Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 14, No. 3, 2026, 381-393.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Ensemble of EfficientNet-B1 and ResNet-101 with Attention Mechanism for Brain Tumor Classification in MRI Images

Amirhossein Zare Kordkheili[1], Amirreza Zare Kordkheili[2] and Sekineh Asadi Amiri[1*]

1. Department of Computer Engineering, University of Mazandaran, Babolsar, Iran.
2. Department of Computer Engineering, University of Tehran, Tehran, Iran.

Article Info

Article History:
Received 14 November 2025
Revised 23 February 2026
Accepted 02 May 2026

DOI:10.22044/jadm.2026.17084.2845

Keywords:
Brain Tumor Classification, EfficientNet-B1, ResNet-101, Transfer learning, Soft Voting.

**Corresponding author: s.asadi@umz.ac.ir (S. Asadi Amiri).*

Abstract

Brain tumor detection is a critical task in medical imaging, requiring accurate and reliable methods. Recent advancements in deep learning have shown great potential in this field. In this article, we present a novel method for brain tumor detection based on a Convolutional Block Attention Module (CBAM)-enhanced attention ensemble of deep learning networks. Initially, image augmentation is applied to increase data diversity. We utilize two deep neural network models, EfficientNet-B1 and ResNet-101, for tumor detection. First, we enhance the performance of these models by integrating the CBAM attention module into their architectures. Then, we ensemble the two networks using a soft voting strategy to achieve higher detection accuracy. The proposed method is evaluated on the three-class Figshare dataset, achieving an accuracy of 99.09% in detecting tumors in MRI images, which outperforms existing methods. This approach leverages the strengths of an ensemble of models, offering a promising solution for improving the accuracy and reliability of brain tumor detection in medical imaging.

1. Introduction

Brain tumors are the most common type of brain disease and can lead to brain cancer. In this form of cancer, the disease progresses rapidly, emphasizing the necessity of early diagnosis of brain tumors [1]. One of the most frequent methods to classify brain tumors is by dividing them into benign and malignant types. Benign brain tumors can form within the brain tissue or near it and can occasionally be life-threatening despite not being cancerous, due to their potential to cause pressure on the brain or surrounding areas. Meningiomas and pituitary tumors are two common types of benign brain tumors [2]. Meningiomas are slow-growing tumors with a high likelihood of successful surgical removal, as they seldom spread to nearby brain tissue. However, meningiomas have the potential to become malignant. The pituitary gland, located at the base of the brain, controls other endocrine glands. It secretes hormones that regulate growth, metabolism, and reproduction. Pituitary tumors develop within this gland and are generally benign; they do not spread to other parts of the body. Despite their non-cancerous nature, these tumors can cause significant health issues by disrupting normal hormonal balance, leading to long-term hormone deficiencies. Additionally, due to the gland's location, pituitary tumors can exert pressure on surrounding structures, potentially resulting in vision loss. Malignant tumor cells are abnormal cells that grow rapidly; gliomas are such tumors and can damage healthy tissues [3]. The characteristics of brain tumors underscore the critical importance of addressing this disease. Early detection significantly facilitates treatment. Various screening techniques, both invasive and non-invasive, are used to detect tumors in the human brain [4]. Medical imaging is a non-invasive technique used to obtain images for diagnosing various diseases. Common types include X-rays, ultrasonic imaging (UI), positron emission tomography (PET), computed tomography (CT), magnetic resonance spectroscopy (MRS), single-photon emission computed tomography (SPECT), and MRI [5]. Magnetic Resonance Imaging (MRI) is particularly preferred because of its ability to provide detailed information about the location, shape, and size of brain tumors in medical images (MI). It is generally considered faster, more cost-effective, and safer compared to other scanning modalities, making it a less harmful and highly informative option for diagnosing brain tumors [6]. Accurate and timely diagnosis of a brain tumor is crucial for a patient's treatment plan. In recent years, experts have spent a lot of time improving how we classify tumors in brain MRI images. Consequently, computer-aided diagnosis (CADx) has been developed to reduce the burden on and assist radiologists and doctors with assessing MI [7]. Initially, machine learning methods were employed for tumor classification due to their significant role in identifying and categorizing brain tumors, even though they still required human oversight. They transmitted thermal data via a Wireless Sensor Network (WSN). Their analysis combined multiple techniques, including feature vectors, gray level co-occurrence matrix, support vector machine (SVM), statistical features, and a backpropagation neural network [8]. Recently, many deep learning-based models have been explored for detecting brain tumors. These networks have become highly popular in medical diagnostics. They efficiently handle large and complex datasets, such as medical images. Additionally, they accurately recognize patterns and extract features that are indicative of diseases.

In this study, we propose an ensemble method for brain tumor classification using deep learning models based on EfficientNet-B1 and ResNet-101. Through extensive experimentation and considering specific requirements, we enhanced the fully connected layers of these models by adding new dense and dropout layers. Our approach leverages the strengths of both models by using soft voting to combine their classification results. This method capitalizes on the capabilities of EfficientNet-B1 and ResNet-101 in tumor classification, resulting in more accurate and reliable brain tumor categorization.

The rest of this study is organized as follows: Section 2 reviews relevant literature on tumor classification. Section 3 presents a comprehensive overview of the dataset. Section 4 provides a detailed explanation of the proposed algorithm. Section 5 presents the experimental results and discussions. Finally, Section 6 concludes with key insights and recommendations for future research.

2. Related work

Advancements in machine learning and deep learning techniques have led to their widespread application across numerous domains. Their use in the medical field has become increasingly prominent. One important application is the detection of brain tumors. If these tumors are not identified and treated promptly, they can lead to serious health issues. Due to the vast amount of research conducted, it is not feasible to discuss all works comprehensively here. However, we will review some of the key studies to illustrate the progress made in brain tumor classification. Kharrat et al. employed a 2D wavelet transform to extract features and utilized a genetic algorithm to select the most relevant features. Using an SVM (Support Vector Machine) classifier, they achieved an accuracy of 97.59% on a dataset comprising 83 images [9]. Similarly, Ismael et al. applied DWT (Discrete Wavelet Transform) and the Gabor filter to extract pertinent statistical features and performed classification using a backpropagation neural network (BPNN). Their work on the Figshare dataset [10] yielded an average accuracy of 94.63% [11]. Paul et al. introduced CNNs to achieve a 91.43% accuracy on the same dataset without relying on time-consuming feature extraction methods, which risk losing important features. However, one challenge with using CNNs is their reliance on large and diverse datasets to achieve optimal performance [12]. To address this issue, Afshar et al. explored the use of Capsule Networks, which require less data for training and better capture spatial relationships and object positioning. Unlike CNNs, Capsule Networks overcome the limitations of pooling layers that often lead to the loss of spatial information. Their approach achieved an accuracy of 86.56% on the Figshare dataset [13]. A notable drawback of their proposed model was its sensitivity to the background of images, necessitating segmentation to remove irrelevant regions. However, this segmentation process inadvertently discarded surrounding textures that could be beneficial for the task. To mitigate this issue, the authors added a vector representing the tumor boundary to the input of the fully connected layers. This modification improved the model's accuracy by approximately 4% on the same dataset [14]. In another attempt, Anaraki et al. utilized Genetic Algorithms (GA) to optimize the hyperparameters of their CNN model, achieving an accuracy of 90.9% on the referenced dataset [15]. Using transfer learning, Ghassemi et al. leveraged the Generative Adversarial Network

(GAN) discriminator to train a model capable of extracting critical features from MRI images and learning their structure [16]. They employed this model to train their classification model, achieving an accuracy of 95.60% on the benchmark Figshare dataset. In a different study, Zhou et al. worked with 3D images, treating them as sequences of 2D slices to train an RNN-based model [17]. They utilized a deep DenseNet-based autoencoder to extract key high-level features from each 2D slice. These sequential features were then fed into an LSTM (Long Short-Term Memory) as input. To evaluate and compare their work on the publicly available Figshare dataset, they transformed the 2D images into sequences of 20 consecutive frames, comprising the original image and 19 zero matrices. Using this approach, they achieved an accuracy of 91.09%. In a more recent study, Aamir et al. first enhanced the image quality in the Figshare dataset. They then utilized pre-trained architectures, EfficientNet and ResNet50, to extract important features from the images. Subsequently, these high-dimensional hybrid features were reduced to a low-dimensional feature vector using Partial Least Squares (PLS). Finally, these features, along with typical tumor locations identified through agglomerative clustering, were fed into a network. Their method achieved an accuracy of 98.95% [18]. Similarly, Demir et al. trained a Residual-CNN on the dataset [19]. From all layers, the top 100 features were extracted using a novel algorithm called L1-Norm SVM ReliefF (L1NSR), which combines the strengths of the L1-norm SVM and the ReliefF algorithm. These selected features were then used for classification with SVM. Their model achieved an accuracy of 96.60% on the 4-class Bhuvaji dataset [20].

Akter et al. proposed a model consisting of two blocks and 39 layers. Each block featured three pathways, allowing for the extraction of diverse features that were subsequently merged [2]. The architecture included convolution, max pooling, batch normalization, dropout to reduce overfitting, flatten, and several dense layers. Their model achieved an accuracy of 98.7% on a merged dataset comprising Figshare and the Bhuvaji dataset and 96.70% on the Figshare dataset. This accuracy increased by 0.1% with manual segmentation. Liu et al. proposed a model based on MobileNetV2, where they unfroze and utilized its pre-trained layers to train the model on a specific dataset [21]. They also modified the model by adding a series of layers in place of the final layer, which included a fully connected layer followed by a linear layer, each accompanied by a dropout layer. Finally, a linear layer with four output classes was added. Their model achieved an accuracy of 99.20% on the Brain Tumor MRI Dataset. Mohammed et al. [30] proposed a BoF-SURF-based brain-tumor classification method on the Figshare dataset that achieved an accuracy of 98.7%. Their approach combined SURF feature extraction, multiple feature-selection techniques, and classifiers such as kNN, SVM, and ensemble models. Mavaddati et al. proposed a Hybrid CNN-ViT model to accurately classify brain tumors from MRI images [31]. The model integrates Convolutional Neural Networks to extract local spatial features alongside Vision Transformers to capture global contextual relationships within the scans. When evaluated on a four-class dataset, the framework achieved a classification accuracy of 98.37%. Zahoor et al. introduced a novel deep residual and region-based convolutional neural network (CNN) architecture, called Res-BRNet, which used three spatial blocks and four modified residual blocks to extract both homogeneity and boundary features, as well as local and global texture variations [7]. They achieved an accuracy of 98.22% by working on a combination of datasets, including Figshare and Br35H [22]. Given the significant advancements in transformers for sequence data over recent years, their application in vision tasks is also developing. Reddy et al. worked on fine-tuning a Vision Transformer model and achieved an accuracy of 98.7% in the best case [23]. Manikandan et al. proposed a method called TriCANet, which integrates axial, sagittal, and coronal MRI views using a shared EfficientNet-B0 backbone. The model fuses cross-plane features through a Cross-Plane Attention module and a Swin Transformer, achieving 95.3% accuracy on the Kaggle dataset and outperforming classical architectures [24]. Through attention-guided CNN processing, erosion–dilation filtering, and LBP encoding, ShallowMRI was proposed by S.U.R. Khan et al. [25]. A simple CNN enhanced by attention to classify brain tumors. The approach achieves an accuracy of 98.16% on the BR35H dataset. Table 1 summarizes the methods discussed in related work, including their key features and performance metrics.[3]

3. CE-MRI Figshare dataset

The Contrast Enhanced-MRI (CE-MRI) Figshare dataset was used for this study. It comprises 3,064 T1-Weighted Contrast Enhanced-MRI images collected from 233 patients diagnosed with three types of brain tumors: meningioma (708 slices), glioma (1,426 slices), and pituitary tumor (930 slices). The distribution of the three classes in this dataset is illustrated in Figure 1.

Table 1. Summary of methods employed in brain tumor classification using MRI images, highlighting key features and performance metrics.

Reference	Year	Method	Dataset	Accuracy
Kharrat et al., [9]	2010	GA + SVM	Simple 83 images	97.59
Ismael et al., [11]	2018	Statistical features + SVM	Figshare	94.63
Paul et al., [12]	2017	CNN	Figshare	91.43
Afshar et al., [13]	2018	CapsNet	Figshare	86.56
Afshar et al., [14]	2018	CapsNet with tumor boundary	Figshare	90.90
Anaraki et al., [15]	2019	CNN + GA	Figshare	95.60
Ghassemi et al., [16]	2020	GAN-based model	Figshare	91.43
Zhou et al., [17]	2019	Auto-encoder + LSTM	Figshare	91.09
Aamir et al., [18]	2022	Pre-train models + PLS	Figshare	98.95
Demir et al., [19]	2022	L1NSR-SVM	Bhuvaji	96.60
Akter et al., [2]	2024	Deep CNN	Figshare	96.70
Mohammed et al. [30]	2024	BoF-SURF	Figshare	98.70
Liu et al., [21]	2024	Fine-Tuned MobileNetV2	Brain Tumor MRI	99.20
Zahoor et al., [7]	2024	Res-BRNet	Figshare + Br35H	98.22
Reddy et al., [23]	2024	Fine-Tuned ViT	Figshare + Br35H + SARTAJ	98.70
Manikandan et al. [24]	2025	TriCANet	Kaggle Brain Tumor	95.30
S.U.R. Khan et al. [25]	2026	ShallowMRI	Br35H	98.16

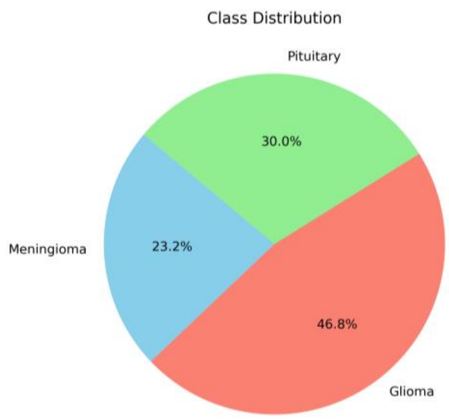

Figure 1. Class distribution of CE-MRI Figshare Brain Tumor Dataset.

The MRI images in this dataset come from three different views: axial, coronal, and sagittal. Some examples are shown in Figure 2. The data was gathered from Nanfang Hospital in Guangzhou, China, and General Hospital, Tianjin Medical University in China, between 2005 and 2010. The dataset is publicly available on Figshare, making it accessible for further research and comparative studies. The images were preprocessed to standardize the intensity values and remove any artifacts. Detailed information about the dataset is provided in Table 2. This table provides a comprehensive overview of the MRI dataset, detailing the number of patients, MRI images, and the different views for each tumor category. This table highlights the distribution and variety of the dataset, which is crucial for training robust models.

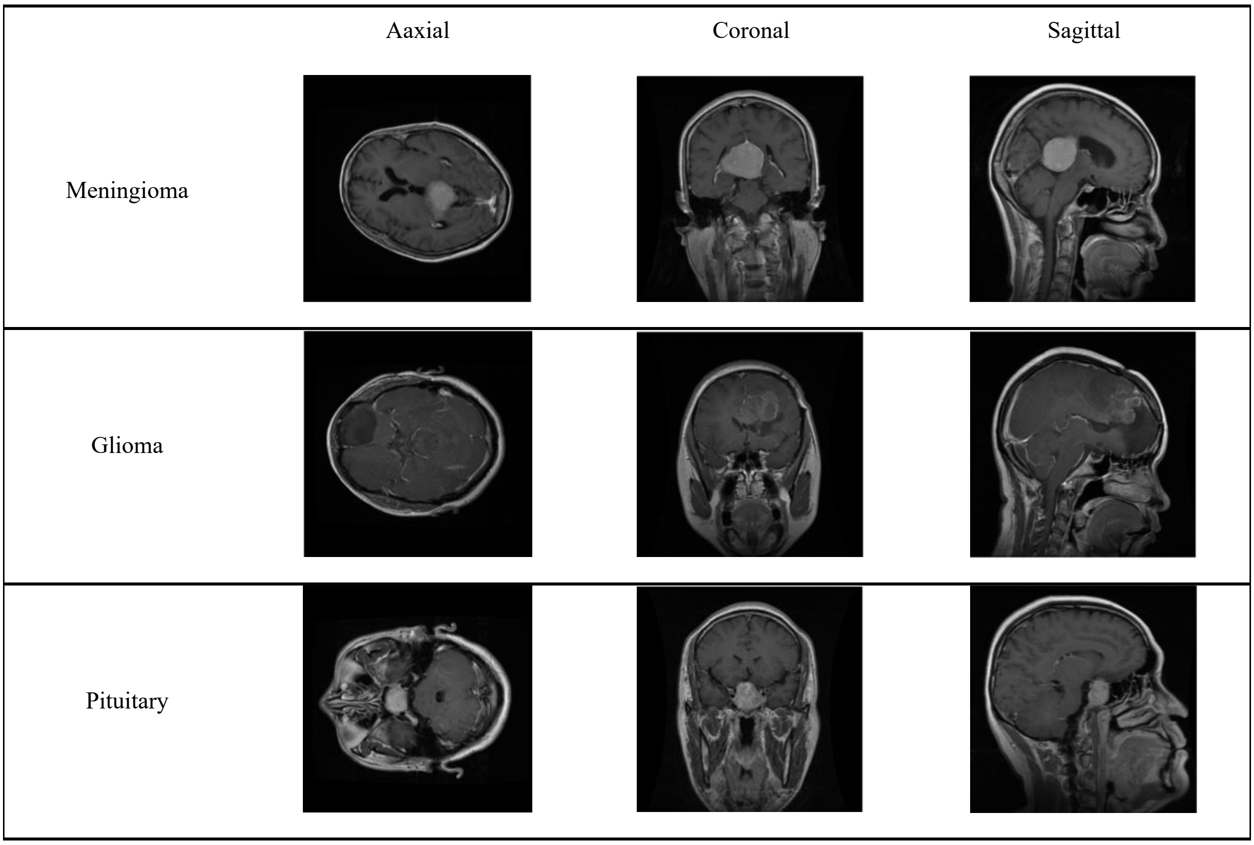

Figure 2. Examples of MRI Images from Various Categories and Views.

Table 2. A comprehensive overview of the MRI dataset.

Tumor's categories	Number of patients per category	Number of MRI images	Different views of MRI images
Meningioma	82	708	Aaxial: 209 Coronal: 268 Sagittal: 231
Glioma	89	1426	Axial: 494 Coronal: 437 Sagittal: 495
Pituitary	62	930	Axial: 291 Coronal: 319 Sagittal: 320
Overall	233	3064	Axial: 994 Coronal: 1024 Sagittal: 1046

4. Proposed Method

In this study, we proposed a CBAM-enhanced attention ensemble of deep learning models for brain tumor classification. We used a 5-fold cross-validation strategy on the entire dataset to ensure a reliable and comprehensive evaluation. In each fold, one subset of the data was designated as the test set, while the remaining subsets were used for training. This process was repeated five times so that every image in the dataset served as test data exactly once, providing a robust assessment of the model's generalization ability. Within each training fold, 10% of the training samples were further separated as a validation set. Both the training and validation subsets were enriched through data augmentation techniques to increase sample diversity and reduce overfitting

To construct the ensemble, we modified and enhanced two deep neural network architectures: EfficientNet-B1 and ResNet-101, and strengthened these models by integrating a CBAM-based attention block. These improvements enabled the

models to better capture discriminative tumor-related patterns while maintaining strong generalization performance across all folds.
Each model not only predicts a class for each test image but also provides the probability of the image belonging to the predicted class. In the proposed method, to further enhance brain tumor detection, the probabilities from both improved models are combined using a soft voting method to determine the winning class for each test image. The proposed method demonstrates high accuracy in detecting brain tumors, making it a promising approach in medical imaging. The flowchart of the proposed method is shown in Figure 3. The details of the proposed method are explained below.

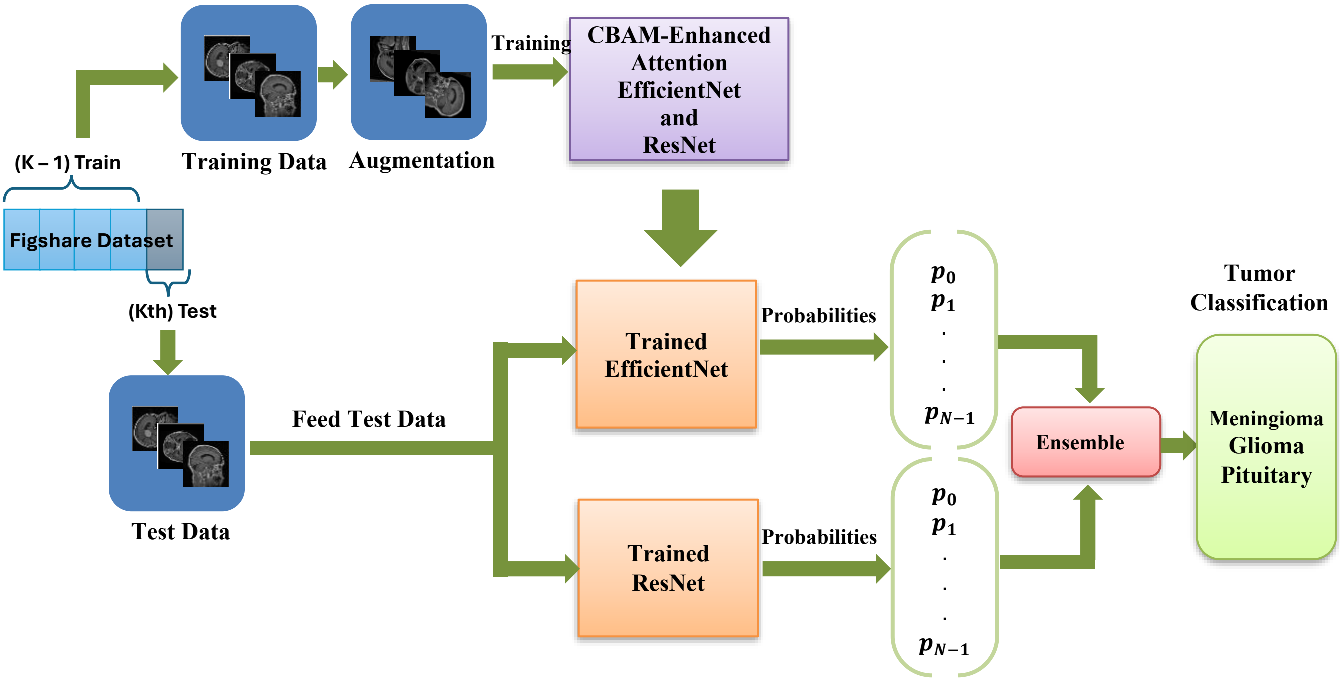

Figure 3. Overview of the proposed brain tumor classification method.

4.1. Preprocessing

Preprocessing is a crucial step in any machine learning pipeline, especially for image classification tasks, as it improves data quality and ensures that models are trained on diverse and representative samples. For the MRI Figshare dataset, which contains three tumor classes—Meningioma, Glioma, and Pituitary—we used 5-fold cross-validation to split the data and evaluate our proposed method. In each fold, 10% of the training data was set aside for validation to fine-tune the model and monitor its performance. To further enhance the dataset, we applied data augmentation techniques summarized in Table 3, which increased the number of training samples for each class to approximately 2,300 images and expanded the validation set to 200 images per class. The augmentation process significantly improved data variability, enhanced model generalization, and reduced overfitting.

Table 3. Overview of augmentation settings.

Augmentation Type	Setting
Rotation	±15°
Width Shift	±10%
Height Shift	±10%
Zoom	0.9–1.1
Horizontal Flip	Enabled

Table 4 summarizes the number of images before and after augmentation across the training, validation, and testing splits for a single representative fold, providing a clear overview of the data distribution among the three tumor classes. This table highlights the impact of augmentation in creating a diverse and balanced dataset, ultimately improving model performance.
As illustrated in Figure 4, the augmented images from the dataset provide a diverse range of variations.

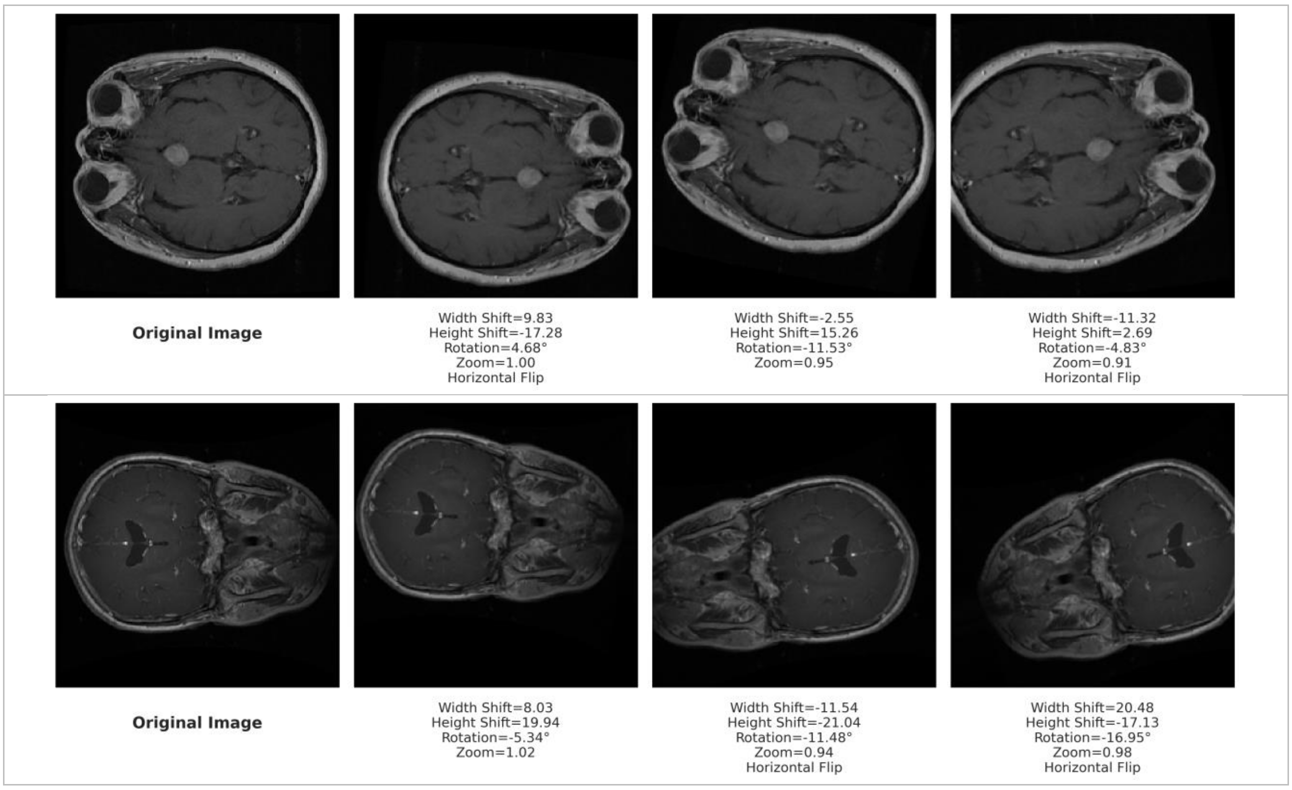

Figure 4. Examples of Augmented MRI Images.

Table 4. Overview of per-fold image distribution before and after data augmentation.

Class Name	Train		Validation		Test
	Before Augmentation	After Augmentation	Before Augmentation	After Augmentation	
Meningioma	509	2186	57	164	142
Glioma	1027	2304	114	214	285
Pituitary	670	2257	74	200	186
Total	2206	6747	245	578	613

4.2. Convolutional Block Attention Module

We employed a custom, lightweight sequential attention mechanism inspired by the Convolutional Block Attention Module [26] in our proposed method. This attention block was integrated before the global average pooling layer to enhance the discriminative capacity of our architecture and suppress irrelevant background artifacts. The feature maps extracted by the deep learning models were refined using a combination of 1D channel attention and 2D spatial attention. In the first phase, channel attention was applied to emphasize the most meaningful information in the extracted features across all channels.

To achieve this, average pooling was used to aggregate spatial information. This aggregated information was then forwarded to a multi-layer perceptron (MLP) to generate the channel attention weights. Finally, these weights were multiplied element-wise with the original feature maps to refine the importance of each channel.

In the second phase, a spatial attention map was generated to highlight the locations of informative regions within the feature maps. To compute spatial attention, we first applied average pooling and max-pooling along the channel axis and concatenated the resulting descriptors.

A convolutional layer was then used to convert these descriptors into a spatial attention map. Finally, we multiplied the spatial attention map by the channel-refined descriptor from the first phase, producing the final attention-enhanced representation.

4.3. EfficientNet-B1

The EfficientNet architecture was introduced by Tan and Le [27]. It uses a compound coefficient to evenly scale the network's structure in all dimensions: depth, width, and resolution. This

method helps improve the network's overall performance in a balanced and consistent way. This network leverages the Mobile Inverted Bottleneck Convolution (MBConv) layers, like those in MobileNet [28], and employs depth-wise separable convolutions. These architectural elements contribute to a significant reduction in the network's parameter count while preserving its ability to capture complex features. This design strategy ensures that the model remains computationally efficient and effective in its performance. We utilize EfficientNet-B1, which is part of the EfficientNet series, where each model is a scaled version of the others. It accepts input images with a size of 380x380 pixels.
In our proposed approach for the classification head, we integrate a lightweight, CBAM-Enhanced attention block immediately after the final convolutional stage of the EfficientNet-B1 backbone. This attention mechanism adaptively emphasizes informative channels and spatial regions, enabling the network to focus on tumor-relevant structures while suppressing background noise and irrelevant anatomical variations. Following that CBAM block, the spatial dimensions of the feature maps are reduced using a global average pooling layer to aggregate the learned features into a compact descriptor and keep their semantic meaning. We then add two fully connected layers with 1024 and 256 neurons. Both fully connected layers are regularized with dropout and a batch normalization layer to reduce overfitting and improve generalization. Finally, a softmax output layer with three units produces class probabilities corresponding to the three tumor classes. The architecture of our proposed model is depicted in Figure 5.

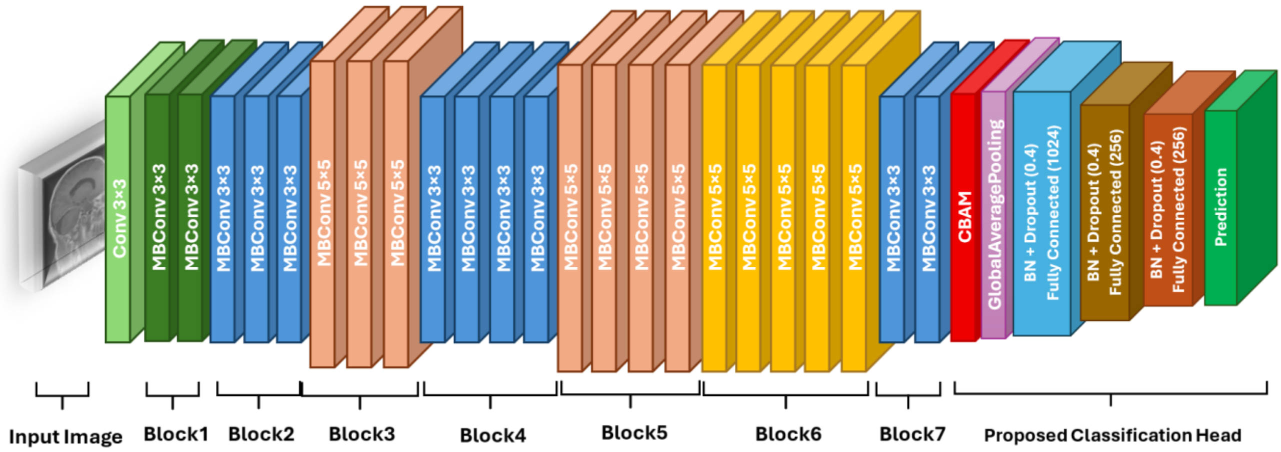

Figure 5. Architecture of the EfficientNet-B1 model with proposed classification head.

4.4. ResNet-101

ResNet, which stands for Residual Networks, was introduced by Kaiming He et al. [29]. ResNet networks utilize residual connections, allowing gradients to flow through directly, preventing them from diminishing to zero after applying the chain rule.
The architecture of ResNet-101 is made up of five distinct convolution modules, comprising 104 convolutional layers in total. It includes 33 blocks, with 29 of them using the output of the previous block as residual connections. This output is based on the output of preceding layers, which comes from the layer denoted by x_{l-1}, as given in Figure 6.

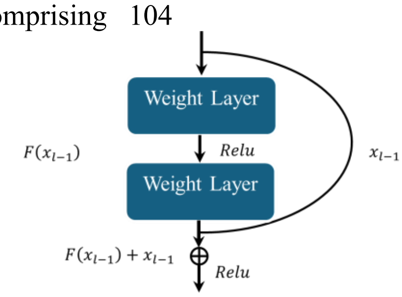

Figure 6. Residual block in ResNet.

These residuals serve as the first operand of the summation operator at the end of each block, providing the input for the subsequent blocks. The remaining 4 blocks use the output of the previous block in a convolution layer with a 1×1 filter size and a stride of 1. This is followed by a batch normalization layer, and the resulting output is fed to the summation operator at the block's output. The depths of the dense blocks vary, as depicted in Figure 7.

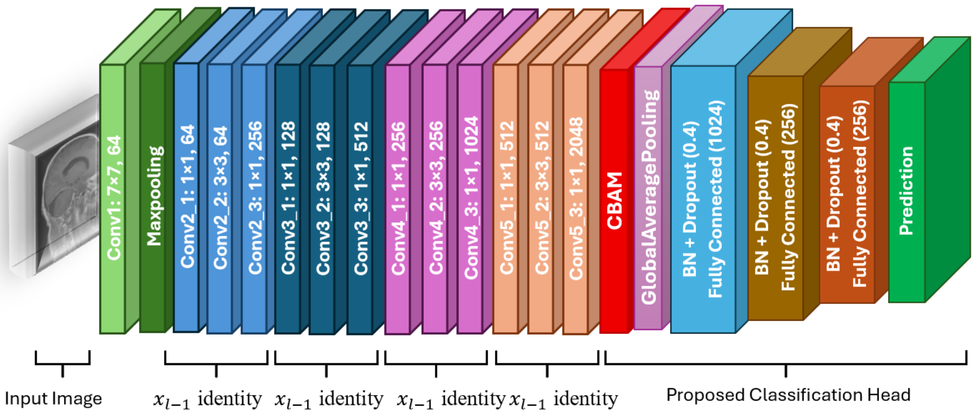

Figure 7. Architecture of the ResNet-101 model with proposed classification head.

We fine-tuned ResNet-101 as the base model for our proposed approach. This process leverages transfer learning, which allows us to adapt the pretrained ResNet-101 model to our specific task by refining it with our own dataset.

Similar to the EfficientNet-B1 configuration, the ResNet-101 model was equipped with the same redesigned classification head with a CBAM-based attention module. This integration enables the network to selectively emphasize informative channel-wise and spatial patterns within the extracted feature maps. The modification enhances the representational capacity of the model and improves its effectiveness in the tumor classification task.

4.5. Soft Voting

This method is used in ensemble learning, where multiple models are trained on the same problem and then combined to make a final prediction. There are two types of voting: hard voting and soft voting. In hard voting, the class of each instance is predicted separately by all models, and the final class label is determined by majority vote. Despite its simplicity, hard voting does not consider the confidence of the predictions. In contrast, soft voting aggregates the probabilities assigned to each class by the models. The final class label is determined by the class with the highest average probability, allowing for a more fine-tuned and confident prediction.

5. Results

This section provides details of the experimental setup and findings, including a comparison of our proposed method with existing approaches. The implementation specifics, evaluation metrics, and results of our method are discussed. The results achieved by our method demonstrate a significant improvement over existing approaches.

5.1. Evaluation Metrics

A confusion matrix is an essential tool for evaluating the performance of classification models. It provides detailed information on the classification results for each class. In a confusion matrix, the values along the diagonal represent correctly classified instances for each class. The off-diagonal values indicate misclassifications, showing where the model incorrectly classified instances. The performance of the proposed model was evaluated by analyzing these values, which form the basis for calculating essential metrics including accuracy, precision, recall, and F1-score. The equations for these metrics are provided below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score = \frac{2TP}{2TP + FN + FP} \tag{4}$$

where TP, FN, FP, and TN are the number of true positive (positive labeled samples classified correctly), false negative (positive labeled samples classified as negative), false positive (negative labeled samples classified as positive), and true negative responses, respectively (negative labeled samples classified correctly).

5.2. Detail of simulation

In this study, experiments were conducted within a Kaggle environment, which provides a robust platform for model training. We utilized a 16 GB NVIDIA Tesla T4 GPU with CUDA to accelerate the training process. The proposed method was implemented using Python, with key libraries including TensorFlow, NumPy, Pandas, Scikit-learn, and Matplotlib. We employed the EfficientNet and ResNet backbones, and their classification heads were fine-tuned with a CBAM-based attention module. The models were trained using the Adam optimizer with a learning rate of 0.0001. To improve generalization and prevent overfitting, we employed a learning rate reduction strategy using ReduceLROnPlateau to lower the learning rate when validation performance plateaued. Each model was trained for up to 50 epochs within a 5-fold cross-validation framework. Early stopping was applied with a patience of 8 epochs and the restoration of the best-performing weights. Table 5 provides a summary of the hyperparameters configured for our proposed models.

5.3. CE-MRI classification performance evaluation

Our proposed ensemble method was evaluated on the CE-MRI dataset using a 5-fold cross-validation strategy to provide a reliable and unbiased assessment of model performance.

Table 5. Hyperparameters and their values for the proposed models.

Hyperparameters	Values
Batch-size	16
Number of epochs	50
Optimizer	Adam
Learning rate	0.0001
Loss Function	Sparse Categorical Crossentropy
Activation Function	Softmax

Table 6 summarizes the results of the ensemble method across all five folds, using precision, recall, and F1-score metrics. Performance remained robust across all folds, achieving accuracies between 0.9821 and 0.9951, with a mean accuracy of 0.9909. The precision, recall, and F1-score metrics also demonstrated strong stability, with mean values of 98.93, 99.00, and 98.96 across all folds, indicating superior classification performance.

Table 6. Performance of the proposed ensemble method across all five cross-validation folds.

Fold Number	Accuracy	Precision	Recall	F1-Score
Fold 1	99.51	99.47	99.41	99.44
Fold 2	99.18	99.01	99.24	99.12
Fold 3	98.21	97.69	98.23	97.93
Fold 4	99.35	99.23	99.23	99.23
Fold 5	99.18	99.23	98.88	99.05
Mean	**99.09**	**98.93**	**99.00**	**98.96**

Figure 8 illustrates the average training and validation accuracy and loss curves obtained through 5-fold cross-validation for both EfficientNet-B1 and ResNet-101.
These curves represent the mean performance across all folds and provide the learning behavior of these two models.

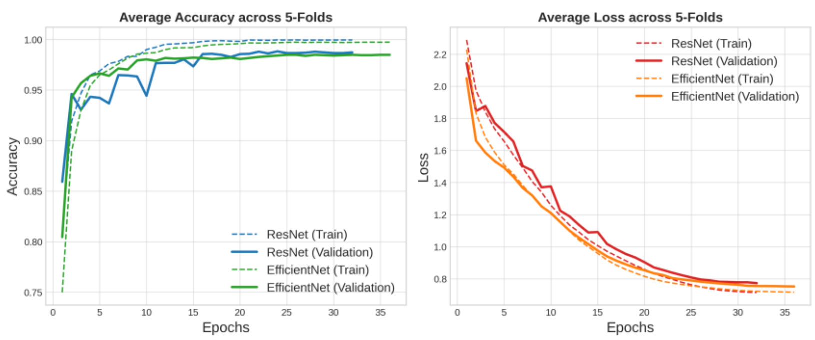

Figure 8. The average accuracy and loss curves for training ResNet and EfficientNet models.

The training accuracy for both models shows a steady and continuous improvement as the training epochs increase. The validation accuracy also follows an overall upward trend, although it has slight fluctuations during the early epochs due to initial adaptation to unseen data, specifically for the

ResNet model. However, after some epochs, it has steady growth to high validation accuracy. The loss curves also support this training behavior; both training and validation loss decrease consistently as training progresses, with EfficientNet showing marginally lower loss values compared to ResNet. This pattern confirms that both models converge effectively and maintain a stable learning procedure throughout the training process. Figure 9 shows the confusion matrix for the proposed method applied to the Figshare dataset. The matrix illustrates the performance of the ensemble model, EfficientNet-B1, and ResNet-101, in classifying brain tumors into three categories: Meningioma, Glioma, and Pituitary. The high percentages along the diagonal indicate that the model accurately predicts the correct class for most instances, with 98.45% for Meningioma, 99.30% for Glioma, and 99.25% for Pituitary tumors. This demonstrates the effectiveness of the proposed method in achieving high accuracy and reliability in brain tumor detection.

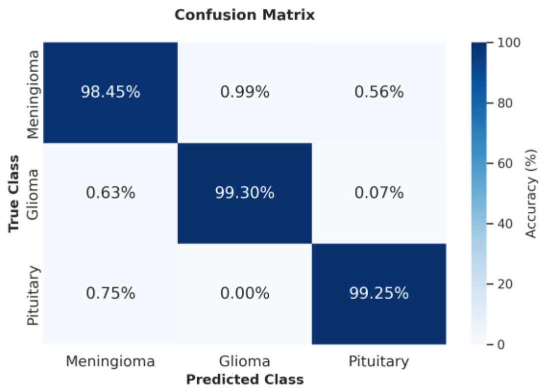

Figure 9. Confusion matrix of the proposed method.

Figure 10 shows the Receiver Operating Characteristic (ROC) curve for the proposed method.

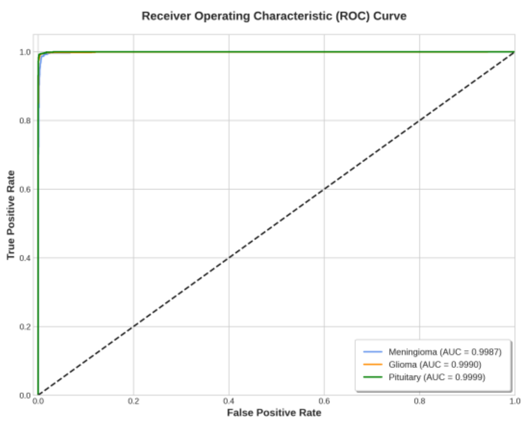

Figure 10. The ROC curve for the proposed method.

The ROC curve illustrates the diagnostic ability of the ensemble model, EfficientNet-B1 and ResNet-101, in distinguishing between different types of brain tumors. The Area Under the Curve (AUC) values for Meningioma, Glioma, and Pituitary tumors are 0.9987, 0.9999, and 0.9999, respectively, indicating excellent classification performance across all classes.

Table 7 presents a comparative analysis of the proposed method with existing approaches on the Figshare dataset. To have a fair comparison with these studies, we used the same 5-fold cross-validation strategy. The table highlights various approaches used for brain tumor detection, including Deep CNN, BoF-SURF, and Pre-train models + PLS methods. Our proposed method, which leverages a CBAM-enhanced attention ensemble of EfficientNet-B1 and ResNet-101, demonstrates superior performance across all evaluated metrics.

Specifically, our model attains an accuracy of 99.09%, outperforming the Deep CNN of Akter et al. with an accuracy of 96.70, the BoFSURF method of Mohammed et al. with an accuracy of 98.70, and the pretrained PLS-based approach of Aamir et al. with an accuracy of 98.95. These results indicate that the proposed ensemble consistently delivers superior performance compared to existing techniques, demonstrating more effective feature representation and stronger generalization across the dataset. Our method also outperformed these prior methods on other evaluation metrics, attaining a precision of 98.93%, a recall of 99.00%, and an F1-score of 98.96%. Achieving improvements within such a high-accuracy domain is inherently challenging. At this level, even marginal gains typically arise from correctly resolving a limited number of difficult and clinically critical cases that prior models often misclassified. The superior precision, recall, and F1-score of the proposed method indicate its effectiveness in addressing these challenging diagnostic instances and demonstrate enhanced robustness and generalization across all tumor categories.

6. Conclusion

In this study, we proposed a novel method for brain tumor detection using a CBAM-enhanced attention ensemble of deep learning networks, specifically EfficientNet-B1 and ResNet-101. The proposed method achieved an impressive accuracy of 99.09% on the Figshare dataset, demonstrating its effectiveness in accurately detecting brain tumors in MRI images. This high level of accuracy makes the method highly applicable in medical imaging, providing a reliable tool for clinicians to diagnose brain tumors. For future work, we suggest exploring the integration of additional deep learning models and incorporating advanced feature extraction techniques to further enhance the model's performance. Additionally, expanding the dataset to include more diverse and larger samples could improve the generalizability of the method. This approach offers a promising solution for improving the accuracy and reliability of brain tumor detection in medical imaging.

Table 7. Comparison of the Proposed Method with Existing Approaches on the Figshare dataset.

Authors	Approaches	Accuracy	Precision	Recall	F1-Score
Akter et al. [2]	Deep CNN	96.70	-	93.60	96.70
Mohammed et al. [30]	BoF-SURF	98.70	98.6	98.4	98.5
Aamir et al. [18]	Pre-train models + PLS	98.95	-	-	-
Our Proposed Method	**CBAM-Enhanced Attention Ensemble**	**99.09**	**98.93**	**99.00**	**98.96**

References

[1] L. Arokia Jesu Prabhu and A. Jayachandran, "Mixture model segmentation system for parasagittal meningioma brain tumor classification based on hybrid feature vector," *Journal of medical systems*, vol. 42, no. 12, p. 251, 2018.

[2] A. Akter et al., "Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor," *Expert Systems with Applications*, vol. 238, p. 122347, 2024.

[3] E. Irmak, "Multi-classification of brain tumor MRI images using deep convolutional neural network with fully optimized framework," *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, vol. 45, no. 3, pp. 1015–1036, 2021.

[4] E.-S. A. El-Dahshan, H. M. Mohsen, K. Revett, and A.-B. M. Salem, "Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm," *Expert systems with Applications*, vol. 41, no. 11, pp. 5526–5545, 2014.

[5] M. Nawaz et al., "Analysis of brain MRI images using improved cornernet approach," *Diagnostics*, vol. 11, no. 10, p. 1856, 2021.

[6] K. M. Iftekharuddin, J. Zheng, M. A. Islam, and R. J. Ogg, "Fractal-based brain tumor detection in multimodal MRI," *Applied Mathematics and Computation*, vol. 207, no. 1, pp. 23–41, 2009.

[7] M. M. Zahoor et al., "Brain tumor MRI classification using a novel deep residual and regional CNN," *Biomedicines*, vol. 12, no. 7, p. 1395, 2024.

[8] P. M. Shakeel, T. E. E. Tobely, H. Al-Feel, G. Manogaran, and S. Baskar, "Neural network based brain tumor detection using wireless infrared imaging sensor," *IEEE Access*, vol. 7, pp. 5577–5588, 2019.

[9] A. Kharrat, K. Gasmi, M. B. Messaoud, N. Benamrane, and M. Abid, "A hybrid approach for automatic classification of brain MRI using genetic algorithm and support vector machine," *Leonardo journal of sciences*, vol. 17, no. 1, pp. 71–82, 2010.

[10] J. Cheng, "brain tumor dataset," ed: figshare, 2017.

[11] M. R. Ismael and I. Abdel-Qader, "Brain tumor classification via statistical features and back-propagation neural network," in *2018 IEEE international conference on electro/information technology* (EIT), 2018: IEEE, pp. 0252–0257.

[12] J. S. Paul, A. J. Plassard, B. A. Landman, and D. Fabbri, "Deep learning for brain tumor classification," in Medical Imaging 2017: *Biomedical Applications in Molecular, Structural, and Functional Imaging*, 2017, vol. 10137: SPIE, pp. 253–268.

[13] P. Afshar, A. Mohammadi, and K. Plataniotis, "Brain tumor type classification via capsule networks. In2018 25th IEEE international conference on image processing (ICIP) 2018 Oct 7 (pp. 3129-3133)," *arXiv preprint arXiv:1801.09597*.

[14] P. Afshar, K. N. Plataniotis, and A. Mohammadi, "Capsule networks for brain tumor classification based on MRI images and coarse tumor boundaries," in *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2019: IEEE, pp. 1368–1372.

[15] A. K. Anaraki, M. Ayati, and F. Kazemi, "Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms," *biocybernetics and biomedical engineering*, vol. 39, no. 1, pp. 63–74, 2019.

[16] N. Ghassemi, A. Shoeibi, and M. Rouhani, "Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images," *Biomedical Signal Processing and Control*, vol. 57, p. 101678, 2020.

[17] Y. Zhou et al., "Holistic brain tumor screening and classification based on densenet and recurrent neural network," in *International MICCAI brainlesion workshop*, 2018: Springer, pp. 208–217.

[18] M. Aamir et al., "A deep learning approach for brain tumor classification using MRI images," *Computers and Electrical Engineering*, vol. 101, p. 108105, 2022.

[19] F. Demir and Y. Akbulut, "A new deep technique using R-CNN model and L1NSR feature selection for brain MRI classification," *Biomedical Signal Processing and Control*, vol. 75, p. 103625, 2022.

[20] A. K. Sartaj Bhuvaji, Prajakta Bhumkar, Sameer Dedge, Swati Kanchan. *Brain Tumor Classification (MRI)*, doi: https://doi.org/10.34740/KAGGLE/DSV/12745533.

[21] X. Liu and Z. Wang, "Deep learning in medical image classification from mri-based brain tumor images," in *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, 2024: IEEE, pp. 840–844.

[22] A. Hamada. Br35H :: *Brain Tumor Detection 2020*. [Online]. Available: https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection

[23] C. K. K. Reddy et al., "A fine-tuned vision transformer based enhanced multi-class brain tumor classification using MRI scan imagery," *Frontiers in oncology*, vol. 14, p. 1400341, 2024.

[24] J. Manikandan and S. Snega, "TriCANet: Cross-Plane Attention and Swin Transformer Fusion for Multi-View Brain MRI Tumor Classification," in *2025 10th International Conference on Research in Intelligent Computing in Engineering (RICE)*, 2025: IEEE, pp. 1–6.

[25] S. U. R. Khan, S. Asif, M. Zhao, W. Zou, Y. Li, and C. Xiao, "ShallowMRI: A novel lightweight CNN with novel attention mechanism for Multi brain tumor classification in MRI images," *Biomedical Signal Processing and Control*, vol. 111, p. 108425, 2026.

[26] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[27] M. Tan and Q. Le, "Efficientnet: Rethinking model scalingfo convolutional neural networks," in *International conference on machine learning*, 2019: PMLR, pp. 6105–6114.

[28] A G. Howard et al., "Mobilenets: Efficient convolutonal neural networks for mobile vision applicatins," *arXiv preprint arXiv*:1704.04861, 2017.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[30] Z. Mohammed and D. J. Mussa, "Brain tumour classificaion using BoF-SURF with filter-based feature selection methods," *Multimedia Tools and Applications*, vol. 83, no. 25, pp. 65833–65855, 2024.

[31] S. Mavaddati, "A Hybrid Approach for Brain Tumor Classification: Enhancing MRI-Based Diagnosis with CNN-Transformer Synergy," *Journal of AI and Data Mining*, vol. 14, no. 1, pp. 37–49, 2026.