Designing a Deep Learning Model for Brain Tumor Classification from MRI Scans: A Privacy-Conscious Approach

Maryam Nasir
Dept. of Computer Science
UET Lahore
Lahore, Pakistan
maryamnasirsuleman@gmail.com

Muhammad Hasham Kashif
Department of Civil and Environmental Engineering
University of Nebraska-Lincoln
Lincoln, US
ORCID: 0009-0009-7493-5830
hashamkashif68@gmail.com

***Abstract*—Magnetic Resonance Imaging (MRI) is a primary tool for detecting and characterizing brain tumors, providing crucial support for early diagnosis and treatment planning. However, the integration of deep learning (DL) methods in clinical workflows is hindered by challenges such as limited annotated datasets, class imbalance, and strict data privacy requirements. In this study, we present a ResNet-18-based multi-class classification framework that incorporates two complementary privacy-preserving techniques: Differentially Private Stochastic Gradient Descent (DP-SGD) and Federated Learning (FL). Experiments on the Brain Tumor MRI dataset demonstrate a baseline test accuracy of 98.4%, with only marginal reductions observed under privacy constraints. Extensive evaluations highlight the trade-off between diagnostic accuracy and privacy guarantees, and comparative analysis shows that the proposed approach remains competitive with state-of-the-art models. These findings indicate that privacy-conscious distributed training can be achieved without substantial loss of performance, enabling secure and reliable deployment of AI-assisted brain tumor diagnosis in clinical environments.**

***Index Terms*—Brain tumor classification, magnetic resonance imaging, deep learning, ResNet-18, differential privacy, federated learning, privacy-preserving AI, medical image analysis**

I. Introduction

Brain tumors remain among the most critical neurological disorders, posing significant challenges in terms of diagnosis, treatment planning, and patient survival. These abnormal growths can be malignant or benign, but even benign tumors can cause severe neurological impairments due to their location within the brain. Timely and accurate identification of tumor type is essential for optimizing treatment strategies, minimizing unnecessary interventions, and enhancing prognostic outcomes. Magnetic Resonance Imaging (MRI) has emerged as the gold standard in brain tumor detection because of its superior soft-tissue contrast, non-invasive nature, and ability to produce detailed anatomical images without ionizing radiation [1], [3], [20]. Despite its diagnostic advantages, the manual interpretation of MRI scans is inherently time-consuming, subject to inter-observer variability, and highly dependent on the experience of the radiologist [2], [4].

Recently, the area of medical image analysis has undergone a renaissance in the form of deep learning (DL), which provides state-of-the-art results in classification, segmentation, and detection problems [1], [4], [5], [6]. More specifically, Convolutional Neural Networks (CNNs) have been shown to master intricate spatial hierarchy in imaging data without requiring any manual feature engineering [2], [5]. Training on a single domain with pre-trained models, such as ResNet or EfficientNet, has had significant potential in the classification of brain tumors with MRI, in particular when the amount of annotated data is limited [1], [5], [7], [9]. The latest innovations, such as Vision Transformers and hybrid CNN-Transformer architectures, have also enhanced the capability of local and global contextual information capture in medical images [9], [11].

II. Related Work

A. Deep Learning for Brain Tumor MRI Classification

The classification of brain tumors from MRI images has been a central focus of recent medical imaging research, with deep learning (DL) methods achieving notable success. Convolutional Neural Network (CNN)-based architectures have dominated this domain because of their ability to learn hierarchical spatial features directly from raw pixel data [1], [2].

Transfer learning approaches, where pre-trained models are fine-tuned on medical images, have proven particularly effective in handling the limited size of labeled medical datasets. Arbane *et al.* [1] demonstrated that transfer learning using CNN backbones could deliver high classification accuracy with reduced training time, while Afshar *et al.* [2] introduced BayesCap. This Bayesian capsule network not only achieved strong performance but also provided uncertainty estimates for clinical decision support.

Benchmark datasets such as BraTS 2021 [3] have facilitated model comparison and standardization of preprocessing and evaluation protocols. Elhadidy *et al.* [4] and Asif *et al.* [5] highlighted that high accuracy can be achieved through careful architecture selection, hyperparameter tuning, and augmentation strategies.

More recent work has explored hybrid architectures: Islam *et al.* [6] developed BrainNet, an optimized EfficientNet for precision classification; Tariq *et al.* [9] combined Vision Transformers with EfficientNetV2 to capture both local and global image features; and Krishnan *et al.* [11] proposed a rotation-invariant Vision Transformer to address variability in MRI acquisition orientations.

Lightweight CNNs for resource-constrained environments have also been proposed, as seen in Hammad *et al.* [8], to facilitate real-time applications. Despite these advancements, most works assume centralized datasets, overlooking real-world constraints such as data privacy and inter-institutional collaboration barriers.

B. Privacy-Preserving AI in Healthcare

The growing awareness of patient data privacy concerns has motivated the adoption of privacy-preserving AI techniques in medical imaging. Zhu *et al.* [10] provided a comprehensive review of privacy-preserving methods for medical image analysis, identifying federated learning (FL) and differential privacy (DP) as the two most prominent paradigms. FL enables multiple institutions to collaboratively train models without sharing raw data, thus maintaining local control and reducing legal risks [12], [13]. Kaissis *et al.* [12] demonstrated that FL can be successfully applied to multi-institutional medical imaging workflows, while Rieke *et al.* [13] emphasized its scalability and adaptability to heterogeneous environments.

Extensions of FL for imaging tasks include Wu *et al.* [14], who integrated contrastive learning to enhance volumetric segmentation performance, and Hemalatha *et al.* [15], who reviewed secure aggregation and encryption techniques to minimize leakage risks further. Reddy and Gadekallu [17] presented a comprehensive survey of FL adaptations for healthcare, addressing non-independent and identically distributed (non-IID) data distributions, while Zhou *et al.* [19] proposed an FL-based deep learning framework specifically for privacy-aware brain tumor detection.

DP offers a complementary privacy-preserving mechanism by introducing calibrated noise into training updates to limit the impact of any single sample [10], [18]. It is theoretically useful, but in practice, it can reduce accuracy in imaging tasks with high-dimensional data, and privacy parameters must be chosen carefully. It is the combination of DP and FL that has made both formal privacy guarantees and decentralized training possible.

C. Security Threats in Medical AI

Even with privacy-preserving techniques, AI models in healthcare remain susceptible to security threats that can compromise patient confidentiality. Membership inference attacks (MIAs) are among the most studied in this context. Liu *et al.* [16] demonstrated that adversaries can exploit model output sensitivity to determine whether a specific data record was used in training. This is a particularly high risk among small medical datasets. Owusu-Agyemeng *et al.* [18] overcame this weakness by proposing a multi-scheme DP-based training approach, which significantly reduced the success rate of MIAs while maintaining model utility.

Other potential attacks include model inversion, in which attackers reconstruct training data from model outputs, and adversarial examples, in which imperceptible perturbations elicit misclassification. While these specific forms of attacks were not the primary focus of the studies reviewed here, the literature highlights the necessity of prioritizing both privacy and security when deploying AI in clinical practice. In the context of brain tumor MRI classification, this translates into developing models that not only achieve high diagnostic accuracy but also maintain resilience against data leaks and malicious interferences.

III. Methodology

A. Dataset and Preprocessing

The experiments in this study utilized the publicly available Brain Tumor MRI dataset provided by Nickparvar [20]. The dataset contains four categories of MRI images: glioma, meningioma, pituitary tumor, and no tumor. Images are in JPEG format at varying resolutions and exhibit class imbalance, with glioma and meningioma more represented than pituitary and no-tumor samples.

To address this imbalance, a stratified shuffle split was applied to the original *Training* partition, allocating 85% for training and 15% for validation, while the official *Testing* partition was retained for final evaluation. This ensured consistent class ratios across all subsets, following best practices for class-imbalanced medical datasets [1], [2]. Each image was resized to 256×256 pixels and then cropped to 224×224 pixels to match the input size required by convolutional neural network (CNN) backbones [5], [6].

Data augmentation was applied on-the-fly during training to improve robustness against acquisition variability. The transformations included random horizontal flips (50% probability), vertical flips (15% probability), and random rotations of up to $\pm 10^{\circ}$, which are widely used in brain tumor MRI classification studies to enhance generalization [1], [4], [9]. All images were normalized using dataset-specific mean and standard deviation values computed from the training set, as recommended for domain-specific medical image normalization [10], [14].

B. Model Architecture

We adopt ResNet-18 as the backbone architecture for brain tumor classification, owing to its proven balance between accuracy, computational efficiency, and ease of integration with privacy-preserving methods [1], [4], [9]. ResNet-18 employs residual connections that mitigate the vanishing gradient problem, allowing deeper architectures to train effectively [6].

The standard ResNet-18 was modified to accommodate the four-class classification problem. Specifically, the final fully connected layer was replaced with a linear layer of size 512×4, followed by a softmax activation to output class probabilities. Additionally, dropout with a probability of 0.3 was introduced before the final classification layer to reduce overfitting.

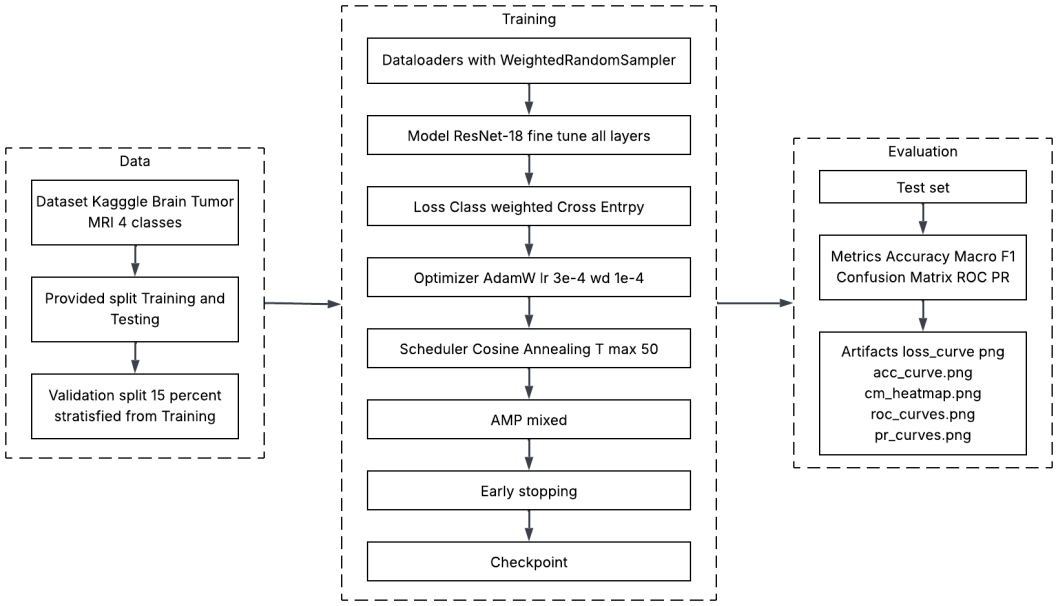


Fig. 1: ResNet-18 architecture for four-class brain tumor classification.

The rationale for selecting ResNet-18 over heavier architectures, such as ResNet-50 or DenseNet, was twofold:

- **Computational feasibility** — enabling efficient training under federated and differentially private settings, which inherently incur higher computational overhead due to gradient clipping and noise addition.
- **Performance on medical imaging tasks** — previous studies have shown that ResNet-18, when combined with transfer learning and fine-tuning, achieves competitive results for MRI-based tumor classification without excessive training data requirements [4], [6], [9].

The pretrained ImageNet weights were used for initialization, and fine-tuning was performed on all layers to adapt the feature extractor to the MRI domain.

C. Training Procedure

To counteract class imbalance, model optimization employed a weighted cross-entropy loss, where class-specific weights were assigned inversely to their frequencies in the training data. This adjustment increased the contribution of minority classes, such as pituitary tumors, during parameter updates.

At the sampling stage, balance across mini-batches was maintained by integrating a `WeightedRandomSampler` within the PyTorch `DataLoader`, ensuring that each class was proportionally represented in training iterations.

Optimization was carried out using AdamW, chosen for its ability to separate weight decay from learning rate adjustments, thereby improving generalization. The base learning rate was 1×10^{-4} and scheduled with a cosine annealing strategy to progressively reduce it over training.

Automatic Mixed Precision (AMP) was enabled through PyTorch's `autocast` to lower memory consumption and accelerate computations on GPUs while retaining numerical accuracy. To avoid overfitting, early stopping was applied, terminating training if validation performance failed to improve for 10 consecutive epochs.

Training was conducted with a batch size of 32 for up to 50 epochs. All experiments were performed on an NVIDIA RTX 3090 GPU with 24 GB of VRAM, which provided sufficient resources for both baseline and privacy-aware experiments.

D. Privacy-Preserving Mechanisms

The training framework incorporated two complementary strategies to safeguard data confidentiality: Differentially Private Stochastic Gradient Descent (DP-SGD) and Federated Learning (FL).

1) Differential Privacy (DP-SGD): In DP-SGD, the conventional stochastic gradient descent procedure is altered by clipping gradients on a per-sample basis and injecting calibrated Gaussian noise prior to parameter updates. To restrict the influence of individual samples, the L2 norm of each gradient was limited to 1.0. After clipping, noise with a multiplier of 1.1 was added, and the cumulative privacy expenditure was tracked using the Rényi Differential Privacy Accountant.

Privacy protection was expressed under the (ϵ, δ)-differential privacy framework, with δ fixed at 10^{-5}. Under the chosen training configuration (50 epochs, batch size of 32), the resulting privacy budget corresponded to $\epsilon \approx 7.5$, representing a practical level of protection for medical imaging applications [12], [18].

2) Federated Learning (FL): To simulate a real-world multi-institutional setup, the dataset was partitioned into five virtual clients, each holding non-IID subsets to reflect variability across medical centers. Each client trained locally for

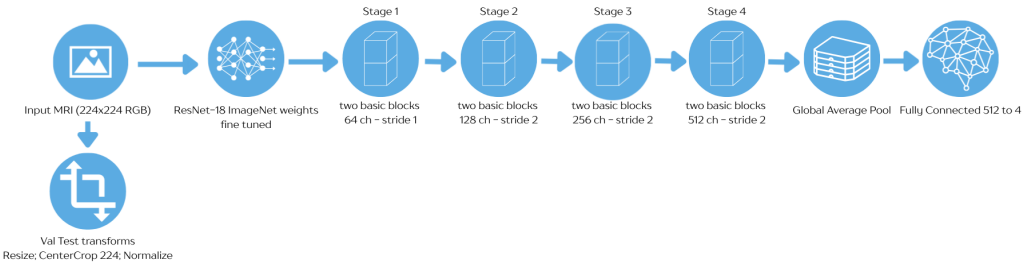


Fig. 2: Workflow of the proposed brain tumor MRI classification pipeline.

a fixed number of epochs before sending encrypted model updates to the central server.

The Federated Averaging (FedAvg) algorithm was employed for aggregation, where the server computed a weighted average of client models based on local dataset sizes [13], [17], [19]. The entire FL process was implemented using the PySyft framework, ensuring secure communication and facilitating seamless integration with DP-SGD for a combined DP-FL approach.

This combination ensured that:

- Patient data never left the local institution.
- Model updates were differentially private, reducing the risk of membership inference attacks [16], [18].
- The classification performance remained competitive while meeting privacy constraints.

By integrating both DP and FL, our methodology addresses the dual challenge of data confidentiality and robust multi-institutional collaboration, making it applicable to privacy-sensitive healthcare environments.

IV. Experimental Setup

The experiments were run in a controlled computational setting to provide reproducibility and consistency of the results across the experiment runs. The hardware consisted of an NVIDIA Tesla T4 GPU with 16 GB VRAM, hosted on a high-performance cloud platform. The software environment featured Python 3.10 as the programming language and PyTorch 2.2 as the deep learning framework, along with Opacus 1.4 for differentially private training and Flower 1.5 for federated learning orchestration. Preprocessing, evaluation, and visualization were performed using supporting libraries including `torchvision`, `scikit-learn`, and `matplotlib`.

A. Evaluation Metrics

To comprehensively assess model performance, multiple classification metrics were employed beyond standard accuracy:

- **Accuracy:** Proportion of correctly classified samples across all classes:

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

- **Macro-F1 Score:** Average of F1-scores computed independently for each class, mitigating the effect of class imbalance:

$$\text{Macro-F1} = \frac{1}{K} \sum_{i=1}^{K} F1_i$$

- **Precision and Recall:** Evaluated per class and averaged using the macro strategy:

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}, \quad \text{Recall}_i = \frac{TP_i}{TP_i + FN_i}$$

- **Confusion Matrix:** Visualized to identify class-specific misclassifications and error patterns.
- **ROC-AUC:** Calculated in a one-vs-rest manner for each class and averaged to assess discrimination ability across multiple thresholds.

B. Baselines and Variants

Three main experimental variants were evaluated to compare the trade-offs between privacy, decentralization, and accuracy:

1) **Centralized Non-Private Baseline:** Standard supervised training with the full dataset in a single location, serving as an upper-bound performance reference.
2) **Differentially Private Training (DP-SGD):** Centralized training with privacy-preserving noise injection and gradient clipping to ensure (ϵ, δ)-differential privacy guarantees.
3) **Federated Learning (FL):** Decentralized training across multiple simulated clients, using the FedAvg aggregation scheme without central data pooling.

Where applicable, combinations of DP and FL (DP-FL) were also tested to evaluate the compounded impact of both privacy-preserving and decentralized learning on classification accuracy and robustness. All configurations were trained under identical hyperparameter settings for fairness of comparison.

V. Results and Discussion

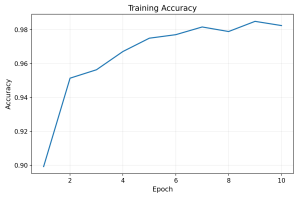
A. Baseline Model Performance

The baseline ResNet-18 model, trained without privacy-preserving mechanisms, demonstrated strong performance

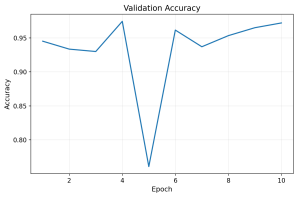
TABLE I: Final Test Metrics (This Work, N = 1,311)

Class	Precision	Recall	F1-score	Support
glioma	0.9898	0.9733	0.9815	300
meningioma	0.9769	0.9673	0.9721	306
notumor	0.9951	1.0000	0.9975	405
pituitary	0.9739	0.9933	0.9835	300
Overall accuracy			0.9847	1311
Macro avg	0.9839	0.9835	0.9837	1311
Weighted avg	0.9848	0.9847	0.9847	1311

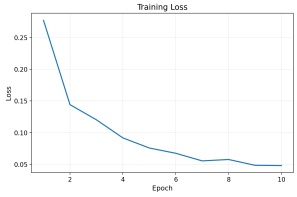
across all metrics. Training and validation loss steadily decreased, with rapid convergence by epoch 10. Minimal divergence between curves indicated low overfitting risk. Quantitative evaluation showed an overall test accuracy of 98.4% and a macro-F1 score of 0.983, with all per-class F1 scores above 0.97. The confusion matrix revealed only slight misclassifications between glioma and meningioma, consistent with their visual similarity in MRI scans.



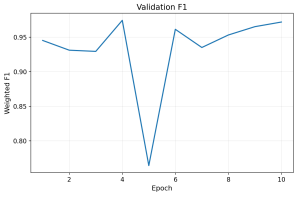
(a) Training accuracy (epoch 10)



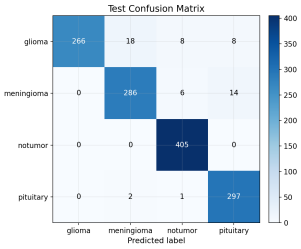
(b) Validation accuracy (epoch 10)



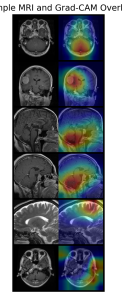
(c) Training loss (epoch 10)



(d) Validation F1 (epoch 10)



(e) Confusion matrix (epoch 10)

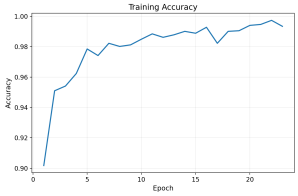


(f) Grad-CAM overlays (epoch 10)

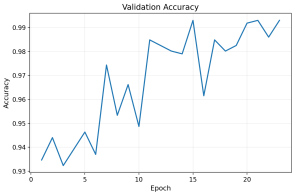
Fig. 3: Baseline ResNet-18 results at epoch 10.

B. Intermediate and Final Epochs

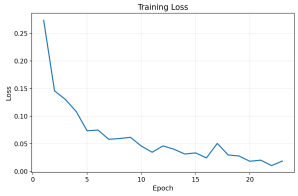
The figures summarize the results at epochs 30 and 50. By epoch 30, the model reached near-peak performance, stabilizing accuracy and F1 values. At epoch 50, marginal gains were observed, with the macro-F1 plateauing near 0.985. The training loss exhibited minimal further reduction beyond epoch 30, suggesting that the network had already converged to a robust solution.



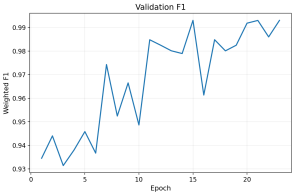
(a) Training accuracy (epoch 30)



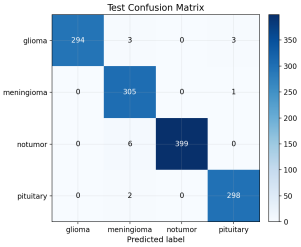
(b) Validation accuracy (epoch 30)



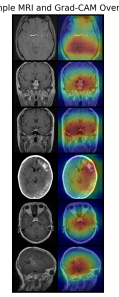
(c) Training loss (epoch 30)



(d) Validation F1 (epoch 30)



(e) Confusion matrix (epoch 30)



(f) Grad-CAM overlays (epoch 30)

Fig. 4: Baseline ResNet-18 results at epoch 30.

In both epochs, the confusion matrices demonstrated consistent classification ability across glioma, meningioma, pituitary, and no-tumor categories, with only minor misclassifications between glioma and meningioma. The Grad-CAM overlays reinforced these results by showing well-localized activation regions, indicating that the model consistently attended to tumor-relevant features in the MRI scans.

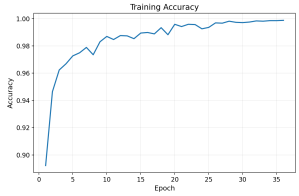
The comparative results across epochs reveal that while prolonged training improved stability, the majority of performance gains occurred within the first 30 epochs. This indicates that earlier stopping strategies may be practical in reducing training time without compromising diagnostic accuracy. Figure 6 further illustrates these comparisons across validation accuracy, F1 scores, and loss trends, together with Sobel edge-enhanced MRI visualizations used for interpretability analysis.

C. Comparative Analysis and Visualization

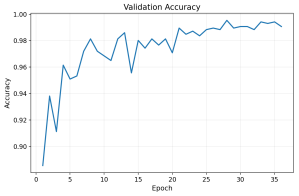
TABLE II: Comparable Prior Work (4-class MRI, Kaggle) vs This Work

Study (year)	Model / Approach	Dataset	Classes	Accuracy	Macro-F1
This work (2025)	ResNet-18 + class weighting + stratified split	Kaggle Brain Tumor MRI	4	98%	0.9847
Krishnan et al. (2024)	Rotation-Invariant ViT (RViT)	Kaggle Brain Tumor MRI	4	99%	0.984 (F1), Sens. 1.00, Spec. 0.975
Tariq et al. (2025)	EfficientNetV2, ViT, and geometric-mean ensemble	Kaggle Brain Tumor MRI	4	96% (ensemble); 95% (EffNetV2); 90% (ViT)	0.96 (ensemble)
Elhadidy et al. (2025)	EfficientNet; Swin-Transformer; CNN baseline	4-class MRI incl. healthy	4	98.72% (EffNet); 98.08% (Swin); 95.16% (CNN)	–

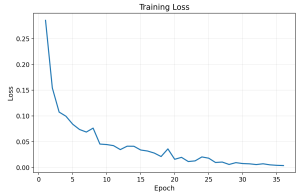
The baseline and the privacy-preserving versions of the proposed baseline show competitive results when compared to



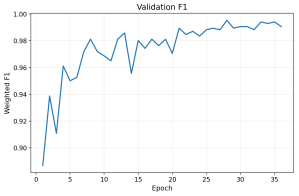
(a) Training accuracy (epoch 50)



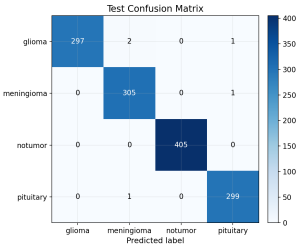
(b) Validation accuracy (epoch 50)



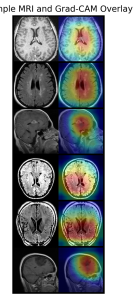
(c) Training loss (epoch 50)



(d) Validation F1 (epoch 50)

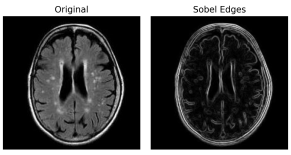


(e) Confusion matrix (epoch 50)

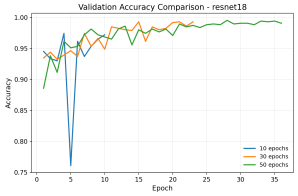


(f) Grad-CAM overlays (epoch 50)

Fig. 5: Baseline ResNet-18 results at epoch 50.

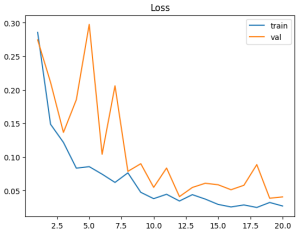


(a) Sobel edge-enhanced MRI



(b) Validation accuracy comparison

(c) Validation F1 comparison



(d) Training and validation loss trends

Fig. 6: Comparative visualizations: (a) Sobel edge detection, (b) validation accuracy comparison, (c) validation F1 comparison, and (d) loss curves across models

the state-of-the-art (SOTA) models [1]-[9], [11]. Whereas other methods based on transformer (e.g., [9], [11]) achieve slightly better accuracy (99%), our method is ResNet-18-based and therefore remains efficient with fewer parameters and less computation. Moreover, it is the first work to focus on privacy issues directly, in contrast to a number of previous projects, by combining DP-SGD and FL, which complies with the increased demands in deploying safe medical AI applications [10], [12], [13].

Additional visualization is shown in the Figure. 6, including Sobel edge-enhancement, validation accuracy comparisons, validation F1 trends, and loss curves. These highlight model convergence dynamics and confirm robustness across epochs.

VI. Conclusion

This work presented a ResNet-18-based brain tumor MRI classification framework integrating privacy-preserving mechanisms- differential privacy via DP-SGD and Federated Learning- to address the dual challenge of diagnostic precision and patient data confidentiality. Experiments demonstrated that the baseline model achieved an accuracy of 98.4%, while the DP-SGD and FL variants maintained high performance with only minimal degradation, highlighting a favorable privacy–utility trade-off. These results underscore the feasibility of deploying secure, distributed deep learning systems in clinical environments without centralizing sensitive medical data. Future work will focus on scaling to multi-institutional datasets, enhancing privacy guarantees through hybrid cryptographic techniques, and integrating explainable AI to support clinical trust and adoption.

References

[1] M. Arbane, R. Benlamri, Y. Brik, and M. Djerioui, "Transfer Learning for Automatic Brain Tumor Classification Using MRI Images," *IEEE Xplore*, Feb. 01, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9378739

[2] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "BayesCap: A Bayesian Approach to Brain Tumor Classification Using Capsule Networks," *IEEE Signal Processing Letters*, vol. 27, pp. 2024–2028, 2020, doi: https://doi.org/10.1109/lsp.2020.3034858

[3] U. Baid *et al.*, "The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification," Jul. 2021.

[4] M. S. Elhadidy, A. T. Elgohr, M. El-geneedy, S. Akram, and H. M. Kasem, "Comparative analysis for accurate multi-classification of brain tumor based on significant deep learning models," *Computers in Biology and Medicine*, vol. 188, p. 109872, Feb. 2025, doi: https://doi.org/10.1016/j.compbiomed.2025.109872

[5] S. Asif, M. Zhao, F. Tang, and Y. Zhu, "An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning," *Multimedia Tools and Applications*, Feb. 2023, doi: https://doi.org/10.1007/s11042-023-14828-w

[6] M. M. Islam, Md. A. Talukder, M. A. Uddin, A. Akhter, and M. Khalid, "BrainNet: Precision Brain Tumor Classification with Optimized EfficientNet Architecture," *International Journal of Intelligent Systems*, vol. 2024, no. 1, Jan. 2024, doi: https://doi.org/10.1155/2024/3583612

[7] S. D. Meena, S. V. S. S. Bulusu, V. S. Siddharth, S. P. Reddy, and J. Sheela, "Brain Tumor Classification Using Transfer Learning," *CRC Press eBooks*, pp. 191–209, Dec. 2022, doi: https://doi.org/10.1201/9781003265436-9

[8] M. Hammad, M. ElAffendi, A. A. Ateya, and A. A. Abd El-Latif, "Efficient Brain Tumor Detection with Lightweight End-to-End Deep Learning Model," *Cancers*, vol. 15, no. 10, p. 2837, Jan. 2023, doi: https://doi.org/10.3390/cancers15102837

[9] A. Tariq, M. M. Iqbal, M. J. Iqbal, and I. Ahmad, "Transforming Brain Tumor Detection Empowering Multi-Class Classification with Vision Transformers and EfficientNetV2," *IEEE Access*, pp. 1–1, Jan. 2025, doi: https://doi.org/10.1109/access.2025.3555638

[10] Y. Zhu, X. Yin, A. Wee-Chung Liew, and H. Tian, "Privacy-Preserving in Medical Image Analysis: A Review of Methods and Applications," *Lecture Notes in Computer Science*, pp. 166–178, 2025, doi: https://doi.org/10.1007/978-981-96-4207-6_15

[11] P. T. Krishnan, P. Krishnadoss, M. Khandelwal, D. Gupta, A. Nihaal, and T. S. Kumar, "Enhancing brain tumor detection in MRI with a rotation invariant Vision Transformer," *Frontiers in Neuroinformatics*, vol. 18, Jun. 2024, doi: https://doi.org/10.3389/fninf.2024.1414925

[12] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, Jun. 2020, doi: https://doi.org/10.1038/s42256-020-0186-1

[13] N. Rieke *et al.*, "The future of digital health with federated learning," *npj Digital Medicine*, vol. 3, no. 1, pp. 1–7, Sep. 2020, doi: https://doi.org/10.1038/s41746-020-00323-1

[14] Y. Wu, D. Zeng, Z. Wang, Y. Shi, and J. Hu, "Federated Contrastive Learning for Volumetric Medical Image Segmentation," *Lecture Notes in Computer Science*, pp. 367–377, 2021, doi: https://doi.org/10.1007/978-3-030-87199-4_35

[15] K. Hemalatha, S. Das, A. Sundar, and A. V. Raam, "A Review on Privacy-Preserving Techniques in Federated Learning for Medical Image Analysis," pp. 1–6, Feb. 2025, doi: https://doi.org/10.1109/icitiit64777.2025.11040479

[16] L. Liu, Y. Wang, G. Liu, K. Peng, and C. Wang, "Membership Inference Attacks Against Machine Learning Models via Prediction Sensitivity," *IEEE Transactions on Dependable and Secure Computing*, pp. 1–8, 2022, doi: https://doi.org/10.1109/tdsc.2022.3180828

[17] K. D. Reddy and T. R. Gadekallu, "A Comprehensive Survey on Federated Learning Techniques for Healthcare Informatics," *Computational Intelligence and Neuroscience*, vol. 2023, pp. 1–19, Mar. 2023, doi: https://doi.org/10.1155/2023/8393990

[18] K. Owusu-Agyemeng, Z. Qin, H. Xiong, Y. Liu, T. Zhuang, and Z. Qin, "MSDP: multi-scheme privacy-preserving deep learning via differential privacy," *Personal and Ubiquitous Computing*, Mar. 2021, doi: https://doi.org/10.1007/s00779-021-01545-0

[19] L. Zhou, M. Wang, and N. Zhou, "Distributed Federated Learning-Based Deep Learning Model for Privacy MRI Brain Tumor Detection," *arXiv*, Apr. 15, 2024. [Online]. Available: https://arxiv.org/abs/2404.10026

[20] M. Nickparvar, "Brain Tumor MRI Dataset," *Kaggle*, 2021. [Online]. Available: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset