Scientific Research Publishing

Journal of Computer and Communications, 2026, 14(3), 149-173
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

A Lightweight MobileViT with a Dual-Path Attention Mechanism for MRI Image Classification

Youji Xu, Siyu Xiang, Huifang Feng

College of Mathematics and Statistics, Northwest Normal University, Lanzhou, China
Email: hffeng@nwnu.edu.cn

How to cite this paper: Xu, Y.J., Xiang, S.Y. and Feng, H.F. (2026) A Lightweight MobileViT with a Dual-Path Attention Mechanism for MRI Image Classification. *Journal of Computer and Communications*, **14**, 149-173.
https://doi.org/10.4236/jcc.2026.143008

Received: March 4, 2026
Accepted: March 23, 2026
Published: March 26, 2026

Abstract

Deep learning has been successfully applied in the field of medical diagnosis, and improving the accurate classification of MRI images through deep learning is important for early treatment and patient prognosis. Aiming at the current deep learning-based MRI image classification algorithms with large parameter counts and high computational complexity, a lightweight MobileViT with a dual-path attention mechanism for MRI image classification is proposed in this paper. Embedding the Convolutional Block Attention Module (CBAM) in the original MobileViT network enhances the extraction of key feature information by attending to both the channel and spatial dimensions of the feature map. A Dual-Path Attention Module (DPAM) is constructed by integrating CSPNet with the CBAM mechanism to further enhance the potential of feature extraction of the proposed model while maintaining a minimal parameter count. The proposed model also employs a transfer learning method to accelerate the learning speed of the network model on the MRI image datasets, and uses a cosine annealing algorithm to optimize the learning rate of the model during the model training process to help the model converge better. The state-of-the-art performance of the proposed model is validated on the Alzheimer's disease and brain tumor MRI datasets, respectively. We evaluate the performance of our proposed model with the latest deep learning models. The experimental results show that the model not only substantially enhances the accuracy of MRI image classification but also exhibits reduced computational complexity, making it highly suitable for mobile devices with constrained computing resources.

Keywords

MRI Image Classification, MobileViT, Attention Mechanism, Data Enhancement, Transfer Learning, Lightweight

1. Introduction

Alzheimer's Disease (AD) is a degenerative brain disease, which means it worsens over time [1]. One of the most notable symptoms of Alzheimer's disease is memory loss and a progressive decline in cognitive function. Patients may gradually forget familiar people, places, and daily activities, which can seriously affect quality of life. There is no effective cure for Alzheimer's disease, and existing medications and treatments are primarily designed to slow the progression of the disease and relieve symptoms.

Brain tumors are a group of abnormal tissue growths that form in the central nervous system, such as the brain, brainstem, and spinal cord, and they are one of the top ten malignant tumors in terms of morbidity and mortality today [2]. The main types of brain tumors are meningiomas, pituitary tumors, and gliomas, of which the most important and common is glioma, which not only affects the neuroglial cells but also invades other surrounding tissues. Brain tumor is a serious problem that endangers human health. As the tumor grows, the patient's intracranial pressure increases, sometimes leading to brain damage or even death. Therefore, timely detection and accurate determination of brain tumor type play an important role in treatment planning and patient care.

Magnetic Resonance Imaging (MRI) of the head is a medical imaging technique that allows MRI to obtain three-dimensional images of the skull without the use of X-rays, and it has an important role in the diagnosis of brain disorders. The diagnosis of both brain tumors and Alzheimer's disease can be determined by the patient's MRI images. In recent years, deep learning has been widely used in the auxiliary diagnosis of medical images and has achieved good results. The use of computer vision technology to assist doctors in reading medical images can reduce the burden on doctors and improve diagnostic efficiency. Therefore, it is of great application value to carry out research on MRI image classification based on deep learning.

Convolutional Neural Networks (CNNs) are among the commonly used deep learning methods for MRI image classification. Asgharzadeh-Bonab *et al.* [3] proposed two feature fusion schemes, decision-level and feature-level, to combine different input information and used different CNN network models to classify MRI images under different features. The experimental results showed that EfficientNet-B7 had a better classification effect. Zhang *et al.* [4] introduced an augmented neural network, ADnet, built upon VGG16. This model employed depthwise separable convolutions in place of conventional convolutions to decrease the number of parameters, and incorporated the ELU activation function instead of the ReLU to mitigate the risk of gradient explosion. Experimental results confirmed that these enhancements led to improved accuracy across various classification tasks. Yang *et al.* [5] proposed a new region-to-sample graph convolutional neural network framework based on graph convolutional neural networks. Qian *et al.* [6] proposed a 3D residual network with multi-scale and an attention module for multi-task learning, which used a 3D network to avoid subjectivity when

manually selecting slices, and also preserved the spatial structure information of the 3D data. Ait Amou *et al.* [7] considered the complexity of hyperparameter tuning for CNN networks. An efficient hyperparameter optimization technique for CNN based on Bayesian optimization was proposed, and the model showed excellent classification results on three categories of brain tumor image datasets. Ozdemir [8] devised a novel deep convolutional neural network architecture to address the brain tumor classification challenge, achieving successful classification across three distinct types of brain tumors.

Although CNN-based network models have excellent performance in classification accuracy on MRI image datasets, the prevalent CNN models have large parameter counts and are not favorable for practical applications. Therefore, more and more researchers are devoted to the study of lightweight MRI image classification models [9]. Zhang *et al.* [10] proposed a lightweight neural network approach based on ShuffleNet and introduced the ECA attention mechanism to achieve efficient and scalable automatic Alzheimer's disease detection. Khatri *et al.* [11] combined the convolutional attention mechanism and the Transformer to design a lightweight Alzheimer's disease diagnostic model, in which lightweight multi-head attention was used instead of multi-head attention, which improved model performance without consuming too many computational resources. Liu *et al.* [12] introduced a lightweight automated 3D algorithm featuring an attention mechanism for segmenting brain tumor images. Specifically, this study used hierarchical decoupled convolution instead of standard convolution to reduce the number of parameters in the model. Dilation convolution was incorporated to augment the network's capability to capture multi-scale information within the bottom convolution module, and an attention mechanism was also introduced to improve model accuracy. Vaiyapuri *et al.* [13] used an integrated model of EfficientNet, DenseNet, and MobileNet for feature extraction of brain tumor images. Luo *et al.* [14] tackled the challenge posed by the extensive parameter count and computational complexity inherent in CNN models by introducing a lightweight brain tumor segmentation network. This network incorporated multi-view extraction and dense attention mechanisms, mitigating the issues associated with the large parameter set and computational demands. Egaz *et al.* [15] developed a lightweight CNN-LSTM model for the diagnosis of Alzheimer's disease. Nizamani *et al.* [16] proposed a lightweight deep fusion model for brain tumor classification tasks, integrating the Lightweight Feature Extraction Module (LEM), Cross-Stream Attention (CSA), Feature Fusion Module (FFM), and Attention Prediction Head (APH).

Although deep convolutional neural networks have achieved some results in the task of image classification, on the whole, traditional convolutional neural networks still have some limitations, such as excessive model complexity, low classification accuracy, and loss of key information about the lesion region in the images. To address these shortcomings, we propose a lightweight MobileViT with a dual-path attention mechanism for MRI image classification. The model is an im-

proved version based on the MobileViT model, which maintains excellent image classification results while keeping a compact structure and small computational overhead, striking an optimal balance between performance and efficiency.

The main contributions of this paper are outlined as follows:

1) A lightweight MobileViT with a dual-path attention mechanism is proposed to solve the problem of large parameter counts and high computational complexity in current deep learning-based MRI image classification.

2) A dual-path attention module is constructed by integrating CSPNet and CBAM mechanisms to further extract detailed features of lesion regions in MRI images and improve the classification accuracy without increasing the computational cost.

3) The transfer learning method is employed to pre-train the model on the ImageNet dataset, which not only discovers more features but also accelerates the learning speed of the network model on brain tumor images. Meanwhile, to improve the robustness and generalization of the model, the cosine annealing algorithm is used to optimize the learning rate of the model during the training process.

Extensive experiments are conducted on Alzheimer's disease and brain tumor MRI image datasets, and the state-of-the-art performance of the model is evaluated by comparing it with the latest deep learning models.

The structure of this paper is organized as follows: In Section 2, we introduce the proposed MRI image classification model, detailing its architecture and key components. Section 3 describes the dataset used in the experiments, along with the data preprocessing techniques and the experimental setup. Section 4 provides an overview of the experimental environment, detailing the hardware and software configurations used. It also presents a comprehensive analysis of the experimental results, comparing the performance of our model with that of existing approaches. Section 5 addresses the limitations of the proposed model, highlighting areas that could benefit from further improvement. Finally, Section 6 provides a summary of the key findings and outlines potential directions for future research to enhance the model's performance and applicability.

2. Methodology

In order to solve the problem of large parameter counts and high computational complexity in current deep learning-based MRI image classification, we propose a lightweight MobileViT with a dual-path attention mechanism for MRI image classification.

2.1. Overall Framework of the Proposed Model

Figure 1 depicts the architecture of our proposed model. The specific parameter information for each module in the proposed model is shown in Table 1. The key components of the proposed model include the MobileViT, MV2, CBAM, and dual-path attention modules. These modules are described in the following sections.

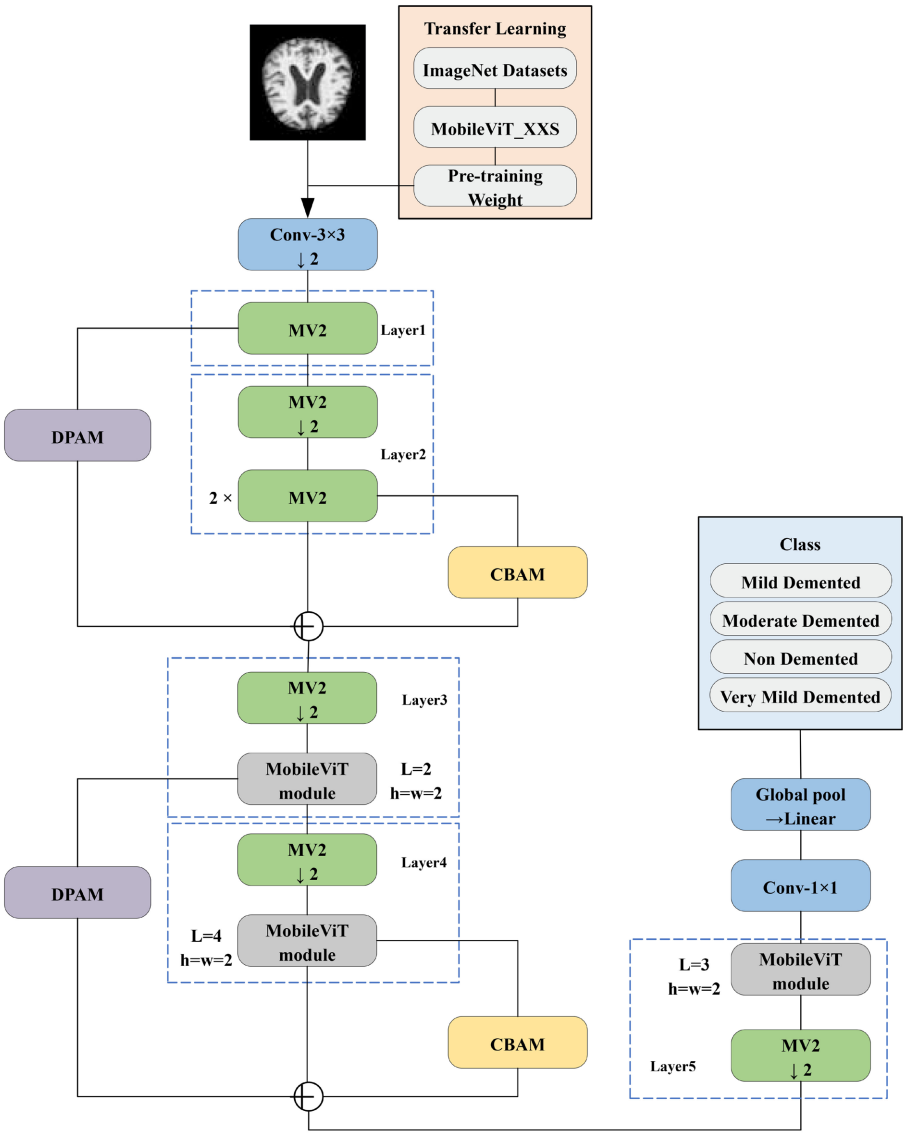

Figure 1. The structure of the proposed model.

Table 1. Parameter information for each module.

Module		Input Feature Matrix	Input Size	Output Feature Matrix	Output Size
Conv-3 × 3 ↓2		X_0^{in}	[224, 224, 3]	X_0^{out}	[112, 112, 16]
Layer1	MV2 ↓2	X_0^{out}	[112, 112, 16]	X_1^{out}	[112, 112, 16]
DPAM		X_1^{out}	[112, 112, 16]	X_{D1}^{out}	[56, 56, 24]
Layer2	MV2 ↓2	X_1^{out}	[112, 112, 16]	$\tilde{X}_2^{out}$	[56, 56, 24]
	MV2 × 2	$\tilde{X}_2^{out}$	[56, 56, 24]	X_2^{out}	[56, 56, 24]
CBAM		X_2^{out}	[56, 56, 24]	X_{C1}^{out}	[56, 56, 24]
Add⊕		X_2^{out}, X_{D1}^{out}, X_{C1}^{out}	[56, 56, 24]	X_3^{in}	[56, 56, 24]
Layer3	MV2 ↓2	X_3^{in}	[56, 56, 24]	$\tilde{X}_3^{out}$	[28, 28, 48]
	MobileViT	$\tilde{X}_3^{out}$	[28, 28, 48]	X_3^{out}	[28, 28, 48]

Continued

DPAM		X_3^{out}	[28, 28, 48]	X_{D2}^{out}	[14, 14, 64]
Layer4	MV2 ↓2	X_3^{out}	[28,28,48]	$\tilde{X}_4^{out}$	[14,14,64]
	MobileViT	$\tilde{X}_4^{out}$	[14,14,64]	X_4^{out}	[14,14,64]
CBAM		X_4^{out}	[14, 14, 64]	X_{C2}^{out}	[14, 14, 64]
Add⊕		X_4^{out}, X_{D2}^{out}, X_{C2}^{out}	[14, 14, 64]	X_5^{in}	[14, 14, 64]
Layer5	MV2 ↓2	X_5^{in}	[14,14,64]	$\tilde{X}_5^{out}$	[7,7,80]
	MobileViT	$\tilde{X}_5^{out}$	[7,7,80]	X_5^{out}	[7,7,80]
Conv-1 × 1		X_5^{out}	[7, 7, 80]	X_6^{out}	[7, 7, 320]
Global pool → Linear		X_6^{out}	[7, 7, 320]	X_7^{out}	[1, 4]

2.2. MobileViT Module

The MobileViT module is the core of the MobileViT model [17] and is depicted in Figure 2. The main calculation process in the MobileViT module can be summarized in the following four steps:

1) An input feature matrix $X \in R^{H\times W\times C}$ is applied to a $n\times n$ convolution layer and followed by a 1 × 1 convolution layer. The $n\times n$ convolution layer is employed to capture local spatial information within the feature map, while the 1 × 1 convolution is utilized to project the feature map into a higher-dimensional feature space. After two convolution operations on the input feature map X, the local representations $X_L \in R^{H\times W\times d}$ are obtained.

2) X_L is transformed into a sequence of non-overlapping flattened patches denoted as $X_U \in R^{N\times P\times d}$, where $P = wh$, $N = HW/P$, (w,h) are the height and width of image patches.

3) The global inter-patch relationship representation $X_G \in R^{N\times P\times d}$ is obtained as follows:

$$X_G(p) = \text{Transformer}\left(X_U(p)\right), 1 \le p \le P \tag{1}$$

4) $X_F \in R^{H\times W\times d}$ is obtained by folding X_G. Then X_F is mapped back to the original feature space using a 1 × 1 convolution layer and is combined with X. Subsequently, the combined features are integrated through a 3 × 3 convolution layer.

2.3. MV2 Module

The MV2 module is the inverted residual module in MobileNetV2 [18], and its structure is shown in Figure 3. First, the traditional convolution layer is replaced with a depthwise separable convolution in MobileNetV2. The depthwise separable convolution consists of two stages: the first stage is depthwise convolution followed by pointwise convolution. This technique effectively decomposes the convolution operation and improves computational efficiency while maintaining expressive power. The depthwise convolution (DWConv) is responsible for extracting features within each channel, and pointwise convolution (Conv) fuses features

between channels. Depthwise separable convolution significantly diminishes the number of parameters and computations needed to attain lightweight networks. The inverted residual with linear bottleneck used in the MV2 is the opposite of the bottleneck structure in ResNet. Initially, we expand the feature maps using a convolution, followed by the extraction of features through a depthwise convolution. Subsequently, a convolution is employed to reduce the number of channels.

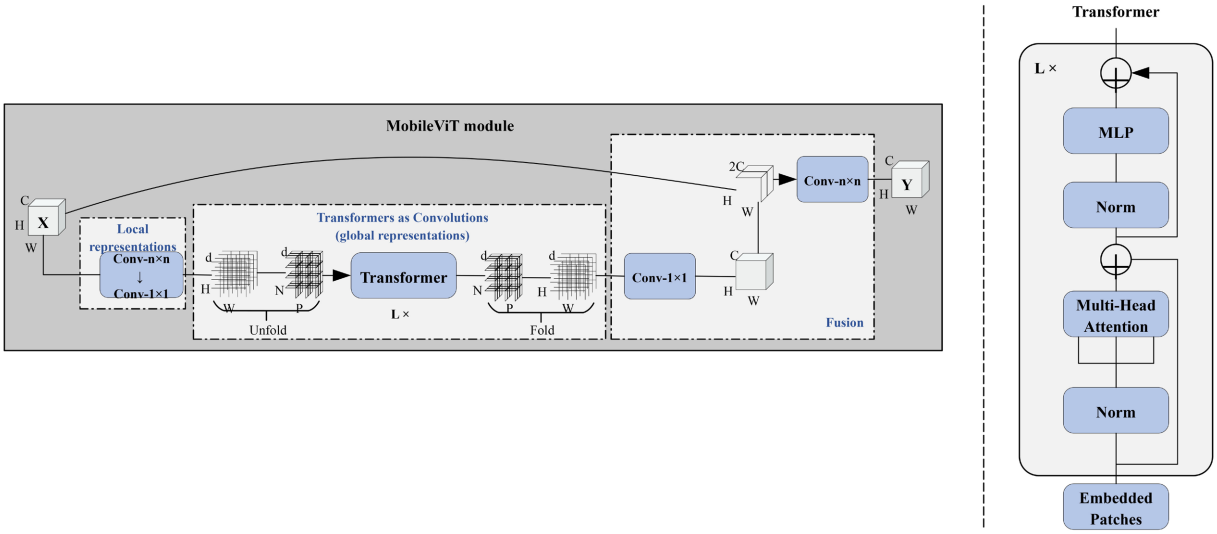

Figure 2. The structure of the MobileViT module.

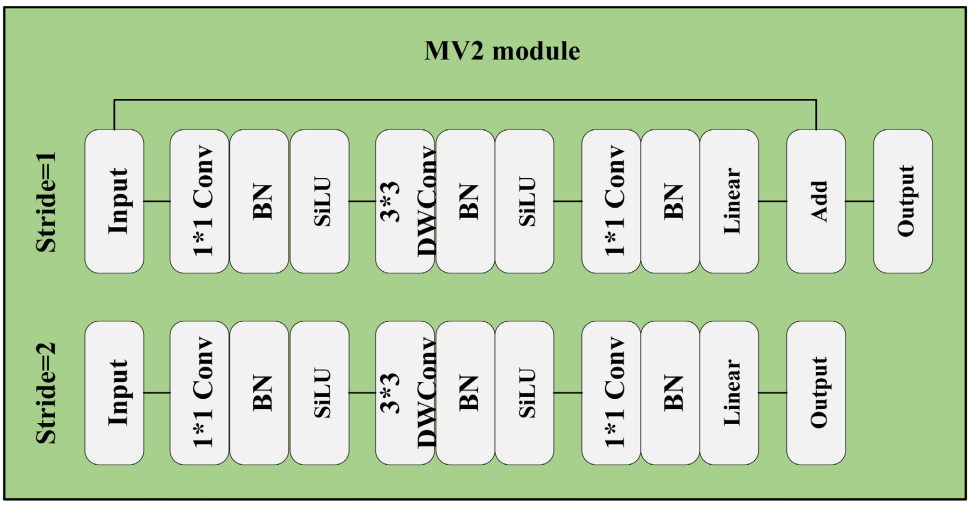

Figure 3. The structure of the MV2 module.

Within the MV2 module, the activation function utilized is ReLU6, a modified version of the Rectified Linear Unit (ReLU). The ReLU function solves the gradient vanishing problem for positive inputs; however, it encounters the challenge of having a constant derivative of 0 for negative inputs, leading to a gradient vanishing problem in negative intervals, which inhibits the updating of many neurons. Therefore, in this paper, we replace the commonly used ReLU activation function

with the SiLU (Sigmoid Linear Unit) activation function, which is an enhanced version combining the advantages of both Sigmoid and ReLU. SiLU offers several important properties, including smoothness, the absence of an upper bound (while still maintaining a lower bound), and non-monotonicity. These characteristics make SiLU more suitable for deeper networks, as it helps to mitigate issues like the vanishing gradient problem and improves the flow of gradients during training. The SiLU activation function is mathematically represented in Equation (2).

$$\mathrm{SiLU}(x) = \frac{x}{1+\mathrm{e}^{-x}} \tag{2}$$

2.4. Convolutional Block Attention Module

The Convolutional Block Attention Module (CBAM) [19] is a powerful composite attention mechanism designed to improve the model's representation capability by incorporating both channel and spatial attention. CBAM operates in two stages: the first stage focuses on channel attention, which enables the model to weigh the importance of different feature channels, and the second stage applies spatial attention, which allows the model to focus on important regions in the feature map. This dual attention mechanism helps the model to selectively enhance useful features while suppressing irrelevant ones, making it highly effective for various vision tasks. The architecture of CBAM, including both its channel and spatial attention sub-modules, is depicted in Figure 4.

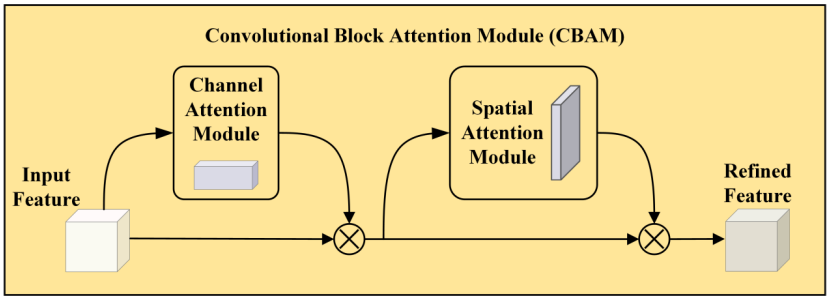

Figure 4. The structure of the Convolutional Block Attention Module (CBAM).

The channel attention module is designed to assist the network in discerning the significance of various channels. The calculation procedure is shown in Equation (3) and Equation (4), and the main steps include:

1) Extracting global information from the feature map via global Average Pooling (AvePool) and global Maximum Pooling (MaxPool) operations.

2) These two types of information are then processed through the Multilayer Perceptron (MLP) respectively, and the obtained results are then summed up.

3) The channel attention weights are obtained through a Sigmoid activation function.

4) Multiply the input features with the channel attention weight vector $M_c(F)$ to obtain the channel attention feature map. Similarly, we can obtain the spatial

attention feature map according to Equation (5) and Equation (6).

$$M_c(F) = \sigma\left(\mathrm{MLP}\left(\mathrm{AvgPool}(F)\right) + \mathrm{MLP}\left(\mathrm{MaxPool}(F)\right)\right) \tag{3}$$

$$F1 = F \otimes M_c(F) \tag{4}$$

$$M_s(F) = \sigma\left(\mathrm{Conv}\left(\left[\mathrm{AvgPool}(F)\right]; \mathrm{MaxPool}(F)\right)\right) \tag{5}$$

$$F2 = F1 \otimes M_s(F) \tag{6}$$

where F, $F1$, and $F2$ denote the input features, channel attention feature, and spatial attention feature map, respectively. $\mathrm{AvgPool}(\cdot)$ denotes global average pooling, and $\mathrm{MaxPool}(\cdot)$ denotes global maximum pooling. $\mathrm{MLP}(\cdot)$ denotes the multi-layer perceptron. $\sigma(\cdot)$ denotes the Sigmoid activation function. $\otimes$ denotes point-by-point multiplication. $\mathrm{Conv}(\cdot)$ is a convolution operation.

2.5. Dual-Path Attention Module

Inspired by CSPNet [20] and CSPAttention [21], we propose a Dual-Path Attention Module (DPAM), which is specifically designed to enhance attention on the lesion foreground in MRI images. By integrating the CSPNet and CBAM mechanisms, the DPAM is able to efficiently capture and refine detailed features of the lesion regions, leading to improved classification accuracy. Importantly, this approach achieves higher performance without significantly increasing computational cost, making it both effective and efficient for medical image analysis. The structure of DPAM is shown in **Figure 5**. Firstly, the input feature map of size [H, W, C_in] is split into two sub-feature maps of size [H, W, C_in/2] along the channel direction. Then, these two sub-feature maps are adjusted to size [H/2, W/2, C_out] through a Conv Block. The adjusted two feature maps are respectively passed through the CBAM attention module followed by a multiplication operation, and finally an output feature map of size [H/2, W/2, C_out] is obtained. The Conv Block contains a Pointwise Convolution (P_Conv) layer, Batch Normalization (BN) layer, SiLU activation function layer, and Avgpool layer. In this context, H, W, and C_in symbolize the height, width, and the quantity of channels in the input feature map, respectively. Additionally, C_out signifies the number of channels in the output feature map.

2.6. Transfer Learning

Transfer Learning (TL) is a machine learning approach that accelerates and improves learning and problem-solving in new domains by leveraging knowledge and experience gained from related domains. TL has been widely used in natural language processing, computer vision, and other fields. For example, it can enhance model performance in tasks such as text categorization, image recognition, target detection, and semantic segmentation.

In our prediction model, designed for image classification in computer vision, we pretrained the MobileViT model parameters using the ImageNet dataset,

which includes 1000 classes and 1.26 million natural images. Although natural images may differ from MRI images, they are still relevant. Through transfer learning, the model can learn features such as corners, edges, colors, and textures from the ImageNet dataset. These learned features can assist in classifying MRI images, thereby improving the effectiveness of convolutional neural networks.

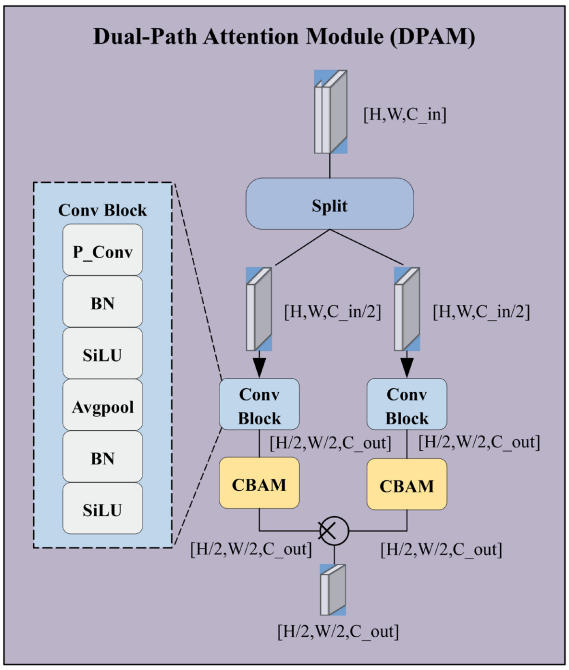

Figure 5. The structure of the Dual-Path Attention Module (DPAM).

3. Datasets and Pre-Processing

3.1. The Alzheimer's Disease Dataset

The dataset used in this paper is the Alzheimer dataset, a publicly available dataset from the Kaggle website [22]. The dataset contains a total of 6400 MRI images of Alzheimer's disease, including 896 for mild demented, 64 for moderate demented, 3200 for non demented, and 2240 for very mild demented, each of which is 128 × 128 in size, and a sample of the Alzheimer's disease dataset is shown in Figure 6.

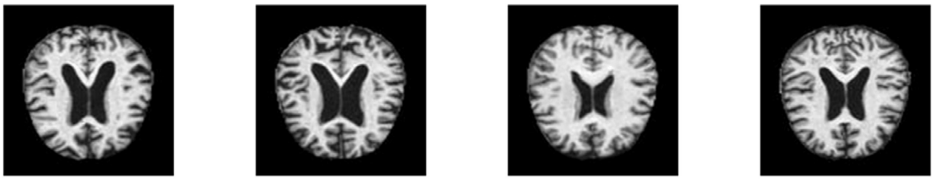

Figure 6. The Alzheimer's disease dataset.

The distribution of these cases is as follows: 14% are classified as mild dementia, 1% as moderate dementia, 50% as non-dementia, and 35% as very mild dementia.

Table 2 shows the distribution of samples in the Alzheimer's disease dataset. Compared to the other three classes, moderate dementia exhibits significant class imbalance. Imbalanced classes can lead to the classifier tending to predict the majority class and neglect the learning of minority class samples during subsequent model training. Even though the overall correctness of the model performs well, the prediction for the minority class is not good, and it happens that the lesion samples in the minority class are exactly what need to be focused on for learning and prediction. In this paper, traditional data augmentation methods such as flipping, cropping, panning, and rotating are used to expand the mild demented samples and moderate dementia samples. Before performing any data augmentation, the Alzheimer's disease dataset is divided into 70% for training and 30% for testing. The augmentation is applied only to the training set, thereby preventing data leakage where augmented variants of the same original image appear simultaneously in both the training and test sets.

Table 2. The Alzheimer's disease datasets.

Dataset	Mild Demented	Moderate Demented	Non Demented	Very Mild Demented	Totals
Training set	628	45	2240	1568	4481
Test set	268	19	960	672	1919
Totals	896	64	3200	2240	6400

3.2. The Brain Tumor Dataset

The brain tumor dataset used in this paper is sourced from Kaggle, a platform that offers publicly accessible datasets. Specifically, it is a four-class classification dataset of brain tumor MRI scans, made available in July 2020 by Sartaj Bhuvaji and colleagues from the National Institute of Technology, Durgapur, India [23]. The dataset includes a total of 3264 MRI images, categorized into four distinct classes: 926 images of glioma tumors, 937 images of meningioma tumors, 500 images representing normal (no tumor) cases, and 901 images of pituitary tumors. These images provide a comprehensive representation of different types of brain tumors, offering a valuable resource for training and evaluating machine learning models in medical image analysis. A selection of images from this dataset is displayed in Figure 7. Table 3 shows the distribution of samples in the brain tumor dataset.

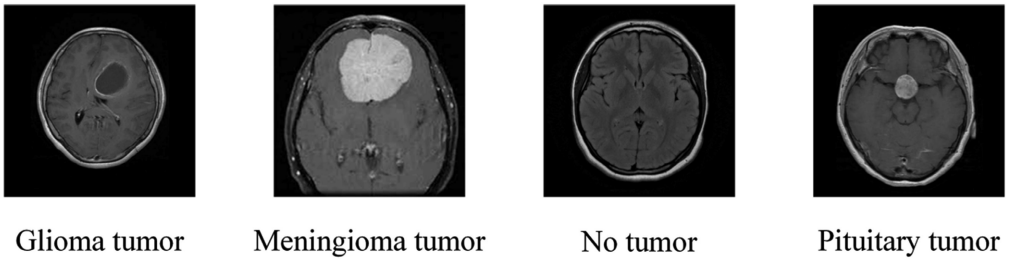

Figure 7. The brain tumor dataset.

Table 3. Distribution of samples in the brain tumor dataset.

	Glioma tumor	Meningioma tumor	No tumor	Pituitary tumor	Totals
Training set	649	656	350	631	2286
Test set	277	281	150	270	978
Totals	926	937	500	901	3264

For better model training, the images are processed to a uniform size of 224 × 224, and the images in the training set are randomly flipped horizontally with a probability of 0.5. Then, the images are converted to a tensor type and normalized. The normalized image X_n can be represented as:

$$X_n = \frac{X_i - \mu}{\sigma} \tag{7}$$

where X_i denotes a pixel value in each channel of the image, μ denotes the mean value of the pixel value in each channel of the image, and σ denotes the standard deviation of the pixel value in each channel of the image.

4. Experiment

4.1. Experimental Environment and Parameter Settings

The experimental environment used in this paper is the PyTorch 2.2.2 deep learning framework based on Windows 10, using the Python 3.8.5 language to build the network model. The processor is an Intel(R) Xeon(R) Gold 6226R CPU @ 2.90 GHz, and the memory is 16.0 GB.

To ensure the rigor of comparative experiments, all baseline models were retrained or fine-tuned under identical training configurations. All models uniformly employed the AdamW optimizer with a weight decay rate of 0.01. A cosine annealing decay strategy was adopted, with an initial learning rate uniformly set to 1e−3 and a minimum learning rate of 1e−5. The experiments are set up with the number of iterations as 40, and the BatchSize is 30. All models utilized the same early stopping mechanism. Transfer learning is uniformly applied across all models to augment their training efficacy. All baseline models utilize the same pre-trained checkpoint and follow a unified phased fine-tuning strategy: first freezing the backbone network while training only the classification head, then progressively unfreezing higher layers for joint optimization.

4.2. Evaluation Metrics

The following performance metrics are used to evaluate the classification performance of the model:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{8}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{10}$$

$$\text{Fl} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

TP (True Positive) refers to instances that were actually positive and were correctly classified as positive by the model. TN (True Negative) refers to instances that were originally negative and were accurately predicted as negative. FP (False Positive) represents cases that were originally negative but were incorrectly classified as positive, and FN (False Negative) refers to instances that were actually positive but were mistakenly predicted as negative by the model. These four metrics provide a comprehensive way to assess a model's ability to correctly and incorrectly identify positive and negative samples.

The Kappa coefficient is a statistic used to assess the performance of a classification model. It accounts for randomness in classification prediction results and corrects for potential biases in accuracy. Its value range is [−1, 1], but in practical applications, it is usually between [0, 1]. A higher Kappa coefficient indicates better classification accuracy. The calculation formulas are shown in (12), (13), and (14).

$$p_0 = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{12}$$

$$p_e = \frac{(\text{TP} + \text{FN})(\text{TP} + \text{FP}) + (\text{TN} + \text{FN})(\text{TN} + \text{FP})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})^2} \tag{13}$$

$$\text{Kappa} = \frac{p_0 - p_e}{1 - p_e} \tag{14}$$

4.3. Results and Analysis

4.3.1. Prediction Accuracy and Loss of the Alzheimer's Disease Dataset

The model proposed in this paper is an enhanced version of MobileViT, designed to improve performance on the given dataset. To assess the effectiveness of the proposed model, we conduct experiments comparing its training results with those of the MobileViT model, both before and after data augmentation. Specifically, **Figure 8(a)** and **Figure 8(b)** present the loss and accuracy curves for both the MobileViT and the proposed model when evaluated on the dataset prior to any data augmentation. This allows for a direct comparison of how the two models perform under the same initial conditions. Both models were trained for 40 epochs using cross-entropy as the loss function. As shown in **Figure 8(a)**, the proposed model has less loss on both the training and test sets and converges to zero faster compared to the MobileViT. **Figure 8(b)** shows that the accuracies of the proposed model surpass those of the MobileViT on both the training and test sets. **Figure 9(a)** and **Figure 9(b)** display the loss and accuracy of both models on the dataset after data augmentation, respectively. These figures reveal that the proposed model performs markedly better on the more balanced dataset.

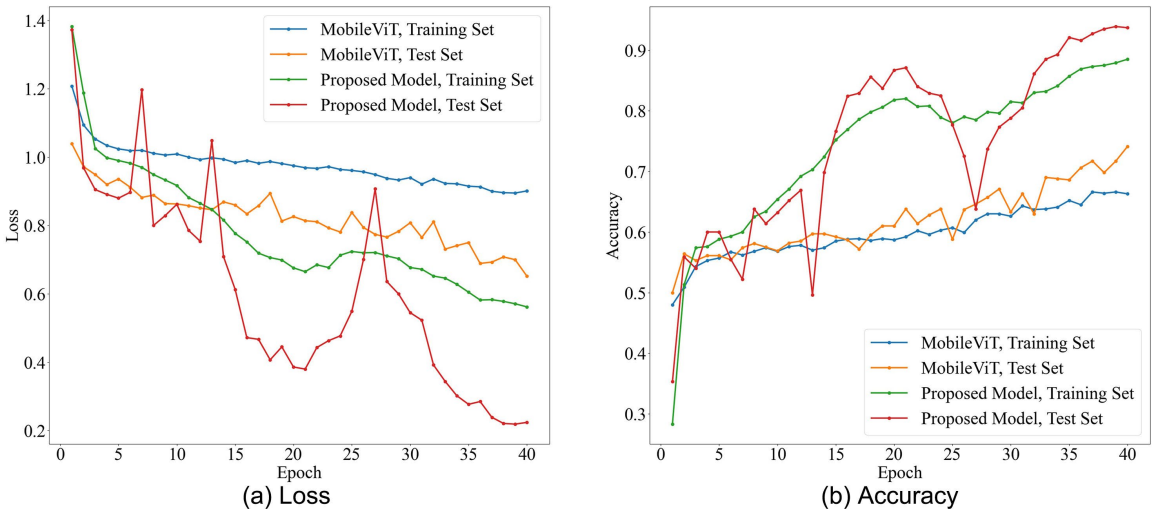

Figure 8. Loss and accuracy of the two models before data augmentation.

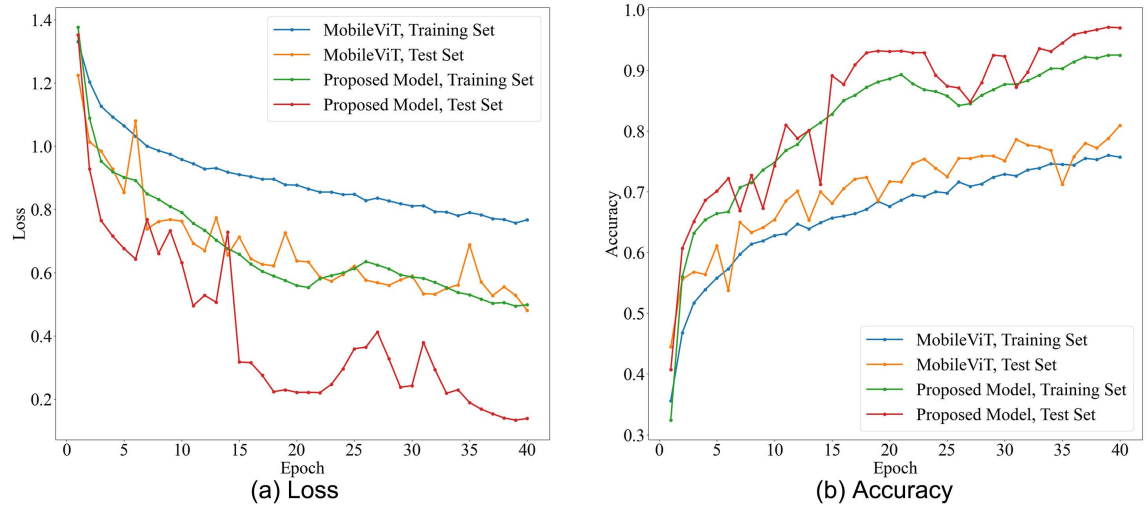

Figure 9. Loss and accuracy of the two models after data augmentation.

Table 4 shows the classification results of the proposed model and MobileViT on the dataset before and after data augmentation. From Table 4, on the original dataset, the classification accuracies of MobileViT and the proposed model are 74.1% and 93.9%, respectively. The Kappa coefficients for both are 0.562 and 0.900, respectively. The values of the two performance metrics have increased by 26.7% and 60.1%, respectively.

From Table 4, we can also conclude the effect of data augmentation on the model classification performance. The precision and Kappa coefficient of the MobileViT increase from 0.741 and 0.562 to 0.809 and 0.746, respectively. The precision and Kappa coefficient of the proposed model show significant improvements, increasing from 0.939 and 0.900 to 0.971 and 0.961, respectively. These results indicate that applying data augmentation techniques to address class imbalance has a notable positive impact on the model's classification performance. Specifically, the increase in precision reflects a better ability of the model to correctly

identify positive instances, while the higher Kappa coefficient suggests improved agreement between the model's predictions and the actual labels. Overall, the evaluation results on the test dataset show that the proposed model is effective in terms of Precision, Recall, F1, Accuracy, and Kappa.

Table 4. Classification results of MobileViT and the proposed model on the Alzheimer's disease dataset.

Augmentation	Model	Class	Precision	Recall	F1	Accuracy	Kappa
Before	MobileViT	Mild Demented	0.697	0.575	0.630	0.741	0.562
		Moderate Demented	0.000	0.000	0.000		
		Non Demented	0.775	0.857	0.814		
		Very Mild Demented	0.700	0.662	0.680		
	Proposed model	Mild Demented	0.875	0.966	0.918	0.939	0.900
		Moderate Demented	0.950	1.000	0.974		
		Non Demented	0.966	0.943	0.954		
		Very Mild Demented	0.928	0.920	0.924		
After	MobileViT	Mild Demented	0.813	0.861	0.836	0.809	0.746
		Moderate Demented	0.979	0.997	0.988		
		Non Demented	0.878	0.675	0.763		
		Very Mild Demented	0.609	0.763	0.678		
	Proposed model	Mild Demented	0.984	0.987	0.985	0.971	0.961
		Moderate Demented	0.997	1.000	0.999		
		Non Demented	0.953	0.974	0.963		
		Very Mild Demented	0.960	0.924	0.942		

4.3.2. Prediction Accuracy and Loss of the Brain Tumor Dataset

Figure 10 shows the loss values and accuracy of MobileViT and the proposed model on the brain tumor dataset, respectively. It is obvious from the figures that, compared with the original MobileViT, the proposed model exhibits lower loss values on both the training and test sets, which indicates a more significant optimization effect. Meanwhile, the classification accuracy of the proposed model on the test set also reaches the optimal level, further proving its superiority and effectiveness.

Table 5. Classification results of the original and improved models on the brain tumor dataset.

Model	Class	Precision	Recall	F1	Accuracy	Kappa
MobileViT	Glioma tumor	0.953	0.881	0.916	0.933	0.908
	Meningioma tumor	0.893	0.918	0.905		
	No tumor	0.973	0.947	0.959		
	Pituitary tumor	0.934	0.993	0.962		
Proposed model	Glioma tumor	0.971	0.960	0.966	0.971	0.961
	Meningioma tumor	0.948	0.968	0.958		
	No tumor	1.000	0.960	0.980		
	Pituitary tumor	0.982	0.993	0.987		

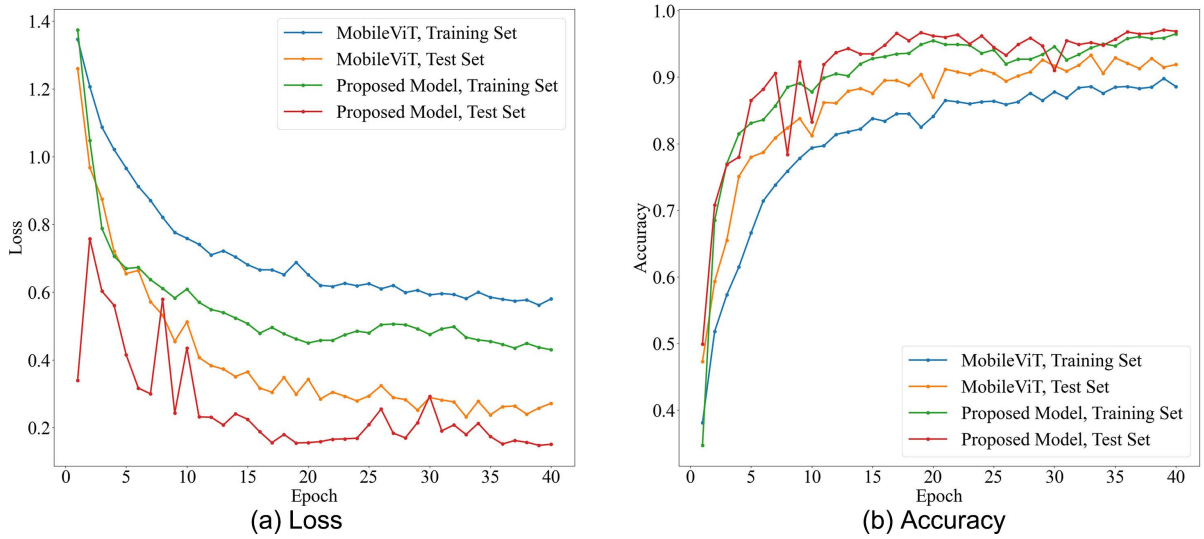

Figure 10. Loss and accuracy of MobileViT and the proposed model on the brain tumor dataset.

Table 5 shows the classification results of the two models on the brain tumor dataset. Based on **Table 5**, it can be observed that the Kappa coefficient of the MobileViT model is 0.908, while that of this paper's model is 0.961. The original model has an overall classification accuracy of 93.3% on the brain tumor dataset, but the classification accuracy for the meningioma category is only 89.3%. The improved MobileViT model not only achieves an overall classification accuracy of 97.1%, but also outperforms the original model in each category. Based on the experimental results in **Table 4** and **Table 5**, it is shown that the proposed model in this paper has better classification results than the original MobileViT model on both class-imbalanced and class-balanced datasets and is more adaptable to the complex and variable dataset situations in practical applications.

4.3.3. Comparative Experiments

In order to validate the state-of-the-art of the proposed model, nine different deep learning models were used to classify the expanded Alzheimer's dataset and brain tumor dataset. These comparison models include: ResNet50 [24], ResNet34 [24], DenseNet121 [25], ShuffleNetV2_x2_0 [26], EfficientNet [27], MobileNetV3 _large [28], MobileNetV3_small [28], MobileNetV2 [18], and MobileViT [17]. Among these baselines, ResNets are deep convolutional neural networks. DenseNet, ShuffleNet, EfficientNet, and MobileNet are lightweight deep learning models.

1) Results of the Alzheimer's disease dataset

Using the elements of the confusion matrix, several performance metrics—such as accuracy, precision, recall, and F1 score—are calculated for all models. These metrics are summarized in **Table 6**. Based on the data in **Table 6**, the following conclusions can be drawn:

a) The proposed model achieves a classification accuracy of 97.1% on the expanded dataset, with precision, recall, and F1 scores all reaching 97.1% and an F1

score of 0.971. Its classification accuracy is just 0.3 percentage points lower than that of ResNet50 and 0.2 percentage points below ResNet34.

b) In the baseline models, the classification performance of ResNet (ResNet50 and ResNet34) is significantly better than that of other lightweight deep learning models (DenseNet, ShuffleNet, EfficientNet, MobileNet, and MobileViT). This can be attributed to the fact that ResNet, as a deep residual network, demonstrates outstanding performance in various computer vision tasks, including image classification, object detection, and semantic segmentation.

c) In the lightweight network model of the baseline model, MobileViT has a classification accuracy of 80.9%, a precision of 82.5%, a recall of 80.9%, and an F1 score of 81.1%. Although the classification accuracy is 5.4% lower than that of DenseNet121, DenseNet121 has more than 7 times the number of parameters as MobileViT. Our proposed model is not only about 16% higher than the original MobileViT model in all four metrics, but its experimental results are also the best among the lightweight network models, including DenseNet121. The above results show that our improvement of MobileViT is effective.

Table 6. Classification results of each model on the Alzheimer's disease dataset.

Model	Accuracy	Precision	Recall	F1
ResNet50	0.974	0.974	0.974	0.974
ResNet34	0.973	0.973	0.973	0.973
DenseNet121	0.863	0.865	0.863	0.863
ShuffleNetV2_x2_0	0.718	0.712	0.718	0.714
EfficientNet	0.663	0.658	0.663	0.659
MobileNetV3_large	0.717	0.710	0.717	0.707
MobileNetV3_small	0.617	0.606	0.617	0.605
MobileNetV2	0.600	0.588	0.600	0.581
MobileViT	0.809	0.825	0.809	0.811
Proposed Model	0.971	0.971	0.971	0.971

2) Results of the brain tumor dataset

Table 7 summarizes the classification performance of different deep learning models on the brain tumor dataset. From it, it can be observed that the MobileViT model, as a lightweight network model, exhibits excellent performance, with an accuracy of 93.3% on brain tumor images. Its prediction performance ranks second among lightweight models. The accuracy of the proposed model is 3.8% higher than the second-ranked MobileViT. Also, the accuracy of the proposed model is higher than the heavyweight network ResNet series. These results show the proposed model's leading position and significant advantages in the brain tumor image classification task.

Table 7. Classification results of each model on the brain tumor dataset.

Model	Accuracy	Precision	Recall	F1
ResNet50	0.970	0.971	0.970	0.970
ResNet34	0.968	0.969	0.968	0.968
DenseNet121	0.926	0.927	0.926	0.926
ShuffleNetV2_x2_0	0.843	0.846	0.843	0.842
EfficientNet	0.814	0.815	0.814	0.814
MobileNetV3_large	0.906	0.907	0.906	0.905
MobileNetV3_small	0.860	0.862	0.860	0.860
MobileNetV2	0.822	0.821	0.822	0.820
MobileViT	0.933	0.933	0.933	0.932
Proposed Model	0.971	0.972	0.971	0.971

4.3.4. Computational Complexity Analysis

Computational complexity is also an important metric for deep learning models. Higher computational complexity may lead to long training times and slow inference, especially in application scenarios with limited resources or high real-time requirements, which may become a major factor limiting the application of the model. We use Params (parameters) and FLOPs (floating point operations per second) to assess computational complexity. FLOPs are commonly employed to gauge the computational load of a system or the model training process, with the number of FLOPs indicating the amount of computational resources needed during model training and inference. Params refers to the number of trainable parameters in the model, which directly impacts the memory and computational resource demands. More parameters usually allow the model to fit the details and complex relationships of the training data more accurately, but an increase in the number of parameters may increase the complexity of the model, which increases the risk of overfitting. The comparative analysis of computational complexity between the proposed model and the baseline model on two MRI datasets is shown in Table 8.

From Table 8, the following conclusions can be drawn:

1) The heavyweight network ResNet with deeper layers has far more Params and FLOPs than the lightweight network. The number of parameters of ResNet50 is about 3.38 times more than that of DenseNet121, the lightweight network with the most parameters, and about 24.69 times more than that of MobileViT, the lightweight network with the least parameters. As seen in conjunction with Table 7, while ResNet50 and ResNet34 achieve higher classification accuracy compared to lightweight networks, this comes at the expense of significantly higher computational complexity.

2) The proposed model has the second smallest number of parameters (957,761) after MobileVIT (952,308). Compared with MobileVIT, although the number of parameters in the proposed model increases by 0.57%, the accuracy improves

from 93.3% to 97.1%. In addition, the FLOPs of our proposed model (0.314778007 G) are larger than those of lightweight models such as MobileNetV3 (0.057191512 G, 0.220317112G, and 0.306178784G) and MobileViT (0.304916367 G). Compared to the accuracy, these costs are worthwhile. All in all, although the complexity of our proposed model is slightly higher than that of MobileViT, it still belongs to the lightweight network, which meets the requirements for real-time use or deployment in edge devices and also has relatively high classification accuracy.

Table 8. Results of computational complexity on two MRI datasets.

Model	Parameters	FLOPs
ResNet50	23,516,228	4.09826048G
ResNet34	21,286,724	3.66699008G
DenseNet121	6,957,956	2.848985856G
ShuffleNetV2_x2_0	5,353,192	0.584959696G
EfficientNet	4,012,672	0.393804448G
MobileNetV3_large	4,207,156	0.220317112G
MobileNetV3_small	1,521,956	0.057191512G
MobileNetV2	2,228,996	0.306178784G
MobileViT	952,308	0.304916367G
Proposed model	957,761	0.314778007G

4.4. Ablation Experiment

The proposed model is an improvement of the original MobileViT. We examine the impact of these enhanced modules on the model's classification performance through ablation experiments. The results of the ablation experiments are presented in Table 9. As shown in Table 9, when the improved MV2 module and cosine annealing module are added to MobileViT separately, both show different degrees of improvement in model metrics on all three datasets compared to the baseline MobileViT. The fusion model of MobileViT + Improved MV2 + Cosine Annealing achieved the best performance on the Alzheimer's Disease Dataset (Before augmentation) dataset. After adding the Dual-Path Attention module to MobileViT + Improved MV2 + Cosine Annealing, the best metrics have been achieved on the Alzheimer's Disease Dataset (After augmentation) dataset and Brain Tumor Dataset. It is proved that the given designed model can improve the performance of the baseline and attain better outcomes than the baseline techniques.

4.5. Interpretability Analysis of Our Proposed Model

In order to demonstrate the interpretability or the rationale behind the decision-making for our proposed model, we used the Grad-CAM method [29] to visualize

the last convolutional layer of both the original model MobileViT and our proposed model. Grad-CAM generates a type of heat map of spatial weights by back-propagating the gradient of the network output and multiplying the weights of the category activations with the feature map. The Grad-CAM algorithm provides a localization map on a given target image.

Table 9. Results of ablation experiment.

Dataset	MobileViT	Improved MV2	Cosine Annealing	Dual-Path Attention	Accuracy	Precision	Recall	F1
Alzheimer's Disease Dataset (Before augmentation)	√	×	×	×	0.741	0.730	0.741	0.733
	√	√	×	×	0.923	0.928	0.923	0.923
	√	√	√	×	0.944	0.946	0.944	0.945
	√	√	√	√	0.939	0.940	0.939	0.939
Alzheimer's Disease Dataset (after augmentation)	√	×	×	×	0.809	0.825	0.809	0.811
	√	√	×	×	0.949	0.949	0.949	0.949
	√	√	√	×	0.969	0.969	0.969	0.969
	√	√	√	√	0.971	0.971	0.971	0.971
Brain Tumor Dataset	√	×	×	×	0.933	0.933	0.933	0.932
	√	√	×	×	0.964	0.964	0.964	0.964
	√	√	√	×	0.966	0.966	0.966	0.966
	√	√	√	√	0.971	0.972	0.971	0.971

Figure 11 and Figure 12 show the heatmaps of each module of the proposed model. In the visualized experiments, the warmer the color of a region, the greater the attention allocated by the model; conversely, the cooler the color of the region, the less attention it receives from the model. From the heatmap, compared to the MobileViT-based variants, the proposed model (MobileViT + Improved MV2 + Cosine Annealing + Dual-Path Attention) demonstrates superior accuracy in identifying single or multiple targets within an image, with less emphasis on the background compared to the baseline MobileViT. Namely, it can precisely cover the lesion information in the MRI images whether they are small, medium, or large. The above results show that the introduction of transfer learning, improved activation function, Dual-Path Attention, and optimized learning rate helps the proposed model to learn more accurate features than MobileViT, thus improving the representation ability of the model.

5. Limitations and Discussions

Although the proposed model has achieved competitive performance in tumor image classification, the current model still has some obvious limitations.

Firstly, in order to improve the model accuracy, we introduced the CBAM module, but also increased the model parameters by a small amount, which resulted in slightly larger FLOPs for the model than for the MobileViT's.

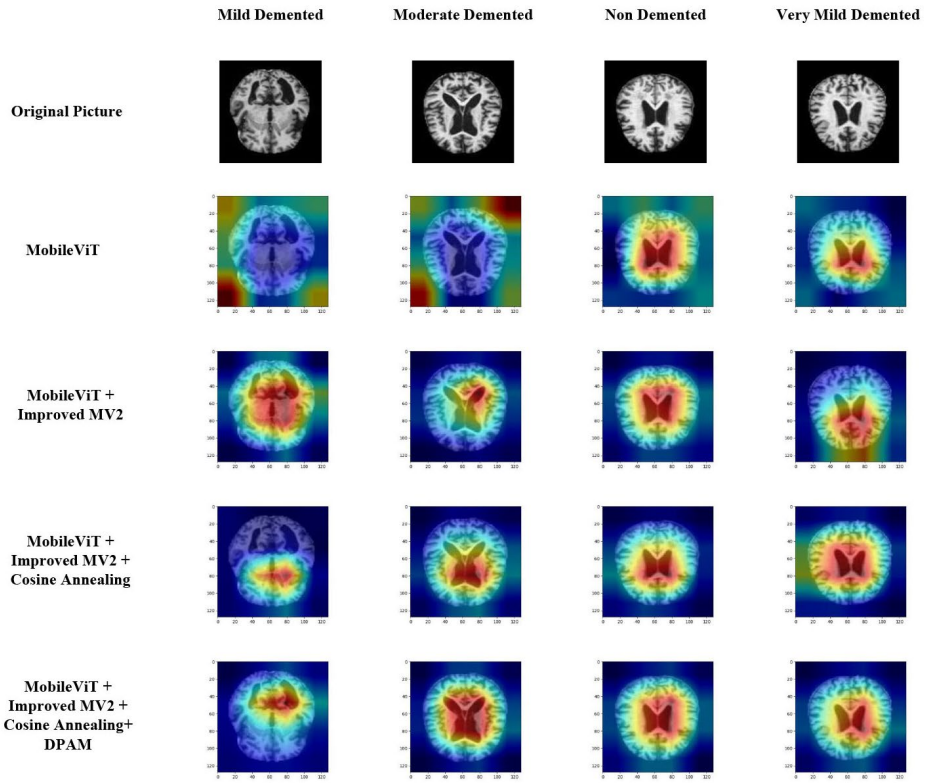

Figure 11. Grad-CAM-based model visualization (Alzheimer's disease).

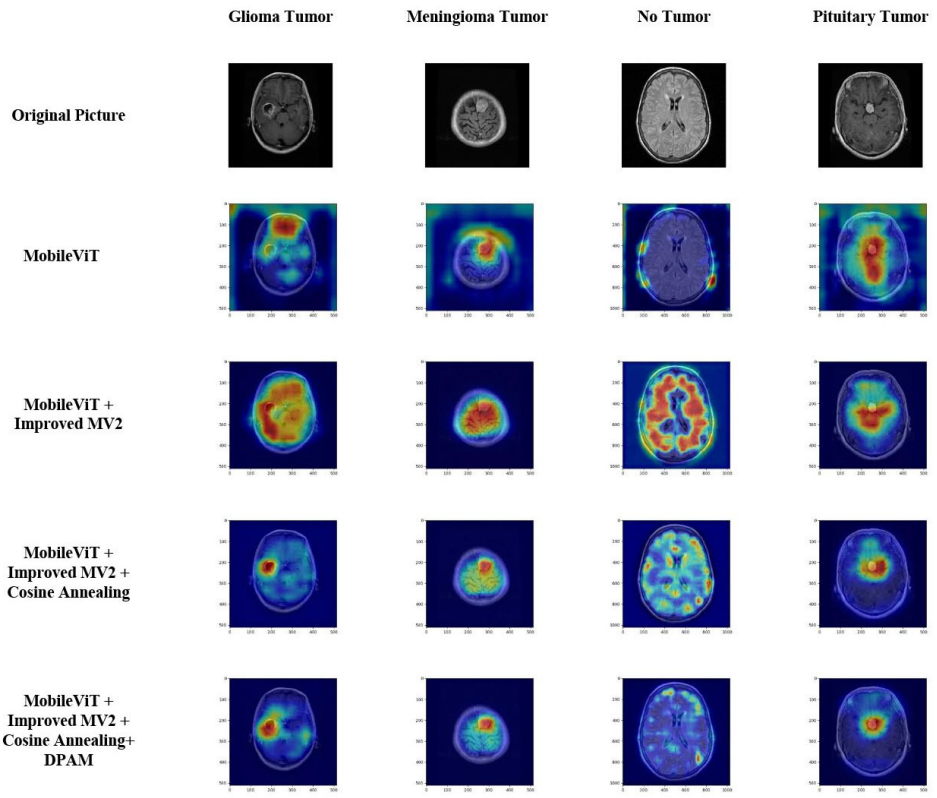

Figure 12. Grad-CAM-based model visualization (brain tumor disease).

Secondly, there exists a large diversity of MRI images, including different scanning parameters, resolutions, types and sizes, and other factors. The proposed model may encounter challenges in dealing with this diversity. The proposed model necessitates an extensive dataset for effective learning and generalization, and its performance could be constrained when dealing with a small dataset.

Thirdly, optimization algorithms play a crucial role in deep learning, as their performance directly impacts both the efficiency of model training and its overall performance. In this paper, we only used cosine annealing to optimize the learning rate of deep neural networks and did not consider the optimization of the other hyperparameters. The model necessitates extensive experimentation and fine-tuning to attain optimal results. In the future, we will investigate better optimization algorithms to determine the parameter set that is most likely to achieve the best results.

Finally, various types of medical image data exist in the field, such as skin disease images, colorectal cancer images, lung disease images, and retina OCT images, among others. Image data exhibit distinct characteristics across various medical domains. Whether our proposed MRI image classification model can be used in these fields remains to be further investigated.

6. Conclusions

In order to enable the MRI image classification model to be applied to mobile and embedded devices, we proposed an improved lightweight MobileViT model based on the MobileViT network model. Firstly, CBAM and Dual-Path Attention were employed to improve the model's ability to capture local information as well as to fuse global information. Secondly, the activation function ReLU6 in the MV2 module was replaced with the more robust activation function SiLU. Thirdly, a cosine annealing algorithm was used to update the learning rate of the proposed model to prevent the model from falling into local optimal points. Finally, the proposed model employs a transfer learning approach, where models are pretrained on the ImageNet dataset to better capture the features of MRI images. Extensive experiments demonstrate that our model delivers competitive performance in MRI image classification. We have developed a low-cost intelligent diagnostic tool that not only assists medical specialists and radiographers in providing early diagnosis of brain disease, but is also suitable for deployment in edge devices.

In future studies, experiments will be conducted using a larger data set. We plan to validate the effectiveness of our model on different medical image domains, such as skin disease images, colorectal cancer images, lung disease images, retinal images, etc. Additionally, the model proposed in this paper is a lightweight deep learning model. While it achieves high classification accuracy, its computational complexity is not the lowest. Therefore, a potential direction for future research could be to further reduce the computational time of the model without sacrificing accuracy.

Acknowledgements

This research is supported in part by the National Natural Science Foundation of China (NSFC) (Grant No. 72461030).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] (2024) 2024 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia*, **20**, 3708-3821.

[2] Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., *et al.* (2021) The 2021 WHO Classification of Tumors of the Central Nervous System: A Summary. *Neuro-Oncology*, **23**, 1231-1251. https://doi.org/10.1093/neuonc/noab106

[3] Asgharzadeh-Bonab, A., Kalbkhani, H. and Azarfardian, S. (2023) An Alzheimer's Disease Classification Method Using Fusion of Features from Brain Magnetic Resonance Image Transforms and Deep Convolutional Networks. *Healthcare Analytics*, **4**, Article ID: 100223. https://doi.org/10.1016/j.health.2023.100223

[4] Zhang, X., Gao, L., Wang, Z., Yu, Y., Zhang, Y. and Hong, J. (2024) Improved Neural Network with Multi-Task Learning for Alzheimer's Disease Classification. *Heliyon*, **10**, e26405. https://doi.org/10.1016/j.heliyon.2024.e26405

[5] Yang, Z., Liu, W., Gan, H., Huang, Z., Zhou, R. and Shi, M. (2024) Alzheimer's Disease Classification Based on Brain Region-to-Sample Graph Convolutional Network. *Biomedical Signal Processing and Control*, **96**, Article ID: 106589. https://doi.org/10.1016/j.bspc.2024.106589

[6] Qian, C. and Wang, Y. (2024) Mmanet: A Multi-Task Residual Network for Alzheimer's Disease Classification and Brain Age Prediction. *IRBM*, **45**, Article ID: 100840. https://doi.org/10.1016/j.irbm.2024.100840

[7] Ait Amou, M., Xia, K., Kamhi, S. and Mouhafid, M. (2022) A Novel MRI Diagnosis Method for Brain Tumor Classification Based on CNN and Bayesian Optimization. *Healthcare*, **10**, Article No. 494. https://doi.org/10.3390/healthcare10030494

[8] Ozdemir, C. (2023) Classification of Brain Tumors from MR Images Using a New CNN Architecture. *Traitement du Signal*, **40**, 611-618. https://doi.org/10.18280/ts.400219

[9] Attallah, O. and Pacal, I. (2026) Comparative Evaluation of Lightweight Convolutional Neural Network and Vision Transformer Models for Multi-Class Brain Tumor Classification Using Merged Large MRI Datasets. *Chemometrics and Intelligent Laboratory Systems*, **269**, Article ID: 105609. https://doi.org/10.1016/j.chemolab.2025.105609

[10] Zhang, Q., Long, Y., Cai, H. and Chen, Y. (2024) Lightweight Neural Network for Alzheimer's Disease Classification Using Multi-Slice sMRI. *Magnetic Resonance Imaging*, **107**, 164-170. https://doi.org/10.1016/j.mri.2023.12.010

[11] Khatri, U. and Kwon, G. (2024) Diagnosis of Alzheimer's Disease via Optimized Lightweight Convolution-Attention and Structural MRI. *Computers in Biology and Medicine*, **171**, Article ID: 108116. https://doi.org/10.1016/j.compbiomed.2024.108116

[12] 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 Systems with Applications*, **214**, Article ID: 119166. https://doi.org/10.1016/j.eswa.2022.119166

[13] Vaiyapuri, T., Mahalingam, J., Ahmad, S., Abdeljaber, H.A.M., Yang, E. and Jeong, S. (2023) Ensemble Learning Driven Computer-Aided Diagnosis Model for Brain Tumor Classification on Magnetic Resonance Imaging. *IEEE Access*, **11**, 91398-91406. https://doi.org/10.1109/access.2023.3306961

[14] Luo, H., Zhou, D., Cheng, Y. and Wang, S. (2024) MPEDA-Net: A Lightweight Brain Tumor Segmentation Network Using Multi-Perspective Extraction and Dense Attention. *Biomedical Signal Processing and Control*, **91**, Article ID: 106054. https://doi.org/10.1016/j.bspc.2024.106054

[15] Haq, E.U., Yong, Q., Yuan, Z., Huarong, X. and Haq, R.U. (2025) Multimodal Fusion Diagnosis of the Alzheimer's Disease via Lightweight CNN-LSTM Model Using Magnetic Resonance Imaging (MRI). *Biomedical Signal Processing and Control*, **104**, Article ID: 107545. https://doi.org/10.1016/j.bspc.2025.107545

[16] Nizamani, A.H., Chen, Z. and Bhatti, U.A. (2026) Deep-Fusion: A Lightweight Feature Fusion Model with Cross-Stream Attention and Attention Prediction Head for Brain Tumor Diagnosis. *Biomedical Signal Processing and Control*, **111**, Article ID: 108305. https://doi.org/10.1016/j.bspc.2025.108305

[17] Mehta, S. and Rastegari, M. (2021) Mobilevit: Light-Weight, General-Purpose, and Mobile-Friendly Vision Transformer.

[18] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L. (2018) Mobilenetv2: Inverted Residuals and Linear Bottlenecks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 4510-4520. https://doi.org/10.1109/cvpr.2018.00474

[19] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision* (*ECCV*), Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[20] Wang, C.Y., Liao, H.Y.M., Wu, Y.H., *et al.* (2020) CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*), 14-19 June 2020, 1571-1580. https://doi.org/10.1109/cvprw50498.2020.00203

[21] Shen, L. and Wang, Y. (2022) TCCT: Tightly-Coupled Convolutional Transformer on Time Series Forecasting. *Neurocomputing*, **480**, 131-145. https://doi.org/10.1016/j.neucom.2022.01.039

[22] Pasnoori, N., Flores-Garcia, T. and Barkana, B.D. (2024) Histogram-Based Features Track Alzheimer's Progression in Brain MRI. *Scientific Reports*, **14**, Article No. 257. https://doi.org/10.1038/s41598-023-50631-1

[23] Muezzinoglu, T., Baygin, N., Tuncer, I., Barua, P.D., Baygin, M., Dogan, S., *et al.* (2023) Patchresnet: Multiple Patch Division-Based Deep Feature Fusion Framework for Brain Tumor Classification Using MRI Images. *Journal of Digital Imaging*, **36**, 973-987. https://doi.org/10.1007/s10278-023-00789-x

[24] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[25] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 4700-4708.

https://doi.org/10.1109/cvpr.2017.243

[26] Zhang, X., Zhou, X., Lin, M. and Sun, J. (2018) Shufflenet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 6848-6856. https://doi.org/10.1109/cvpr.2018.00716

[27] Tan, M. and Le, Q. (2019) Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks. *International Conference on Machine Learning*, *PMLR*, Long Beach, 9-15 June 2019, 6105-6114.

[28] Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L., Tan, M., *et al.* (2019) Searching for MobileNetV3. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27-28 October 2019, 1314-1324. https://doi.org/10.1109/iccv.2019.00140

[29] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 618-626. https://doi.org/10.1109/iccv.2017.74