Wiley
Complexity
Volume 2025, Article ID 1644859, 14 pages
https://doi.org/10.1155/cplx/1644859

WILEY

Research Article

Convolutional Neural Network and Channel Attention Mechanism for Multiclass Brain Tumor Classification

Ali Naderi,[1] **Akbar Asgharzadeh-Bonab**,[2] **Farid Ahmadi**,[3] **and Hashem Kalbkhani**[4]

[1]*Department of Mechatronics Engineering, Urmia University of Technology, Urmia, Iran*
[2]*Department of Electrical and Computer Engineering, Urmia University, Urmia, Iran*
[3]*Department of IT and Computer Engineering, Urmia University of Technology, Urmia, Iran*
[4]*Department of Electrical Engineering, Urmia University of Technology, Urmia, Iran*

Correspondence should be addressed to Farid Ahmadi; f.ahmadi@uut.ac.ir

Received 30 August 2024; Accepted 7 March 2025

Academic Editor: Hassan Zargarzadeh

The complexity of brain tumors highlights the critical need for advanced computer-aided diagnosis (CAD) tools to support surgeons in clinical decision-making and improve patient outcomes. This paper introduces a novel deep learning model for the multiclass classification of brain tumors using magnetic resonance imaging (MRI), offering significant advancements in feature extraction and classification accuracy. The proposed model comprises three key components: (1) a fine-tuned EfficientNetB7 convolutional neural network (CNN), adapted through transfer learning by freezing the initial layers and retraining subsequent layers to optimize feature extraction from MR images; (2) a channel attention module that refines extracted feature maps, emphasizing essential features for accurate tumor detection; and (3) a fully connected classifier, optimized through grid search, to achieve precise multiclass tumor classification. Additionally, hyperparameter tuning and data augmentation techniques enhance generalization and model robustness. Experimental results confirm the model's superior performance, outperforming recent approaches in multiclass and binary classification scenarios, underscoring its potential to advance brain tumor diagnosis and treatment.

Keywords: brain tumor classification; channel attention; CNN; MRI; transfer learning

1. Introduction

A brain tumor is an abnormal growth of cells in the brain or the central part of the spinal cord. These tumors are either benign (noncancerous) or malignant (cancerous). They can arise from different types of cells within the brain, such as glial cells, which support nerve cells, or neurons (nerve cells). The specific treatment plan depends on factors such as the type of tumor and its location. It is vital for individuals experiencing symptoms that may indicate a brain tumor to seek medical attention promptly for diagnosis and treatment. Early detection and intervention can significantly improve outcomes for individuals with brain tumors [1–3].

Generally, brain tumors are categorized into benign and malignant [4]. Malignant tumors are further classified into meningiomas, pituitary tumors, and gliomas [5]. Meningiomas originate in the meninges, the membranes surrounding the brain and spinal cord. Typically slow-growing and benign (noncancerous), these tumors can occasionally be malignant (cancerous). Meningiomas can occur at any age but are more commonly found in adults, particularly women. Pituitary tumors, medically termed pituitary adenomas, develop within the pituitary gland, a small endocrine organ located at the base of the brain, behind the nose. The pituitary gland is the master gland. This small organ produces hormones that regulate many bodily functions, from growth and metabolism to reproduction and the stress

response. Gliomas are a type of brain tumor originating from glial cells, supportive cells in the central nervous system. There are three main types of glial cells: astrocytes, oligodendrocytes, and ependymal cells. Gliomas can occur anywhere in the brain or spinal cord and vary widely in their aggressiveness and prognosis [6, 7].

Brain tumor classification is crucial as it directly influences diagnosis, treatment planning, and patient outcomes. Accurate classification helps differentiate between benign and malignant tumors, guiding the urgency and type of medical intervention. It enables tailored treatment strategies by identifying specific tumor subtypes and their molecular or genetic markers, paving the way for precision medicine. Moreover, classification aids in predicting prognosis, understanding disease progression, and facilitating research into novel therapies. Advanced methods, such as those integrating imaging, histopathological, and genetic data, ensure consistency and objectivity in classification. Brain tumor classification is fundamental for effective clinical decision-making and advancing oncological research.

Therefore, brain tumors must be detected as early as possible. It is well-known that manual detection and classification of brain tumors is challenging and error-prone; therefore, the classification of these tumors requires a specialized radiologist. The automatic classification of brain tumors using MRI has become a prominent area of research, with numerous machine learning techniques being widely studied and implemented in this field. Using machine learning and deep learning methods, computer-aided diagnosis (CAD) tools for brain tumors have made tremendous progress. They are invaluable for medical experts in detecting and classifying brain tumors. These highly reliable techniques offer greater precision and save valuable time [8]. Deep learning—a subfield of machine learning—is the ideal approach for automatic brain tumor classification using MRI. It is a superior alternative to traditional machine learning approaches because it does not require manual functions and offers excellent results [9]. The success of deep learning approaches in medical imaging is undeniable. Their superior performance in tasks such as segmentation and classification is unparalleled. However, their effectiveness hinges on a crucial factor: the size of the dataset. This poses a significant obstacle in the medical field, where datasets are often limited. Transfer learning is a game-changing technique that addresses this challenge head-on [10].

This paper proposes a novel deep architecture for classifying brain tumors. At first, the transfer learning approach is utilized to fine-tune the EfficientNetB7 pretrained model [11]. The first four blocks of this model are frozen, and the following three blocks are to be learned for feature extraction from MR images. After the last layer of the EfficientNetB7, a channel attention module (CAM) [12] is added to refine the extracted feature map to highlight the features needed for the classification task. Finally, a fully connected classifier is designed. The main contributions of this paper can be summarized as follows:

- Designing a feature extraction framework based on transfer learning and fine-tuning
- Refining the extracted features by adding a CAM to the EfficientNetB7
- Designing a fully connected classifier and optimizing its architecture by grid search to accurately classify MR images into two- and four-class scenarios
- Employing data augmentation and hyperparameter tuning to improve the generalization and training phase
- Performing extensive simulations and ablation studies on some datasets to demonstrate the efficiency of the proposed method in four-class and binary scenarios

The remainder of the paper is organized as follows. Section 2 reviews the related works. Section 3 introduces the proposed deep architecture. Section 4 presents the dataset and simulation setup. Results are given in Section 5, and Section 6 concludes the paper.

2. Related Works

In recent years, the automatic classification of brain tumors has been extensively studied using traditional machine learning and deep learning techniques. Traditional machine learning techniques consist of three basic steps: preprocessing, feature extraction, and classification. Based on recent studies, researchers have proposed several methods for automatically classifying brain tumors. Some works used the transfer learning approach and pretrained networks. Dedicated convolutional neural networks (CNNs) and vision transformers (ViTs) are also designed for feature extraction, and then fully connected layers or other traditional classifiers are used for classification. Also, the attention mechanism is utilized to obtain more helpful feature maps.

To address poor MRI image contrast, the system proposed in [1] first applies optimal dual-threshold wavelet-constrained histogram equalization to enhance image contrast. It then utilizes deep transfer learning with a pretrained Inception-V3 model. The authors in [13] presented a comparative performance analysis of CNN-based models, including VGG-16, residual network (ResNet)-50, and Inception-v3, for predicting tumor cells based on transfer learning. Authors in [14] showed that transfer learning methods achieve high accuracy and low computational cost with pretrained networks. However, the study highlights the need for a more precise and error-free model for the high-risk medical field. A new CNN, an improved ResNet50, was presented in [15] to classify brain tumors in MRI images into benign or malignant classes. Authors in [16] introduced an approach for brain tumor classification using a CNN based on the EfficientNetV2b0 architecture, which employs transfer learning to enhance performance and reduce training time, addressing the challenge of limited medical image data.

The researchers in [17] achieved 99.25% accuracy in classifying brain tumor images using a dedicated CNN. The research in [18] presents M-Net, a deep learning architecture

based on an encoder-decoder network designed for medical image segmentation and classification, explicitly targeting the diagnosis of Alzheimer's disease using structural MRI (sMRI) data. In [19], an optimized histogram equalization enhances the contrast, and then, a new nine-layer CNN model is trained. The features were selected from the last fully connected using differential evolution and flame optimization and subsequently were classified using the multiclass support vector machine (SVM). In [20], the authors first increase the image's visual quality using morphological analysis. Subsequently, an adaptive fusion network generates a consolidated feature vector that multiclass SVM classifies.

In [21], feature fusion blocks, attention modules, and CNNs were utilized to develop a robust multipath network. The attention feature fusion VGG19 (AFF-VGG19) network has been demonstrated to outperform other advanced networks. Furthermore, in [22], deep convolutional generative adversarial networks, the unsupervised learning method, have been employed to generate expansion datasets on four types of MRI brain tumor images. Recently, the method of [23] presents a novel 3D hybrid model named 3DUV-NetR + for segmenting brain tumors using multimodal MRI data. The architecture combines 3D U-Net and V-Net models, enhanced with transformers, to capture detailed semantic features and spatial relationships within medical images. A hybrid quantum dilated CNN-deep maxout network (QDCNN-DMN) was proposed in [24] for brain tumor classification using MRI images. The method involves image enhancement using logarithmic transformations, skull-stripping with deep embedded clustering, and tumor segmentation through a structure-correcting adversarial network. Statistical and critical features are then extracted for classification. The pyramid QDCNN (PyQDCNN) was proposed in [25] for multilevel brain tumor classification using MRI images. The model includes a preprocessing stage using a median filter to remove noise segmentation with U-Next optimized by exponential Tasmanian Devil anticoronavirus optimization.

Self-supervised learning (SSL) using the DINOV2 ViT-S14 model was utilized in [26] to enhance the accuracy of brain tumor diagnosis. This approach leverages pretrained ViT models, including self-distillation with no labels (DINO), SimCLRv2-ResNet50, and supervised models such as ResNet50, VGG-19, and EfficientNetB0, to achieve improved diagnostic performance. The study of [27] presents a self-supervised semantic segmentation framework for medical images with limited annotations. It combines segmentation and contrastive losses to distinguish anatomical regions, using a parallel transformer module with multiview multiscale feature fusion for enhanced performance. A multiencoder transformer is trained on unlabeled data and fine-tuned with minimal annotations, achieving superior results in brain tumor segmentation (Brats) and other tasks. In [28], a SSL method was proposed to automatically extract meaningful characteristics of Alzheimer's disease using the ViT architecture. This approach employs a pretrained feature extractor based on DINO and utilizes an extreme learning machine (ELM) as the classifier. In [29], a generative SSL model comprising two stages was introduced for brain tumor classification. The first stage involves pretraining a Res ViT (ResViT) model for MRI synthesis as a pretext task. The second stage focuses on fine-tuning a ResViT-based classifier model for the downstream classification task.

Even though all the works in this category have good accuracy and have tried to reduce computing costs, they still require high computing power compared to other works, and the time to perform and use these models is long. The motivation behind this work was the success of using pretrained models, attention mechanisms, and dense layers as feature extractors and classifiers in various object recognition problems.

3. Proposed Methodology

This paper aims to design a deep learning model capable of automatically categorizing brain MR images into either two (no-tumor and tumor) or four categories (no-tumor, glioma, meningioma, and pituitary tumor). Categorizing MR images into no-tumor or tumor categories involves binary classification tasks. Conversely, categorizing brain MR images into four categories constitutes a multiclass classification task. The procedure of the proposed method is shown in Figure 1. The training pipeline for the proposed method for brain MRI classification begins with data augmentation. The medical datasets usually contain a small number of samples, while the deep networks require a more significant number of samples for training. Hence, it is necessary to utilize data augmentation techniques to increase the number of training samples. This paper considers the following operations: (1) zero-mean Gaussian noise with variances 0.01 and 0.02 is added to images, (2) gamma correction with gamma value ranging from 0.7 to 1.3 is applied on images, (3) images are rotated with a step size of five degrees using from the angle of -45° to 45°, and (4) images are flipped in different vertical and horizontal directions. After that, in the preprocessing step, the MR image is resized to 224×224 dimensions using bicubic interpolation.

3.1. Feature Extraction. The proposed model uses the pretrained CNN, EfficientNetB7, for the first phase of feature extraction. EfficientNetB7 was chosen as the pretraining model due to its exceptional balance between accuracy and efficiency, leveraging a scalable compound coefficient framework that optimizes network width, depth, and resolution. As the largest and most powerful variant in the EfficientNet family, it achieves state-of-the-art performance on benchmarks like ImageNet while maintaining computational efficiency with fewer parameters and operations than other leading models. Its pretrained weights, derived from extensive training on large datasets, provide a strong foundation for fine-tuning specific tasks, enabling high accuracy with reduced data requirements. This scalability, performance, and efficiency combination makes EfficientNetB7 ideal for demanding and complex tasks. As shown in Table 1 and Figure 2, this

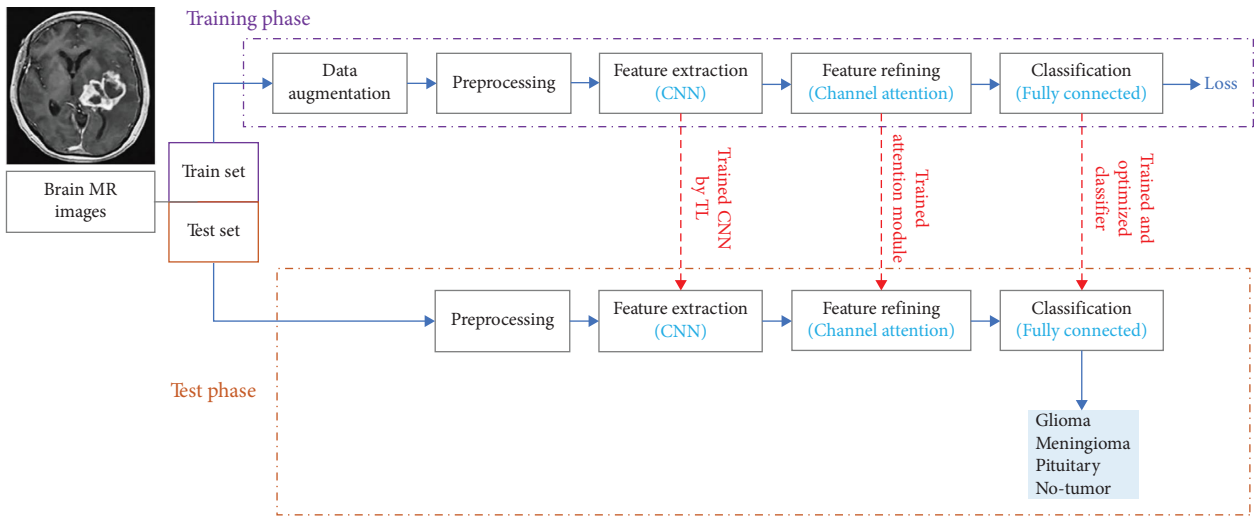


Figure 1: The procedure of the proposed method for multiclass brain tumor classification.

Table 1: Specifications of the EfficientNetB7 pretrained model used in this study.

	Block	Operator	Resolution	Number of channels	Number of layers
Freezed	0	Conv, 3 × 3	224 × 224	64	1
	1	MBConv1, 3 × 3	112 × 112	32	4
	2	MBConv6, 3 × 3	56 × 56	192	7
	3	MBConv6, 5 × 5	28 × 28	288	7
	4	MBConv6, 3 × 3	14 × 14	480	10
Fine-tuned	5	MBConv6, 5 × 5	14 × 14	960	10
	6	MBConv6, 5 × 5	7 × 7	1344	13
	7	MBConv6, 3 × 3	7 × 7	3840	4

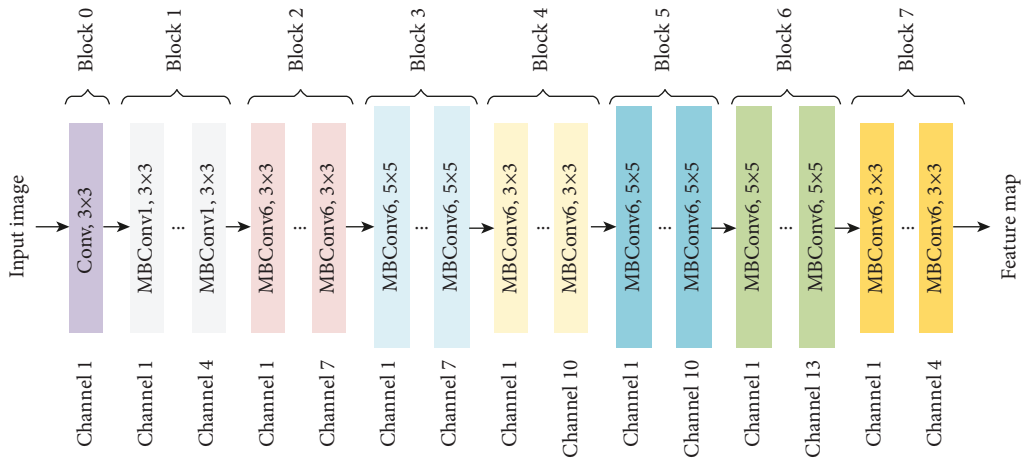


Figure 2: Architecture of EfficientNetB7.

model consists of multiple MBConv blocks (mobile inverted bottleneck convolution), as illustrated in Figure 3, which serve as the fundamental building blocks of EfficientNet [30]. The architecture is organized into seven blocks, each distinguished by a unique color. Each MBConvX block is annotated with its respective filter size, where $X = 1$ and $X = 6$ indicate standard ReLU and ReLU6 activation functions, respectively.

These layers assist the model in extracting features from the input MR images set. Furthermore, in addition to the conventional CNN, the proposed model employs high-level feature chaining through skip connections utilizing the global average pooling layer [31]. By concatenating high-level features, the network can learn more sophisticated features, resulting in more accurate classification of normal brain and various brain tumors MR images.

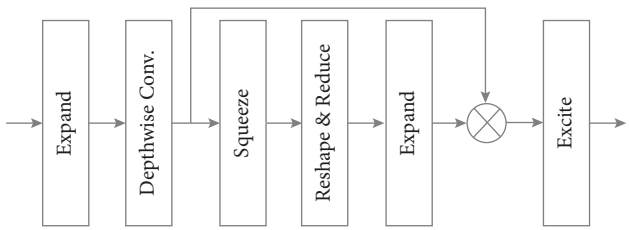


Figure 3: MBConv: the basic building block of EfficientNet.

Freezing Partial Network: A freezing partial network is one in which a subset of nodes is frozen, preventing further communication between them. In machine learning, freezing refers to setting the weights of a specific layer or the entire network during the fine-tuning phase. The network freezing enables the retention of the knowledge captured by the pretrained model while updating only specific layers, allowing for the model's adaptation to the target task. Therefore, this step must not be overlooked when refining a pretrained model [32].

If the pretrained model was trained on a substantial dataset analogous to the target task, freezing the entire network can assist in preserving the learned representations, thereby preventing their overwriting. Conversely, the freezing of only a subset of the network may prove advantageous if the lower layers of the model of pretrained have captured general features that are likely to be pertinent to the novel task. Together with freezing these lower layers, it is possible to utilize the knowledge acquired by the pretrained model during the process of updating the upper layers to enhance their task-specific feature extraction capabilities. This approach is utilized in the proposed model due to the limited size of the brain tumor dataset, and the dataset in question differs from the one on which the pretrained model was trained. As shown in Table 1, the first four MBConv blocks of the pretrained model EfficientNetB7 are frozen.

3.2. Feature Refining. Attention modules make models learn and focus more on important information instead of learning useless background information. A CAM, shown in Figure 4, is a module for channel-based attention in CNNs. A CAM is built by exploiting the feature relationship between channels. Since each channel of a feature map is considered a feature detector, the channel's attention is focused on "what" makes sense given an input image. Therefore, after extracting the feature map from the pretrained model EfficientNetB7, it is given into a channel attention layer. Furthermore, the channel attention block in the proposed model helps the network to focus only on the relevant area of the brain for better classification [33].

Given an intermediate feature map $\mathbf{X} \in \mathbb{R}^{c\times h\times w}$ as input, the CAM infers a one-dimensional attention map $\mathbf{M}_c \in \mathbb{R}^{c\times 1\times 1}$ as shown in Figure 2, can be summarized as [33]

$$\mathbf{X}' = M_c(\mathbf{X}) \otimes \mathbf{X}, \tag{1}$$

where $\otimes$ denotes element-wise multiplication. At first, the spatial information of a feature map is aggregated by using average-pooling and max-pooling information. As a result, two different spatial context descriptors, $\mathbf{X}^c_{\text{avg}}$ and $\mathbf{X}^c_{\text{max}}$, are generated, which denote the average-pooled and max-pooled features, respectively. The CAM $\mathbf{M}_c \in \mathbb{R}^{c\times 1\times 1}$ is produced by forwarding both descriptors to a shared multilayer perceptron (MLP) network with one hidden layer. The activation size of the hidden layer is set to $\mathbb{R}^{c/r\times 1\times 1}$ to reduce the parameter overhead, where r is the reduction ratio. The output feature vectors, obtained by applying shared MLP to each descriptor, are merged using element-wise summation and passed through the sigmoid function. This process can be formulated as [33]

$$\begin{aligned}\mathbf{M}_c(\mathbf{X}) &= \sigma(\text{MLP}(\text{AvgPool}(\mathbf{X})) + \text{MLP}(\text{MaxPool}(\mathbf{X})))\\ &= \sigma\left(\mathbf{W_1}\left(\mathbf{W_0}\left(\mathbf{X}^c_{\text{avg}}\right)\right) + \mathbf{W_1}\left(\mathbf{W_0}\left(\mathbf{X}^c_{\text{max}}\right)\right)\right),\end{aligned} \tag{2}$$

where σ denotes the sigmoid function, $\mathbf{W_0} \in \mathbb{R}^{c/r\times c}$ and $\mathbf{W_1} \in \mathbb{R}^{c/r\times c}$. Note that the MLP weights, $\mathbf{W_0}$ and $\mathbf{W_1}$, are shared for both inputs, and the rectified linear unit (ReLU) activation function is followed by $\mathbf{W_0}$ [33].

3.3. Classification. The structure of the considered classifier and related parameters is given in Table 2. The designed fully connected classifier comprises a batch normalization layer, two dense layers, and a dropout layer. A dense layer is one in which every neuron is connected to every neuron in the previous layer. In other words, the output of each neuron in a dense layer is calculated as the weighted sum of the inputs of all neurons in the previous layer. In the context of artificial neural networks, the activation function is a mathematical operation employed to transform the input values of the neurons [34]. It introduces nonlinearity into artificial neural networks, enabling them to discern the relationship between output and input values. In the first dense layer, we employed a type of activation function called the ReLU activation function. The CNN employs dense layers to transform the concatenated features into high-level features by applying the nonlinear ReLU activation function. Moreover, the second dense layer employed a softmax activation function to transform a vector of values into a probability distribution. The number of neurons in the first dense layer is obtained by grid-search optimization, while the number of neurons in the second dense layer depends on the number of classes; hence, there are two or four neurons in the second dense layer.

The dropout layer [35] is used when building neural networks to prevent overfitting. Individual nodes are excluded in several training runs with a 45% probability as if they were not part of the network architecture. Batch normalization is a technique used in machine learning to normalize the inputs to a level for each mini-batch. Batch normalization is a technique employed during the training

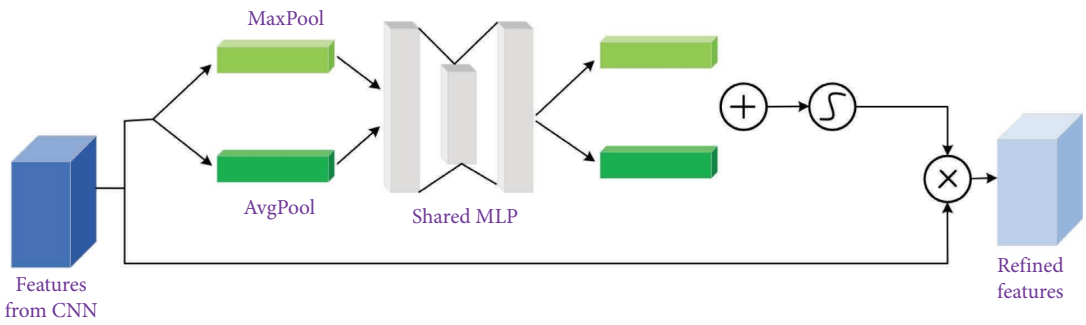


Figure 4: Channel attention module architecture.

Table 2: Structure of the fully connected classifier used in this study.

Layer	Parameters (value)
Batch normalization	
Dense 1	Units (256), activation (ReLU)
Dropout	Rate (0.45)
Dense 2	Units (class counts, 2 or 4), activation (softmax)

of extremely deep neural networks. It entails standardizing the inputs to a consistent level for each mini-batch [36]. This has the effect of stabilizing the learning process and significantly reducing the number of training epochs that are required to train deep networks.

4. Simulation Setup

4.1. Dataset. This section presents the datasets of brain tumor MR images utilized in this study. These datasets were downloaded from the Kaggle website belonging to Brats challenges. These datasets are Brats-4C [37] (the term 4C indicates four classes), Brats-2C-large [38] (the term 2C indicates two classes), and Brats-2C-small [39]. In Brats-2C-small and Brats-2C-large, the MR images were saved in different sizes, dpi values, and JPEG format and belonging to tumor or no-tumor classes. The Brats-4C contains the MRI of normal brain tissue and three different types of brain tumors (glioma, pituitary, and meningioma). This dataset comprises 3264 MR images in JPEG format and standard sizes (512×512) with 96 dpi. The training and test procedures were performed separately for each dataset. A detailed description of each brain tumor dataset is provided in Table 3 and Figures 5, 6, 7 illustrate representative examples of each class.

4.2. Performance Metrics. Simulations were conducted on Python using TensorFlow and Keras libraries using specific hardware specifications, including Quad-Core 2.9 GHz Intel i7 processor, 32 GB of RAM, and NVIDIA GeForce GTX 1070 Ti GPU. The efficacy of the proposed model is assessed by utilizing some evaluation metrics, including recall, precision, specificity, accuracy, and F1-score that are calculated as follows:

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

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}}, \tag{4}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}, \tag{5}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}}, \tag{6}$$

$$F_1 = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{7}$$

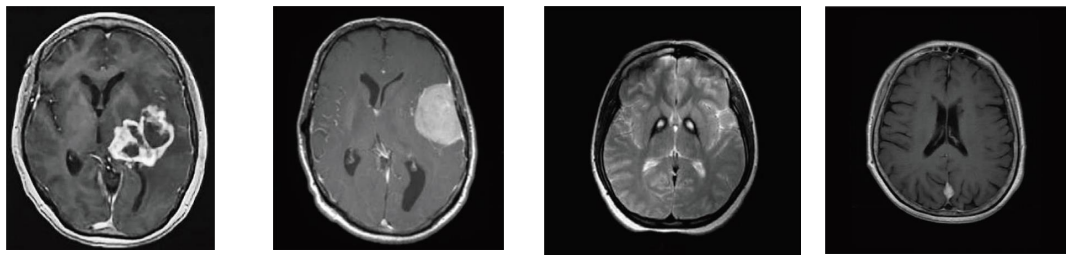
In this context, TP, TN, FN, and FP represent true positive, true negative, false negative, and false positive, respectively.

4.3. Data Partitioning. Stratified sampling is significant when the number of images in each class is not uniform. The "no-tumor" class in Brats-4C and Brats-2C-small has significantly fewer images than the other classes, which could lead to its underrepresentation in any random split. Stratified sampling ensures that each class is proportionally represented in the training, validation, and test sets. This is especially crucial in multiclass classification problems to avoid bias toward the classes with more samples. Stratified sampling can be used with five-fold cross-validation and hold-out (training-validation-test) splits, making it a versatile choice for ensuring a balanced representation of all classes.

Five-fold cross-validation with stratified sampling would be the most robust option in our simulations. This method splits the dataset into five folds, ensuring each fold contains an appropriate proportion of each class. It then trains and evaluates the model five times, each with a different fold as the validation set. This approach reduces the risk of overfitting and ensures that the model's performance is evaluated on different subsets of the data, giving a more reliable measure of generalization. Given the class imbalance, this method provides a good balance between using the data for training while still testing on unseen data. The repeated evaluation across multiple folds also helps mitigate potential biases from a single validation split. While the hold-out (training-validation-test) method could be faster and computationally simpler, it is not as robust as cross-

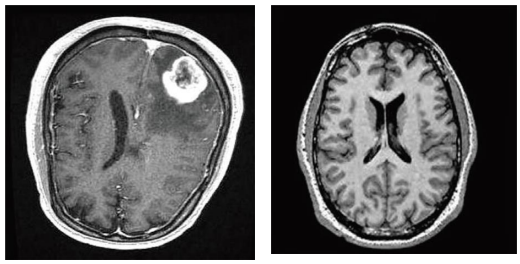
Table 3: Details of datasets used in the simulations.

Dataset	Class name	Number of images	
Brats-2C-small	Tumor	155	Total = 253
	No-tumor	98	
Brats-2C-large	Tumor	1500	Total = 3000
	No-tumor	1500	
Brats-4C	Glioma	926	Total = 3264
	Meningioma	937	
	Pituitary	901	
	No-tumor	500	



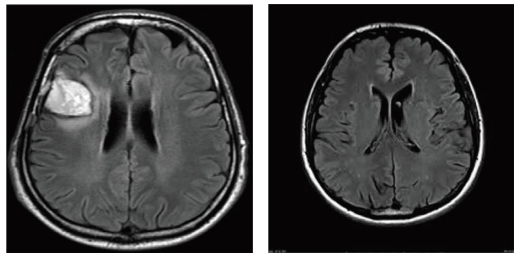
(a) (b) (c) (d)

Figure 5: Samples from Brats-4C. (a) Glioma, (b) meningioma, (c) pituitary, and (d) no-tumor (normal).



(a) (b)

Figure 6: Samples from Brats-2C-small. (a) Tumor and (b) no-tumor (normal).



(a) (b)

Figure 7: Samples from Brats-2C-large. (a) Tumor and (b) no-tumor (normal).

validation, especially with imbalanced datasets. Also, hold-out methods are more likely to suffer from high variance in performance, as the evaluation is based on a single training-validation-test split.

Hence, we considered the stratified five-fold cross-validation. It ensures that each fold maintains the class distribution, provides more reliable performance metrics, and helps to mitigate the impact of class imbalance. This method will likely give the most robust and generalized model performance evaluation. To implement this cross-validation, first, the data shuffling is performed on the dataset to ensure randomness. This step is important for preventing any ordering bias in the data. Then, the dataset is divided into five equal-sized subsets (or folds). After that, the stratification ensures that the class distribution within each fold mirrors the class distribution of the entire dataset. In each of the five iterations, the model will be trained using four folds and validated using the remaining fold. Since the data is stratified, the class distribution in the training and validation sets will be similar in each iteration. After the five iterations, the averages of the performance metrics are calculated from each fold to get an overall performance estimate.

4.4. Hyperparameter Setting. As evidenced in [40], fine-tuning necessitates iterative tuning and meticulous experimentation of parameters and hyperparameters to attain the optimal equilibrium and the desired level of performance in the bespoke model. By fine-tuning, the capabilities of the proposed architecture can be adapted according to the requirements of a specific task, thus saving time and valuable computing resources. The hyperparameters used in the training process of the proposed architecture are given in Table 4. The learning rate is a hyperparameter that determines the step size in each iteration during optimization [41]. This parameter determines the number of model parameters updated in response to the gradients calculated during back-propagation. The loss function measures the discrepancy between the model's predicted outcomes and the correct values [42]. In fine-tuning a pretrained model, selecting an appropriate loss function that aligns with the specific

Table 4: Hyperparameters used in training.

Parameter	Value
Optimizer	SGDM (stochastic gradient descent with momentum)
Loss function	Cross-entropy
Momentum	0.85
Learning rate	0.0001
Epsilon	0.001
Number of epochs	50
Batch size	16
Kernel regularizer	L2 (0.016)
Activity regularizer	L1 (0.006)
Bias regularizer	L1 (0.006)

task we are working on is paramount. Selecting an appropriate loss function is crucial for ensuring the model is optimized for the desired outcome during the training phase. Integrating regularization methods, such as L1 or L2, into the loss function during fine-tuning can prevent overfitting and enhance model generalization. As elucidated in [43], L2 regularization can be applied by setting the value of weight decay in the optimizer, whereas L1 regularization necessitates a somewhat distinct methodology.

5. Results

5.1. Accuracy of Different Pretrained Models. Transfer learning is a technique that employs the knowledge derived from previously trained models to assist in the generalization of a created model to new datasets, thereby enhancing its performance without necessitating a substantial amount of training data. This paper explores their impact and potential applications by examining five of the most effective pretrained models, including VGG16, ResNet50, EfficientNetB0, EfficientNetB3, and EfficientNetB7. The specifics of each of these models are provided in Table 5. It should be noted that the input size of all CNNs is 224×224.

The Visual Geometry Group at the University of Oxford developed the VGG16 CNN model, which is widely used due to its simplicity and effectiveness [44]. The model's architecture comprises 16 convolutional and fully connected layers with a fixed input size. In classifying MR images using this network, an accuracy of 0.9249 is obtained, the lowest accuracy compared to other pretrained models. In addition, the amount of classification loss is more considerable than in other models. Therefore, it is not a suitable model for continuing the experiments and building our brain tumor classification model. The ResNet is a CNN architecture introduced by [37] to mitigate the vanishing gradient problem. ResNet represents a revolutionary CNN architecture that introduced skip connections to mitigate the vanishing gradient problem. Various ResNet models have been trained on large-scale image classification datasets. Considering Figure 8, in our experiments, this model achieved an accuracy of 0.9723 and a loss of 0.179. Tan et al. [45] provide a solution to the problem of model scaling with EfficientNet, which represents a significant advancement in deep learning. By employing a systematic approach to scaling width, depth, and resolution, EfficientNet has achieved state-of-the-art performance while maintaining efficiency. Among the examined models for the classification of brain tumor images, there are three models of EfficientNet, including EfficientNetB0, EfficientNetB3, and EfficientNetB7, which have been able to record accuracies of 0.9723, 0.9754, and 0.9816, respectively, with close performance. According to the results, model EfficientNetB7, with the lowest loss rate and the highest accuracy, is the best pretrained model to build our brain tumor classification model. Therefore, we continue our experiments to improve its performance.

5.2. The Impact of CAM and Layer Freezing. After extracting the feature map from the EfficientNetB7, a CAM layer helps focus on valuable weights. Table 6 shows the effect of this layer in improving the performance of the proposed model for brain tumor detection. It can be seen that this layer increases the accuracy and reduces the loss of the model.

Since the model EfficientNetB7 is trained on the ImageNet dataset, which is different from brain tumor images, we freeze the model's first four blocks to maintain the model's general weights and retrain the model's last three blocks. The required weights for brain tumor classification are extracted, and the required features are taught to the model. According to Table 6, it can be seen that this technique achieves 0.9816 accuracy and 0.110 loss, which has grown compared to other models.

5.3. Performance Metrics on Datasets. Table 7 and Figure 6 present the performance metrics of the proposed model on different datasets. In the Brats-4C dataset, a multiclass classification scenario, the proposed model achieved an accuracy of 0.9816. The proposed model demonstrated the precision, recall, F_1-score, and specificity of 0.998, 0.998, 0.998, and 0.99, respectively. Also, the proposed model for binary scenarios, datasets Brats-2C-large and Brats-2C-small, reaches the accuracy of 0.994 and 0.992, respectively. These results reveal the proposed method's efficiency for classifying brain tumors from MR images.

5.4. Confusion Matrix. A confusion matrix is a table used to define a classification algorithm's performance. This matrix visualizes and summarizes the performance of a classification algorithm. The confusion matrices for datasets Brats-4C, Brats-2C-large, and Brats-2C-small are given in Tables 8, 9, and 10. According to Table 8, the accuracy of the proposed method is 0.9708, 0.9765, 0.9933, and 0.99 in classifying brain tumor

Table 5: Specifications and performance of different CNNs.

CNN	Depth	No. Conv	No. MP	Parameters (m)
VGG16	16	13	5	138.4
ResNet50	107	48	2	25.6
EfficientNetB0	132	65	1	5.3
EfficientNetB3	210	104	1	112.3
EfficientNetB7	438	218	1	66.7

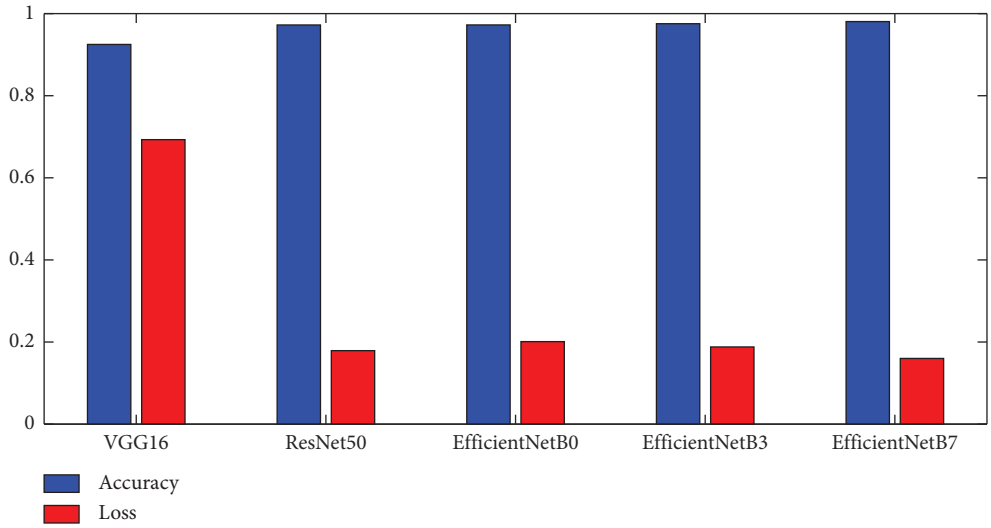


Figure 8: The accuracy and loss of the pretrained models.

Table 6: Ablation study on the impact of attention module and layer freezing.

Architecture	Accuracy	Loss
EfficientNetB7	0.9706	0.160
EfficientNetB7 + layer freezing	0.9715	0.154
EfficientNetB7 + channel attention	0.9811	0.142
EfficientNetB7 + channel attention + layer freezing	**0.9816**	**0.110**

Note: Bold values represent the best performance.

Table 7: Performance metrics of the proposed model on different datasets.

Dataset	Accuracy	Loss	Precision	Recall	F1-score	Specificity
Brats-4C	0.9816	0.110	0.998	0.998	0.998	0.99
Brats-2C-large	0.994	0.062	0.992	0.996	0.994	0.992
Brats-2C-small	0.992	0.084	0.994	0.994	0.994	0.989

images belonging to the glioma, meningioma, pituitary, and no-tumor categories, respectively. A total of 2.16% of the brain tumor images from the glioma class were misclassified and detected as meningioma. Furthermore, 1.71% of samples from the meningioma class are misclassified as glioma. Also, 0.67% of samples from the pituitary class and 1% from the no-tumor class were misclassified as meningioma and glioma. Tables 9 and 10 show that the maximum misclassification in binary scenarios is 1.02%. These results indicate the efficiency of the proposed method.

The reliance on specific datasets (Brats-4C, Brats-2C-Large, Brats-2C-Small) may not generalize to real-world clinical data, which often includes variability in noise, resolution, and patient demographics. Glioma and meningioma classes showed notable misclassification (2.16% of glioma misclassified as meningioma). This may arise from overlapping features or insufficient differentiation in the MR images. Also, while the CAM boosts performance, it could inadvertently amplify noise or irrelevant features in noisy datasets.

Expanding the dataset to include a broader range of MR image conditions, patient demographics, and image acquisition techniques can enhance the accuracy. Also, incorporating synthetic data generation using advanced generative models can simulate rare tumor subtypes or edge cases. Introducing region-specific attention mechanisms helps better focus on tumor-relevant areas, and ensemble

Table 8: Confusion matrix for Brats-4C dataset.

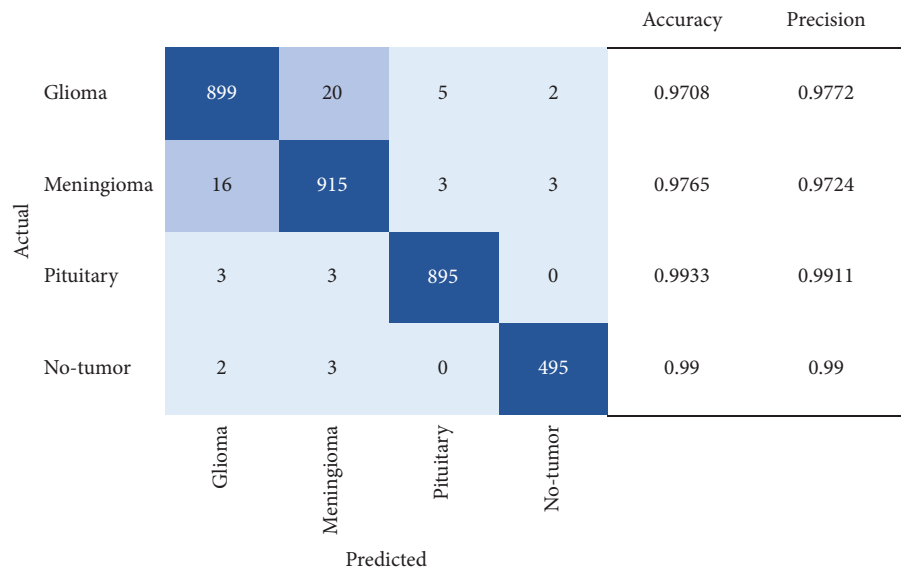


Table 9: Confusion matrix for Brats-2C-large dataset.

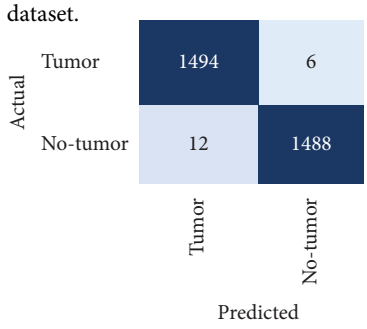
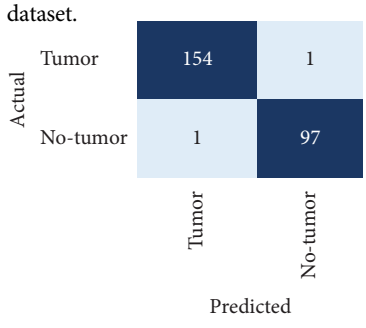


Table 10: Confusion matrix for Brats-2C-small dataset.



methods that combine predictions from multiple architectures can reduce errors in challenging cases. The explainable AI techniques can be integrated to visualize the model's decision-making process and improve clinical trust.

5.5. Performance Comparison. Here, the performance of the proposed method is compared with the recently introduced methods in Table 11. To facilitate a fair comparison, only state-of-the-art methodologies whose algorithms were applied to the identical datasets utilized in the present study are selected. The introduced scheme in [5] comprised two modules. At first, a simple, unsupervised convolutional PCANet extracts the features; then, a supervised CNN classifies the features. This model reached the accuracy of 0.9673, 0.9796, 0.9883, and 0.9402 on Cheng, Brats-2C-small, Brats-2C-large, and Brats-4C datasets, respectively. In [46], the features were extracted from brain MRI using some pretrained CNNs. The features are extracted from each CNN separately and concatenated as an ensemble of deep features. A classifier then classified these features. Their results show that combining extracted features by DenseNet-169, ShuffleNet, and MnasNet and their classification by a fully connected classifier reaches the highest accuracy. The accuracy of 0.9216, 0.9816, and 0.9158 was obtained for Brats-2C-small, Brats-2C-large, and Brats-4C, respectively. Authors in [47] designed their dedicated CNN. This CNN achieves an accuracy of 0.9699 on the Brats-2C-small dataset. The fine-tuned two-layer stacked sparse autoencoder model was introduced in [48] and yielded an accuracy of 0.95 on the Brats-2C-small dataset. In [49], convolutional and fully connected layers of a residual-CNN (R-CNN) model extract the deep features. After selecting the 100 highest distinctiveness features, SVM with Gaussian kernel was used for classification. The results show that this method reaches the accuracy of 0.98 and 0.966 on Brats-2C-small and Brats-4C datasets. The dedicated CNN designed in [50] reached an accuracy of 0.9266 on the Brats-4C dataset.

The proposed EfficientNetB7-based model with a CAM demonstrates significant advancements in brain tumor classification, achieving superior accuracy and performance metrics compared to state-of-the-art methods. The model identifies tumor-relevant features by fine-tuning EfficientNetB7 and leveraging the CAM for feature refinement, achieving a remarkable accuracy of 98.16% on the Brats-4C dataset. Its computational efficiency and robust performance on limited datasets further underscore its potential for clinical applications.

Table 11: Performance comparison between the proposed model and recently introduced methods.

Dataset	Reference	Method	Acc. (%)
Brats-2C-small	Kang et al. [46]	(DenseNet-169 + ShuffleNet + MnasNet) + fully connected	0.9216
	Kalaiselvi et al. [47]	Designed CNN by authors	0.9699
	Amin et al. [48]	Stacked sparse autoencoder model	0.95
	Demir et al. [49]	R-CNN	0.988
	Shahin et al. [5]	MPCANet	0.9796
	Proposed method	EfficientNetB7 + CAM + fully-connected	0.992
Brats-2C-large	Kang et al. [46]	(DenseNet-169 + ShuffleNet + MnasNet) + fully connected	0.9867
	Shahin et al. [5]	MPCANet	0.9883
	Proposed method	EfficientNetB7 + CAM + fully-connected	0.994
Brats-4C	Kang et al. [46]	(DenseNet-169 + ShuffleNet + MnasNet) + fully connected	0.9158
	Irmak [50]	Designed CNN by authors	0.9266
	Demir and Akbulut [49]	R-CNN	0.966
	Shahin et al. [5]	MPCANet	0.9402
	Proposed method	EfficientNetB7 + CAM + fully-connected	0.9816

Table 12: The effect of the number of neurons on the accuracy.

Number of neurons	2	4	8	16	32	64	128	256	512	1024
Accuracy	0.8741	0.8943	0.9066	0.9216	0.9387	0.9531	0.9654	0.9816	0.9819	0.9822

Table 13: The effect of training optimizer on the accuracy of the proposed method.

Optimizer	RMSPROP	ADAM	SGDM
Accuracy	0.8805	0.9326	0.9816

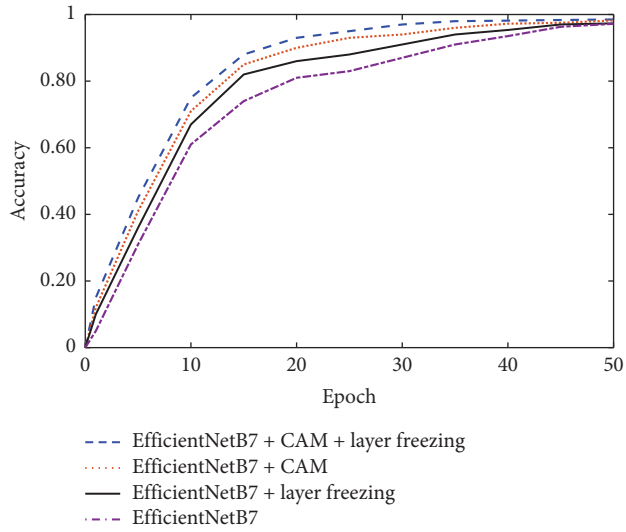


Figure 9: Convergence of the proposed method.

While the proposed deep learning architecture demonstrates impressive performance in multiclass brain tumor classification, several limitations warrant discussion. One major constraint is the dependence on specific datasets (Brats-4C, Brats-2C-large, and Brats-2C-small), which may limit the generalizability of the results to other datasets or real-world clinical scenarios with varied imaging conditions, noise levels, and patient demographics. The reliance on augmentation, fine-tuning of EfficientNetB7, and optimization for hyperparameter tuning adds complexity to the training phase. Moreover, like many deep learning models, the proposed architecture functions as a black box, lacking transparency in decision-making. This could impede its adoption by clinicians who require interpretable results for critical decision-making.

5.6. Hyperparameter Tuning. As mentioned, the fine-tuning of hyperparameters was adopted to maximize the accuracy of the proposed method. The fine-tuning process assessed the performance of the proposed method for the set of considered values and selected the value that yielded the

maximum accuracy. Table 12 presents the accuracy of the proposed method for different numbers of neurons in the hidden layer of the fully-connected classifier. It is observed that increasing the number of neurons from two to 256 increases the accuracy; however, increasing from 256 to 1024 does not considerably affect the accuracy. Table 13 also presents the accuracy of different optimizers. According to the results, the optimizer stochastic gradient descent with momentum (SGDM) achieves the highest accuracy than the adaptive moment estimation (ADAM) and root mean squared propagation (RMRProp).

5.7. Convergence. Figure 9 illustrates the proposed model's convergence speed over 50 epochs, demonstrating no overfitting or underfitting. It also compares the effect of different modules and layer freezing schemes on the accuracy. The CAM module and layer freezing enhance convergence speed and accuracy, as observed.

6. Conclusion

This study introduced a new deep learning architecture combining EfficientNetB7, a CAM, and a fully-connected classifier for multiclass brain tumor classification using MR images. The EfficientNetB7 model was fine-tuned, augmented with channel attention to refine extracted features, and integrated with a four-layer classifier with 256 neurons optimized for performance. The proposed method achieved remarkable accuracy of 0.9816, 0.994, and 0.992 on the Brats-4C, Brats-2C-large, and Brats-2C-small datasets, surpassing state-of-the-art models such as VGG16 and ResNet50. These findings highlight its effectiveness and superiority in brain tumor classification tasks.

Building on the promising results of this study, future research could focus on evaluating the proposed architecture across broader and more diverse datasets to ensure its generalizability to real-world clinical scenarios. Integrating multimodal data, such as CT scans or genetic information may enhance diagnostic accuracy and robustness. Advancing model interpretability could improve clinical trust and adoption. Also, extending the framework to longitudinal studies could enable tracking of tumor progression, aiding in personalized treatment and monitoring strategies.

Data Availability Statement

The datasets utilized in the present study are available in the following public domains:

- Brats-4C: https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri
- Brats-2C-large: https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection
- Brats-2C-small: https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

No funding was received for this research.

References

[1] M. Agarwal, G. Rani, A. Kumar, P. Kumar, R. Manikandan, and A. H. Gandomi, "Deep Learning for Enhanced Brain Tumor Detection and Classification," *Results in Engineering* 22 (2024): 102117, https://doi.org/10.1016/j.rineng.2024.102117.

[2] M. Aljohani, W. M. Bahgat, H. M. Balaha, et al., "An Automated Metaheuristic-Optimized Approach for Diagnosing and Classifying Brain Tumors Based on a Convolutional Neural Network," *Results in Engineering* 23 (2024): 102459, https://doi.org/10.1016/j.rineng.2024.102459.

[3] T. Agrawal, P. Choudhary, V. Kumar, P. Singh, M. Diwakar, and S. Kumar, "A Comparative Study of Brain Tumor Classification on Unbalanced Dataset Using Deep Neural Networks," *Biomedical Signal Processing and Control* 94 (2024): 106256, https://doi.org/10.1016/j.bspc.2024.106256.

[4] Q. T. Ostrom and J. S. Barnholtz-Sloan, "Current State of Our Knowledge on Brain Tumor Epidemiology," *Current Neurology and Neuroscience Reports* 11, no. 3 (2011): 329–335, https://doi.org/10.1007/s11910-011-0189-8.

[5] A. I. Shahin, S. Aly, and W. Aly, "A Novel Multi-Class Brain Tumor Classification Method Based on Unsupervised PCANet Features," *Neural Computing & Applications* 35, no. 15 (2023): 11043–11059, https://doi.org/10.1007/s00521-023-08281-x.

[6] H. Mzoughi, I. Njeh, M. BenSlima, N. Farhat, and C. Mhiri, "Vision Transformers (ViT) and Deep Convolutional Neural Network (D-CNN)-Based Models for MRI Brain Primary Tumors Images Multi-Classification Supported by Explainable Artificial Intelligence (XAI)," *The Visual Computer* 41, no. 4 (2024): 2123–2142, https://doi.org/10.1007/s00371-024-03524-x.

[7] O. Mete and M. B. Lopes, "Overview of the 2017 WHO Classification of Pituitary Tumors," *Endocrine Pathology* 28, no. 3 (2017): 228–243, https://doi.org/10.1007/s12022-017-9498-z.

[8] S. Saurav, A. Sharma, R. Saini, and S. Singh, "An Attention-Guided Convolutional Neural Network for Automated Classification of Brain Tumor From MRI," *Neural Computing & Applications* 35, no. 3 (2023): 2541–2560, https://doi.org/10.1007/s00521-022-07742-z.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature* 521, no. 7553 (2015): 436–444, https://doi.org/10.1038/nature14539.

[10] L. Torrey and J. Shavlik, "Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques," *IGI global* (2010).

[11] B. Koonce and B. Koonce, "EfficientNet, Convolutional Neural Networks With Swift for *Tensorflow: Image Recognition and Dataset Categorization*" (2021), 109–123.

[12] M.-H. Guo, T. X. Xu, J. J. Liu, et al., "Attention Mechanisms in Computer Vision: A Survey," *Computational Visual Media* 8, no. 3 (2022): 331–368, https://doi.org/10.1007/s41095-022-0271-y.

[13] C. Srinivas, "Deep Transfer Learning Approaches in Performance Analysis of Brain Tumor Classification Using MRI Images," *Journal of Healthcare Engineering* 2022, no. 1 (2022): 1–17, https://doi.org/10.1155/2022/3264367.

[14] O. Özkaraca, O. İ. Bağrıaçık, H. Gürüler, et al., "Multiple Brain Tumor Classification With Dense CNN Architecture Using Brain MRI Images," *Life* 13, no. 2 (2023): 349, https://doi.org/10.3390/life13020349.

[15] S. Kumar, S. Choudhary, A. Jain, K. Singh, A. Ahmadian, and M. Y. Bajuri, "Brain Tumor Classification Using Deep Neural Network and Transfer Learning," *Brain Topography* 36, no. 3 (2023): 305–318, https://doi.org/10.1007/s10548-023-00953-0.

[16] E. Hassan and H. Ghadiri, "Advancing Brain Tumor Classification: A Robust Framework Using EfficientNetV2 Transfer Learning and Statistical Analysis," *Computers in Biology and Medicine* 185 (2025): 109542, https://doi.org/10.1016/j.compbiomed.2024.109542.

[17] I. Abd El Kader, G. Xu, Z. Shuai, S. Saminu, I. Javaid, and I. Salim Ahmad, "Differential Deep Convolutional Neural Network Model for Brain Tumor Classification," *Brain Sciences* 11, no. 3 (2021): 352, https://doi.org/10.3390/brainsci11030352.

[18] S. Sreelakshmi, G. Malu, E. Sherly, and R. Mathew, "M-Net: An Encoder-Decoder Architecture for Medical Image Analysis Using Ensemble Learning," *Results in Engineering* 17 (2023): 100927, https://doi.org/10.1016/j.rineng.2023.100927.

[19] M. I. Sharif, J. P. Li, M. A. Khan, S. Kadry, and U. Tariq, "M3BTCNet: Multi Model Brain Tumor Classification Using Metaheuristic Deep Neural Network Features Optimization," *Neural Computing & Applications* 36, no. 1 (2024): 95–110, https://doi.org/10.1007/s00521-022-07204-6.

[20] M. Aamir, Z. Rahman, W. Ahmed Abro, U. Aslam Bhatti, Z. Ahmed Dayo, and M. Ishfaq, "Brain Tumor Classification Utilizing Deep Features Derived From High-Quality Regions in MRI Images," *Biomedical Signal Processing and Control* 85 (2023): 104988, https://doi.org/10.1016/j.bspc.2023.104988.

[21] I. D. Apostolopoulos, S. Aznaouridis, and M. Tzani, "An Attention-Based Deep Convolutional Neural Network for Brain Tumor and Disorder Classification and Grading in Magnetic Resonance Imaging," *Information* 14, no. 3 (2023): 174, https://doi.org/10.3390/info14030174.

[22] M. Liu and J. Yang, "Image Classification of Brain Tumor Based on Channel Attention Mechanism," *Journal of Physics: Conference Series* 2035, no. 1 (2021): 012029, https://doi.org/10.1088/1742-6596/2035/1/012029.

[23] I. Aboussaleh, J. Riffi, K. el Fazazy, A. M. Mahraz, and H. Tairi, "3DUV-NetR+: A 3D Hybrid Semantic Architecture Using Transformers for Brain Tumor Segmentation With Multi-Modal MR Images," *Results in Engineering* 21 (2024): 101892, https://doi.org/10.1016/j.rineng.2024.101892.

[24] P. Kanchanamala, V. Kuppusamy, and G. Ganesan, "QDCNN-DMN: A Hybrid Deep Learning Approach for Brain Tumor Classification Using MRI Images," *Biomedical Signal Processing and Control* 101 (2025): 107199, https://doi.org/10.1016/j.bspc.2024.107199.

[25] C. Jetlin, "PyQDCNN: Pyramid QDCNNet for Multi-Level Brain Tumor Classification Using MRI Image," *Biomedical Signal Processing and Control* 100 (2025): 107042, https://doi.org/10.1016/j.bspc.2024.107042.

[26] A. Y. Paulindino, G. N. Elwirehardja, and B. Pardamean, "Evaluating Self-Supervised Pre-Trained Vision Transformer for Brain Tumor Classification With Histogram Equalization Image Enhancement," *ICIC Express Letters, Part B: Applications* 15, no. 11 (2024).

[27] A. Qayyum, I. Razzak, M. Mazher, T. Khan, W. Ding, and S. Niederer, "Two-Stage Self-Supervised Contrastive Learning Aided Transformer for Real-Time Medical Image Segmentation," *IEEE Journal of Biomedical and Health Informatics* (2023): 1–10, https://doi.org/10.1109/jbhi.2023.3340956.

[28] U. Khatri and G.-R. Kwon, "Explainable Vision Transformer with Self-Supervised Learning to Predict Alzheimer's Disease Progression Using 18F-FDG PET," *Bioengineering* 10, no. 10 (2023): 1225, https://doi.org/10.3390/bioengineering10101225.

[29] M. A. Karagoz, O. U. Nalbantoglu, and G. C. Fox, "Residual Vision Transformer (ResViT) Based Self-Supervised Learning Model for Brain Tumor Classification," (2024), arXiv preprint arXiv:2411.12874.

[30] H. Gholamalinezhad and H. Khosravi, "Pooling Methods in Deep Neural Networks: A Review," (2020), arXiv preprint arXiv:2009.07485.

[31] X. Zhang, X. Zhang, and W. Wang, "Convolutional Neural Network," in *Intelligent Information Processing with Matlab* (Springer, 2023), 39–71.

[32] C. D. Rios-Urrego, J. C. Vásquez-Correa, J. R. Orozco-Arroyave, and E. Nöth, "Transfer Learning to Detect Parkinson's Disease From Speech in Different Languages Using Convolutional Neural Networks With Layer Freezing," in *International Conference on Text, Speech, and Dialogue* (Springer, 2020), 331–339.

[33] Z. Niu, G. Zhong, and H. Yu, "A Review on the Attention Mechanism of Deep Learning," *Neurocomputing* 452 (2021): 48–62, https://doi.org/10.1016/j.neucom.2021.03.091.

[34] S. Sharma, S. Sharma, and A. Athaiya, "Activation Functions in Neural Networks," *International Journal of Engineering Applied Sciences and Technology* 04, no. 12 (2020): 310–316, https://doi.org/10.33564/ijeast.2020.v04i12.054.

[35] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks From Overfitting," *Journal of Machine Learning Research* 15, no. 1 (2014): 1929–1958.

[36] V. Thakkar, S. Tewary, and C. Chakraborty, "Batch Normalization in Convolutional Neural Networks: A Comparative Study With CIFAR-10 Data," in *2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT)* (IEEE, 2018), 1–5.

[37] S. K. Bhuvaji, P. Bhumkar, S. Dedge, and S. Kanchan, "Brain Tumor Classification (MRI) Dataset," (2024), https://www.kaggle.com/sartajbhuvaji/brain-tumor-classification-mri.

[38] A. Hamada, "Br35H Brain Tumor Detection 2020 Dataset," (2024), https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection.

[39] N. Chakrabarty, "Brain MRI Images for Brain Tumor Detection Dataset," (2024), https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection.

[40] M. Iman, H. R. Arabnia, and K. Rasheed, "A Review of Deep Transfer Learning and Recent Advancements," *Technologies* 11, no. 2 (2023): 40, https://doi.org/10.3390/technologies11020040.

[41] Y. Li, C. Wei, and T. Ma, "Towards Explaining the Regularization Effect of Initial Large Learning Rate in Training Neural Networks," *Advances in Neural Information Processing Systems* 32 (2019).

[42] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss Functions for Neural Networks for Image Processing," (2015), arXiv preprint arXiv:1511.08861.

[43] I. Salehin and D.-K. Kang, "A Review on Dropout Regularization Approaches for Deep Neural Networks Within the Scholarly Domain," *Electronics* 12, no. 14 (2023): 3106, https://doi.org/10.3390/electronics12143106.
[44] K. Simonyan, "Very Deep Convolutional Networks for Large-Scale Image Recognition," (2014), arXiv preprint arXiv: 1409.1556.
[45] M. Tan and Q. Le, "Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks," in *International Conference on Machine Learning* (PMLR, 2019), 6105–6114.
[46] J. Kang, Z. Ullah, and J. Gwak, "MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers," *Sensors* 21, no. 6 (2021): 2222, https://doi.org/10.3390/s21062222.
[47] T. Kalaiselvi, S. Padmapriya, P. Sriramakrishnan, and K. Somasundaram, "Deriving Tumor Detection Models Using Convolutional Neural Networks From MRI of Human Brain Scans," *International Journal of Information Technology* 12, no. 2 (2020): 403–408, https://doi.org/10.1007/s41870-020-00438-4.
[48] J. Amin, M. Sharif, N. Gul, et al., "Brain Tumor Detection by Using Stacked Autoencoders in Deep Learning," *Journal of Medical Systems* 44, no. 2 (2020): 32–12, https://doi.org/10.1007/s10916-019-1483-2.
[49] 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* 75 (2022): 103625, https://doi.org/10.1016/j.bspc.2022.103625.
[50] 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* 45, no. 3 (2021): 1015–1036, https://doi.org/10.1007/s40998-021-00426-9.