Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

A Dual-Backbone CNN Framework with Gated Cross-Attention Fusion and Mixup Focal Loss for Robust and Explainable Classification of Brain Tumors

Cmak Zeelan Basha[1], Prasanth Yalla[2]

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.

E-mail: [1]zeelanbashaklu@gmail.com, [2]prasanthyalla@kluniversity.in

Abstract

Diagnosing and classifying brain tumors based on MRI scans is a difficult and complicated process due to differences in tumor types, low contrast, and class imbalance. In this study, we propose a framework that uses a dual-backbone convolutional neural network (CNN) featuring EfficientNetV2-S and ConvNeXt-Tiny for multi-scale feature extraction and classification. The gated cross-attention fusion module enables adaptive bidirectional interaction among the feature representations. In addition, Mixup regularization, focal loss, and label smoothing improve robustness, calibration, and class imbalance handling. The model employs a two-stage transfer learning framework, supported by exponential moving average (EMA)-based stabilization. The framework achieves 95.73% classification accuracy, a macro-F1 score of 95.7%, and a macro-average ROC–AUC of 0.9967 when evaluated on a multi-class MRI brain tumor dataset containing 7,023 images. The proposed fusion and training strategy demonstrate effectiveness through comparative and ablation analyses. Furthermore, Grad-CAM visualizations show that the model focuses on tumor-related regions. The proposed dual-backbone gated cross-attention framework for brain tumor classification demonstrates both high performance and good interpretability. The model also shows potential for clinical decision-support systems but requires additional multi-center validation.

Keywords: Brain Tumor Classification, MRI, Dual-Backbone CNN, Cross-Attention, Focal Loss, Grad-CAM.

1. Introduction

Artificial intelligence has improved clinical neuroimaging. Brain tumors are abnormal cellular growths occurring within the brain. They are serious neurological disorders due to the brain's complex structure and the surgical risks involved. The gold standard for the diagnosis of these malignancies is Magnetic Resonance Imaging (MRI), because it provides high soft-tissue contrast and is completely free of ionizing radiation [1]. Manual analysis of multimodal MRI scans, including T1-weighted (T1-w), T2-weighted (T2-w), and fluid-attenuated inversion recovery (FLAIR), is time-consuming and may lead to human error and subjective variability [2].

DOI: https://doi.org/10.36548/jiip.2026.2.014
Received: 18.03.2026, received in revised form: 22.04.2026, accepted: 09.05.2026, published: 22.05.2026

Classification of brain tumors using automation is rather complex owing to the variability among various types of tumors. For example, gliomas, meningiomas, and pituitary tumors have comparable intensity profiles and geometrical features that make it difficult to discriminate between them using computer vision techniques [3]. Moreover, the edges of the tumor core with respect to peritumoral edema and normal brain tissue are often vague, particularly on low-field MR images. These imaging artifacts, combined with class imbalance, often reduce the performance and generalization capability of conventional machine learning models [4].

During the last decade, CNNs have become the state-of-the-art technique for the classification of medical images [5]. Although they are effective in learning hierarchical feature representations, single-backbone architectures often struggle to balance local textures against global context [6]. For instance, deep networks can learn complex patterns, but they can suffer from vanishing gradients or over-parameterization when trained on small medical databases [7]. Multi-scale architectures have thus become essential for aggregating global semantic information with fine-grained local features to maintain diagnostic quality [8].

This paper proposes a robust dual-backbone CNN model for addressing these clinical challenges. We propose an improved model based on EfficientNetV2-S and ConvNeXt-Tiny, trained on a brain MRI dataset. We introduce a Gated Cross-Attention Fusion method to integrate complementary feature representations, guiding the model toward informative spatial regions and suppressing irrelevant noise. We extend the work in [9] by proposing a Mixup Focal Loss that enhances robustness against class imbalance and improves calibration. Finally, we visualize the model predictions using Grad-CAM to enhance interpretability and clinical trust in the model predictions by highlighting the regions responsible for classification [10].

Although much progress has been made using multi-backbone and attention-based models, existing methods are still limited in several ways. Most of the fusion strategies rely on static operations like concatenation or addition which do not capture inter-feature dependencies. In addition, many attention mechanisms are unidirectional, limiting the effective exchange of information between the feature streams. Moreover, little research has jointly addressed class imbalance, probability calibration, and interpretability in a single framework. As a result of these challenges, the robustness and clinical applicability of existing models of brain tumors have been undermined.

The key contributions of this work are as follows:

1. A combination of EfficientNetV2-S and ConvNeXt-Tiny backbones for multi-scale feature extraction.
2. A new gated cross-attention fusion mechanism to facilitate adaptive bidirectional feature interaction.
3. Mixup, focal loss, and label smoothing are combined for enhanced robustness and calibration.
4. A transfer learning strategy with EMA stabilization to stabilize convergence.
5. Comprehensive evaluation through ablation, calibration, and interpretability analyses.

2. Related Work

The automatic brain tumor classification started with classical image processing and hand-crafted features. Previous works have used methods like the Gray-Level Co-occurrence Matrix (GLCM) and Gabor filter for texture extraction. Afterwards, they have used Support Vector Machine (SVM) or K-Nearest Neighbor (KNN) for classification [11]. Although these methods offered a baseline, they were heavily sensitive to noise and did not generalize well across different MRI scanners and protocols [12]. The transition to Deep Learning (DL) was a major milestone as these architectures enabled end-to-end feature learning from raw pixel data instead of engineering them manually [14]. The recent literature has shown an increased use of ensemble and multi-stream networks for enhancing classification accuracy. Recent studies have investigated the integration of different architectures, such as ResNet and Inception to leverage their diverse receptive fields [15]. However, majority voting or feature concatenation are relatively simple ensemble methods that may not fully capture complex interactions between different feature maps [16]. This limitation has motivated the development of attention mechanisms, inspired by human visual perception that enable the network to focus selectively on significant pathological features while avoiding healthy anatomical structures [17], [18].

Apart from architectural design, there is a focus on the development of methods to tackle data-centric issues such as class imbalance and overconfidence. The Focal Loss function allows models to focus more attention on difficult samples, as it is important in rare tumor subtypes [19]. Likewise, data augmentation techniques such as Mixup and CutMix are capable of regularizing models through the generation of artificial training instances, which averts overfitting in medical datasets with restricted instances [20]. These methods are increasingly used in transfer learning, where models trained on large-scale datasets (e.g., ImageNet) are retrained or fine-tuned for domain-specific medical tasks [21]. Lastly, Explainable AI (XAI) has gained popularity in the medical imaging community. Layer-wise Relevance Propagation (LRP) and Gradient-weighted Class Activation Mapping (Grad-CAM) are widely used techniques for visualizing the decision-making process of CNNs [22]. Recent analysis has indeed shown that the heatmaps produced by such tools are significantly correlated with the physiological location of tumors according to expert radiologists [23]. In this research, we combine both a gated attention mechanism and an individual loss function, which has the ability to achieve explainable results along with competitive classification accuracy.

Recently, ViT-based and Swin Transformer-based architectures in classifying medical images can better capture global features; however, they may require large datasets and high computational resources, which limit their translational applicability in clinics. In contrast, our framework is competitive and computationally inexpensive.

Table 1. Comparison of Fusion Strategies and Attention Mechanisms

Method	Fusion Type	Attention	Bidirectional	Gating	Calibration
CNN and Attention	Static	Single	X	X	X
Dual Backbone	Concatenation	X	X	X	X
Transformer-based	Self-attention	✓	X	X	X
Proposed Method	Dynamic Fusion	Cross-Attention	✓	✓	✓

Table 1 presents our method in comparison with other methods concerning fusion strategy, attention mechanisms, and learning characteristics. Approaches based on traditional CNNs rely heavily on static fusion techniques, such as concatenation, and are unable to capture complex interdependencies. Conversely, models based on transformers engage self-attention in a single feature stream, while neglecting cross-backbone interactions in the modeling

process. In contrast, our framework proposes dynamic fusion using bidirectional cross-attention so that useful information can be exchanged between feature encoders. In addition, a learnable gating mechanism is incorporated which adaptively regulates feature contribution. Furthermore, Mixup regularization, focal loss, and label smoothing are used to enhance robustness and calibration. In summary, the proposed method stands out from previous research due to its unique combination of bidirectional attention and adaptive gating with a calibration-aware optimizer, leading to improved performance and generalization capability on image classes.

3. Problem Statement

We have a labeled dataset $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^{N}$, where the input $X_i \in \mathbb{R}^{H\times W\times C}$ represents the brain MRI scan and the corresponding class label, which can be glioma, meningioma, pituitary tumor, or no tumor represented by $y_i \in \{1,2, \dots, K\}$ where K=4. Hence, the goal is to learn a discriminative and interpretable model $f_\theta: \mathbb{R}^{H\times W\times C} \rightarrow [0,1]^K$, such that provides an accurate posterior estimate of class probability $P(y \mid X)$ accurately. The poor contrast between the tumor and surrounding brain tissue in MRI images presents a number of challenges to this task. To mitigate these problems, we formulate the problem as optimize the model parameters θ that minimize the expected classification loss $\mathbb{E}_{(X,y)\sim\mathcal{D}}[\mathcal{L}(y, f_\theta(X))]$, while being robust to data variation and providing clinically relevant visual explanations. The multi-class brain tumor classification is a non-trivial learning problem that requires the model to combine the complementary representations extracted from several feature encoders while being robust to noisy labels and few-shot samples.

4. Methodology

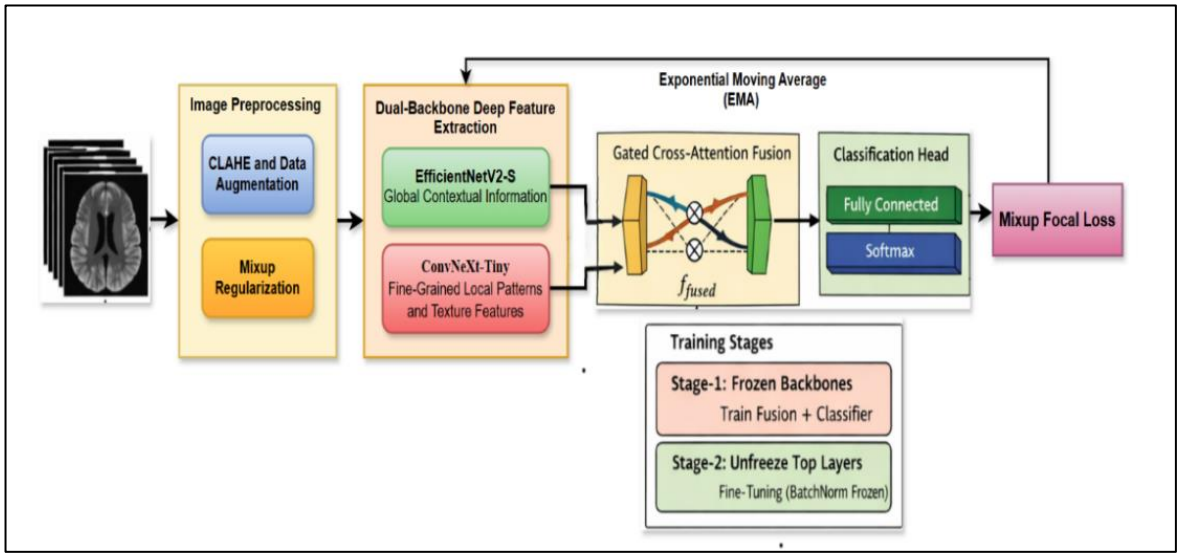


Figure 1. Architecture Diagram of the Proposed Model

The brain tumor classification pipeline was designed to be accurate, robust, and clinically reliable. It incorporates contrast enhancement, dual backbone feature extraction, gated cross-attention fusion, and customized training. Figure 1 shows the architecture of the proposed framework. The pipeline starts with CLAHE-based preprocessing and data augmentation, followed by a dual backbone using EfficientNetV2-S and ConvNeXt-Tiny for

feature extraction. A gated cross-attention mechanism is used to fuse the extracted features, and a fully connected layer trained with Mixup focal loss is utilized for final classification.

4.1 Image Preprocessing Using CLAHE and Data Augmentation

Automatic boundary detection in MRI images is challenging due to low contrast, intensity non-uniformity, and scanner-related variations. To address these challenges, Contrast Limited Adaptive Histogram Equalization (CLAHE) is used to improve local contrast while limiting noise amplification. For a given input MR image, the image is first converted to HSV color space. Let X_i^v be the value (intensity) channel. CLAHE is applied to small local regions and shifts the intensity as in Eq (1).

$$X_i^{v'} = \text{CLAHE}(X_i^v; \alpha, \tau), \tag{1}$$

where α is the contrast clip limit and τ is the tile grid size. The modified luma channel is then appended with the original chrominance channels to reconstruct the RGB image. A sample image after CLAHE preprocessing is shown in Figure 2 and intensity before preprocessing and after is shown in Figure 3.

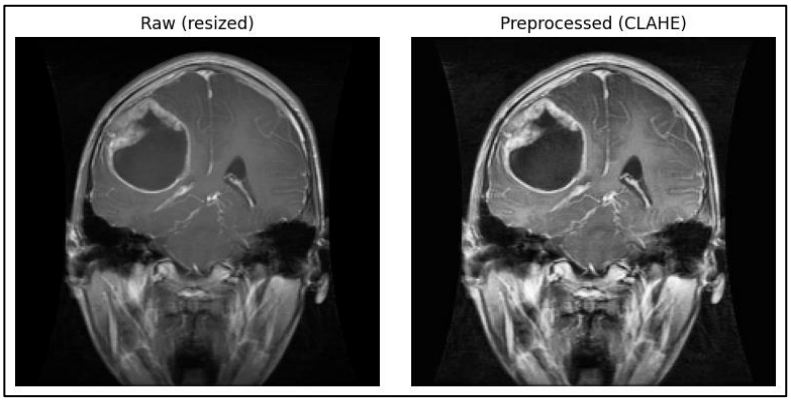


Figure 2. Sample Preprocessing Image

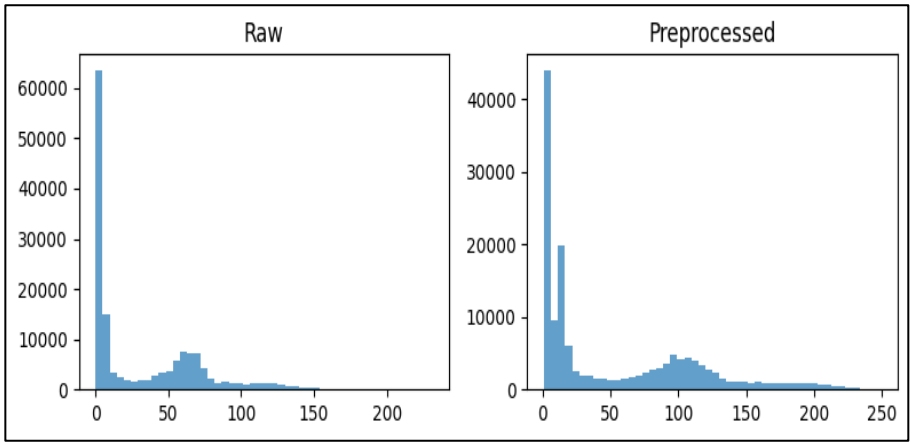


Figure 3. Intensities of an Image before and after Preprocessing

The augmented image is resized to a fixed spatial resolution of 224×224 pixels and normalized by ImageNet statistics for improved transfer learning performance, as in Eq (2).

$$X_i' = \frac{X_i - \mu_{\text{ImageNet}}}{\sigma_{\text{ImageNet}}}. \tag{2}$$

To enhance generalization and overcome overfitting due to inadequate medical imaging data, online data augmentation techniques are implemented during training. Our stochastic augmentation operator $\mathcal{T}(\cdot)$ involving horizontal and vertical flipping, random brightness variation and contrast perturbation consists of $X_i^{\text{aug}} = \mathcal{T}(X_i')$. The sample image (and augmented views) is depicted in Figure 4.

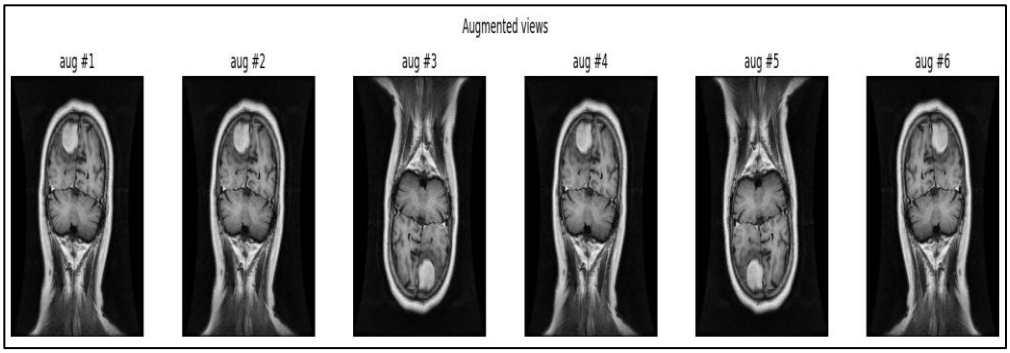


Figure 4. Augmented Views

4.2 Mixup Regularization

To further regularize the model and encourage smoother decision boundaries, Mixup is incorporated after preprocessing and standard augmentation but prior to feature extraction. As shown in Figure 5, given two randomly sampled training instances $\left(X_i^{\text{aug}}, y_i\right)$and $\left(X_j^{\text{aug}}, y_j\right)$, Mixup generates a virtual training sample as in Eq. (3).

$$\tilde{X} = \lambda X_i^{\text{aug}} + (1-\lambda)X_j^{\text{aug}}, \tilde{y} = \lambda y_i + (1-\lambda)y_j, \tag{3}$$

where $\lambda \sim \text{Beta}(\alpha, \alpha)$and $\alpha > 0$ controls the interpolation strength. This formulation produces soft labels $\tilde{y} \in [0,1]^K$, encouraging the model to learn linear behavior between samples, thereby reducing overfitting, improving robustness to label noise, and enhancing probability calibration.

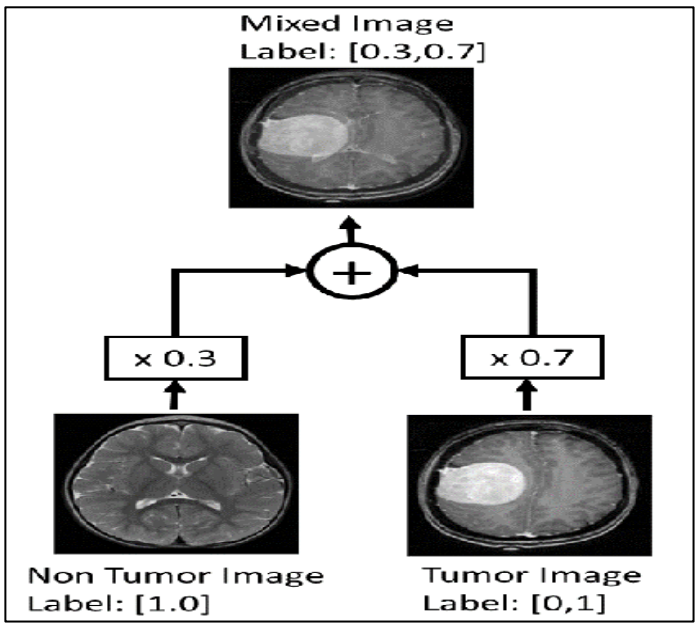


Figure 5. Mixup Regularization [24]

4.3 Dual-Backbone Deep Feature Extraction

To learn complementary tumor features, a dual-backbone convolutional neural network is adopted with EfficientNetV2-S and ConvNeXt-Tiny. Each backbone independently extracts feature representations from the same mixed input $\tilde{X}$. EfficientNetV2-S is further scaled using compound scaling and uses depth-wise separable convolutions to capture global context efficiently. The resulting feature map is defined as in Eq (4).

$$F_1 = \phi_1(\tilde{X}), F_1 \in \mathbb{R}^{H_1 \times W_1 \times D_1}, \tag{4}$$

where $\phi_1(\cdot)$denotes the EfficientNetV2-S feature extraction function.

Meanwhile, ConvNeXt-Tiny uses upgraded convolutional blocks that are similar to the Transformer architecture, allowing it to capture more fine-grained local patterns and texture-based information. Its feature map can be represented as in Eq (5).

$$F_2 = \phi_2(\tilde{X}), F_2 \in \mathbb{R}^{H_2 \times W_2 \times D_2}, \tag{5}$$

where $\phi_2(\cdot)$represents the ConvNeXt-Tiny feature extractor. To obtain compact and translation-invariant representations, global average pooling (GAP) is applied:

$$f_1 = \text{GAP}(F_1), f_2 = \text{GAP}(F_2), \tag{6}$$

resulting in fixed-length feature vectors $f_1, f_2 \in \mathbb{R}^d$. This dual representation enables the model to jointly exploit global tumor context and fine-grained structural details.

4.4 Gated Cross-Attention Feature Fusion

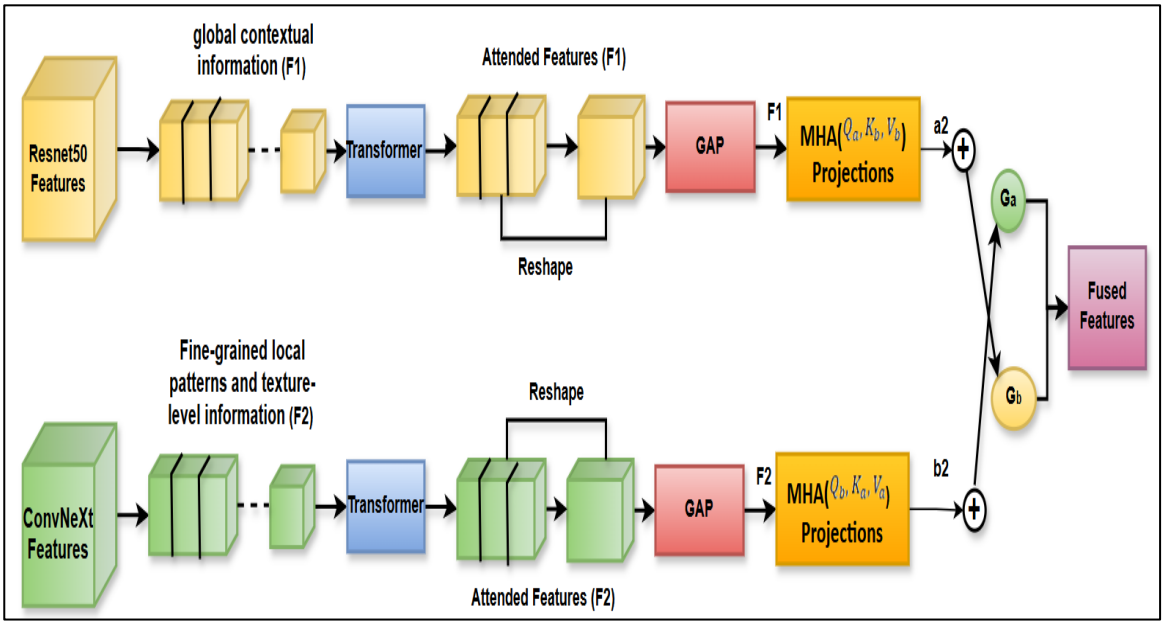


Figure 6. Gated Cross-Attention Feature Fusion

The gating mechanism functions as an adaptive feature selector that dynamically balances the contribution of cross-attended features from both backbones. Unlike plain concatenation, cross-attention enables one feature stream to selectively attend to informative parts of the other to enhance inter-backbone dependency learning as illustrated in Figure 6.

Given the pooled feature vectors $f_1, f_2 \in \mathbb{R}^d$, query, key, and value projections are computed as in Eqs. (7) and (8).

$$Q_a = W_Q^a f_1, K_b = W_K^b f_2, V_b = W_V^b f_2 \tag{7}$$

$$Q_b = W_Q^b f_2, K_a = W_K^a f_1, V_a = W_V^a f_1. \tag{8}$$

where $\odot$ denotes element-wise multiplication. The sigmoid-based gating function learns importance weights conditioned on the joint feature representation, enabling the model to emphasize discriminative tumor regions while suppressing redundant or noisy activations. This is particularly useful in MRI analysis where irrelevant background structures can affect classification performance. The gating mechanism we propose differs from fixed fusion methods like concatenation, and allows features to be modulated in a way that depends on the input. As a result, features become more informative, and thereby improving the overall robustness of the model.

The gating variables g_a and g_b regulate the contribution of cross-attended features from each backbone, and ensuring adaptive feature selection. This dual gating design ensures balanced bidirectional information exchange rather than favoring a single backbone. The gate function is conceptually seen as working like a selection operator that can filter out important features from less important ones depending on the nature of the MRI being used. This has helped ensure that the fusion is data-driven. This design differentiates the proposed method from existing fusion approaches by enabling dynamic, input-dependent feature selection through bidirectional attention and gating, which is not achieved by conventional concatenation or single-direction attention mechanisms.

4.5 Focal Loss with Label Smoothing

Brain tumor datasets often suffer from class imbalance, which can bias the learning process toward majority classes. To mitigate this problem, we use focal loss with label smoothing. Focal loss is expressed mathematically in Equation (14).

$$\mathcal{L}_{\text{focal}} = -\alpha(1 - p_t)^{\gamma} \log(p_t), \tag{14}$$

The predicted probability of the true class is p_t, α is the class-balancing factor, and γ is the focusing parameter that emphasizes hard-to-classify samples.

The ground-truth labels are optimized using label smoothing to enhance generalization and calibration:

$$y_i^* = (1 - \varepsilon) y_i + \frac{\varepsilon}{K}, \tag{15}$$

where $\varepsilon$is the smoothing factor and $K$is the number of classes. This approach helps avoid over-confident predictions and improves the calibration of the model.

Essentially, the smoothed labels y_i^* are used to compute p_t, which is the effective target probability used for the focal loss. This helps prevent the model from assigning the full probability mass to just one class, even for a correctly classified sample.

Combining focal loss and label smoothing offers distinct advantages. While focal loss focuses on harder predictions, label smoothing balances predictions and curtails overconfidence. Together, they induce stable optimizations and greater generalization, and especially so with unbalanced medical datasets, produce better calibrated probabilities.

It should be noted that the proposed loss formulation is fundamentally different from Dice loss or class-balanced loss, which are commonly used. Dice loss is designed for segmentation problems and focuses on maximizing the spatial overlap between predicted and ground truth regions so it is not suitable for the image-level classification problems considered in this work. In contrast, class-balanced loss reweights samples according to the effective number of instances to alleviate class imbalance, which uses static weighting without explicitly focusing on hard samples.

The proposed Mixup focal loss with label smoothing mitigates over-confidence by using soft labels through Mixup, jointly optimizing on misclassified examples using focal loss, and further regularizing with label smoothing. It improves generalization and stability, and it is also better calibrated than existing loss functions.

4.6 Two-Stage Transfer Learning with Exponential Moving Average

The proposed framework is optimized using a two-stage transfer learning strategy to ensure stable convergence and effective adaptation to MRI brain tumor data. Let $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^{N}$denote the training dataset. During each iteration, images are preprocessed, augmented, and mixed using Mixup to generate soft-labeled samples $(\tilde{X}_i, \tilde{y}_i)$. In the first stage, both backbone networks are frozen, and only the gated fusion and classification layers are optimized:

$$\theta^{(1)} = \arg\min_{\theta} \sum_{i=1}^{N} \mathcal{L}_{\text{focal}}\left(y_i^*, f_\theta(X_i)\right). \tag{16}$$

In Stage-2 fine-tuning, the top 20% of layers from both EfficientNetV2-S and ConvNeXt-Tiny backbones were unfrozen, excluding batch normalization layers to maintain stable statistics. This corresponds to the final convolutional blocks responsible for high-level semantic feature extraction. The earlier layers remained frozen to preserve general visual representations learned from ImageNet.

$$\theta^{(2)} = \arg\min_{\theta,\theta_b} \sum_{i=1}^{N} \mathcal{L}_{\text{focal}}\left(y_i^*, f_{\theta,\theta_b}(X_i)\right). \tag{17}$$

To stabilize training and reduce parameter oscillations, an exponential moving average (EMA) of model weights is maintained:

$$\theta_{\text{EMA}}^{(t)} = \beta\theta_{\text{EMA}}^{(t-1)} + (1-\beta)\theta^{(t)}, \tag{18}$$

where $\beta \in [0.9, 0.999]$is the decay factor. EMA-smoothed weights are incorporated during evaluation to enhance generalization performance. Algorithm 1 summarizes the entire training process, including preprocessing, Mixup generation, feature extraction, gated fusion, computing loss, updating parameters, and EMA stabilizing.

Algorithm 1: Training Procedure of the Proposed Dual-Backbone Gated Cross-Attention Network

Input: Training dataset $\mathcal{D}$, number of classes K, learning rates η_1, η_2, Mixup parameter α, focal loss parameters (γ, α_k)

1. Initialize EfficientNetV2-S and ConvNeXt-Tiny backbones with ImageNet pretrained weights
2. Initialize gated cross-attention fusion module and classification head
3. Freeze all backbone parameters

Stage-1: Fusion and Classifier Training

4. for epoch $=$ 1to E_1do
5. Sample mini-batch $\{(X_i, y_i)\}_{i=1}^{B}$from $\mathcal{D}$
6. Apply preprocessing and data augmentation
7. Generate Mixup samples: $\tilde{X} = \lambda X_i + (1-\lambda)X_j, \tilde{y} = \lambda y_i + (1-\lambda)y_j$
8. Extract features $f_1 = \phi_1(\tilde{X})$, $f_2 = \phi_2(\tilde{X})$
9. Fuse features using gated cross-attention to obtain f_{fused}
10. Predict class probabilities $\hat{y} = \text{softmax}(W f_{\text{fused}})$
11. Compute Mixup focal loss $\mathcal{L}_{\text{focal}}(\tilde{y}, \hat{y})$
12. Update fusion and classifier parameters using Adam optimizer
13. end for

Stage-2: Selective Fine-Tuning

14. Unfreeze top layers of both backbones (excluding BatchNorm layers)
15. for epoch $=$ 1to E_2do
16. Repeat Steps 5–11
17. Update all trainable parameters using Adam optimizer with learning rate η_2
18. end for
19. Return final optimized parameters θ^*

Output: Optimized model parameters θ^*

5. Results

5.1 Dataset Description

The experiments in this study used a composite Brain Tumor MRI dataset [25] compiled from three open-source datasets: Figshare, SARTAJ, and Br35H. The combined dataset includes 7,023 T1-weighted, contrast-enhanced MRI images categorized into four classes: glioma, meningioma, pituitary tumor, and no-tumor, as shown in Figure 7. An unbiased evaluation of the results was accomplished by dividing the dataset into training, validation, and testing sets in a ratio of 80:10:10, resulting in 5,618 training images, 702 validation images, and 703 testing images. The model training and evaluation bias were minimized by maintaining the ratio of each class using a stratified splitting method. Table 2 shows the distribution of the training, validation, and test sets by class. Using a dataset from multiple sources resulted in imaging with varied conditions, and scanner types, and resolutions, making the model more robust and flexible.

Table 2. Class-wise Dataset Distribution after Stratified Split

Class	Training	Validation	Testing	Total
Glioma	1297	162	162	1621
Meningioma	1316	164	165	1645
No Tumor	1600	200	200	2000
Pituitary	1405	176	176	1757
Total	5618	702	703	7023

Through the use of three-way partitioning of data (train, validate, test), model evaluation on fresh data can be achieved, thus reducing overfitting and improving the reliability of the results produced.

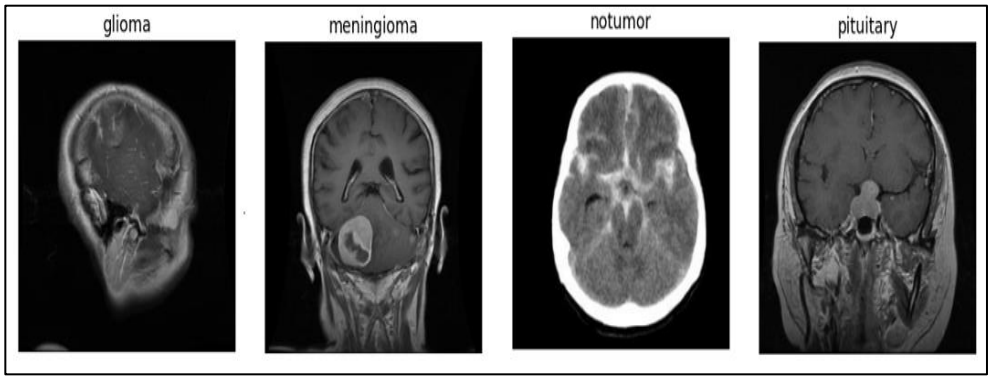


Figure 7. Representative MRI Samples from the Dataset

5.2 Experimental Setup

Experiments were conducted using the Python programming language with TensorFlow (version v2.x) and Keras. For image preprocessing and augmentation, we used OpenCV and TensorFlow libraries. The evaluation process was conducted using Scikit-learn, Matplotlib, and Seaborn libraries. It is important to note that a CUDA-enabled NVIDIA GPU was employed in all experiments to speed up training. Image rescaling included resizing all MRI images to 224x224 pixels in accordance with pre-trained ImageNet models. Such dimensions provide a good balance between computational power restrictions and the amount of structural information related to the tumor. Batch size was chosen empirically to provide a tradeoff between convergence stability and generalization performance.

The Adam optimizer trained the model with a two-stage strategy. In this approach the backbone layers were frozen during initial training, followed by selective fine-tuning. Furthermore, early stopping and learning rate scheduling were employed to prevent overfitting and ensure stable convergence. Several strategies including Mixup, focal loss, and label smoothing were used to improve the models' robustness, handle class imbalance, and improve model calibration. Finally, to improve training stability and generalization, the Exponential Moving Average (EMA) strategy was employed. Each experiment was repeated using fixed random seeds to establish reproducibility. Table 3 summarizes the hyperparameters used in the experiments, and their sensitivity analysis is provided in Section 5.8.

Table 3. Summary of Hyperparameter Settings

Parameter	Symbol	Value	Description
Input Size	–	224 × 224	Image resolution
Batch Size	–	32	Selected for stable convergence
Learning Rate (Stage-1)	–	3×10^{-4}	Fusion and classifier training

Learning Rate (Stage-2)	–	1×10^{-5}	Fine-tuning of backbones
Mixup Parameter	α_{mixup}	0.2	Controls interpolation strength
Focal Loss Gamma	γ	2	Focuses on hard samples
Class Balance Factor	α_{focal}	Inverse frequency	Handles class imbalance
Label Smoothing	ε	0.05	Reduces overconfidence
EMA Decay	–	0.999	Stabilizes training
Epochs (Stage-1 / Stage-2)	–	20 / 10	Training duration

The hyperparameters of the focal loss were tuned manually. The values of focusing parameter, class-balancing factor and label smoothing parameter were set to γ in [1, 3], α in [0.25, 0.75], ϵ in [0.05, 0.1]. The optimal performance in terms of validation accuracy was attained with γ=2, α=0.5, and ϵ=0.1.

5.3 Training Convergence Analysis

The optimization convergence characteristics of the suggested framework are demonstrated via the validation and training accuracy and loss curves of both training stages. According to Figure 8, fast and stable convergence was observed during Stage 1 training with frozen backbone networks. The rise in training accuracy and the accompanying validation accuracy show the learning of the gated cross-attention fusion and classification layers. Validation accuracy follows a closely matching curve. At the same time, the training and validation losses decrease continuously without noticeable oscillations, indicating stable optimization and no overfitting in the initial phases.

Further improvements observed during Stage-2 fine-tuning are shown in Figure 9. Unfreezing the top layers of both backbones makes overall task-specific feature adaptation easier, which leads to a measurable refinement of performance. The training accuracy reaches around 98-99 %, while the validation accuracy is around 95-96%, indicating improved generalization performance. Importantly, the validation loss continues to decrease and follows the training loss closely, suggesting that fine-tuning increases the discriminative capacity without overfitting. The reason for the smooth convergence over both stages can be attributed to the combination of Mixup regularization, focal loss with label smoothing and EMA based weight stabilization to promote stable gradients, reduce variance and define well-calibrated learning dynamics.

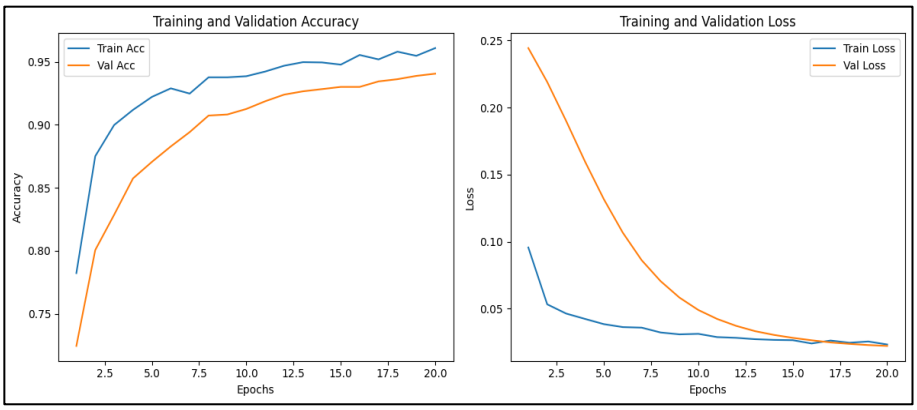


Figure 8. Training and Validation Accuracy and Loss Curves at Stage 1

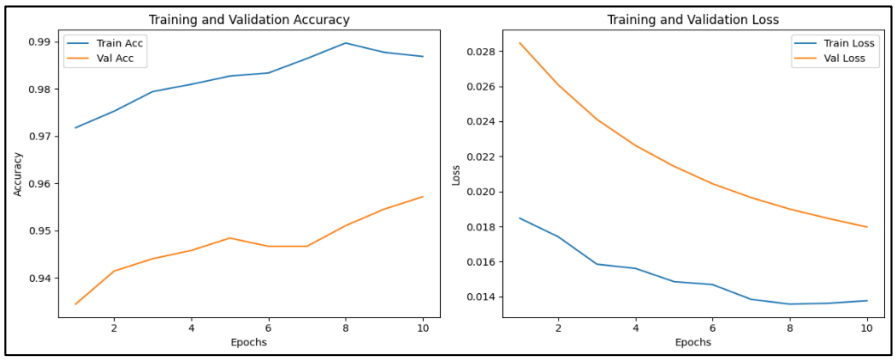


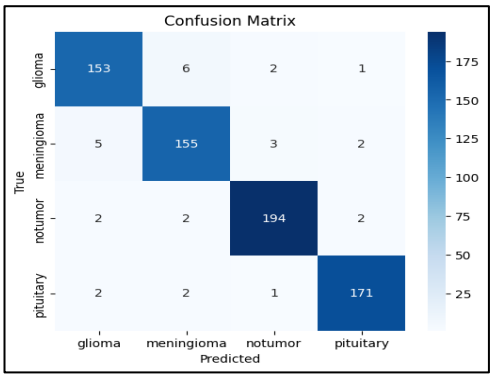
Figure 9. Training and Validation Accuracy and Loss Curves at Stage 2

5.4 Classification Performance

The classification results of the proposed framework are evaluated using precision, recall, F1-score, and overall accuracy. Class-wise performance details are shown in Figure 10(a), and the confusion matrix is shown in Figure 10(b) The proposed framework achieves 95.73% overall classification accuracy on the testing set, with strong discriminative capability across all four classes. The macro-averaged F1-score of 95.7% demonstrates balanced performance despite moderate class imbalance, which further verifies the effectiveness and robustness of our learning strategy.

	precision	recall	f1-score	support
glioma	0.9444	0.9444	0.9444	162
meningioma	0.9394	0.9394	0.9394	165
notumor	0.9700	0.9700	0.9700	200
pituitary	0.9716	0.9716	0.9716	176
accuracy			0.9573	703
macro avg	0.9564	0.9564	0.9564	703
weighted avg	0.9573	0.9573	0.9573	703

(a)



(b)

Figure 10. a) Classification Report b) Confusion Matrix

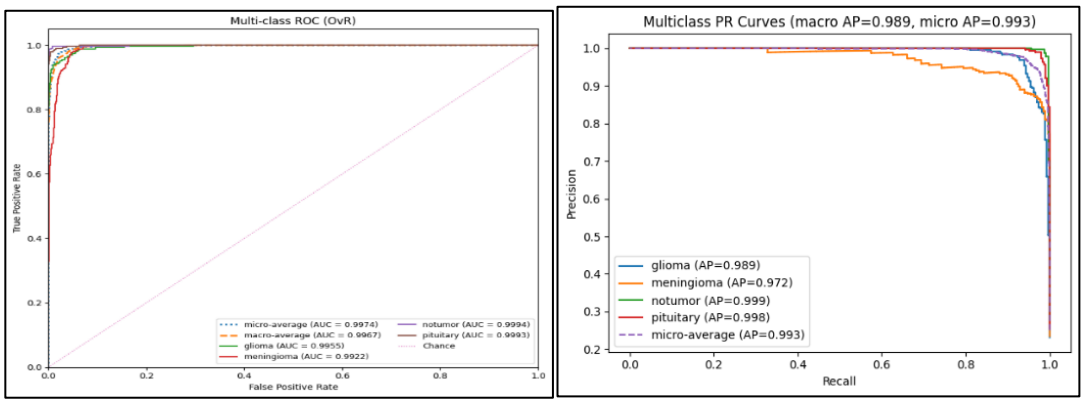
As seen from Figure 10(a), the recall of the no-tumor class reaches the highest value at 97.00%, indicating accurate detection of healthy subjects. This characteristic is highly important for clinical application purposes. The F1 score of the pituitary tumor class is also quite high, with a value of 97.10%, showing consistency in the detection of tumor regions with sharp boundaries. Regarding the glioma class, both precision and recall reach 94.44%, indicating consistency in the output results. The meningioma class performs slightly worse compared to the glioma class, with a precision value of 94.32% and a recall of 93.94%.

The confusion matrix provided in Figure 10(b) offers further insight into prediction performance per class. Most samples fall correctly along the diagonal, with a few off-diagonal errors. Due to some confusion between the glioma and meningioma classes resulting from overlapping radiological features, the model predicts the no-tumor class nearly perfectly, with few false positives. The results indicate that the gated cross-attention fusion mechanism can increase intra-class discrimination while decreasing inter-class confusion. The similar imaging features of glioma and meningioma on MRI contribute to confusion between the two.

Both types of tumors may exhibit similar intensity distributions, irregular forms, and heterogeneous textures, most notably in T1-weighted images following contrast enhancement. Additionally, differences in tumor site and size increase intra-class variability. Nevertheless, the gated cross-attention mechanism proposed in this paper reduces their misclassification by focusing on the discriminative features of the tumors; however, confusion still exists due to the clinical similarity of the tumor types.

5.5 ROC–AUC and Precision–Recall Analysis

We also conducted the one-vs-rest (OvR) ROC–AUC and Precision–Recall (PR) analyses to evaluate the discriminative ability of our proposed framework, presented in Figure 11(a) and Figure 11(b), respectively. The model achieved a high macro-averaged ROC–AUC of 0.9967 and a micro-averaged ROC–AUC of 0.9974, as shown in Figure 11(a), implying high separability among all tumor categories. The class-wise ROC–AUC values are also high, no-tumor (0.9994) and pituitary (0.9993) show perfect discrimination followed by glioma (0.9955) and meningioma (0.9922), which share radiological similarities with each other. The PR curves in Figure 11(b) further demonstrate robustness against moderate class imbalance, with a macro-averaged average precision (AP) of 0.989 and a micro-averaged AP of 0.993 showing that high precision is retained over a broader range of recall values. We observe that both no-tumor (AP = 0.999) and pituitary (AP = 0.998) classes have PR behavior very close to their ideal, while glioma (AP = 0.989) and meningioma (AP = 0.972) exhibit strong performance as well. Altogether, our results validate that the dual-backbone gating cross-attention framework achieves reliable discrimination performance for multi-class MRI brain tumor classification.



(a) (b)

Figure 11. a) Multiclass ROC Curves b) Multiclass PR Curves

5.6 Ablation and Comparative Results

A complete ablation and comparison study was performed to investigate the role of components in the proposed system. The ablation results in Table 4 show that the performance is worse if inter-feature interaction does not exist, with only simple feature concatenation. Adding gating fusion to the combination process successfully enhances the classification results by removing redundant features, and cross-attention strengthens the performance by allowing different feature interactions across backbones. The best results are achieved by adopting the proposed Gated Cross-Attention Fusion (GCAF), verifying the necessity of adaptive feature re-weighting.

Table 4. Ablation Study on Feature Fusion Strategies

Fusion Strategy	Accuracy (%)	Macro F1	Macro ROC–AUC
Feature Concatenation	92.84	0.926	0.982
Gated Fusion (No Attention)	94.12	0.940	0.989
Cross-Attention (No Gating)	94.68	0.945	0.991
Proposed Gated Cross-Attention	95.73	0.957	0.997

Moreover, the influence of Mixup regularization and EMA stabilization is summarized in Table 5, illustrating that their combination yields the best accuracy, macro-F1 score, as well as the lowest calibration error.

Table 5. Effect of Mixup and EMA on Model Performance

Training Configuration	Accuracy (%)	Macro F1	ECE ↓
Without Mixup & EMA	93.47	0.934	0.083
With Mixup Only	94.82	0.947	0.061
With EMA Only	95.01	0.949	0.053
Proposed Model	95.73	0.957	0.044

Performance comparison to state-of-the-art single-backbone CNN models and attention-based baselines is given in Table 6 which verifies that the proposed dual-backbone framework outperforms existing methods consistently. More importantly, it brings explain ability, indicating it is potentially useful for clinical decision-support, subject to further clinical validation. All baseline methods were reimplemented under identical preprocessing, augmentation, and training conditions to ensure a fair comparison.

Table 6. Comparative Performance with Existing Methods

Method	Backbone Type	Accuracy (%)	Macro F1
VGG19 [13]	Single CNN	94.90	0.939
ResNet50[6]	Single CNN	90.59	0.907
DenseNet	Single CNN	94.40	0.946
EfficientNet-B0	Single CNN	91.73	0.929
CNN with Attention	Single CNN and Attn	94.21	0.941
Proposed Method	Dual CNN and GCAF	95.73	0.957

The proposed model performs well on the used dataset, but its generalization to multi-center or external datasets must be considered. Differences in MRI acquisition protocols, scanners, and imaging conditions in different clinical centers might induce domain shifts that affect model performance. The proposed framework utilizes Mixup regularization, label smoothing, and EMA-based stabilization to improve robustness to such variations. Moreover, variability in training sensitivity is incorporated into our training methodology. Nonetheless, further validation on independent multi-center datasets is necessary to determine the model's generalizability and clinical applicability. In the future, we will investigate cross-dataset evaluations and domain adaptation for improved performance in other clinical datasets.

Although transformer-based methods can achieve better performance on large-scale datasets, the proposed framework achieves comparable accuracy with better efficiency and stability on relatively small and heterogeneous medical datasets.

5.7 Statistical Reliability

We used various statistical agreement and calibration metrics, in addition to some accuracy-based metrics, to evaluate the robustness and stability of the proposed framework. The findings presented in Table 7 indicate that the Cohen's Kappa score of the model is 0.9415 which means the predicted label had almost perfect agreement above chance with the ground truth. The Matthews Correlation Coefficient (MCC) of 0.9418 also indicates that a strong correlation is observed among the predictions and true classes over all categories, even in a moderate class imbalance situation. The model exhibited an Expected Calibration Error (ECE) of 0.0441, which shows good calibration. The improved calibration may stem from Mixup regularization and label smoothing, as well as the weight stabilization from EMAs, which encourages more consistent behavior across data points. Collectively, these statistics validate that the proposed model offers not only high predictive accuracy but also robust and clinically reliable predictions of high consistency, which are also decision-support applicable.

Table 7. Statistical Reliability Metrics

Metric	Value	Interpretation
Cohen's Kappa	0.9415	Almost perfect agreement
Matthews Correlation Coefficient (MCC)	0.9418	Strong prediction–ground truth correlation
Expected Calibration Error (ECE) ↓	0.0441	Well-calibrated probability estimates

To examine the validity of these numbers, a calibration test was conducted using ECE. The ECE of the proposed model is very small, at 0.0441, suggesting that the estimated probability values agree well with the true probability values. Through Mixup and label smoothing, overconfidence can be avoided, lead to the improvement in terms of ECE. Although we did not perform any threshold tuning to optimize the predictive accuracy (e.g., decision curve analysis), this low ECE suggests that the model provides valid probabilistic outputs that can be used for future validations.

To validate the significance of the improvements, a paired t-test comparing the proposed method to the baseline models was conducted using accuracy scores from multiple experiments. The improvement is statistically significant, as all the results show statistically significant differences ($p<0.05$).

5.8 Hyperparameter Sensitivity Analysis

To validate the robustness of the selected hyperparameters in Table 8, a sensitivity analysis was performed by varying the key parameters and evaluating their effect on classification performance. This analysis provides numerical validation of the selected configuration.

Table 8. Sensitivity Analysis of Key Hyperparameters

Parameter	Value	Accuracy (%)	Observation
Batch Size	16	94.82	Noisy gradients and unstable convergence
	32	95.73	Best balance between stability and generalization
	64	95.10	Slight reduction in generalization performance
Mixup α	0.1	94.95	Insufficient regularization
	0.2	95.73	Optimal trade-off between mixing and separability
	0.4	95.21	Moderate performance degradation
	0.6	94.88	Over-regularization effect
Focal γ	1	95.02	Limited focus on hard samples
	2	95.73	Best emphasis on difficult samples

	3	95.18	Over-emphasis on hard samples
Label Smoothing ε	0.01	95.10	Slight overconfidence in predictions
	0.05	95.73	Best calibration and performance
	0.1	95.05	Over-smoothing reduces confidence

The analysis included evaluating the class-balancing factor α in focal loss, in the range [0.25,0.75]. The best performance was obtained at α=0.5, which offers a balanced weighting between majority and minority classes. When the value was lower, it under-emphasized minority classes. Conversely, a higher value introduced instability during training.

All hyperparameters were selected based on grid search using validation accuracy as the primary metric. The last combination always provided the best trade-off between accuracy, stability, and generalization performance.

The sensitivity analysis shows that the selected hyperparameter configuration (with batch size=32, α=0.5, γ=2, ϵ=0.05, and Mixup α=0.2) achieves the maximum classification accuracy of 95.73%. If these values deviate from what is intended, we suffer performance degradation as a result. This result either from unstable optimization (smaller batch sizes) or a lack of regularization (lower Mixup α) or over-regularization (higher γ and ϵ). To be specific, moderate parameters obtain a good trade-off between hard-sample emphasis, class balancing and calibration. The modeling results show that being sensitive to hyperparameter selection, the proposed model also has stable performance in a properly defined optimal interval.

5.9 Model Predictions and Visualization

In this section, we assess the predictability and interpretability of our proposed model by displaying predicted images and Grad-CAM in Figures 12 and 13. As shown in Figure 12 the model makes accurate class predictions with large confidence scores on all tumor classes (i.e., glioma, meningioma, pituitary tumor, and no tumor) obtained from multiple anatomical views. The high and steady confidence indicators show that the learned feature representations are stable, which is corroborated by our quantitative results that demonstrate competitive performance. The results indicate that the dual-backbone network can learn tumor representations that effectively discriminate and generalize across different MRI morphologies.

Figure 13 demonstrates further interpretability analysis using Grad-CAM visualizations that consist of original MRI images along with their activation maps and heatmaps. The outcomes clearly highlight tumor-relevant regions while suppressing background structures, suggesting that the model appropriately attends to clinically relevant areas for making decisions. Through backpropagation of gradients from the classification layer, Grad-CAM reveals the important regions of the image contributing to the prediction of classification used in the model. This helps determine whether the network has learned to differentiate between tumor and non-tumor characteristics. In cases of tumor identification with detectable lesions, the activation maps are consistently concentrated in the tumor region of the training and validation samples. For no-tumor cases, however, the activations cover more diffuse regions of the normal brain and exhibit no false or spurious focus. This behavior indicates that the model does not rely on spurious features, confirming that the proposed gated cross-attention fusion mechanism effectively emphasizes discriminative tumor features.

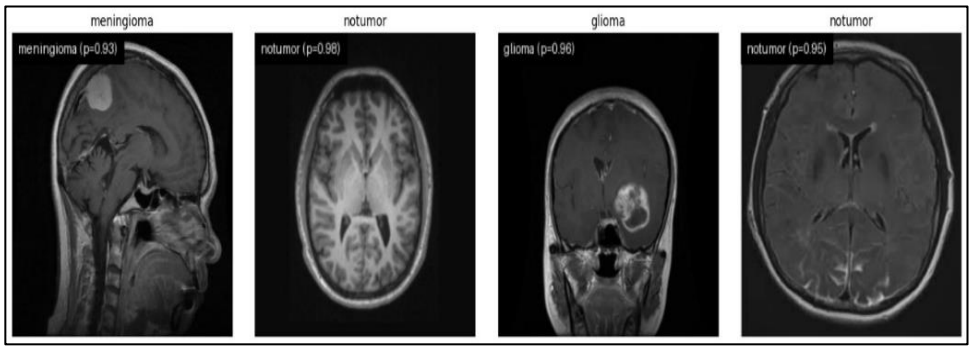


Figure 12. Model Predictions with Class-wise Confidence Scores

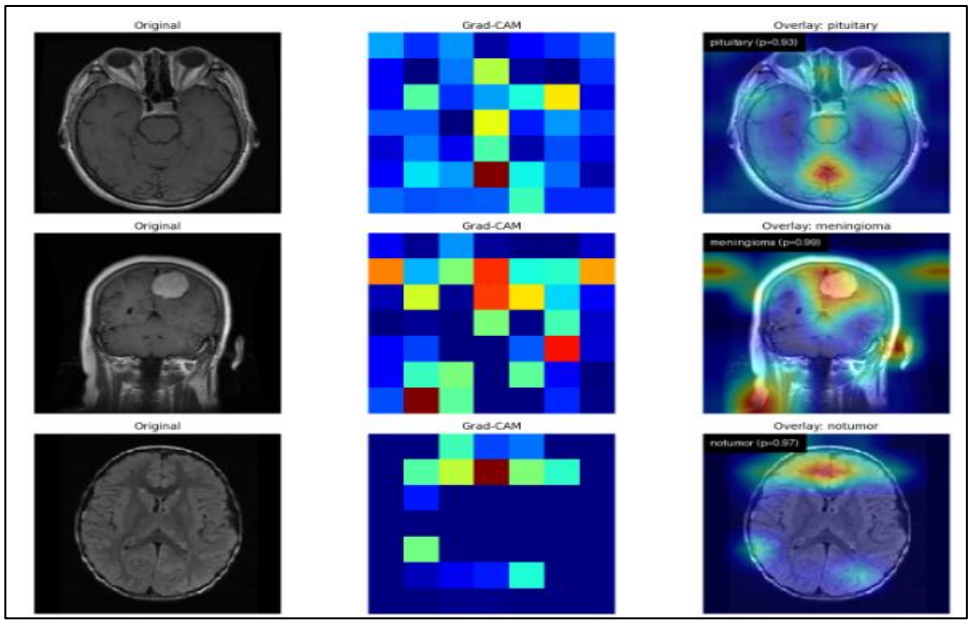


Figure 13. Grad-CAM Interpretability Maps for the Proposed Dual-Backbone Gated Cross-Attention Framework

We conducted this experiment further, changing the initial randomization seed for every training procedure while keeping the hyperparameters fixed. We observed that the Grad-CAM visualizations were highly consistent in showing relevant areas for the tumors, particularly in clearly defined categories such as the pituitary gland or the absence of a tumor. Differences occurred mainly at the boundaries, but these two classes, glioma and meningioma, look alike in terms of MRI, indicating an ambiguity in their appearance. Additionally, the Mixup technique and the EMA technique help reduce variance and create smooth decision boundaries. This further explains why the attention visualization was consistent across training seeds. Th We conducted this experiment further, changing the initial randomization seed for every training procedure while keeping the hyperparameters fixed. We observed that the Grad-CAM visualizations were highly consistent in showing relevant areas for the tumors, particularly in clearly defined categories such as the pituitary gland or the absence of a tumor. Differences occurred mainly at the boundaries, but these two classes, glioma and meningioma, look alike in terms of MRI, indicating an ambiguity in their appearance. Additionally, the Mixup technique and the EMA technique help reduce variance and create smooth decision boundaries. This further explains why the attention visualization was consistent across training seeds. Therefore, this demonstrates that the proposed gated cross-attention generates consistent results that improve the interpretability of the models, making them more trustworthy. Nevertheless, application in clinical settings requires further validation.

Therefore, this demonstrates that the proposed gated cross-attention generates consistent results that improve the interpretability of the models, making them more trustworthy. Nevertheless, application in clinical settings requires further validation.

6. Discussion

The dual-backbone gated cross-attention architecture is found to be consistent across all evaluation metrics. The bidirectional attention mechanism and adaptive gating facilitate the integration of global and local features, thus enhancing classification performance and the macro-F1 metric. This method enables dynamic interactions between the features, unlike fixed fusion techniques, which improve discriminative ability. The use of mixup regularization, focal loss, and label smoothing enhances the robustness and calibration of the model in cases of class imbalance. Grad-CAM also improves model interpretation and reveals that the model attends to important tumor locations. However, model validation is done using a single dataset and needs further multi-center analysis for generalization. Furthermore, the dual-backbone architecture increases computation costs compared to lightweight architectures for clinical use in real-time applications.

However, this technique also has its limitations, even though its performance was quite satisfactory. First, the evaluation is restricted to only one dataset, without testing generalization for the multi-center clinical datasets. Second, there is increased computational complexity associated with the dual backbone architecture to lightweight models. Future work concerning this model includes cross-dataset validation and computational optimization.

7. Conclusion

The proposed model in this paper is an approach based on a dual backbone CNN architecture with gated cross attention fusion that can perform multi-class brain tumor classification using MRI image data. With techniques such as CLAHE pre-processing, Mixup regularization, focal loss with label smoothing, and two stage transfer learning with an EMA stabilizer, the proposed model is successful in addressing challenges like class imbalance, data variability, and inter-class similarity. Experimental results demonstrate a classification accuracy of 95.73% with a macro-F1 score of 95.7% and a macro-averaged ROC-AUC value of 0.9967. The performance of the fusion and training technique is validated using ablation analysis and comparative analysis. Grad-CAM analysis offers insight into the interpretability by identifying regions pertaining to tumors only. The proposed model performs well and provides interpretability for the classification of brain tumor classes using MRI data. Even though these results are encouraging, further evaluation using a multi-center dataset is required to prove the applicability of the proposed technique in clinical practice.

Author Contributions

CMAK Zeelan Basha: Conceptualization, Methodology, Writing – original draft.

Prasanth Yalla: Supervision, Review & editing.

Acknowledgements

The authors would like to thank Koneru Lakshmaiah Education Foundation for providing the necessary infrastructure and support to conduct this research. No external funding was received for this study.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Louis, David N., Arie Perry, Pieter Wesseling, Daniel J. Brat, Ian A. Cree, Dominique Figarella-Branger, Cynthia Hawkins et al. "The 2021 WHO Classification of Tumors of the Central Nervous System: A Summary." Neuro-oncology 23, no. 8 (2021): 1231-1251.

[2] Menze, Bjoern H., Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)." IEEE transactions on medical imaging 34, no. 10 (2014): 1993-2024.

[3] Havaei, Mohammad, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. "Brain Tumor Segmentation with Deep Neural Networks." Medical image analysis 35 (2017): 18-31.

[4] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778. 2016.

[5] Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I. Sánchez. "A Survey on Deep Learning in Medical Image Analysis." Medical image analysis 42 (2017): 60-88.

[6] Isty, Maherun Nessa, Raiyan Gani, Jubaer Ahmed, Tasmia Islam, and Shamim Ripon. "Deep Learning Techniques for Early Brain Tumor Detection: A Comparative Study on Models Performance Utilizing Dataset Enhancement." In 2024 IEEE International Conference on Computing, Applications and Systems (COMPAS), IEEE, 2024, 1-6.

[7] Lakineni, Prasanna Kumar, D. Jayanarayana Reddy, M. Chitra, R. Umapriya, L. Vadivel Kannan, and S. R. Barkunan. "Optimal Feature Selection and Classification Using Convolutional Neural Network-Based Plant Disease Prediction." In 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS), IEEE, 2023, 1-6.

[8] Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv preprint arXiv:2010.11929 (2020).

[9] Zhang, Hongyi, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. "Mixup: Beyond Empirical Risk Minimization." arXiv preprint arXiv:1710.09412 (2017).

[10] Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-Cam: Visual Explanations from Deep Networks Via Gradient-Based Localization." In Proceedings of the IEEE international conference on computer vision, 2017, 618-626.

[11] Haralick, Robert M., Karthikeyan Shanmugam, and Its' Hak Dinstein. "Textural Features for Image Classification." IEEE Transactions on systems, man, and cybernetics 6 (1973): 610-621.

[12] Cheng, Jun, Wei Huang, Shuangliang Cao, Ru Yang, Wei Yang, Zhaoqiang Yun, Zhijian Wang, and Qianjin Feng. "Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition." PloS one 10, no. 10 (2015): e0140381.

[13] Ulli, Sai Shanmukh, Hanish Akkineni, Uma Santhosh Vuyyuru, Sai Ram Pathyala, Uttej Kumar Nannapaneni, and B. Suvarna. "Brain Tumor Detection Using Modified VGG-19 Model." In 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT), vol. 1, IEEE, 2024, 1-5.

[14] Khan, Amjad Rehman, Siraj Khan, Majid Harouni, Rashid Abbasi, Sajid Iqbal, and Zahid Mehmood. "Brain Tumor Segmentation Using K-Means Clustering and Deep Learning with Synthetic Data Augmentation for Classification." Microscopy Research and Technique 84, no. 7 (2021): 1389-1399.

[15] Chollet, François. "Xception: Deep Learning with Depthwise Separable Convolutions." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 1251-1258.

[16] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-Excitation Networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, 7132-7141.

[17] Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. "Cbam: Convolutional Block Attention Module." In Proceedings of the European conference on computer vision (ECCV), 2018, 3-19.

[18] Baseer, K. K., K. Sivakumar, Duggineni Veeraiah, Gunjan Chhabra, Prasanna Kumar Lakineni, M. Jahir Pasha, Ramu Gandikota, and Gopakumar Harikrishnan. "Healthcare Diagnostics with an Adaptive Deep Learning Model Integrated with the Internet of Medical Things (IoMT) for Predicting Heart Disease." Biomedical Signal Processing and Control 92 (2024): 105988.

[19] Cui, Yin, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. "Class-Balanced Loss Based on Effective Number of Samples." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, 9268-9277.

[20] Yun, Sangdoo, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. "Cutmix: Regularization Strategy to Train Strong Classifiers with Localizable Features." In Proceedings of the IEEE/CVF international conference on computer vision, 2019, 6023-6032.

[21] Yahyaoui, Hela, Fethi Ghazouani, and Imed Riadh Farah. "Deep Learning Guided by an Ontology for Medical Images Classification Using a Multimodal Fusion." In 2021 International congress of advanced technology and engineering (ICOTEN), IEEE, 2021, 1-6.

[22] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "" Why Should I Trust You?" Explaining the predictions of any classifier." In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, 1135-1144.

[23] Lundberg, Scott M., and Su-In Lee. "A Unified Approach to Interpreting Model Predictions." Advances in neural information processing systems 30 (2017).

[24] El-Assiouti, Omar S., Ghada Hamed, Hadeer El-Saadawy, Hala M. Ebied, and Dina Khattab. "RegionInpaint, Cutoff and RegionMix: Introducing Novel Augmentation Techniques for Enhancing the Generalization of Brain Tumor Identification." IEEE Access 11 (2023): 83232-83250.

[25] Chaki, J. “Brain Tumor MRI Dataset”. IEEE Dataport, 2023.