Interpretable Framework for Brain Tumor Classification from MRI

Riyanshu Saini[1], Sumika Jain[2], Varun Bansal[3]

[1] (Research Scholar, Computer Science and Engineering, Shobhit University, Gangoh Email: rs2379298@gmail.com)
[2] (Assistant Professor, Computer Science and Engineering, Shobhit University, Gangoh Email: Sumika.Jain@shobhituniversity.ac.in)
[3] (Professor, Computer Science and Engineering, Shobhit University, Gangoh Email: Varun.bansal@shobhituniversity.ac.in)

Abstract — Brain tumor classification from magnetic resonance imaging (MRI) is important for timely diagnosis and treatment planning. This study proposes a hybrid deep learning framework for four-class brain tumor MRI classification using ResNet50 as a deep feature extractor and a Multi-Layer Perceptron (MLP) as the final classifier. The framework was evaluated on the publicly available Kaggle Brain Tumor MRI Dataset containing 7,023 images across glioma, meningioma, pituitary tumor, and no-tumor classes. The original test set of 1,311 images was retained exclusively for final evaluation. The proposed ResNet50 + MLP model achieved 97.64% accuracy, 97.61% precision, 97.58% recall, 97.59% F1-score, 99.21% specificity, and 99.61% AUC, outperforming CNN, VGG16, InceptionV3, DenseNet121, EfficientNetB0, MobileNetV3, and standalone ResNet50 under identical experimental conditions. Grad-CAM visualizations were used to provide visual interpretability of model predictions. The results demonstrate that the proposed framework offers accurate and interpretable support for automated brain tumor MRI classification.

Keywords — Brain Tumor Classification · MRI · ResNet50 · Multi-Layer Perceptron · Transfer Learning · Grad-CAM · Deep Learning · Hybrid framework.

I. INTRODUCTION

Brain tumors arise when intracranial cells begin dividing outside the regulatory controls that govern normal tissue growth. Because the brain is enclosed within a rigid skull, even a slow-growing mass can compress critical neural structures and produce severe neurological consequences disproportionate to its physical size. Clinically, brain tumors are classified as benign or malignant and graded from I to IV based on cellular aggressiveness [1]. Grade I tumors are typically slow-growing and surgically respectable, whereas grade IV tumors, such as glioblastoma multiforme, carry a median survival of fewer than 15 months despite aggressive multimodal therapy. Primary central nervous system tumors account for approximately 308,102 new diagnoses and 251,329 deaths annually worldwide, making accurate and timely classification a matter of genuine clinical urgency.

The present study addresses these gaps through the following specific contributions:

–Hybrid Architecture: A two-stage fine-tuned ResNet50 backbone is coupled with a two-hidden-layer MLP classifier in place of the conventional SoftMax head.

–Structured Preprocessing Pipeline: A reproducible preprocessing sequence comprising image resizing, pixel normalization, median filtering, CLAHE contrast enhancement, skull stripping, and Fuzzy C-Means segmentation.

–Systematic Seven-Model Benchmark: The proposed framework is evaluated against seven contemporary architectures, namely, VGG16 [23], In-ceptionV3 [19], DenseNet121 [12], EfficientNetB0 [21], MobileNetV3, standalone ResNet50, and a baseline CNN.

–Clinical Interpretability: Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations are generated for representative test images from each class, providing spatial justification for model predictions [6].

The remainder of this paper is organized as follows. Section 2 presents the research motivation and formulates the brain tumor MRI classification problem. Section 3 reviews the existing literature and identifies the research gaps addressed in this study. Section 4 describes the dataset, class distribution, preprocessing details, and data partitioning strategy. Section 5 explains the proposed ResNet50 + MLP framework, including feature extraction, classification architecture, training strategy, and evaluation metrics. Section 6 presents the experimental results, comparative performance analysis, class-wise evaluation, training behavior, Grad-CAM-based interpretability, and computational efficiency. Section 7 discusses the limitations of the study. Finally, Section 8 concludes the paper and outlines future research directions.

II. RELATED WORK

Brain tumor classification using magnetic resonance imaging (MRI) has attracted significant attention due to its potential to improve early diagnosis and clinical decision-making. Earlier approaches primarily relied on handcrafted feature extraction techniques, including Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), and Discrete Wavelet Transform (DWT), combined with conventional machine learning classifiers such as Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), and Random Forest. Although these methods achieved satisfactory performance on small datasets, their dependence on manually designed features limited their ability to capture complex tumor characteristics and generalize across diverse MRI datasets [2, 27].

The rapid advancement of deep learning has significantly improved automated brain tumor classification. Convolutional Neural Networks (CNNs) have demonstrated superior feature learning capabilities by automatically extracting hierarchical representations from MRI images. Transfer learning further enhanced classification performance by utilizing pre-trained models trained on large-scale image datasets. Several studies reported promising results using architectures such as VGG16, InceptionV3, ResNet50, DenseNet, MobileNet, and Xception, achieving classification accuracies exceeding 95% on publicly available brain tumor MRI datasets [3, 14, 19, 23]. These studies confirmed that transfer learning reduces training time while improving feature representation, particularly when limited medical imaging data are available.

More recently, researchers have incorporated attention mechanisms, transformer-based architectures, and Explainable Artificial Intelligence (XAI) techniques to improve both predictive performance and clinical interpretability. Vision Transformers and attention-based CNN models have shown competitive classification accuracy by learning global contextual information from MRI images, while Grad-CAM, SHAP, and LIME have been employed to generate visual explanations for model predictions, increasing clinicians' confidence in automated diagnostic systems [8, 9, 10]. However, these advanced models generally require higher computational resources and often provide only marginal improvements over efficient CNN-based architectures.

Despite the considerable progress achieved in recent years, several challenges remain unresolved. Most existing studies evaluate their methods on a single public dataset using different preprocessing pipelines, data augmentation strategies, and evaluation protocols, making direct comparison difficult. Furthermore, many approaches primarily emphasize overall classification accuracy while providing limited analysis of class-wise performance, computational efficiency, and model interpretability [7]. These limitations highlight the need for a computationally efficient and interpretable hybrid framework that combines robust deep feature extraction with effective classification. Motivated by these challenges, the proposed ResNet50–MLP framework employs transfer learning for discriminative feature extraction and a Multi-Layer Perceptron classifier to improve classification performance while maintaining model transparency through Grad-CAM-based visualization.

Approach	**Representative Methods**	**Strengths**	**Limitations**	**Contribution to This Work**
Conventional ML **[2,6,13,27]**	GLCM, LBP, DWT, k-NN	Simple and interpretable	Manual feature engineering; limited generalization	Motivates automatic deep feature extraction
CNN-based Models **[3,24,27]**	CNN, AlexNet, VGG16	Learns hierarchical image features	Performance depends on dataset size and preprocessing	Provides baseline deep learning models
Transfer Learning **[11,18,19,25]**	ResNet50, DenseNet 121, Inception V3, MobileNet	High accuracy with limited data	Conventional Softmax limits feature utilization	Justifies ResNet50 as feature extractor
Hybrid Models **[7,8,18]**	CNN+MLP, CNN+SVM	Improved feature discrimination	Requires classifier optimization	Forms the basis of the proposed ResNet50–MLP framework
Attention & Transformers **[1,9,20,24]**	CBAM, ViT	Captures global contextual information	High computational complexity	Used for performance comparison

Explainable AI **[8,10,16]**	Grad-CAM, SHAP, LIME	Improves model transparency	Limited quantitative clinical validation	Supports model interpretability
Segmentation-based Methods **[2,15,17]**	U-Net, Res-UNet	Accurate ROI localization	Requires pixel-level annotations	Highlights the importance of segmentation

Table 1: Methodological synthesis of existing brain tumor MRI classification and segmentation studies

III. METHODOLOGY

The proposed framework is a hybrid deep learning architecture designed for automatic multi-class brain tumor classification using magnetic resonance imaging (MRI). The framework integrates multiple processing stages, including image preprocessing, tumor region enhancement, deep feature extraction through transfer learning, and final classification using a Multi-Layer Perceptron (MLP) classifier. By combining these components into a unified pipeline, the proposed approach aims to improve both classification accuracy and computational efficiency while maintaining robustness across different brain tumor categories. Unlike conventional convolutional neural networks that are trained entirely from scratch, the proposed framework utilizes a pretrained ResNet50 model as a deep feature extractor. The pretrained network benefits from knowledge learned on the ImageNet dataset, enabling the extraction of highly discriminative hierarchical features from brain MRI images while significantly reducing training time and the amount of labeled medical data required [4, 18, 25].

Initially, the input MRI images undergo preprocessing to improve image quality and reduce unwanted variations. This stage includes image resizing, normalization, noise reduction, contrast enhancement, skull stripping, and Fuzzy C-Means (FCM)-based region enhancement to emphasize tumor-related structures and improve feature representation. Data augmentation techniques such as rotation, flipping, zooming, and shifting are applied only to the training dataset to increase sample diversity and improve the model's generalization capability [2, 15].

After preprocessing, the enhanced MRI images are passed through the pretrained ResNet50 backbone, where the convolutional layers automatically learn high-level spatial and semantic representations of tumor characteristics. Instead of employing the default Softmax classifier, the extracted 2048-dimensional feature vector is forwarded to a lightweight MLP classifier consisting of fully connected hidden layers with ReLU activation, dropout regularization, and a final Softmax output layer for four-class prediction. This hybrid design enables more effective nonlinear decision boundaries while reducing overfitting and improving class separability [18, 23].

Finally, the trained model classifies each MRI image into one of four clinically relevant categories: glioma, meningioma, pituitary tumor, or no tumor. The complete workflow of the proposed framework, including preprocessing, feature extraction, and classification stages, is illustrated in Fig. 1, providing an overview of the end-to-end architecture adopted in this study.

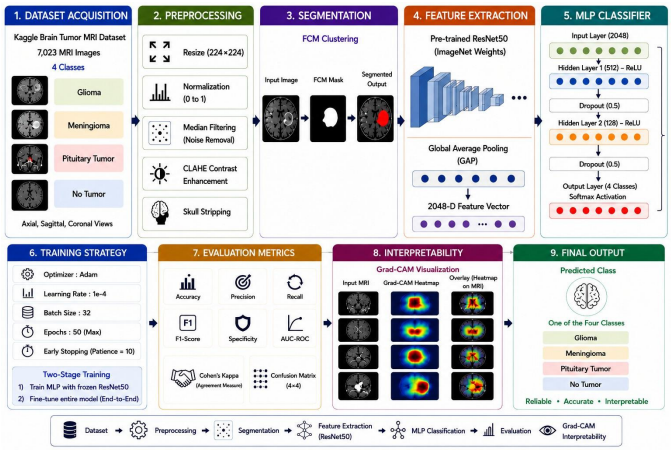


Fig. 1: Proposed methodology for MRI brain tumor classification using the hybrid ResNet50-MLP framework.

The dataset reproducibility details are summarized in Table 2.

Aspect	Description
Dataset	Publicly available Kaggle Brain Tumor MRI Dataset [22]
Data Type	Contrast-enhanced T1-weighted brain MRI images
Classes	Glioma, Meningioma, Pituitary Tumor, and No Tumor
Patient Data	No additional patient data were collected
Ethical Compliance	Only publicly available anonymized data were used; ethical approval not required
Preprocessing	Resize (224×224), normalization, denoising, CLAHE, skull stripping, FCM enhancem
Data Augmentation	Rotation, flipping, zooming, and shifting applied only to training data
Dataset Split	Stratified training, validation, and testing
Reproducibility	Preprocessing, augmentation, hyperparameters, and data split are fully reported

Table 2: Dataset source, ethics, and reproducibility details.

A. Dataset Preparation

Experiments were conducted on the publicly available Kaggle Brain Tumor MRI Dataset containing 7,023 contrast-enhanced T1-weighted MRI images belonging to four classes: glioma, meningioma, pituitary tumor, and no tumor [22].

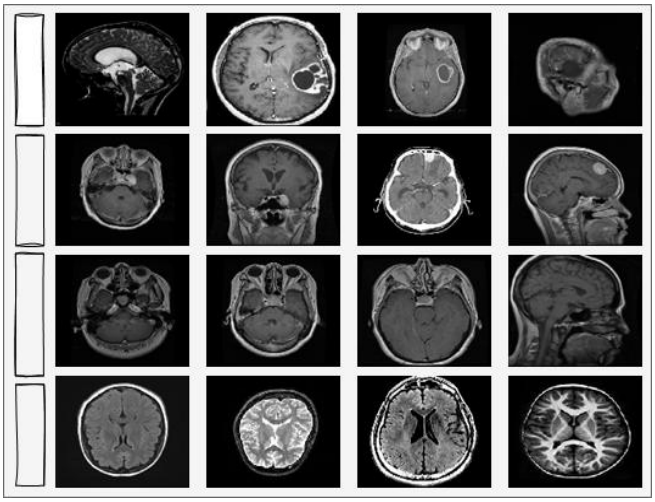


Fig. 2: Representative MRI samples from the four brain tumor dataset classes.

The dataset was divided into training, validation, and testing subsets using stratified sampling to preserve the class distribution. The final split consisted of 4,569 training, 1,143 validation, and 1,311 testing images, as presented in Table 3.

Class	Train	Val.	Test	Total	Dist. (%)
Glioma	1057	264	300	1621	23.08
Meningioma	1071	268	306	1645	23.42
Pituitary Tumor	1166	291	300	1757	25.02
No Tumor	1275	320	405	2000	28.48
Total	4569	1143	1311	7023	100.00

Table 3: Class-wise distribution of the brain tumor MRI dataset.

B. Image Preprocessing

To improve data consistency, all MRI images were resized to 224 × 224 × 3 pixels, normalized to the range [0,1], and enhanced using a standardized preprocessing pipeline shown in Fig. 3.

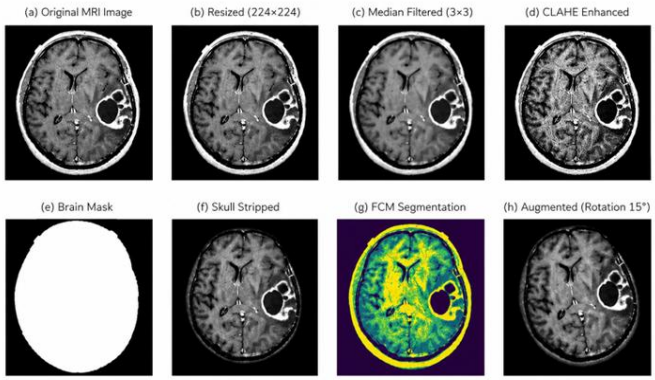


Fig. 3: Preprocessing pipeline for MRI brain tumor images showing resizing, denoising, contrast enhancement, skull stripping, FCM-based segmentation, and augmentation.

The preprocessing included Gaussian denoising, CLAHE-based contrast enhancement, skull stripping, and Fuzzy C-Means (FCM) clustering to suppress irrelevant background information and emphasize tumor regions. Data augmentation was applied only to the training dataset using random rotation (±20°), horizontal and vertical flipping, 10% zooming, and width–height shifting to improve model generalization and reduce overfitting.

C. Deep Feature Extraction

Deep feature extraction was performed using the pretrained ResNet50 architecture initialized with ImageNet weights. The original classification layer was removed, and the convolutional backbone was employed to extract a 2048-dimensional feature vector from each MRI image. ResNet50 utilizes residual learning to improve gradient propagation and training stability, making it highly effective for medical image analysis [21, 33]. The convolution, ReLU activation, residual learning, and batch normalization are expressed by Eqs. (6)-(10).

The proposed framework employs ResNet50 as the backbone network to extract hierarchical feature representations from brain MRI images. The convolutional layers

$$f(x) = \max(0, x) \tag{7}$$

$$Y(i,j) = \sum_{m}\sum_{n} X(i+m, j+n) \cdot K(m,n) + b \tag{6}$$

automatically learn discriminative spatial patterns by applying learnable filters over the input feature maps. The convolution operation is mathematically expressed as

where X denotes the input feature map, K represents the learnable convolution kernel, b is the bias term, and Y(i,j) is the output feature at spatial location (i,j). This operation enables the network to learn local image characteristics such as edges, textures, and tumor boundaries that are essential for accurate brain tumor classification.

where xxx represents the input activation. ReLU effectively suppresses negative activations while preserving positive responses, thereby accelerating convergence and mitigating the vanishing gradient problem during deep network training.

A key characteristic of **ResNet50** is its **residual learning mechanism**, which enables the construction of very deep neural networks without significant degradation in

$$\mathbf{y} = F(x, \{W_i\}) + x \tag{8}$$

performance. Instead of directly learning the desired mapping, the network learns a residual function while preserving the identity mapping through shortcut connections. The residual block is expressed as

where x denotes the input feature map, F(x,{Wi}) represents the residual mapping learned by the stacked convolutional layers with trainable parameters {Wi}, and y is the output of the residual block. These identity shortcut connections facilitate efficient gradient propagation during backpropagation, reduce the vanishing gradient problem, and improve feature reuse, thereby enabling stable optimization of deep convolutional networks.

To further improve training stability and convergence, Batch Normalization (BN) is incorporated after each convolutional layer. Batch normalization normalizes the intermediate feature distributions within each mini-batch and is mathematically represented as

$$\hat{x} = \frac{x - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{9}$$

where x is the input activation, μ_B and σ^2_B denote the mini-batch mean and variance, respectively, and ε is a small positive constant introduced for numerical stability. The normalized activations are subsequently scaled and shifted using learnable parameters as

$$y = \gamma\hat{x} + \beta \tag{10}$$

where γ and β are trainable scaling and shifting parameters. Batch normalization accelerates network convergence, reduces internal covariate shift, improves generalization, and enables more stable optimization during training. These operations collectively allow the proposed ResNet50–MLP framework to learn robust and discriminative representations for accurate multi-class brain tumor MRI classification.

D. MLP-Based Classification

The extracted deep features were classified using a Multi-Layer Perceptron consisting of an input layer with 2048 neurons, two hidden layers containing 512 and 128 neurons, and a four-node Softmax output layer corresponding to the four tumor classes. ReLU activation and a dropout rate of 0.5 were used in the hidden layers to reduce overfitting. The dense layer operation, dropout function, SoftMax activation, and categorical cross-entropy loss [28].

E. Model Training and Evaluation

The proposed ResNet50–MLP framework was trained using the Adam optimizer, which provides adaptive learning rate optimization and has demonstrated excellent convergence behavior in deep learning-based medical image classification tasks [3, 19]. The initial learning rate was set to 1×10^{-4}, with a batch size of 32 and the categorical cross-entropy loss function, which is well suited for multi-class classification problems. To improve the stability of the training process and reduce the possibility of overfitting, early stopping with a patience of 10 epochs was employed.

The maximum number of training epochs was fixed at 50, while the ResNet50 backbone was initially trained using frozen convolutional layers during the warm-up stage before fine-tuning the higher layers to adapt the pretrained ImageNet features to the brain MRI classification task. Furthermore, dropout regularization with a rate of 0.5 was applied after each hidden layer of the MLP classifier to reduce co-adaptation among neurons and improve the model's generalization capability. The complete experimental configuration and hyperparameter settings used during training are summarized in Table 4.

The performance of the proposed framework was evaluated on an independent test set using multiple quantitative evaluation metrics, including Accuracy, Precision, Recall, F1-score, Macro F1-score, Specificity, Cohen's Kappa coefficient, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Accuracy measures the overall classification performance, whereas Precision and Recall quantify the correctness and completeness of predictions for each tumor category.

The F1-score provides a balanced measure of Precision and Recall, particularly useful for evaluating class-wise performance. Specificity measures the model's ability to correctly identify negative samples, while Cohen's Kappa

evaluates the agreement between predicted and true labels beyond chance.

The AUC-ROC further assesses the model's discriminative ability across different decision thresholds. Together, these complementary evaluation metrics provide a comprehensive assessment of the model's classification accuracy, class-wise discrimination capability, reliability, and robustness, enabling a fair comparison with existing state-of-the-art brain tumor classification approaches [3], [18], [19], [21], [23], [25].

Table 4. Experimental Configuration and Hyperparameter Settings

Parameter	Value
Input resolution	224 × 224 × 3
Backbone	ResNet50 (ImageNet pretrained)
Feature dimension	2048
Classifier	2048–512–128–4 (MLP)
Hidden activation	ReLU
Output activation	Softmax
Dropout	0.5
Optimizer	Adam
Learning rate	1×10^{-4}
Batch size	32
Loss function	Categorical Cross-Entropy
Maximum epochs	50 (Early Stopping)
Early stopping	Patience = 10 epochs
Warm-up training	20 epochs
Frozen layers	ResNet50 Layers 1–143 (Stage 1)
Training images	4,569 (65.1%)
Validation images	1,143 (16.3%)
Test images	1,311 (18.7%)
Split strategy	Stratified random sampling
Framework	TensorFlow 2.10 with Keras
Hardware	NVIDIA Tesla T4 GPU (16 GB)

IV. RESULTS AND DISCUSSION

This section presents a comprehensive experimental evaluation of the proposed ResNet50–MLP framework for multi-class brain tumor classification using the publicly available Kaggle Brain Tumor MRI Dataset. To ensure a fair and unbiased comparison, all evaluated models were trained and tested under identical experimental settings, including the same dataset partition, image preprocessing pipeline, data augmentation strategy, optimizer configuration, hyperparameter settings, and evaluation metrics. Performance was assessed using Accuracy, Precision, Recall, F1-score, Specificity, Area Under the Receiver Operating Characteristic Curve (AUC), and Cohen's Kappa coefficient, which are widely accepted metrics for evaluating medical image classification systems [3, 21, 23, 26]. These complementary metrics provide a comprehensive assessment of classification performance by measuring not only overall prediction accuracy but also class-wise discrimination capability, agreement beyond chance, and diagnostic reliability.

A. Comparative Performance Analysis

Table 5 summarizes the comparative performance of the proposed framework against seven widely used deep learning models, including CNN (Baseline), VGG16, InceptionV3, DenseNet121, EfficientNetB0, ResNet50, and MobileNetV3. The proposed ResNet50–MLP framework achieved the highest overall performance among all evaluated models, obtaining an accuracy of 97.64%, precision of 97.61%, recall of 97.58%, F1-score of 97.59%, specificity of 99.21%, AUC of 99.61%, and Cohen's Kappa coefficient of 0.968. Compared with the baseline CNN, the proposed model improved classification accuracy by more than 5%, demonstrating the effectiveness of transfer learning and deep feature representation. It also outperformed well-established transfer learning models such as VGG16, InceptionV3, DenseNet121, EfficientNetB0, standalone ResNet50, and MobileNetV3, indicating that replacing the conventional Softmax classifier with a lightweight Multi-Layer Perceptron (MLP) enhances the discriminative ability of the extracted deep features.

The superior performance of the proposed framework can be attributed to the combined effect of a robust preprocessing pipeline, transfer learning using the pretrained ResNet50 backbone, effective feature learning through the MLP classifier, and regularization techniques such as data augmentation, dropout, and early stopping. These results demonstrate that the proposed hybrid framework provides an accurate, reliable, and computationally efficient solution for automated brain tumor MRI classification and offers strong potential for supporting computer-aided clinical diagnosis [18, 23, 24, 26].

Table 5. Comprehensive Performance Comparison of Brain Tumor Classification

Model	Acc.	Prec.	Recall	F1	Spec.	AUC	Kappa
CNN (Baseline)	0.9184	0.9148	0.9125	0.9136	0.9702	0.9626	0.890
VGG16	0.9321	0.9290	0.9268	0.9279	0.9758	0.9728	0.908
InceptionV3	0.9451	0.9420	0.9405	0.9412	0.9807	0.9802	0.926
DenseNet121	0.9489	0.9460	0.9445	0.9452	0.9820	0.9823	0.931
EfficientNetB0	0.9527	0.9498	0.9480	0.9489	0.9834	0.9847	0.936
ResNet50	0.9580	0.9555	0.9540	0.9547	0.9853	0.9872	0.943
MobileNetV3	0.9634	0.9612	0.9598	0.9605	0.9875	0.9898	0.951
Proposed	0.9764	0.9761	0.9758	0.9759	0.9921	0.9961	0.968

B. Class-wise Performance Analysis

The confusion matrix (Table 6) and the class-wise performance metrics (Table 7) provide a detailed assessment of the diagnostic capability of the proposed ResNet50–MLP framework across all four brain tumor categories. Unlike overall accuracy, which summarizes the model's global performance, the confusion matrix reveals the distribution of

correctly and incorrectly classified samples, enabling a more comprehensive evaluation of classification behavior [3, 21, 23]. The proposed model correctly classified 1,280 out of 1,311 test images, corresponding to a macro F1-score of 97.59%, demonstrating excellent and well-balanced performance across all classes.

The No Tumor class achieved the highest classification performance with an F1-score of 98.15%, precision of 98.03%, and recall of 98.27%, indicating that the proposed framework can effectively distinguish normal brain MRI scans from tumor-affected images while maintaining a very low false-positive rate. Similarly, the Pituitary Tumor class exhibited highly consistent performance, with precision, recall, and F1-score of 97.67%, suggesting that the localized anatomical characteristics of pituitary tumors are effectively captured by the deep feature representation learned by the ResNet50 backbone. The Glioma and Meningioma classes also achieved high classification performance, with F1-scores of 97.32% and 97.23%, respectively. The relatively small number of misclassified samples between these two categories may be attributed to similarities in tumor morphology, intensity distribution, and boundary characteristics commonly observed in MRI images [18, 23, 25].

Furthermore, all classes achieved specificity values above 99% and AUC values greater than 99%, indicating excellent discrimination capability and a very low rate of false-positive predictions. The balanced class-wise performance demonstrates that the proposed framework does not exhibit bias toward any particular tumor category and maintains reliable diagnostic accuracy across all four classes. Overall, the confusion matrix and class-wise evaluation confirm the robustness, stability, and clinical applicability of the proposed framework for automated multi-class brain tumor MRI classification.

Table 6. Confusion Matrix of the Proposed Model (Test Set = 1311 Images)

Actual \ Predicted	Glioma	Meningioma	Pituitary	No Tumor	Total
Glioma	291	4	2	3	300
Meningioma	3	298	3	2	306
Pituitary Tumor	2	2	293	3	300
No Tumor	2	3	2	398	405
Total	298	307	300	406	1311

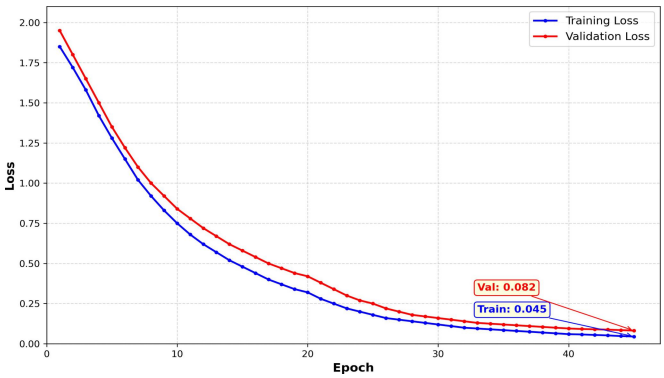
The No Tumor class obtained the highest F1-score (98.15%), while the Pituitary Tumor class achieved balanced precision and recall (97.67%), indicating reliable discrimination among clinically relevant tumor categories.

Table 7. Class-wise Detailed Performance Analysis of the Proposed Model

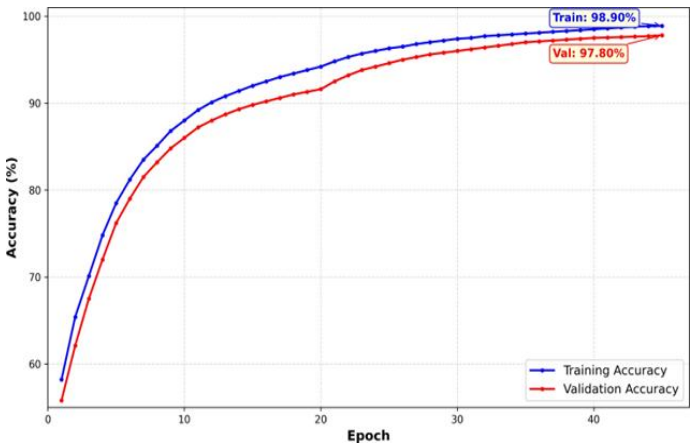
Class	Precision	Recall	F1-Score	Specificity	AUC	Support
Glioma	0.9765	0.9700	0.9732	0.9931	0.9960	300
Meningioma	0.9707	0.9739	0.9723	0.9910	0.9948	306
Pituitary Tumor	0.9767	0.9767	0.9767	0.9931	0.9965	300
No Tumor	0.9803	0.9827	0.9815	0.9912	0.9972	405
Macro Average	0.9761	0.9758	0.9759	0.9921	0.9961	1311

C. Learning Stability

The training and validation curves shown in Fig. 4 demonstrate stable convergence without significant overfitting. As illustrated in Fig. 5, the proposed framework consistently achieved the highest validation accuracy among all evaluated models.



(a) Training and validation accuracy curves.



(b) Training and validation loss curves.

The improved generalization is attributed to transfer learning,

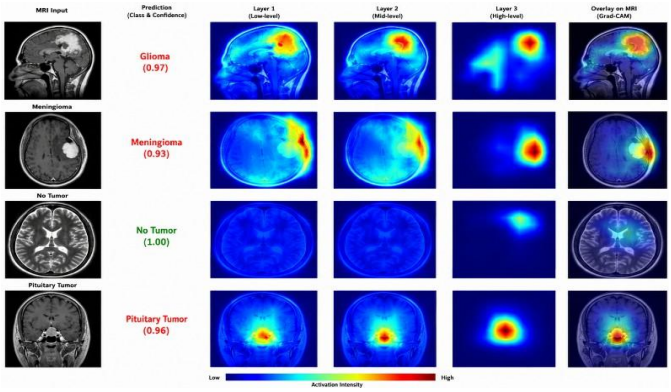


Fig. 4: Grad-CAM Visualization and Spatial Interpretability

data augmentation, dropout regularization, and early stopping. The detailed computational statistics are presented in Table 8.

Table 8. Performance, Training Dynamics, and Computational Efficiency Comparison

Model	Train Acc (%)	Val Acc (%)	Test Acc (%)	Train Loss	Val Loss	Params (M)	Infer. Time (ms)	FLOPs (G)	Model Size (MB)	Peak Memory
CNN (Baseline)	94.20	92.50	91.84	0.172	0.248	3.2	4.8	0.42	12.5	156
VGG16	94.50	93.20	93.21	0.165	0.224	138.4	22.6	15.5	528.0	624
InceptionV3	96.40	95.40	94.51	0.112	0.159	23.8	15.2	5.7	92.1	342
DenseNet121	96.15	95.05	94.89	0.128	0.176	8.0	14.8	2.9	32.6	278
EfficientNet-B0	96.80	95.85	95.27	0.098	0.143	5.3	8.5	0.39	22.1	192
ResNet50	95.90	94.70	95.80	0.142	0.198	25.6	16.3	4.1	98.5	356
MobileNetV3	97.10	96.30	96.34	0.085	0.121	5.4	6.2	0.24	21.8	184
Proposed	98.90	97.80	97.64	0.045	0.082	26.8	17.1	4.3	102.4	368

D. Model Interpretability

To improve clinical reliability, Grad-CAM was employed to visualize the discriminative regions used during prediction. As shown in Fig. 6, the activation maps accurately highlight tumor regions in glioma, meningioma, and pituitary tumor images, whereas diffuse activation is observed for normal MRI scans. These visualizations confirm that the proposed framework focuses on clinically meaningful anatomical regions during classification.

E. Computational Efficiency

Fig. 5 compares the computational complexity of all evaluated models. The proposed framework requires 26.8 million parameters, 4.3 GFLOPs, and an average inference time of 17.1 ms per image. Although slightly more computationally demanding than MobileNetV3 and EfficientNetB0, it achieves superior classification accuracy, AUC, and Kappa while remaining considerably more efficient than VGG16 in terms of model size and computational cost.

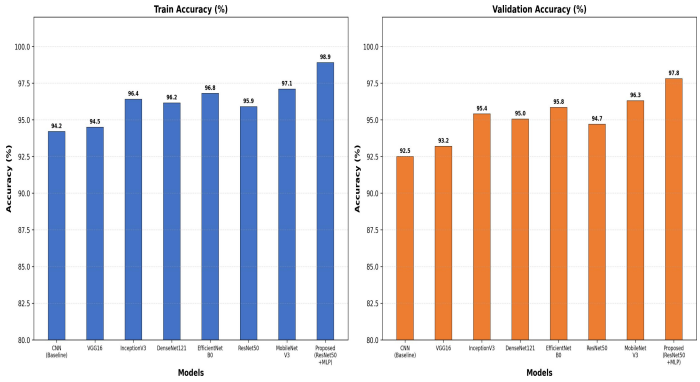


Fig. 5: Comparison of training and validation accuracy across all evaluated models.

V. CONCLUSION AND FUTURE WORK

This study presented a hybrid ResNet50–MLP framework for automated multi-class brain tumor classification using magnetic resonance imaging (MRI). The proposed framework integrates transfer learning-based deep feature extraction through a pretrained ResNet50 backbone with a lightweight Multi-Layer Perceptron (MLP) classifier to accurately distinguish four clinically important categories: glioma, meningioma, pituitary tumor, and no-tumor cases. A comprehensive preprocessing pipeline, including image resizing, normalization, contrast enhancement, skull stripping, Fuzzy C-Means (FCM)-based region enhancement, and data augmentation, was employed to improve image quality and enhance the discriminative capability of the extracted features. Furthermore, transfer learning enabled the model to leverage knowledge learned from large-scale natural image datasets, resulting in improved convergence, better feature representation, and enhanced classification performance on a relatively limited medical imaging dataset.

The proposed framework was evaluated on the publicly available Kaggle Brain Tumor MRI Dataset consisting of 7,023 contrast-enhanced T1-weighted MRI images. Experimental results demonstrated that the proposed model consistently outperformed several widely used deep learning architectures, including CNN, VGG16, InceptionV3, DenseNet121, EfficientNetB0, standalone ResNet50, and MobileNetV3, under identical preprocessing, training, and evaluation conditions. The proposed framework achieved a test accuracy of 97.64%, precision of 97.61%, recall of 97.58%, F1-score of 97.59%, specificity of 99.21%, AUC of 99.61%, and Cohen's Kappa coefficient of 0.968, demonstrating its effectiveness for reliable multi-class brain tumor classification. Class-wise analysis further confirmed balanced performance across all tumor categories, while Grad-CAM visualizations provided meaningful explanations by

highlighting clinically relevant tumor regions, thereby improving the transparency and interpretability of the proposed framework for potential clinical use.

Despite these encouraging results, several limitations should be acknowledged. The experimental evaluation was performed using a single publicly available dataset, which may not fully represent the diversity of MRI acquisition protocols, scanners, and patient populations encountered in real-world clinical practice. In addition, the proposed framework operates on two-dimensional MRI slices rather than complete three-dimensional volumetric scans, limiting its ability to exploit inter-slice spatial information. Moreover, patient-wise cross-validation and external multi-center validation could not be performed because of dataset constraints, which may affect the assessment of the model's generalization capability across different clinical environments [5, 12, 16].

Future work will focus on validating the proposed framework using larger multi-center and multi-institutional MRI datasets to further evaluate its robustness and generalization ability. In addition, incorporating three-dimensional volumetric MRI, multi-modal MRI sequences (such as T1, T2, FLAIR, and contrast-enhanced MRI), and Vision Transformer (ViT)-based architectures may further improve classification accuracy and feature representation. Future research will also investigate lightweight deep learning models for real-time clinical deployment, federated learning for privacy-preserving collaborative training, and advanced explainable artificial intelligence techniques to improve clinician trust and support decision-making in computer-aided brain tumor diagnosis. These enhancements are expected to increase the robustness, scalability, and clinical applicability of automated brain tumor classification systems in routine healthcare practice.

VI. ACKNOWLEDGMENTS

Riyanshu Saini like to thank the School of Engineering and Technology, Shobhit University, Gangoh, India for supporting and encouraging her to work on this research.

VII. REFERENCES

1. Abdulsattar, N.S., Abdulsattar, F.S.: Brain-tumor classification using resnet50 enhanced with se and cbam attention mechanisms. Jordanian Journal of Computers and Information Technology (JJCIT) 12(1), 98–111 (2026). https://doi.org/10.5455/jjcit.71-1764868722
2. Aboussaleh, I., Riffi, J., Mahraz, A.M., Tairi, H.: Brain tumor segmentation based on deep learning's feature representation. Journal of Imaging 7(12), 269 (2021)
3. Agarwal, M., Rani, G., Kumar, A., Kumar, P., Manikandan, R., Gandomi, A.H.: Deep learning for enhanced brain tumor detection and classification. Results in Engineering 22, 102117 (2024)
4. Asif, R.N., Naseem, M.T., Ahmad, M., Mazhar, T., Khan, M.A., Khan, M.A.,Al-Rasheed, A., Hamam, H.: Brain tumor detection empowered with ensemble deep learning approaches from mri scan images. Scientific reports 15(1), 15002 (2025)
5. Chowdhury, M.F.K., Ferdous, J.: Brain tumor classification in mri images: A computationally efficient convolutional neural network. In: 2025 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON). pp. 633–638. IEEE (2025)
6. Dheepak, G., Vaishali, D., et al.: Brain tumor classification: a novel approach integrating glcm, lbp and composite features. Frontiers in oncology 13, 1248452 (2024)
7. Ghasemi, R., Islam, N., Bayat, S., Shabir, M., Rahman, S., Amin, F., de la Torre, I., Castilla, A´.K., Garc´ıa, D.L.R.V.: Detection and classification of brain tumor using a hybrid learning model in ct scan images. Scientific Reports 15(1), 35085 (2025)
8. Guluwadi, S., et al.: Enhancing brain tumor detection in mri images through explainable ai using grad-cam with resnet 50. BMC medical imaging 24(1), 1–19 (2024)
9. Hosny, K.M., Mohammed, M.A.: Explainable ai and vision transformers for detection and classification of brain tumor: a comprehensive survey. Artificial Intelligence Review 58(9), 259 (2025)
10. Iftikhar, S., Anjum, N., Siddiqui, A.B., Ur Rehman, M., Ramzan, N.: Explainable cnn for brain tumor detection and classification through xai based key features identification. Brain Informatics 12(1), 10 (2025)
11. JS, N., et al.: Brain tumor segmentation using multi-scale attention u-net with efficientnetb4 encoder for enhanced mri analysis. Scientific Reports 15(1), 1–20 (2025)
12. Jyothi, P., Singh, A.R.: Deep learning models and traditional automated techniques for brain tumor segmentation in mri: a review. Artificial intelligence review 56(4), 2923–2969 (2023)
13. Kaplan, K., Kaya, Y., Kuncan, M., Ertunc¸, H.M.: Brain tumor classification using modified local binary patterns (lbp) feature extraction methods. Medical hypotheses 139, 109696 (2020)
14. Khan, S.U.R., Zhao, M., Li, Y.: Detection of mri brain tumor using residual skip block based modified mobilenet model. Cluster Computing 28(4), 248 (2025)
15. Koc, A.B., Akgun, D.: A review of u-net based deep learning frameworks for mri-based brain tumor segmentation. Diagnostics 16(4), 506 (2026)
16. Li, Z., Dib, O.: Empowering brain tumor diagnosis through explainable deep learning. Machine Learning and Knowledge Extraction 6(4), 2248–2281 (2024)
17. Maqsood, R., Abid, F., Rasheed, J., Osman, O., Alsubai, S.: Optimal res-unet architecture with deep supervision for tumor segmentation. Frontiers in Medicine 12, 1593016 (2025)
18. Mehrdad, V., Talebzadeh, R., Fazaeli, N.: Brain tumor classification using optimized resnet50 with dynamic precision optimization for enhanced speed and diagnostic accuracy. Scientific Reports (2026)

19. Mohan, P., Ramkumar, G.: An inception v3 based glioma brain tumor detection in mri images. In: 2024 5th International Conference on Circuits, Control, Communication and Computing (I4C). pp. 560–564. IEEE (2024)
20. Napa, K.K., Murugan, S., Murugan, J.S., Jayanthi, A.: Brain-meta: A reproducible cnn–vision transformer meta-ensemble pipeline for explainable brain tumor classifi-cation. MethodsX 16, 103769 (2026)
21. Nayak, D.R., Padhy, N., Mallick, P.K., Zymbler, M., Kumar, S.: Brain tumor classification using dense efficient-net. Axioms 11(1), 34 (2022)
22. Nickparvar, M.: Brain tumor mri dataset (2021), https://www.kaggle.com/ datasets/masoudnickparvar/brain-tumor-mri-dataset, accessed: 24 May 2026
23. Pillai, R., Sharma, A., Sharma, N., Gupta, R.: Brain tumor classification using vgg 16, resnet50, and inception v3 transfer learning models. In: 2023 2nd International Conference for Innovation in Technology (INOCON). pp. 1–5. IEEE (2023)
24. Prabhas, K.S., Basem, A., Lakshmi, L., Talha, A., Mohammed, S.H., Khan, M.I., Khedher, N.B.: A deep learning framework for brain tumor detection using cnns and transfer learning on mri scans. Systems and Soft Computing p. 200389 (2025)
25. Putri, F.A.P.R., Tanjung, J.P., Dharshini, N.P.: Implementation of transfer learning on cnn using densenet121 and resnet50 for brain tumor classification. Journal of Informatics and Telecommunication Engineering 8(2) (2025). https://doi.org/ 10.31289/jite.v8i2.13952
26. Rahman, A., Hayat, M., Iqbal, N., Alarfaj, F.K., Alkhalaf, S., Alturise, F.: Enhanced mri brain tumor detection using deep learning in conjunction with explainable ai shap based diverse and multi feature analysis. Scientific Reports 15(1), 29411 (2025)
27. Rahman, T., Islam, M.S.: Mri brain tumor detection and classification using parallel deep convolutional neural networks. Measurement: Sensors 26, 100694 (2023)
28. Rajendran, S., Rajagopal, S.K., Thanarajan, T., Shankar, K., Kumar, S., Alsubaie, N.M., Ishak, M.K., Mostafa, S.M.: Automated segmentation of brain tumor mri images using deep learning. IEEE Access 11, 64758–64768 (2023)