Federated Learning-Based EfficientNet in Brain Tumor Classification

Baicheng Chen[a]
School of Data Science, The Chinese University of Hong Kong, Shenzhen, China

Keywords: Machine Learning, Federated Learning, Brain Tumor Classification, FedAvg, EfficientNet.

Abstract: The trend of implementing Machine Learning algorithms in the medical diagnosis field is necessary and meaningful. However, data privacy has become a big problem in applications. This paper uses the Federated Learning (FL) architecture to deal with the privacy problem and finds ways to improve the model's performance. The study combines the FedAvg FL Algorithm and the CNN model EfficientNet to train the model on the Brain Tumor Classification (MRI) dataset. Before implementing the algorithm, the study did some preprocessing on the data. Then, the study used EfficientNet to further process and recognize the images and FedAvg to weighted average the models trained by clients. Moreover, the study explored the optimizers and loss functions, choosing the AdamW and Cross-entropy loss which fitted this task better. Finally, the study went deep into parameter tuning work, drawing some curves and tables to visualize the results. After parameter tuning, this paper found a nice testing accuracy of 81.218% and a high training accuracy of almost 99% averaged by all the clients. Also, the paper discusses the conditions for implementing different CNN models and analyses their pros and cons in the medical diagnosis field, providing some ideas for the combination of network models and algorithms.

1 INTRODUCTION

Image Classification is a basic task in the vision recognition field. It trains a model using images with tags, and labels other pre-unknown images. Nowadays, image classification technology has been applied in numerous fields, such as medical images, security and automatic driving (Li, 2024; Liu, 2023; Qiu, 2022; Qiu, 2024). Thereinto, the medical images field has received much attention recently. In the past, it took doctors and researchers a long time to label medical images and diagnose patient conditions. However, with the development of medical image classification technology, doctors can diagnose disease characteristics efficiently and correctly, researchers can discover new disease characteristics and pathological mechanisms. As a result, the treatment and patient survival rates have been greatly improved.

Currently, the industry still mainly uses Centralized Machine Learning (ML) architecture to train medical image classification models. In centralized learning, data are sent to the cloud, where the ML model is built. The model is used by a user through an Application Programming Interface (API) by sending a request to access one of the available services (AbdulRahman et al., 2020). However, patients' image data are very sensitive and scientists have a responsibility to protect the privacy of these data during training. In Centralized ML, the sensitive data are sent to the server, leading to the risk of privacy leakage. Another ML architecture, Distributed On-Site Learning, is also not proper for this important task because in distributed on-site learning, the server sends the model to the users, and the users train models locally. There is no communication among the trained models.

To solve the problem, Federated Learning (FL) can be considered as an effective solution. Federated learning is a machine learning setting where multiple entities (clients) collaborate in solving a machine learning problem, under the coordination of a central server or service provider. Each client's raw data is stored locally and not exchanged or transferred; instead, focused updates intended for immediate aggregation are used to achieve the learning objective

[a] https://orcid.org/0009-0005-7657-9877

Chen, B.
Federated Learning-Based EfficientNet in Brain Tumor Classification.
DOI: 10.5220/0012950900004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 458-462
ISBN: 978-989-758-713-9

(McMahan et al., 2017; Kairouz et al., 2021). Due to the local training and model aggregating, the FL architecture can protect data privacy, fitting the medical aim better. With the proposal of FL, many algorithms based on FL architecture have emerged, like FedAvg (McMahan et al., 2017), FedProx (Li et al., 2020), SCAFFOLD (Karimireddy et al., 2020), FedNova (Wang et al., 2020) etc. However, how to implement FL to solve the privacy problem in brain tumor diagnosis received little attention. This article tries to use the FL architecture to train the medical image dataset “Brain Tumor Classification (MRI) (Bhuvaji et al., 2020)”, choosing a proper Algorithm and exploring the best values of the parameters that lead to a nice test accuracy.

The remainder of this paper is organized as follows. In the Method section, the paper chose the combination of preprocessing methods, FL algorithms, CNN models, optimizers and loss functions, illustrating the implementation details. Then, in the Results and Discussions section, this paper shows the results of the experiments and deeply discusses the impact of each parameter and the performance of different combinations to find the best training strategy. Finally, in the Conclusion part, the paper summarizes the findings of the study and the further problems that need solving.

2 METHOD

2.1 Dataset Preparation

The MRI dataset used in this study contains 3, 260 T1-weighted contrast-enhanced images that have been processed and enhanced (Bhuvaji et al., 2020). The dataset includes two folders, Training and Testing, and each folder contains four subfolders, which store images of glioma tumor (803 images), meningioma tumor (905 images), pituitary tumor (814 images) and no tumor (668 images) respectively. Each image has a resolution of 512×512, using grayscale color mode. The sample images are provided in Figure 1.

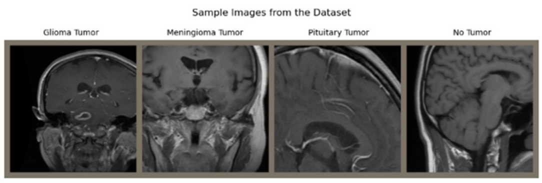

Figure 1: Sample images of brain tumor selected from the dataset (Photo/Picture credit: Original).

This study also implemented some preprocessing to improve the classification accuracy. First, because of the large resolution, this study randomly cropped the image to a size of 224×224 and changed images into RGB mode. Second, the images were flipped horizontally (left-right flip) to increase data diversity. Third, converting the image to a PyTorch tensor, normalizing the image values from integers ranging from 0 to 255 to float numbers between 0 and 1, and changing the image’s dimension format to fit PyTorch models. Finally, normalizing the images, aimed to improve the model’s efficiency and effectiveness. Through these transformations, the model’s generalization ability and the data’s consistency are enhanced.

2.2 Federated Learning-Based EfficientNet for Brain Tumor Classification

Federated Learning is a novel Distributed Machine Learning architecture. It mainly focuses on the privacy problems in machine learning tasks. The basic procedure of Federated Learning is shown as Figure 2.

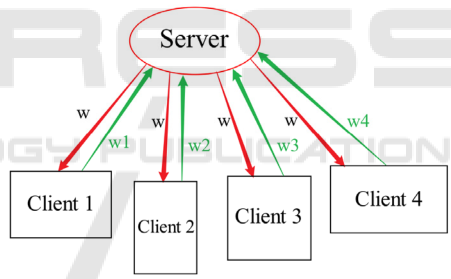

Figure 2: Basic procedure of Federated Learning (Photo/Picture credit: Original).

First of all, the parameter server sends the initial model w_0 to all the clients. Then, each client uses its own data to train the model and get a new trained model w_i. Finally, the clients send the models w_i back to the server. The server aggregates all the models and gets the final version of the model. The procedure guarantees that there is no data exchange between clients and the server, in order to protect data privacy. Meanwhile, the structure of Federated Learning is distributed, increasing efficiency but causing computational heterogeneity.

For the Convolutional Neural Network (CNN), this study chose EfficientNets (Tan and Le, 2019). To increase the accuracy of CNN, increasing width, depth and image resolution are three aspects to mainly consider. EfficientNets have better accuracy through improving these factors. This study used the EfficientNet-B0 baseline network. EfficientNet-B0 baseline network has nine stages, including one normal Conv, seven MobileNetConv (MBConv), and one 1×1 Conv, Pooling Layers & Full Connections (FC), with Batch Normalization (BN) and activation function Swish.

To combine the Federated Learning architecture and EfficientNet-B0 baseline network, the study used the FedAvg Algorithm. FedAvg is a fundamental FL Algorithm. The Algorithm improves the aggregate step in the procedure of FL, adding an averaging step to get a $\overline{w}$ weighted averaged by w_i 's model parameters. So, the study used EfficientNet to process data and detect the features to classify the images. And used FedAvg to aggregate and average every client's trained model to get an accurate model finally.

2.3 Implementation Details

This study set the hyperparameters including global epochs, local epochs, number of clients, number of clients participating in each global round, mini-batch size and learning rate. In terms of optimizer, the study used AdamW (Loshchilov and Hutter, 2017). AdamW inherits the advantages of adaptive learning rate from Adam. Compared with Adam, AdamW adds weight decay regularization after gradient calculation, having better generalization and convergence. Suppose the model weights are represented by θ, λ represents the regularization coefficient and η represents the learning rate, the change of AdamW can be written as (C represents the momentum correction):

$$\theta_t = \theta_{t-1} - \{\eta_t(C + \lambda\theta\}_{t-1}) \quad (1)$$

As for loss function, the study chose Cross-entropy loss. Cross-entropy loss is widely used in the image classification field because it only focuses on the current category and no need to update the weights when the classification is correct. Cross-entropy is used to measure the difference between two possibility distributions. In the machine learning field, if the true possibility distribution is Y(X), and when training, using an approximate distribution P(X) to fit, the Cross-entropy is:

$$H(Y,P) = -\sum_{i}^{n} Y(X = x_i) \log P(X = x_i) \quad (2)$$

In this image classification task, if the number of categories is n, batch size is b, the true distribution is Y, and the trained distribution is $\hat{Y}$, the Cross-entropy loss is:

$$LCE = -\frac{1}{b}\sum_{i}^{b}\sum_{j}^{n} y_{ij} \log \hat{y}_{ij} \quad (3)$$

3 RESULTS AND DISCUSSION

3.1 Parameter Tuning Results and Final Accuracy

After coding and parameter tuning, the study found the best accuracy based on the mentioned methods in the last part. The parameters set are shown in Table 1.

Table 1: Parameters Set.

Index	Value
Dataset	MRI
CNN model	EfficientNet-B0
Number of clients	5
Number of participated clients in each round	3
Number of global epochs	100
Number of local epochs	5
Batch size	32
Learning rate	0.0001

The highest accuracy emerged at the 66th global epoch shown in Figure 3, which was 81.218%, exceeded 80%. And the lowest loss reached 0.657.

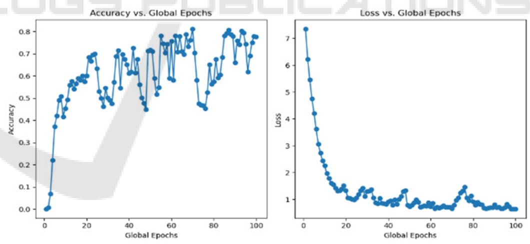

Figure 3: Final Testing Accuracy & Loss (Photo/Picture credit: Original).

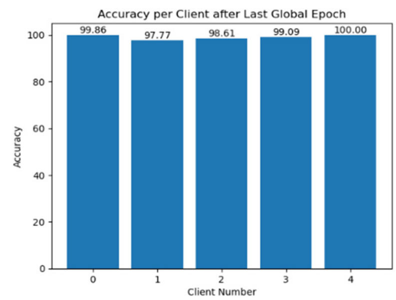

Figure 4: The Training Accuracy of each client with the parameters in the parameters set (Photo/Picture credit: Original).

Figure 4 shows the training accuracy of each client after the last training epoch. Every client was trained with a high accuracy, averaging 99%. The data's heterogeneity makes the curves rough, but the accuracy curve still shows an increasing trend and eventually stabilizes at around 70%. In order to further improve the accuracy, other algorithms' ideas like FedProx and SCAFFOLD will be added to reduce heterogeneity and the impact of data bias. Moreover, the method can well trim the loss value to make the loss curve converge faster.

3.2 Comparison of Different CNN Models

Except for EfficientNet, the study also tests the performance of ResNet (He et al., 2016) and VGG16 (Simonyan and Zisserman, 2014) on the MRI dataset. Table 2 shows the comparison of the two CNN models' performance. And Figure 5 shows the running results using ResNet. Every experiment set other parameters with the same values as Table 1 shows.

Table 2: Comparison of different CNN models using testing accuracy and loss.

CNN Model	Testing Accuracy (Max)	Testing Loss (Min)
EfficientNet-B0	81.218%	0.638
ResNet-50	68.367%	1.026
VGG16	77.157%	0.862

Through the accuracy and loss curves of ResNet, the study found the accuracy, loss and smooth of curves performance worse than EfficientNet. ResNet has been greatly affected by heterogeneity and is very unstable. Also, ResNet model cannot converge well after 100 global epochs.

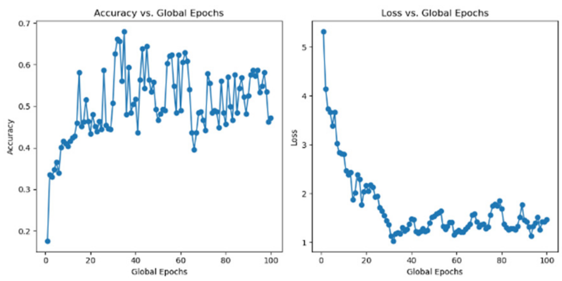

Figure 5: Testing accuracy and loss of using ResNet 50 versus global epochs (Photo/Picture credit: Original)

ResNet is a CNN model which focuses on increasing the depth of model through deep residual learning. Although it can recognize many details of the data, ResNet needs more computing resources and time to train. Compared with ResNet, EfficientNet uses Compound Model Scaling to flexibly adjust the depth, width, and resolution of the data simultaneously. This feature makes it easier to adapt to different types of data, handling data heterogeneity problems more effectively. Moreover, EfficientNet uses the technology of AutoAugment (Cubuk et al., 2018) to get the different operated images for training. Thus, EfficientNet is more efficient than ResNet, and needs fewer computing resources and less time to get a high accuracy and better convergence. To get a better performance using ResNet, the study may do further image preprocessing and use more GPUs to train.

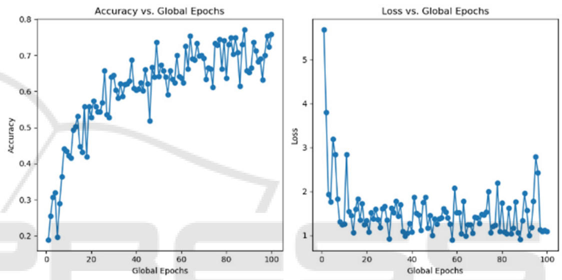

Figure 6: Testing accuracy and loss of using VGG16 versus global epochs (Photo/Picture credit: Original).

Through Figure 6, the whole performance of VGG16 is also worse than EfficientNet. The loss is more unstable than in Figure 3, and there are huge fluctuations in the curve. However, the accuracy curve is smoother and more stable, with a lower accuracy of 77.157% than EfficientNet. Also, because of the large depth of VGG, it needs much more time to train a model. During statistics, on the same GPU and CPU conditions, the running time cost is seven times longer than EfficientNet.

In a word, due to the stability, speed, and high accuracy, the study finally chose EfficientNet as the final CNN model in the experiment.

3.3 Learning Rate

In the parameter tuning process, the study also changed the learning rate to test the impact. The study set learning rates equal to 0.001 and 0.0001 respectively and get the results in Figure 7 and Figure 3.

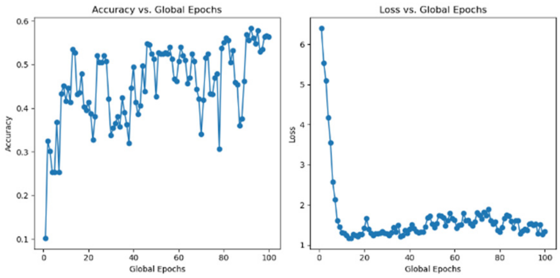

Figure 7: Testing accuracy and loss with learning rate = 0.001, other parameters' values are the same as the parameters set (Photo/Picture credit : Original).

When the learning rate = 0.001, the accuracy dropped a lot and the performance of stability and convergence also dropped. However, in the first few epochs, this model quickly reached a higher accuracy than the model of 0.0001 learning rate. Also, it was about to converge earlier but did not keep converging.

A larger learning rate is not suitable for training such detailed medical data, and it is easy to skip the details and achieve the wrong classification. On the contrary, a smaller learning rate can have better accuracy and convergence because it can focus on more details of the images and use these details to do the right classification.

4 CONCLUSIONS

This article applies Federated Learning to the MRI dataset, aiming to improve data privacy. Combining the EfficientNet-B0 and FedAvg Algorithm, the study developed a flexible and secure classification method compared with recent methods. Through experiments, the study found the best hyperparameters to train the model with high accuracy and fast convergence. Furthermore, the study compared the performance of different CNN models to demonstrate the advantages of the combination. In terms of future study, heterogeneity of the data is a big deal, how to further combine a good method to improve the accuracy in more heterogenous data will be an important research direction. Also, the method should be tested through other complex datasets.

REFERENCES

AbdulRahman, S., Tout, H., Ould-Slimane, H., Mourad, A., Talhi, C., & Guizani, M. 2020. A survey on federated learning: The journey from centralized to distributed on-site learning and beyond. IEEE Internet of Things Journal, 8(7), 5476-5497.

Bhuvaji, S., Kadam, A., Bhumkar, P., & Dedge, S. 2020. Brain Tumor Classification (MRI). Kaggle. https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri/data

Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. 2018. Autoaugment: Learning augmentation policies from data. arXiv preprint arXiv:1805.09501.

He, K., Zhang, X., Ren, S., & Sun, J. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A. T. 2020. Scaffold: Stochastic controlled averaging for federated learning. In International conference on machine learning (pp. 5132-5143). PMLR.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. 2021. Advances and open problems in federated learning. Foundations and trends® in machine learning, 14(1–2), 1-210.

Li, S., Kou, P., Ma, M., Yang, H., Huang, S., & Yang, Z. 2024. Application of Semi-supervised Learning in Image Classification: Research on Fusion of Labeled and Unlabeled Data. IEEE Access.

Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. 2020. Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems, 2, 429-450.

Liu, Y., Yang, H., & Wu, C. 2023. Unveiling patterns: A study on semi-supervised classification of strip surface defects. IEEE Access, 11, 119933-119946.

Loshchilov, I., & Hutter, F. 2017. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101.

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics (pp. 1273-1282). PMLR.

Qiu, Y., Hui, Y., Zhao, P., Cai, C. H., Dai, B., Dou, J., ... & Yu, J. 2024. A novel image expression-driven modeling strategy for coke quality prediction in the smart cokemaking process. Energy, 294, 130866.

Qiu, Y., Wang, J., Jin, Z., Chen, H., Zhang, M., & Guo, L. 2022. Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. Biomedical Signal Processing and Control, 72, 103323.

Simonyan, K., & Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Tan, M., & Le, Q. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

Wang, J., Liu, Q., Liang, H., Joshi, G., & Poor, H. V. 2020. Tackling the objective inconsistency problem in heterogeneous federated optimization. Advances in neural information processing systems, 33, 7611-7623.