



## OPEN ACCESS

## EDITED BY

Deepti Deshwal,  
Maharaja Surajmal Institute of Technology,  
India

## REVIEWED BY

Arun Kumar Sunaniya,  
National Institute of Technology, Silchar, India  
Anjana Subba,  
National Institute of Technology, Silchar, India  
Jane Rubel Angelina Jeyaraj,  
Kalasalingam University, India

## \*CORRESPONDENCE

Wenna Chen  
✉ chenwenna0408@163.com  
Ganqin Du  
✉ dgq99@163.com

RECEIVED 26 April 2025

ACCEPTED 12 June 2025

PUBLISHED 03 July 2025

## CITATION

Chen W, Tan X, Zhang J, Du G, Fu Q and  
Jiang H (2025) MLG: a mixed local and  
global model for brain tumor classification.  
*Front. Neurosci.* 19:1618514.  
doi: 10.3389/fnins.2025.1618514

## COPYRIGHT

© 2025 Chen, Tan, Zhang, Du, Fu and Jiang.  
This is an open-access article distributed  
under the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other forums  
is permitted, provided the original author(s)  
and the copyright owner(s) are credited and  
that the original publication in this journal is  
cited, in accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# MLG: a mixed local and global model for brain tumor classification

Wenna Chen<sup>1\*</sup>, Xinghua Tan<sup>2</sup>, Jincan Zhang<sup>2</sup>, Ganqin Du<sup>1\*</sup>,  
Qizhi Fu<sup>1</sup> and Hongwei Jiang<sup>1</sup>

<sup>1</sup>The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology, Luoyang, China, <sup>2</sup>College of Information Engineering, Henan University of Science and Technology, Luoyang, China

**Introduction:** Brain tumors seriously endanger human health. Therefore, accurately identifying the types of brain tumors and adopting corresponding treatment methods is of vital importance, which is of great significance for saving patients' lives. The use of computer-aided systems (CAD) for the differentiation of brain tumors has proved to be a reliable scheme.

**Methods:** In this study, a highly accurate Mixed Local and Global (MLG) model for brain tumor classification is proposed. Compared to prior approaches, the MLG model achieves effective integration of local and global features by employing a gated attention mechanism. The MLG model employs Convolutional Neural Networks (CNNs) to extract local features from images and utilizes the Transformer to capture global characteristics. This comprehensive scheme renders the MLG model highly proficient in the task of brain tumor classification. Specifically, the MLG model is primarily composed of the REMA Block and the Biformer Block, which are fused through a gated attention mechanism. The REMA Block serves to extract local features, effectively preventing information loss and enhancing feature expressiveness. Conversely, the Biformer Block is responsible for extracting global features, adaptively focusing on relevant sets of key tokens based on query positions, thereby minimizing attention to irrelevant information and further boosting model performance. The integration of features extracted by the REMA Block and the Biformer Block through the gated attention mechanism further enhances the representation ability of the features.

**Results:** To validate the performance of the MLG model, two publicly available datasets, namely the Chen and Kaggle datasets, were utilized for testing. Experimental results revealed that the MLG model achieved accuracies of 99.02% and 97.24% on the Chen and Kaggle datasets, respectively, surpassing other state-of-the-art models. This result fully demonstrates the effectiveness and superiority of the MLG model in the task of brain tumor classification.

## KEYWORDS

classification of brain tumor, CNN, transformer, feature fusion, gated attention mechanism

## 1 Introduction

Brain diseases most commonly manifest as brain tumors, which represent a severe health threat to the human body and necessitate early diagnosis and treatment (Lyu et al., 2024; Akter et al., 2024; Liu et al., 2023). The classification of brain tumors constitutes a significant area of research in medical imaging and artificial intelligence. Classification of brain tumors using Magnetic Resonance Imaging (MRI) is the main technique (Li and Zhou, 2025). This process is critical for accurate diagnosis, treatment planning, and prognosis assessment. Recently, Computer-Aided Detection and Diagnosis (CAD) systems have played a pivotal role in assisting medical professionals with the detection and classification of brain tumors. Traditional manual methods of brain tumor classification rely heavily on experienced specialists and are often time-consuming, labor-intensive, and inefficient (Sharma et al., 2024; Zhou et al., 2024). To address this issue, extensive research has been conducted into automatic classification techniques that can classify brain tumors from MRI, employing CAD technology for tumor classification from MRI, which exhibits high reliability due to its high accuracy.

Traditional machine learning often relies on manually designed features, which places high demands on the user's domain knowledge and experience. The selection and construction of features are complex and time-consuming, having a crucial impact on model performance. When faced with complex, high-dimensional, or nonlinear problems, the generalization ability of traditional machine learning algorithms may be limited (Kaur and Mahajan, 2025). More crucially, when confronted with new, unseen data, their predictive performance may decline, affecting their practical utility (Mehnatkesh et al., 2023; Pandiselvi and Maheswaran, 2019). In contrast, deep learning possesses stronger data representation capabilities, able to automatically learn high-level abstract representations of data, significantly enhancing the performance and effectiveness of machine learning. Deep learning models are not only highly complex but also capable of handling more complex tasks and larger datasets. Consequently, deep learning has found widespread application in the field of medical imaging, providing powerful support for disease diagnosis and treatment (Kshatri and Singh, 2023; Mazurowski et al., 2023; Mukadam and Patil, 2024; Yu et al., 2022).

Convolutional Neural Networks (CNNs), as a type of deep learning algorithm, have demonstrated remarkable prowess in the field of image processing, thanks to their unique advantages. The CNNs not only accept input images, but also adeptly assign varying degrees of importance to different elements or objects within those images through learnable weights and biases, enabling effective differentiation among them. Compared to other classification algorithms, the CNNs significantly reduce the need for preprocessing, greatly enhancing ease of use. In earlier image processing, filters were typically manually designed. However, CNNs can automatically learn these filters or features during training. Consequently, CNNs have seen widespread application in fields such as medical image analysis. Cao et al. (2024) introduced a Multi-branch Spectral Channel Attention Network (MbsCANet) for breast cancer classification. By extracting features in the frequency domain and applying attention mechanisms to the backbone network, MbsCANet achieves more precise

feature extraction and classification, thereby not only improving classification accuracy but also providing robust support for early diagnosis and treatment of breast cancer. Regarding retinal disease classification, Peng et al. (2024) proposed a multi-scale-denoising residual convolutional network (MS-DRCN) model. This model integrates the strengths of Deep Residual Network (ResNet) along with multiscale processing and feature fusion techniques. Aimed at enhancing the accuracy and robustness of Optical Coherence Tomography (OCT) image classification, MS-DRCN offers an effective tool for precise diagnosis of retinal diseases. Moreover, SkinLesNet, a deep learning model specifically designed for skin lesion classification, is built upon a CNN architecture that has undergone meticulous design and optimization (Azeem et al., 2024). Through a series of CNNs, it progressively extracts image features, enabling in-depth understanding and analysis of lesion images. This structure enables the model to precisely capture subtle differences and key features within the images, significantly boosting classification accuracy and reliability. As a result, it provides crucial assistance in the early detection and treatment of skin lesions.

The Transformer, an attention mechanism originating from the field of natural language processing, has demonstrated remarkable performance in computer vision. Its advantages over CNNs are particularly evident in handling long-distance dependencies and global contextual information in images (Liu et al., 2021b; Yan et al., 2023; Huang S. K. et al., 2024). Bofan Song et al. (Song et al., 2024) utilized Vision Transformer (ViT) and Swin Transformer (SwinT) for the classification of oral cancer images. In the literature (Huang L. et al., 2024), Swin-residual transformer (SRT), was proposed for thyroid ultrasound image classification. The SRT model introduces residual blocks and triplet loss into the SwinT structure, aiming to improve sensitivity to both global and local features of thyroid nodules and better identify subtle feature differences. Additionally, Chincholi and Koestler (2024) designed a model combining ViT and Detection Transformer architectures for glaucoma detection. As the application of Transformers in disease detection continues to grow, researchers have begun exploring the integration of CNNs with Transformers to simultaneously extract local and global features. For instance, Fang et al. (2024) employed CNNs to extract local features while utilizing ViT for global feature extraction, designing a deep integrated feature fusion module for feature aggregation. Yan et al. (2023) developed the Transformer based High Resolution Network (TransHRNet) for brain tumor segmentation. TransHRNet initially used CNNs as an encoder for image preprocessing, followed by feeding the extracted features from the CNNs into an Effective Transformer (EffTrans) module, and finally generating segmentation results through a CNNs decoder. Notably, EffTrans incorporates Group Linear Transformations (GLTs) with an expansion-reduction strategy and spatial-reduction attention (SRA) layers, significantly reducing the computational burden and memory consumption of the Transformer.

The classification of brain tumors poses a highly challenging task in computer vision. These tumors vary significantly in size, shape, and location within the brain, and their categorization depends not only on the characteristics of the lesion itself but also on the surrounding tissue environment (ThamilSelvi et al., 2025; Verma and Yadav, 2025). Furthermore, the diversity and

spatial distribution of brain tumors underscore the importance of utilizing both local and global features. In response to these challenges, the Mixed Local and Global (MLG) model is introduced. The uniqueness of the MLG model lies in its utilization of two advanced feature extraction methods. On one hand, Residual Efficient Multi-scale Attention (REMA) block is designed to extract local fine-grained features. On the other hand, the Bi-Level transformer (Biformer) block is used to capture the global context features. The REMA module integrates two layers of convolution and an Efficient Multi-scale Attention (EMA) component (Ouyang et al., 2023), which are interconnected through residual connections. This classical residual connection design ensures that gradients can propagate more effectively throughout the network during training, thereby mitigating gradient vanishing issues (He et al., 2016; Shafiq and Gu, 2022). Channel attention and spatial attention mechanisms have proven to be highly effective in generating more discriminative feature representations (Hu et al., 2018; Woo et al., 2018; Yu et al., 2023). In this block, EMA enhances both spatial and channel-wise features and achieves the ability to capture feature information across different scales by constructing parallel subnetwork structures operating at multiple resolutions. The core of Biformer is its Bi-Level Routing Attention (BRA), which facilitates dynamic and query-based content-aware sparse attention allocation while circumventing the high computational cost of full-space attention. Biformer realizes this pattern by introducing the Bi-Level Routing Attention mechanism, where it first prunes irrelevant key-value pairs at a coarse-grained region level, and subsequently conducts fine-grained token-to-token attention computations only within the selected candidate regions (Zhu et al., 2023). The integration of features from REMA and Biformer via gated attention mechanisms further refines these features, enhancing model performance. To validate the efficacy of the MLG model, two publicly available brain tumor datasets were utilized for experimental evaluation. Experimental results demonstrated that the proposed model outperforms other existing advanced models in terms of performance. In summary, the main contributions of this paper are as follows:

- Development of a brain tumor classification model that integrates both local and global features.
- The innovative application of the REMA module to extract local features and the use of Biformer for capturing global features, with both being effectively fused through a gated attention mechanism.
- Validation of the proposed model on two open datasets, achieving superior results compared to the current state-of-the-art performance.

## 2 Related work

The application of deep learning techniques in medical image analysis is becoming increasingly popular, particularly in the study of brain tumor classification, where it has demonstrated significant value. In recent years, research efforts on brain tumor classification tasks have continued to deepen, and these studies

can be broadly categorized into two camps: one is the CNN-based approach, and the other is the emerging strategy based on the Transformer architecture.

### 2.1 CNN in brain tumor classification

The CNN has been widely used in brain tumor classification tasks. In the task of brain tumor classification, CNNs have been widely employed. Kang et al. (2021) adopted a transfer learning-based framework using a pre-trained deep CNN to extract deep features from MRI data. By fusing features obtained from different levels of the network and integrating them with multiple machine learning classifiers, this method achieved significant results. Alanazi et al. (2022) proposed a 22-layer CNN model, which was initially trained on a binary brain tumor dataset. Subsequently, with the help of transfer learning technique, the model weight was utilized for multi-class data, resulting in promising outcomes. Saurav et al. (2023) designed an Attention-Guided Convolutional Neural Network (AG-CNN) specifically tailored for brain tumor classification tasks. The network incorporates an internal channel attention module, which aids in focusing on processing image regions relevant to tumors, thereby facilitating effective feature extraction and classification. Alturki et al. (2023) proposed an optimization scheme for brain tumor classification performance. The CNNs were utilized to extract deep features from raw brain tumor MRI data and two classification algorithms including logistic regression (LR) and stochastic gradient descent (SGD) were incorporated into a voting ensemble classifier. By inputting these deep features into the ensemble classifier, the model achieved accurate classification of brain tumors. Hossain et al. (2023) conducted a study implementing transfer learning to investigate the performance of various models, including VGG16, InceptionV3, and ResNet50, inceptionResNetv2, Xception, for brain tumor classification. Ultimately, three best performing models were chosen to be used to construct an ensemble model, which was named IVX16. Sachdeva et al. (2024) evaluated multiple pre-trained models such as ResNet50, DenseNet121, EfficientNetB0, and EfficientNetV2L, et al., by incorporating Dropout layers, global average pooling layers, and tuning hyperparameters to enhance model performance. The results show that EfficientNetB0 model achieved a higher classification accuracy.

### 2.2 Transformer in brain tumor classification

Transformer has also been applied in brain tumor classification tasks. Ferdous et al. (2023) proposed a Linear Complexity Data-Efficient Image Transformer (LCDEiT) framework based on a teacher-student mechanism specifically designed for tumor classification from brain MRI images. In the teacher model component, gated pooling techniques were employed to optimize the feature extraction efficiency of CNNs. The pre-trained teacher model was able to extract crucial knowledge pertinent to the tumor classification task. On the other hand, the student model introduced an image transformer equipped with an external attention mechanism, which leveraged the knowledge acquired

from the teacher model for tumor classification in brain MRI. In paper, [Asiri et al. \(2024\)](#) proposed an innovative and robust method based on the SwinT architecture, aiming to improve the accuracy of brain tumor image classification. This method integrated complex preprocessing procedure, sophisticated feature extraction techniques, and a thorough classification system, enabling the SwinT model to effectively analyze and discriminate various types of brain tumors. [Wang et al. \(2024\)](#) employed a pre-trained ViT as the backbone for their brain tumor classification model, named as RanMerFormer. Additionally, to enhance the computational efficiency of the ViT backbone, a Token Merging Algorithm (TMA) was used. Instead of using a traditional linear classification head, Random Vector Functional Link (RVFL) networks were utilized. [Poornam and Angelina \(2024\)](#) proposed the ViT with Attention and Linear Transformation module (VITALT) for brain tumor detection and classification. VITALT primarily consists of a ViT, a Split bidirectional feature pyramid network (S-BiFPN), and a linear transformation module (LTM). ViT was used to capture global and local features, while S-BiFPN fusions the features extracted by ViT. The LTM enhanced the model's linear expressive ability. In paper ([Şahin et al., 2024](#)), the Bayesian Multi-Objective (BMO) optimization method was employed to optimize the hyperparameters of the ViT network in order to improve its performance in brain tumor classification tasks. [Gade et al. \(2024\)](#) proposed the Lite Swin Transformer (OLiST) model for brain tumor detection. This model combined the Lite Swin Transformer's ability to capture global features with the advantage of CNNs in extracting local features. By fusing the features extracted by both, the model leveraged the strengths of both approaches.

In summary, the use of CNNs and Transformers have been used in brain tumor classification tasks with excellent performance. CNNs have the advantage of extracting local features of images, while Transformers have the advantage of exploiting global features of images. Therefore, this paper innovatively introduces a hybrid model, MLG, which effectively integrates the respective strengths of CNNs and Transformers, thus significantly enhancing the performance of brain tumor classification tasks.

## 3 Materials and methods

In this section, the datasets used and the proposed model are described in detail.

### 3.1 Datasets and preprocessing

In this study, two widely used public datasets, namely the Chen dataset and the Kaggle dataset, were adopted. The Chen dataset, provided by [Cheng et al. \(2015\)](#), primarily focuses on three types of brain tumors: gliomas, meningiomas, and pituitary tumors. Comprising a total of 3,064 images, this dataset offers a rich resource for our in-depth research and analysis. On the other hand, the Kaggle dataset is a meticulously compiled and shared public dataset by [Bhuvaji et al. \(2020\)](#). This dataset encompasses four categories of images: glioma tumors, meningioma tumors, pituitary tumors, and normal brain tissues, totaling 3,264 images. For efficient model training and testing, the two datasets were

randomly divided into a training set and a test set. Specifically, 80% of the data was allocated to the training set for model training and optimization, while the remaining 20% was designated as the testing set for evaluating the model's performance. Detailed statistics on the number of images in each dataset are presented in [Table 1](#).

A simple and efficient data preprocessing method is used in the preprocessing phase of the dataset. In the experimental process, to preserve the integrity of image content and stability of features, all images were uniformly resized to  $224 \times 224 \times 3$  pixels. This resizing not only helps maintain the spatial structure and information integrity of the images but also significantly reduces computational burden during network training, thereby enhancing training efficiency. Additionally, normalization was performed, which is a standard preprocessing step in deep learning. This aims to mitigate differences in brightness, contrast, and other attributes among images, enabling the model to focus more acutely on learning the inherent features of the images. For medical images, acquiring a large volume of such data can be challenging ([Dhar et al., 2023](#)). Given that deep neural networks typically require substantial amounts of data for training, and considering the relatively limited scale of the datasets utilized in this study, data augmentation strategies were employed to alleviate overfitting concerns. Specifically, random rotation and random horizontal flipping techniques were utilized, both of which effectively enhance dataset diversity without introducing additional noise, thereby improving the model's generalization capability.

### 3.2 Mixed local and global model

In this section, details of the proposed model are provided. The architecture of the MLG model, which combines both local and global components, is depicted in [Figure 1](#). Initially, brain tumor images undergo preprocessing before being fed into a convolutional layer with a kernel size of  $5 \times 5$  and a stride of 1, designed to enlarge the receptive field. Subsequently, a max pooling layer is applied for downsampling and dimensionality reduction of the extracted features. And then, the features are further processed through five REMA and Biformer (RB) Mixing Blocks to refine the extraction of characteristics specific to brain tumor images. Finally, the resulting features are classified accordingly. The structure of the RB Mixing Block is illustrated in [Figure 2](#).

[Figure 2](#) presents the structure of the RB Mixing Block, primarily consisting of REMA and Biformer units. The REMA

TABLE 1 Details of the datasets.

Dataset name	Classes	Number of each class	Total image count
Chen	Glioma	1,426	3,064
	Meningioma	708	
	Pituitary tumor	930	
Kaggle	Glioma	826	3,264
	Meningioma	822	
	Pituitary tumor	827	
	No tumor	395	

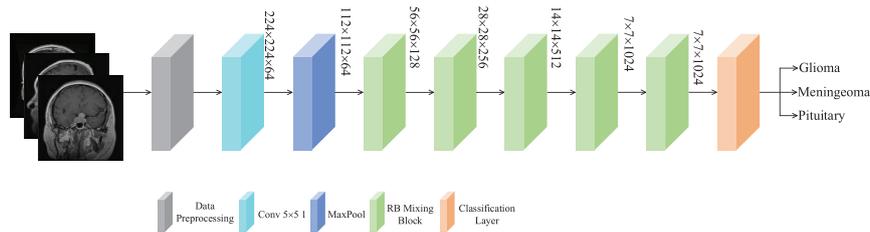


FIGURE 1 Proposed brain tumor classification system.

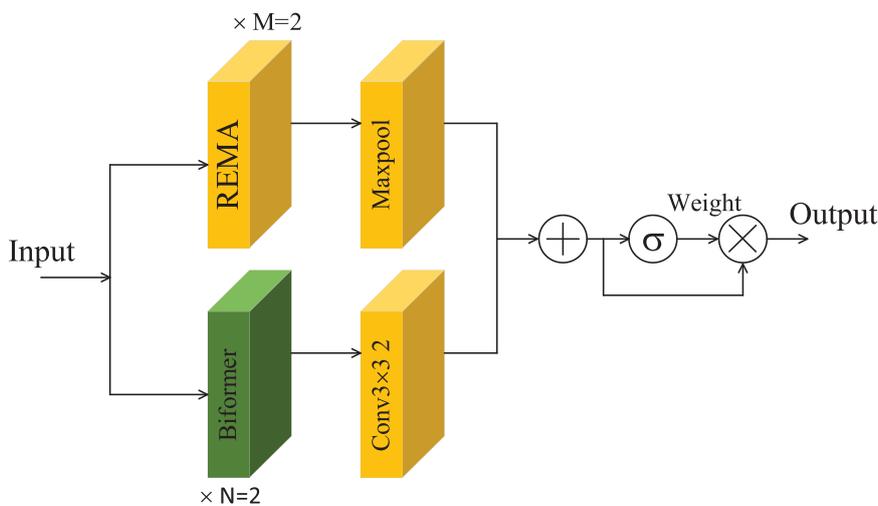


FIGURE 2 The RB Mixing Block structure.

unit is designed to extract local features from the images, while the Biformer unit focuses on extracting global features. After combining the features derived from these two modules, a gating mechanism adjusts the weights of the fused features to better suit the task of brain tumor classification, thereby enhancing the model’s classification performance. Here,  $M$  denotes the number of REMA convolution modules used and  $N$  denotes the number of Biformer modules used,  $M = N = 2$ . REMA utilizes max pooling for downsampling, aiming to broaden the receptive field of the module. On the other hand, Biformer employs convolutions with a stride of 2 for downsampling, intending to derive higher-level feature representations. Subsequently, the features extracted by both REMA and Biformer are merged and subjected to processing by the gating mechanism. Then, the adjusted features are multiplied with the original ones to modulate their significance in influencing the model’s overall performance, effectively filtering out a set of features that have a more substantial impact on the model’s classification results. The output of the RB Mixing module can be expressed as:

$$out_{RB} = sigmoid(f_{REMA} + f_{Biformer}) \times (f_{REMA} + f_{Biformer}) \quad (1)$$

where,  $f_{REMA}$  and  $f_{Biformer}$  represent the features extracted by the modules REMA and Biformer, respectively.

In order to present the structure and parameter characteristics of the REMA module and the Biformer module more clearly. We have detailed the number of parameters, input dimensions and output dimensions of these two modules in Table 2.

The structure and computational complexity of the REMA block and the Biformer block in the MLG model can be understood more specifically through Table 2.

### 3.3 REMA Block

The structure of the REMA block is depicted in Figure 3. This module consists of two convolutional layers and an EMA unit, interconnected via residual connections to facilitate information fusion and propagation. This design aims to enhance the model’s representation learning capacity while alleviating the gradient

TABLE 2 Parameters and dimension information of the REMA block and the Biformer block.

Block	Input size	Output size	No. of parameters
REMA	112 × 112 × 64	112 × 112 × 64	74,160
Biformer	112 × 112 × 64	112 × 112 × 64	10,4576

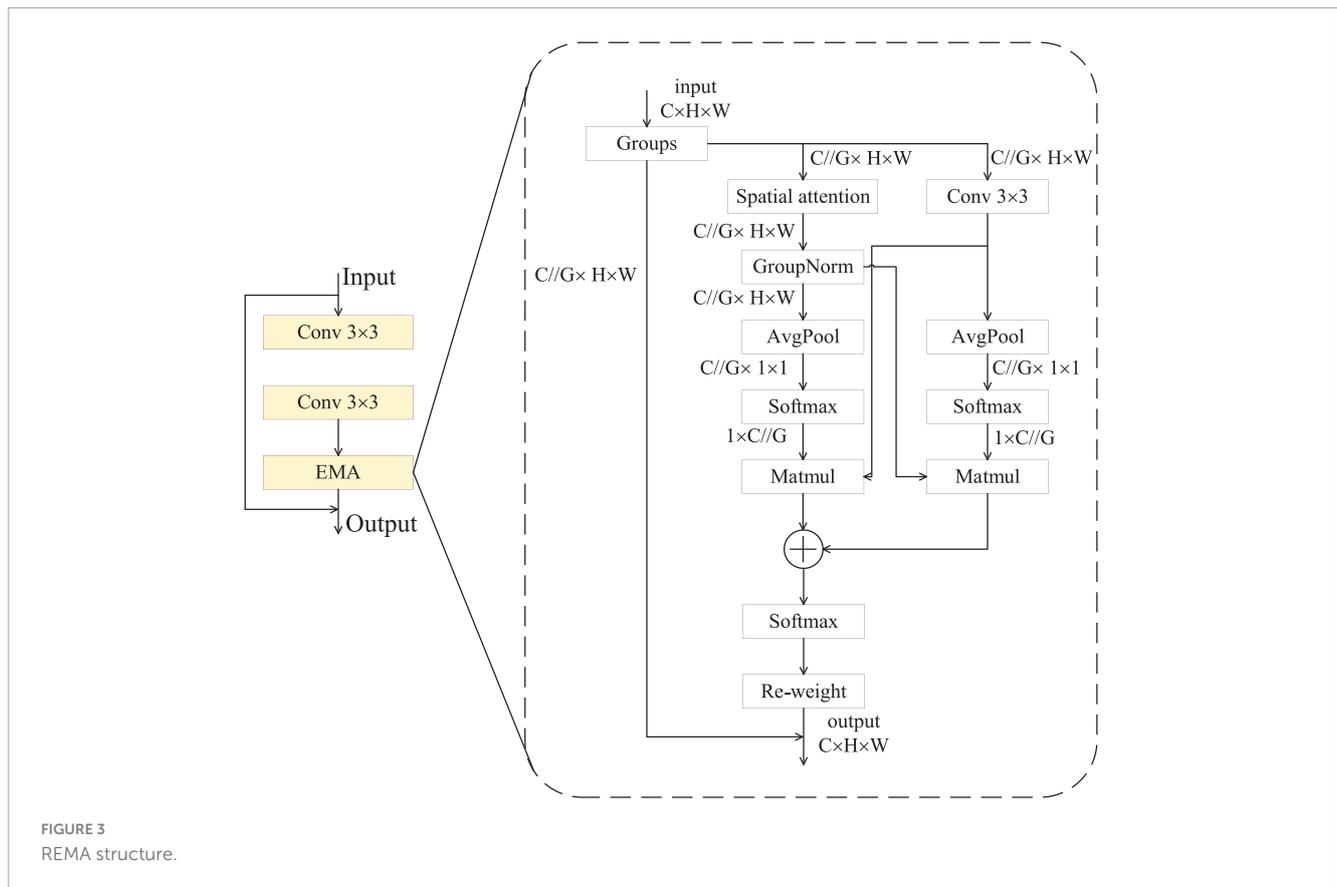


FIGURE 3  
REMA structure.

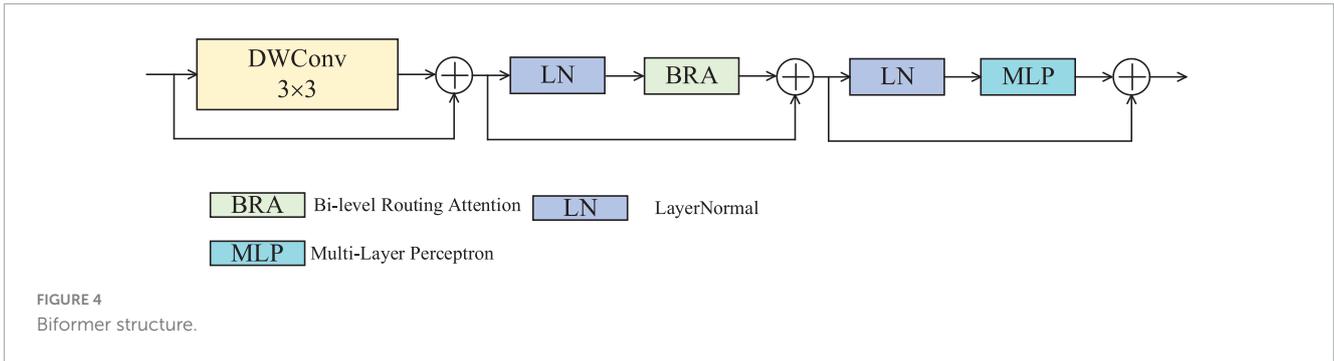
vanishing problem often encountered in deep networks. By incorporating the EMA unit (Ouyang et al., 2023), the REMA block is better equipped to capture inherent data features, thereby boosting the model’s performance. The core idea of the EMA module is to group the channel dimensions into multiple sub-features and ensure good distribution of spatial semantic features within each feature group. This method not only preserves information in each channel but also reduces computational overhead. Specifically, the EMA module recalibrates the channel weights of each parallel branch using global information encoding. Moreover, the output features from the two parallel branches are aggregated through cross-dimensional interaction methods, further enhancing the representational power of the features. Inside the EMA module, there are three parallel paths designed to extract attention weight descriptors for the grouped feature maps. Two of these paths belong to the  $1 \times 1$  branch, while the third one is part of the  $3 \times 3$  branch. Within the  $1 \times 1$  branch, two one-dimension global average pooling operations along two spatial directions are employed to encode channel attention. In contrast, the  $3 \times 3$  branch uses a single  $3 \times 3$  convolutional kernel to capture multi-scale feature representations. The output of the REMA module can be mathematically represented as follows:

$$\text{out} = \text{EMA}(\text{conv}(\text{conv}(x))) + x \tag{2}$$

The structure of the Biformer Block is depicted in Figure 4. The core of the Biformer lies in its BRA, which consists of a deep convolution, two layers of Layer Normalization (LN), and

a Multilayer Perceptron (MLP) interconnected through residual connections (Zhu et al., 2023).

The design principle of BRA revolves around dynamic, query-content based sparsity. Initially, irrelevant key-value pairs are filtered out at a coarse-grained regional level by constructing and pruning a directed graph representing region-level relationships. Subsequently, a fine-grained token-to-token attention mechanism is applied over the joint set of the remaining, or routed, regions to selectively focus on locally relevant information while bypassing globally unrelated data. In BRA process, given a two-dimensional input feature map  $X$ , it is partitioned into  $S \times S$  non-overlapping regions, each containing a specific number of feature vectors. These region-based features undergo linear projections to generate query, key, and value tensors  $Q, K, V$ . An inter-region association matrix  $A^V$  is then constructed by computing average query and key vectors across regions, with its elements indicating semantic relevance between pairs of regions. The critical step involves selecting the top  $k$  most related adjacent regions for each region based on this relevance measure, yielding a routing index matrix  $I^V$  via row-wise top- $k$  operations. Building upon this, the model applies fine-grained token-to-token attention. Specifically, for a query token originating from region  $i$ , it attends to all key-value pairs within the  $k$  routed regions indexed by  $I_{(i,1)}^V$  through  $I_{(i,k)}^V$ . To efficiently execute this, despite these regions potentially being scattered throughout the feature map, the model first employs a gather operation to collect the key and value tensors from these regions, forming aggregated key and value sets  $K_g$  and  $V_g$ . Finally,



attention computation is performed using the gathered key and value tensors:

$$O = \text{soft max}\left(\frac{(QK_g)^T}{\sqrt{C}}\right)V_g + LCE(V) \quad (3)$$

here,  $\sqrt{C}$  is usually a factor that scales the denominator in the formula for calculating the attention score in order to prevent the occurrence of over-concentration of weights and loss of gradients.  $LCE(V)$  represents local context enhancement, which is implemented by depth separable convolution to enhance local information.

### 3.4 Loss function

In classification tasks, the cross-entropy loss function is a commonly used loss function. Originating from the concepts of entropy and mutual information in information theory, it serves to quantify the discrepancy between two probability distributions. Specifically, when training neural networks, it is employed to measure the difference between the model's predicted probability distribution and the true distribution of the observed data. For classification tasks, assuming the true label is  $y$  and the model predicted probability is  $q$ , the cross-entropy loss function can be expressed as:

$$H(y, q) = - \sum_i y_i \log(q_i) \quad (4)$$

where,  $y_i$  represents the true label for the  $i$ -th category and  $q_i$  denotes the model predicted probability that the sample belongs to the  $i$ -th class.

## 4 Results

This section introduces the experimental setup, experimental results, and ablation experiments, collectively serving to comprehensively and rigorously substantiate the proposed model.

### 4.1 Experimental apparatus

A PyTorch implementation is performed for the model proposed by us, while experiments were carried out on a Windows 11 system equipped with a 12GB RTX 4070 GPU and an Intel

i5-13400F processor. The Adam optimizer was utilized, with the initial learning rate set at 0.0001, the batch size fixed at 16, and the number of epochs specified as 50. In our experiments, early stopping was utilized to prevent overfitting. Detailed information about the parameters can be found in Table 3.

### 4.2 Evaluation metrics

In the experiments, the accuracy, recall, precision, and F1-score were employed as evaluation metrics, with their respective calculation methods presented in Formulas (5–8). The accuracy is one of the most commonly used evaluation metrics in classification problems, representing the proportion of correctly classified samples out of the total number of samples. The recall, focuses on the ability of the model to correctly identify positive samples, which refers to the ratio of true positives (correctly identified positive instances) to all actual positive instances in the dataset. The precision measures the proportion of instances predicted by the model as positive that are truly positive, that is, the ratio of true positives to all instances predicted as positive. The F1-score, being the harmonic mean of precision and recall, integrates the performance of both precision and recall, offering a more comprehensive assessment of the model's performance (Zulfiqar et al., 2023; Zebari et al., 2024). When both precision and recall are high, the F1-score will also be high, and conversely, when either of these values is low, so will the F1-score. This implies that a high F1-score indicates strong overall performance in terms of both accurately identifying true positives and minimizing false predictions.

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

TABLE 3 Training Hyper-parameter values of proposed network.

Parameters	Value
Initial learning rate	0.0001
Batch size	16
Optimizer	Adam
Number of epoch	50
Learning rate decays	0.1

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \tag{8}$$

### 4.3 The results of the experiment

Figure 5 illustrates the confusion matrices for the classification results of the model on the test sets of two publicly available datasets, where G, M, and P stand for glioma, meningioma, and pituitary adenoma, respectively, and N stands for normal state, indicating the absence of brain tumor. From the confusion matrices, detailed classification performance metrics for the model were calculated according to Formulas (5–8) and summarized in Table 4. From Table 4, it is evident that, on the test set of the Chen dataset, the average performance metrics for model MLG include a recall of 98.88%, precision of 98.94%, F1-score of 98.91%, and accuracy of 99.02%. On the Kaggle dataset test set, MLG corresponding metrics are 96.89% for recall, 97.21% for precision, 96.89% for F1-score, and 97.24% for accuracy. These indicators demonstrate that across both the Chen and Kaggle datasets, the MLG model exhibits outstanding classification performance, which further validates the effectiveness and generalization capabilities of the MLG model, enabling it to achieve satisfactory performance in brain tumor classification tasks on diverse datasets.

### 4.4 Ablation study

In Section 4.3, performance metrics for the classification results of the proposed model are presented. To further confirm the validity of the proposed model, an ablation study was performed. In this study, different combinations of modules are explored within the framework of the model. This process allows for a meticulous examination of each component’s contribution to the

overall performance, thereby providing deeper insights into the effectiveness and robustness of the proposed model architecture.

In the first part of the study, brain tumor classification was conducted separately using REMA and Biformer independently. Figure 6 presents the testing results of various models in the Chen dataset during the ablation experiment. The accuracies achieved by REMA and Biformer are 98.53 and 98.37%, respectively, both lower than the 99.02% accuracy obtained by MLG. Upon conducting a detailed analysis of the ablation experiment results, it becomes clear that the integration of the strengths of both the REMA and Biformer modules within the MLG model effectively boosts the accuracy rate in brain tumor classification.

In the second part of the study, the performance of the MLG model upon incorporating the gated attention mechanism was meticulously examined. The gated attention mechanism plays a pivotal role within the model, serving to regulate the flow of information by deciding which pieces of information should be emphasized and which should be disregarded. By means of gating, the attention mechanism assigns weights to information based on its importance, thereby enhancing the model performance by focusing on crucial features. Figure 7 shows the performance of the model with and without the gated attention mechanism. Where, GA stands for Gated Attention. It can be observed that when the model does not include the gated attention, its performance lags behind the version with the gated attention mechanism by 2.12%. The results strongly demonstrate the effectiveness of the gated attention in improving the performance of the model.

In the third segment of the investigation, the impact of data augmentation on the MLG model was thoroughly explored, particularly in scenarios involving small sample datasets. Data augmentation is a critical technique that can significantly enhance a model generalization capability while mitigating overfitting issues. In this work, two prevalent data augmentation strategies were employed: random rotation and random flipping. Figure 8 provides a detailed account of the model accuracy rates on both the training and test sets of the Chen dataset when data augmentation is applied. Ar stands for data augmentation. From the figure, it is evident

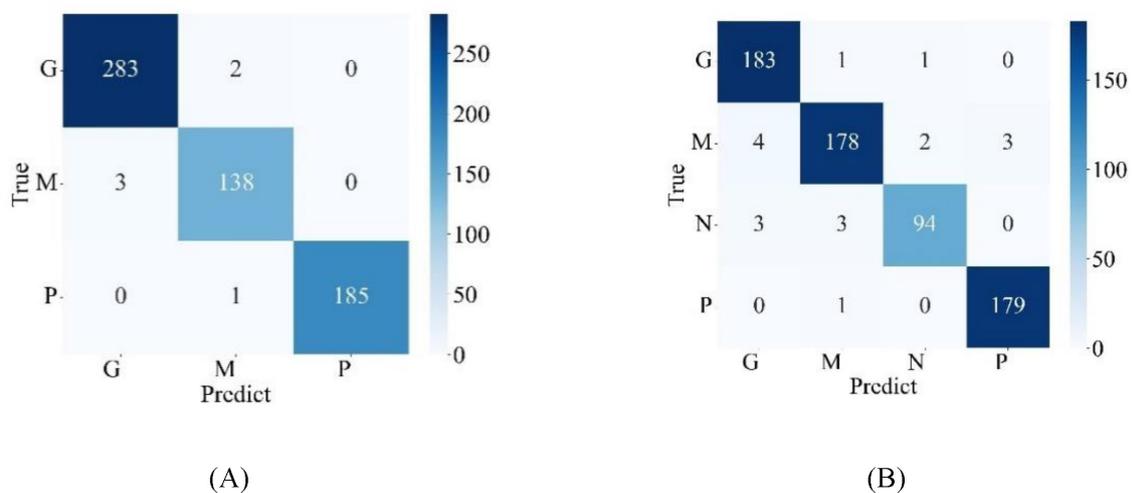


FIGURE 5 Confusion matrix for model classification results (A) Chen dataset (B) Kaggle dataset.

TABLE 4 Detailed values of metrics for the proposed model.

Dataset	Tumor type	Recall (%)	Precision (%)	F1-score (%)	Accuracy (%)
Chen	Glioma	99.30	98.95	99.12	99.02
	Meningioma	97.87	97.87	97.87	
	Pituitary	99.46	1.00	99.73	
	<b>Average</b>	<b>98.88</b>	<b>98.94</b>	<b>98.91</b>	
Kaggle	Glioma	98.92	96.32	97.60	97.24
	Meningioma	95.19	97.27	96.22	
	No Tumor	94.00	96.91	95.43	
	Pituitary	99.44	98.35	98.90	
	<b>Average</b>	<b>96.89</b>	<b>97.21</b>	<b>96.89</b>	

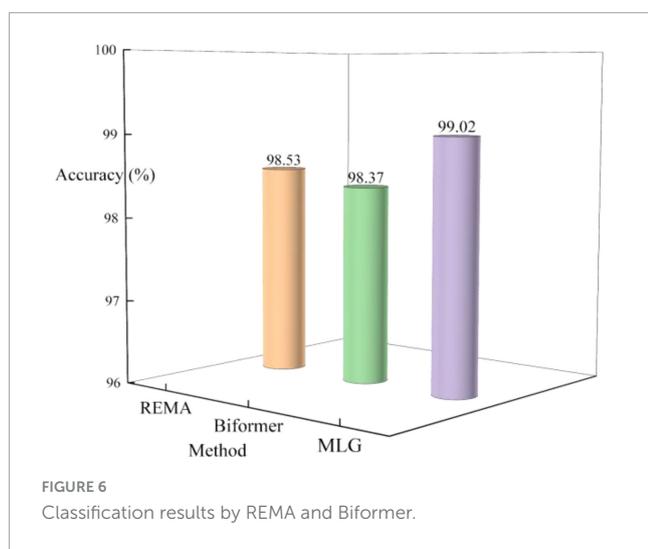


FIGURE 6 Classification results by REMA and Biformer.

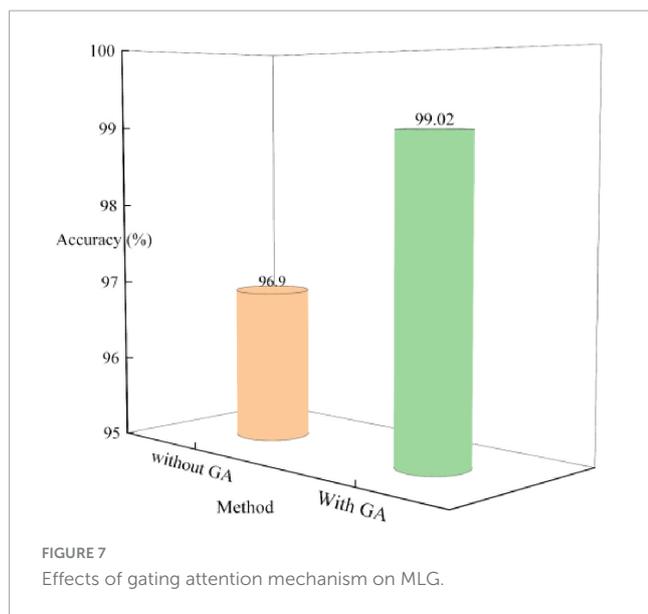


FIGURE 7 Effects of gating attention mechanism on MLG.

that with data augmentation, the training and test set accuracies reach 99.96 and 99.02%, respectively. In contrast, without data augmentation, while the accuracy on the training set reached 100%, the accuracy on the test set notably decreased to 96.73%. This

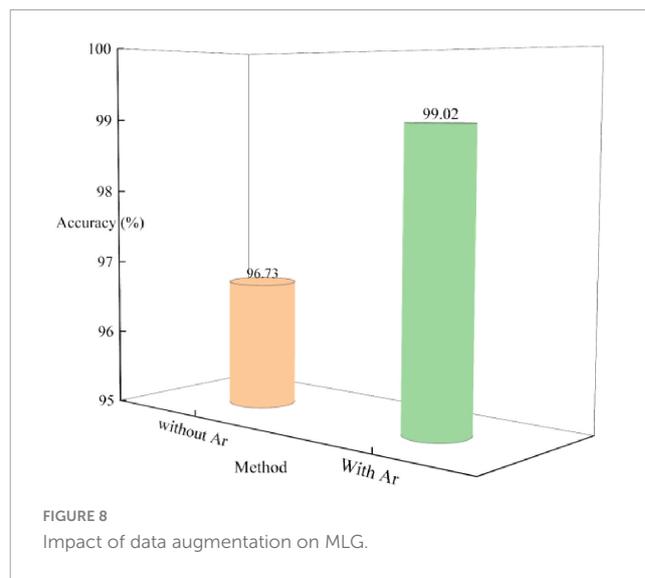


FIGURE 8 Impact of data augmentation on MLG.

comparative outcome vividly demonstrates that data augmentation has a pronounced effect on improving model performance.

## 5 Discussion

According to the data in Table 4, the MLG model achieves impressive accuracies of 99.02% on the Chen dataset and 97.24% on the Kaggle dataset, which attest to its effectiveness and satisfactory performance. Moreover, through ablation studies, the superiority of the MLG model was further substantiated, emphasizing the significant improvements gained by fusing the REMA and Biformer modules via the gated attention mechanism, rather than merely adding them together. Additionally, the application of data augmentation has led to noticeable performance enhancements, further bolstering the model generalization capabilities.

Beyond internal validation, the proposed model was also compared against other advanced methods utilizing the same datasets. Table 5 clearly outlines these comparative results. On the Chen dataset, the MLG model outperforms the current best-performing model, Multimodal-CNN Model (Maqsood et al., 2022), by 0.1% in accuracy. Similarly, on the Kaggle dataset, the MLG model surpasses the previously best-reported model IVX16 (Hossain et al., 2023) by an accuracy margin of 0.3%. When juxtaposed against methodologies outlined in literature sources paper (Alanazi et al., 2022) and paper (Saurav et al., 2023), the MLG model consistently demonstrates higher performance on both the Chen and Kaggle datasets. Precisely, on the Chen dataset, MLG accuracy exceeds that of paper (Alanazi et al., 2022) by 2.13% and that of paper (Saurav et al., 2023) by 1.79%. On the Kaggle dataset, MLG accuracy advantage over paper (Alanazi et al., 2022) is 1.49%, while over (Saurav et al., 2023) it is 1.53%. These comparative results serve as compelling evidence of the MLG model superior performance in the task of brain tumor classification, reinforcing its potential applicability in real-world scenarios.

The Receiver Operating Characteristic Curve (ROC Curve) is a widely used visualization tool in statistics, machine learning, medical diagnostics, and other fields that require categorical judgments for evaluating the performance of classification models.

TABLE 5 Compare with advanced methods on datasets Chen and Kaggle.

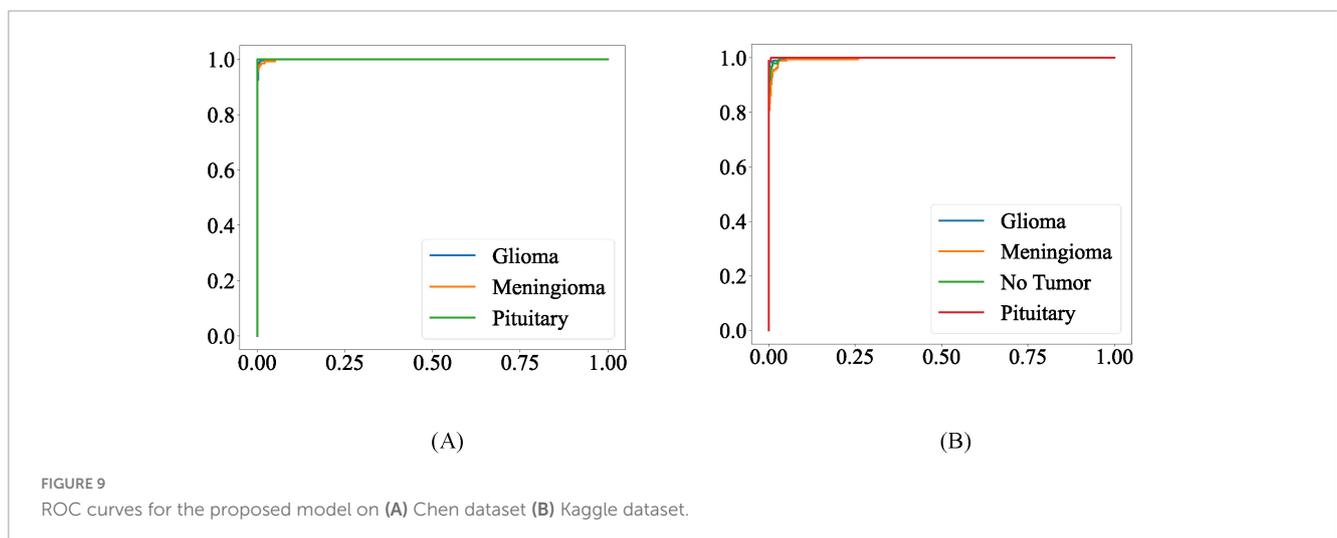
Method category	References	Method	Dataset	Accuracy (%)
CNN	Sachdeva et al. (2024)	Transfer learning	Kaggle	96.25
	Jun and Liyuan (2022)	Attention-Guided	Chen	98.61
	Maqsood et al. (2022)	Multimodal-CNN Model	Chen	98.92b
	Alanazi et al. (2022)	22-layer CNN	Chen	96.89
			Kaggle	95.75
	Saurav et al. (2023)	AG-CNN	Chen	97.23
Transformer			Kaggle	95.71
	Wang et al. (2024)	RanMerFormer	Chen	98.86
	Şahin et al. (2024)	BMO	Chen	98.09
	Hossain et al. (2023)	IVX16	Kaggle	96.94
	Anaya-Isaza et al. (2023)	Cross-Transformer	Chen	97.22
	Dosovitskiy et al. (2021)	Vision Transformer	Chen	97.39
			Kaggle	95.88
	Liu et al. (2021a)	Swin Transformer	Chen	98.69
CNN+transformer			Kaggle	97.10
	Ferdous et al. (2023)	LCDEiT	Chen	98.11
	Chen et al. (2025)	EnSLDe	Chen	98.69
	Proposed model	MLG	Chen	99.02
		Kaggle	97.24	

It graphically illustrates the trade-off relationship between the true positive rate (TPR) and false positive rate (FPR) of the model under different threshold conditions. The area under curve (AUC), indicates better model performance when its value is larger.

Typically, the closer the curve is to the upper left corner (with higher TPR and lower FPR), the better the model performance. The ROC curves of the model on the two datasets are shown in Figure 9. It can be observed that the ROC curves closely adhere to the upper left corner. On the Chen dataset, the AUC values of the MLG model for glioma, meningioma, and pituitary tumors are 0.9996, 0.9993, and 1.00, respectively. Meanwhile, on the Kaggle dataset, the AUC values of the MLG model for glioma, meningioma, normal tissue, and pituitary tumors are 0.9991, 0.9965, 0.9989, and 0.9999, respectively.

## 6 Conclusion

Brain tumors, constituting a severe health issue, pose a significant threat to people's lives. Therefore, timely and accurate identification of brain tumor types, followed by appropriate treatment planning, is critical for patients. The advent of CAD technology has provided substantial support to doctors in diagnosing brain tumors. In this paper, a novel MLG brain tumor classification model is proposed, and the model skillfully integrates local features and global features, and provides a new solution for the classification of brain tumors. The core components of the MLG model are RMEA, Biformer and gated attention. The RMEA Block, through carefully designed convolutional structures, efficiently retains information across channels, emphasizing spatial and channel-wise features, thereby extracting richly informative local features. Conversely, the Biformer employs a unique BRA mechanism to dynamically and contextually select a subset of the most relevant key-value pairs for each query, optimizing the computational process. Meanwhile, BRA can capture remote dependencies across regions and even objects, providing powerful support for extracting global features. The MLG model uses a gated attention to selectively filter and fuse the local features extracted by the RMEA block with the global features extracted by the Biformer block. This significantly enhances the representation capability of the fused features, thereby improving the classification performance of the model. The integration of both local and global features enables the MLG model to exhibit outstanding



performance in brain tumor classification tasks. Experimental results on two public datasets demonstrate that the MLG model achieves satisfactory performance across multiple metrics, including accuracy, precision, recall, and F1-score. Compared with existing advanced methods, the MLG model exhibits marked advantages, fully validating its effectiveness in practical applications. In future work, it is planned to continue exploring other methods of feature fusion first to further improve the performance of the MLG model. Secondly, the introduction of more refined feature detection methods will be explored, or they will be combined with other advanced attention mechanisms to enhance the selection ability for key areas. In addition, efforts will also be made to obtain data on other brain diseases, expand the application scope of the model, and provide more auxiliary diagnostic tools for the medical field.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: the datasets used are free and open. Dataset Chen from figshare ([https://figshare.com/articles/dataset/brain\\_tumor\\_dataset/1512427](https://figshare.com/articles/dataset/brain_tumor_dataset/1512427)). Dataset Kaggle from Kaggle (<https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri>).

## Author contributions

WC: Project administration, Conceptualization, Visualization, Writing – review & editing, Investigation. XT: Formal Analysis, Software, Writing – original draft. JZ: Conceptualization, Project administration, Writing – review & editing, Software. GD: Project administration, Supervision, Writing – review & editing. QF: Writing – review & editing, Validation. HJ: Writing – review & editing, Project administration, Validation.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was

funded by the Henan Province Young Backbone Teachers Training Program (No. 2023GGJS045), the Major Science and Technology Projects of Henan Province (No. 221100210500), the Foundation of Henan Educational Committee (No. 24A320004), the Medical and Health Research Project in Luoyang (No. 2001027A), and the Construction Project of Improving Medical Service Capacity of Provincial Medical Institutions in Henan Province (No. 2017-51).

## Acknowledgments

The provision of these two public datasets by Kaggle and Chen is greatly appreciated by us.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Akter, A., Nosheen, N., Ahmed, S., Hossain, M., Yousof, M., Almoayad, A., et al. (2024). Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor. *Expert Syst. Appl.* 238:122347. doi: 10.1016/j.eswa.2023.122347
- Alanazi, M., Ali, M., Hussain, S., Zafar, A., Mohatram, M., Irfan, M., et al. (2022). Brain tumor/mass classification framework using magnetic-resonance-imaging-based isolated and developed transfer deep-learning model. *Sensors* 22:372. doi: 10.3390/s22010372
- Alturki, N., Umer, M., Ishaq, A., Abuzinadah, N., Alnowaiser, K., Mohamed, A., et al. (2023). Combining CNN features with voting classifiers for optimizing performance of brain tumor classification. *Cancers* 15:1767. doi: 10.3390/cancers15061767
- Anaya-Isaza, A., Mera-Jiménez, L., Verdugo-Alejo, L., and Sarasti, L. (2023). Optimizing MRI-based brain tumor classification and detection using AI: A comparative analysis of neural networks, transfer learning, data augmentation, and the cross-transformer network. *Eur. J. Radiol. Open* 10:100484. doi: 10.1016/j.ejro.2023.100484
- Asiri, A., Shaf, A., Ali, T., Pasha, M., Khan, A., Irfan, M., et al. (2024). Advancing brain tumor detection: Harnessing the Swin Transformer's power for accurate classification and performance analysis. *PeerJ Comput. Sci.* 10:e1867. doi: 10.7717/peerj-cs.1867
- Azeem, M., Kiani, K., Mansouri, T., and Topping, N. (2024). SkinLesNet: Classification of skin lesions and detection of melanoma cancer using a novel multi-layer deep convolutional neural network. *Cancers* 16:108. doi: 10.3390/cancers16010108
- Bhuvaji, S., Kadam, A., Bhumkar, P., Dedge, S., and Kanchan, S. (2020). *Brain tumor classification (MRI)*. San Francisco, CA: Kaggle, doi: 10.34740/KAGGLE/DSV/1183165
- Cao, L., Pan, K., Ren, Y., Lu, R., and Zhang, J. (2024). Multi-branch spectral channel attention network for breast cancer histopathology image classification. *Electronics* 13:459. doi: 10.3390/electronics13020459

- Chen, W., Liu, J., Tan, X., Zhang, J., Du, G., Fu, Q., et al. (2025). EnSLDe: An enhanced short-range and long-range dependent system for brain tumor classification. *Front. Oncol.* 15:1512739. doi: 10.3389/fonc.2025.1512739
- Cheng, J., Huang, W., Cao, S., Yang, Ru, Yang, W., Yun, Z., et al. (2015). Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PLoS One* 10:e0140381. doi: 10.1371/journal.pone.0140381
- Chincholi, F., and Koestler, H. (2024). Transforming glaucoma diagnosis: Transformers at the forefront. *Front. Artif. Intell.* 7:1324109. doi: 10.3389/frai.2024.1324109
- Dhar, T., Dey, N., Borra, S., and Sherratt, R. S. (2023). Challenges of deep learning in medical image analysis—improving explainability and trust. *IEEE Trans. Technol. Soc.* 4, 68–75. doi: 10.1109/TTS.2023.3234203
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv [Preprint]*. doi: 10.48550/arXiv.2010.11929 arXiv:2010.11929.
- Fang, M., Fu, M., Liao, B., Lei, X., and Wu, F.-X. (2024). Deep integrated fusion of local and global features for cervical cell classification. *Comput. Biol. Med.* 171:108153. doi: 10.1016/j.combiomed.2024.108153
- Ferdous, G. J., Sathi, K. A., Hossain, A., Hoque, M. M., and Dewan, M. A. A. (2023). LCDEiT: A linear complexity data-efficient image transformer for MRI brain tumor classification. *IEEE Access* 11, 20337–20350. doi: 10.1109/ACCESS.2023.3244228
- Gade, V. S. R., Cherian, R. K., Rajarao, B., and Kumar, M. A. (2024). BMO based improved Lite Swin transformer for brain tumor detection using MRI images. *Biomed. Signal Process. Control* 92:91. doi: 10.1016/j.bspc.2024.106091
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR)*, (Las Vegas, NV: IEEE), 770–778. doi: 10.1109/CVPR.2016.90
- Hossain, S., Chakrabarty, A., Gadekallu, T. R., Alazab, M., and Piran, J. (2023). Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection and classification. *IEEE J. Biomed. Health Inform.* 28, 1261–1272. doi: 10.1109/JBHI.2023.3266614
- Hu, J., Shen, L., and Sun, G. (2018). “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (Salt Lake City, UT), 7132–7141.
- Huang, S. K., Yu, Y.-T., Huang, C.-R., and Cheng, H.-C. (2024). Cross-scale fusion transformer for histopathological image classification. *IEEE J. Biomed. Health Inform.* 28, 297–308. doi: 10.1109/JBHI.2023.3322387
- Huang, L., Xu, Y., Wang, S., Sang, L., and Ma, H. (2024). SRT: Swin-residual transformer for benign and malignant nodules classification in thyroid ultrasound images. *Med. Eng. Phys.* 124:104101. doi: 10.1016/j.medengphy.2024.104101
- Jun, W., and Liyuan, Z. (2022). Brain tumor classification based on attention guided deep learning model. *Int. J. Comput. Intell. Syst.* 15:35. doi: 10.1007/s44196-022-00090-9
- Kang, J., Ullah, Z., and Gwak, J. (2021). MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers. *Sensors* 21:2222. doi: 10.3390/s21062222
- Kaur, P., and Mahajan, P. (2025). Detection of brain tumors using a transfer learning-based optimized ResNet152 model in MR images. *Comput. Biol. Med.* 188:109790. doi: 10.1016/j.combiomed.2025.109790
- Kshatri, S. S., and Singh, D. (2023). Convolutional neural network in medical image analysis: A review. *Arch. Comput. Methods Eng.* 30, 2793–2810. doi: 10.1007/s11831-023-09898-w
- Li, Z., and Zhou, X. (2025). A global-local parallel dual-branch deep learning model with attention-enhanced feature fusion for brain tumor MRI classification. *CMC Comput. Mater. Contin.* 83, 739–760. doi: 10.32604/cmc.2025.059807
- Liu, H., Huo, G., Li, Q., Guan, X., and Tseng, M. (2023). Multiscale lightweight 3D segmentation algorithm with attention mechanism: Brain tumor image segmentation. *Expert Syst. Appl.* 214:9166. doi: 10.1016/j.eswa.2022.119166
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021b). “Swin transformer: Hierarchical vision transformer using shifted windows,” in *Proceedings of the 2021 IEEE/CVF international conference on computer vision (ICCV)*, (Montreal, QC: IEEE), 9992–10002. doi: 10.1109/ICCV48922.2021.00986
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021a). Swin transformer: Hierarchical vision transformer using shifted windows. *arXiv [Preprint]*. doi: 10.48550/arXiv.2103.14030 arXiv:2103.14030.
- Lyu, I. J., Han, K., Park, K.-A., and Oh, S. Y. (2024). Ocular motor cranial nerve palsies and increased risk of primary malignant brain tumors: South Korean national health insurance data. *Cancers* 16:781. doi: 10.3390/cancers16040781
- Maqsood, S., Damaševičius, R., and Maskeliūnas, R. (2022). Multi-modal brain tumor detection using deep neural network and multiclass SVM. *Medicina* 58:1090. doi: 10.3390/medicina58081090
- Mazurowski, M. A., Dong, H., Gu, H., Yang, J., Konz, N., and Zhang, Y. (2023). Segment anything model for medical image analysis: An experimental study. *Med. Image Anal.* 89:102918. doi: 10.1016/j.media.2023.102918
- Mehnatkesh, H., Jalali, S. M. J., Khosravi, A., and Nahavandi, S. (2023). An intelligent driven deep residual learning framework for brain tumor classification using MRI images. *Expert Syst. Appl.* 213:119087. doi: 10.1016/j.eswa.2022.119087
- Mukadam, S. B., and Patil, H. Y. (2024). Machine learning and computer vision based methods for cancer classification: A systematic review. *Arch. Comput. Methods Eng.* 31, 3015–3050. doi: 10.1007/s11831-024-10065-y
- Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., et al. (2023). “Efficient multi-scale attention module with cross-spatial learning,” in *Proceedings of the ICASSP 2023—2023 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, (Rhodes), 1–5. doi: 10.1109/ICASSP49357.2023.10096516
- Pandiselvi, T., and Maheswaran, R. (2019). Efficient framework for identifying, locating, detecting and classifying MRI brain tumor in MRI images. *J. Med. Syst.* 43:189. doi: 10.1007/s10916-019-1253-1
- Peng, J., Lu, J., Zhuo, J., and Li, P. (2024). Multi-scale-denoising residual convolutional network for retinal disease classification using OCT. *Sensors* 24:150. doi: 10.3390/s24010150
- Poornam, S., and Angelina, J. J. R. (2024). VITALT: A robust and efficient brain tumor detection system using vision transformer with attention and linear transformation. *Neural Comput. Appl.* 36, 6403–6419. doi: 10.1007/s00521-023-09306-1
- Sachdeva, J., Sharma, D., and Ahuja, C. K. (2024). Comparative analysis of different deep convolutional neural network architectures for classification of brain tumor on magnetic resonance images. *Arch. Comput. Methods Eng.* 31, 1959–1978. doi: 10.1007/s11831-023-10041-y
- Şahin, E., Özdemir, D., and Temurtaş, H. (2024). Multi-objective optimization of ViT architecture for efficient brain tumor classification. *Biomed. Signal Process. Control* 91:105938. doi: 10.1016/j.bspc.2023.105938
- Saurav, S., Sharma, A., Saini, R., and Singh, S. (2023). An attention-guided convolutional neural network for automated classification of brain tumor from MRI. *Neural Comput. Appl.* 35, 2541–2560. doi: 10.1007/s00521-022-07742-z
- Shafiq, M., and Gu, Z. (2022). Deep residual learning for image recognition: A survey. *Appl. Sci. Basel* 12:8972. doi: 10.3390/app12188972
- Sharma, P., Nayak, D. R., Balabantaray, B. K., Tanveer, M., and Nayak, R. (2024). A survey on cancer detection via convolutional neural networks: Current challenges and future directions. *Neural Netw.* 169, 637–659. doi: 10.1016/j.neunet.2023.11.006
- Song, B., Kc, D. R., Yang, R. Y., Li, S., Zhang, C., and Liang, R. (2024). Classification of mobile-based oral cancer images using the vision transformer and the Swin transformer. *Cancers* 16:987. doi: 10.3390/cancers16050987
- ThamilSelvi, C., Vinoth Kumar, S., Asaad, R. R., Palanisamy, P., and Rajappan, L. K. (2025). An integrative framework for brain tumor segmentation and classification using neural networks. *Intell. Data Anal.* 29, 435–458. doi: 10.3233/IDA-240108
- Verma, A., and Yadav, A. K. (2025). FusionNet: Dual input feature fusion network with ensemble based filter feature selection for enhanced brain tumor classification. *Brain Res.* 1852:149507. doi: 10.1016/j.brainres.2025.149507
- Wang, J., Lu, S.-Y., Wang, S.-H., and Zhang, Y.-D. (2024). RanMerFormer: Randomized vision transformer with token merging for brain tumor classification. *Neurocomputing* 573:127216. doi: 10.1016/j.neucom.2023.127216
- Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). “CBAM: Convolutional block attention module,” in *Computer vision – ECCV 2018*, Vol. 11211, eds V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss (Cham: Springer International Publishing), 3–19. doi: 10.1007/978-3-030-01234-2\_1
- Yan, Q., Liu, S., Xu, S., Dong, C., Li, Z., Shi, J., et al. (2023). 3D medical image segmentation using parallel transformers. *Pattern Recogn.* 138:109432. doi: 10.1016/j.patcog.2023.109432
- Yu, X., Wang, J., Hong, Q.-Q., Teku, R., Wang, S.-H., and Zhang, Y.-D. (2022). Transfer learning for medical images analyses: A survey. *Neurocomputing* 489, 230–254. doi: 10.1016/j.neucom.2021.08.159
- Yu, Y., Zhang, Y., Cheng, Z., Song, Z., and Tang, C. (2023). MCA: Multidimensional collaborative attention in deep convolutional neural networks for image recognition. *Eng. Appl. Artif. Intell.* 126:107079. doi: 10.1016/j.engappai.2023.107079
- Zebari, N. A., Mohammed, C., Zebari, D., Mohammed, M., Zeebaree, D., Marhoon, H., et al. (2024). A deep learning fusion model for accurate classification of brain tumours in magnetic resonance images. *CAAI Trans. Intell. Technol.* 9:76. doi: 10.1049/cit2.12276
- Zhou, L., Jiang, Y., Li, W., Hu, J., and Zheng, S. (2024). Shape-scale co-awareness network for 3d brain tumor segmentation. *IEEE Trans. Med. Imaging* 43, 2495–2508. doi: 10.1109/TMI.2024.3368531
- Zhu, L., Wang, X., Ke, Z., Zhang, W., and Lau, R. (2023). “BiFormer: Vision transformer with bi-level routing attention,” in *Proceedings of the 2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, (Vancouver, BC: IEEE), 10323–10333. doi: 10.1109/CVPR52729.2023.00995
- Zulficar, F., Bajwa, U. I., and Mehmood, Y. (2023). Multi-class classification of brain tumor types from MR images using EfficientNets. *Biomed. Signal Process. Control* 84:104777. doi: 10.1016/j.bspc.2023.104777