-11. doi:10.1155/2021/5513500

26. Aurna NF, Yousuf MA, Taher KA, Azad AKM, Moni MA. A classification of MRI brain tumor based on two stage feature level ensemble of deep CNN models. *Comput Biol Med*. 2022;146:105539. doi:10.1016/j.compbiomed.2022.105539
27. Kundroo M, Kim T. Federated learning with hyper-parameter optimization. *J King Saud Univ Comput Inf Sci*. 2023;35(9):101740. doi:10.1016/j.jksuci.2023.101740
28. Islam M, Reza MT, Kaosar M, Parvez MZ. Effectiveness of federated learning and CNN ensemble architectures for identifying brain tumors using MRI images. *Neural Process Lett*. 2022;55(4):1-31. doi:10.1007/s11063-022-11014-1
29. Mahlool DH, Abed MH. Optimize weight sharing for aggregation model in federated learning environment of brain tumor classification. *J Al-Qadisiyah Comput Sci Math*. 2022;14(3):76. doi:10.29304/jqcm.2022.14.3.989
30. Viet KLD, Ha K, Quoc TN, Hoang VT. MRI brain tumor classification based on federated deep learning. MRI Brain Tumor Classification Based on Federated Deep Learning 2023. doi:10.1109/zinc58345.2023.10174015
31. Bhatia L, Samet S. A decentralized data evaluation framework in federated learning. *Blockchain: Res Appl*. 2023;100152:100152. doi:10.1016/j.bcra.2023.100152
32. Liu J, Wang JH, Rong C, Xu Y, Yu T, Wang J. Fedpa: an adaptively partial model aggregation strategy in federated learning. *Comput Netw*. 2021;199:108468. doi:10.1016/j.comnet.2021.108468
33. Ren C, Yu H, Yan R, et al. Towards quantum federated learning. arXiv Preprint arXiv:2306.09912. 2023.
34. Liu C, Tian Y, Tang J, Dang S, Chen G. A novel local differential privacy federated learning under multi-privacy regimes. *Expert Syst Appl*. 2023;227:120266. doi:10.1016/j.eswa.2023.120266
35. Ullah N, Khan JA, Khan MS, et al. An effective approach to detect and identify brain tumors using transfer learning. *Appl Sci*. 2022;12(11):5645. doi:10.3390/app12115645
36. Nickparvar M. Brain tumor MRI dataset. 2021. Available online. doi:10.34740/KAGGLE/DSV/2645886 Accessed on April 18, 2023.
37. Biratu ES, Schwenker F, Ayano YM, Debelee TG. A survey of brain tumor segmentation and classification algorithms. *J Imaging*. 2021;7(9):179. doi:10.3390/jimaging7090179
38. Pranitha K, Vurukonda N, Nayak RK. A comprehensive survey on MRI images classification for brain tumor identification using deep learning techniques. *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*. IEEE; 2022:1206-1212.


**How to cite this article:** Ay Ş, Ekinci E, Garip Z. A brain tumour classification on the magnetic resonance images using convolutional neural network based privacy-preserving federated learning. *Int J Imaging Syst Technol*. 2024;34(1):e23018. doi:10.1002/ima.23018