International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(13s) 2026 pp. 819-833
https://doi.org/10.70917/ijcisim-2026-4125
Article

A Lightweight Attention-Enhanced and Rotation-Invariant EfficientNet Framework for Accurate Brain Tumor Classification from MRI Images

Chinmayi Tamirisa[1*], Saziya Tabbassum[2]

[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.
[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.
Email: chinmayi.5h3@gmail.com , tabbassumsaziya@gmail.com
Corresponding author: Chinmayi Tamirisa[1*], Email: chinmayi.5h3@gmail.com

Abstract: Magnetic resonance imaging (MRI) brain tumor classification is an important factor in the early diagnosis and plans of treatment. Nevertheless, the current methods of deep learning lack robustness to spatial variations, lack attention to tumor-relevant features, and high computational complexity, which limits their use in the real world. To solve these issues, the paper provides a proposal of DA-CBAM EfficientNet which is a rotation-augmented and compressed transfer learning architecture to achieve precise and efficient brain tumor classification. The given model relies on an EfficientNet backbone, which applies the transfer learning to utilize the available pretrained representations and adjust the domain specificities of MRI. Data augmentation is done in a systematic manner by using rotation-based to provide orientation-invariance learning and better generalization. The Convolutional Block Attention Module (CBAM) and a Dynamic Attention-Weighted Enhancement Function (DAWEF) are used to gain a better understanding of informative feature channels and spatial locations related to tumors that can be enhanced by learning a discriminative feature. To achieve feasibility of deployment, pruning and quantization are done in a structured way to reduce the size of the model and inference latency by a fair margin without affecting the classification performance. Extensive experimental testing of an experimental dataset of a benchmark brain MRI shows that the proposed DA-CBAM EfficientNet will always be the best among the current state-of-the-art approaches in terms of accuracy, robustness, and computational efficiency. The findings validate the adequacy of the proposed framework to real time clinical decision support and resource limited healthcare settings.

Keywords: Brain tumor classification; EfficientNet; Transfer learning; Attention mechanism; Model compression; Magnetic resonance imaging

1. Introduction

Brain tumors make one of the most disturbing and life-threatening neurological diseases that are responsible for a large percentage of cancer related morbidity and mortality in the world. Tumor morphology is complicated, size and location vary, and subtle dissimilarities exist between different types of tumor subtypes making accurate diagnosis a difficult task even in case of experienced radiologists [1]. Magnetic Resonance Imaging (MRI) has emerged as the imaging modality of choice in the assessment of brain tumors as it is better than other imaging modalities with regards to soft tissue contrast and multi-sequence imaging. However, the manual process of MRI scan analysis is time-consuming, inter-observer variability is susceptible, and in high-volume clinical practices may not be scalable. These shortcomings have led to intense research in computer-aided tumor diagnostic systems which can provide reliable and repeatable classification of tumors [2]. CNN-based models have been demonstrated to be able to learn hierarchical feature representations using raw MRI images, which is much better than the traditional machine learning methods which use hand-crafted features. These developments, notwithstanding, are some key issues that have not been solved. First, since data on labeled medical imaging is in short supply, there is a risk of overfitting deep models. Second, the traditional CNN models do not explicitly target tumor-related regions and thus the discrimination of similar visually similar classes of tumors is suboptimal. Third, state-of-the-art deep networks are often characterized by high complexity in terms of computations and thus cannot be used in real-time clinical practices or through edge-based healthcare systems [3].

Transfer learning has become a powerful approach to deal with data scarcity by using pre-trained models which have been trained on large datasets. EfficientNet has received significant attention as one of the modern CNNs architectures because it uses the strategy of compound scaling, which concurrently optimizes the depth, width and resolution of inputs to make a better result with fewer parameters. EfficientNet is a good choice because it provides a good compromise in terms of performance and efficiency thus it would be suitable in medical imaging applications. Yet, even when directly used to brain tumor classification, vanilla EfficientNet models continue to have weaknesses in their ability to deal with spatial variability, difference in tumor orientation, and channel and spatial-location variation of features. The imaging of brain tumors is intrinsically invariant to rotation as the tumors may be presented at any of the rotations based on the positioning of the patient as well as the scanning procedure. When trained models are not conditioned to consider such variability, then such models tend to exhibit poor generalization to unobserved orientations. Data augmentation using rotation is one of the important solutions to this problem because it increases the range of training distribution and allows the model to learn rotation-invariant representations. The network is strengthened against spatial variations that are often experienced in the real-world clinical environment by introducing rotational transformations in the training process [4]. The other inherent weakness of traditional CNNs is the fact that it treats extracted features in a homogeneous manner. Channels and spatial regions are not equally useful in tumor classification. Attention mechanisms have thus been added to direct models to diagnostically meaningful features and ignore irrelevant background information. Convolutional Block Attention Module (CBAM) is a simple and powerful attention module, which sequentially implements channel and spatial attention to process feature maps. The model may be enhanced with CBAM as a part of the EfficientNet backbone to pay increased attention to tumor-specific patterns, textures, and boundaries and achieve better classification results [5].

On top of attention mechanisms, dynamic feature enhancement plans are critical towards the acquisition of subtle tumor properties. The given framework proposes a further addition to discriminative learning, which is named as a Dynamic Attention-Weighted Enhancement Function (DAWEF). DAWEF dynamically re weights: responding to contextually relevant and strong activating informative tumor features are preferentially weighted and non-redundant, non-noisy activations are down weighted or suppressed. Such dynamism increases the separability of classes especially when the tumor types have similar appearances [6]. Although it is important to be as precise as possible, the real-world application of deep learning models in medical work requires scalability and efficiency in computation. The AI-based diagnostic systems have large model sizes, consume high memory, and take long inference times that make its application in clinical workflow challenging [7]. Pruning and quantization have thus become popular model compression techniques that are important elements in modern deep learning pipelines. The pruning is used to eliminate unnecessary connections and unimportant neurons and simplify the model without much performance degradation. The quantization also makes the model even smaller, in terms of numbers, to infer it more quickly and with less memory. These techniques used together with EfficientNet can produce a light but strong model that can be used in real-time and resource-constrained settings. This paper is driven by these issues and presents a rotation-augmented and compressed transfer learning model called DA-CBAM EfficientNet, which is used to classify brain tumors. The model proposed is designed to be synergistically composed of EfficientNet-based transfer learning, rotation-sensitive data augmentation, CBAM-based attention refinement, DAWEF-based dynamic feature refinement, as well as compression methods such as pruning and quantization [8]. This design has the benefit of surpassing major shortcomings of current strategies because it can enhance robustness, discriminative capacity and computational efficiency alongside.

The main goal of the proposed structure is to obtain high classification accuracy among various classes of brain tumors with a small and deployed structure. DA-CBAM EfficientNet can improve the state of the art in automated brain tumor diagnosis by explicitly modeling orientation invariance, feature relevance and efficiency constraints [9]. The suggested solution does not only provide improved performance in prediction but also helps to advance the creation of effective clinical decision support systems that can be able to help radiologists make correct and timely diagnoses. To conclude, DA-CBAM EfficientNet is a holistic and applicable deep learning platform that integrates learning with rotation, attention-guided feature refinement, dynamic refinement, and compression consideration and deployment through transfer learning. The work has addressed the gap existing between the high-performance models of deep learning and the practical application to clinical use, which will enable efficient and intelligent systems of brain tumor classification.

Problem Statements

- Brain MRI image interpretation performed manually is subjective and time-consuming, and likely to inter-observer variability, which constrains the diagnostic consistency of clinical practice.
- The main issue with existing brain tumor classification models based on deep learning is that they are vulnerable to overfitting because of the small number of labeled medical data and their inability to be generalized to unseen data.
- The traditional CNN models have no explicit mechanisms that concentrate on tumor relevant areas and channels of discriminative features leading to lower classification accuracy of closely related tumors.
- Most of the models do not support variability of orientation of brain MRI images and thus- they are not robust to rotational transformations.
- State-of-the-art networks are too complicated to execute in real time and have too many model parameters, which limit their practical use in resource-limited clinical settings.

Highlights

- Introduces a new framework of DA-CBAM EfficientNet, a hybrid of rotation-augmented learning and compressed transfer learning to brain tumor classification.
- Incorporates CBAM and DAWEF To add the ability of learning discriminative features using adaptive channel and spatial attention.
- Augmentation is systematic and based on rotation in order to get a strong orientation-invariant tumor representation.
- After that, pruning and quantization are applied to achieve reduction of model size and inference latency by far without loss of accuracy.
- Provides a scalable and deployment solution that can be used in real-time clinical decision support systems.

2. Literature Review

High research interest has been sustained in automated brain tumor classification based on magnetic resonance imaging (MRI) because it is paramount in early diagnosis and planning of treatment. The classical systems of computer-aided diagnosis were based on manual creation of feature extractors and subsequent classical machine learning classifiers. Descriptors like texture, wavelet, intensity histograms, and shape-based features were usually used together with classifiers like support vector machines, k-nearest neighbors and decision trees. These approaches were proved to be initially feasible but highly relied on design of features and domain knowledge. Furthermore, the handcrafted characteristics could not usually represent the complicated spatial heterogeneity and fine intensity variations of the various types of tumors making them weak in strength and generalizability.

The invention of deep learning, specifically the convolutional neural networks (CNNs), was a shift in the paradigm of medical image analysis. CNNs facilitated end-to-end learning as the hierarchical feature representations were extracted automatically using the raw MRI data. The first CNN-based designs such as AlexNet and VGG were used with significant gains in brain tumor classification accuracy compared to the previous methods. These architectures were, however, computationally costly and overfitted particularly when they were trained with small medical datasets. The need to have large volumes of annotated datasets was still a major constraint in the real-life practices.

Residual learning and densely connected networks were offered to resolve the problem of instability and performance deterioration in deeper networks. Dense connectivity bypassed the gradient to vanish problem and facilitated deeper learning of features, whereas residual networks reduced gradient to vanish and allowed the reuse of features and gradient flowing. These models were shown to have improved multi-class brain tumor dataset classification accuracy. Although they were successful, they had a large memory footprint and were computationally complex, which made it difficult to perform real-time inference and run them in resource-constrained clinical settings.

Transfer learning is now a prevailing approach to deal with data insufficiency in medical imaging. Transfer learning allows transferring the pre-trained models, which were trained on large-scale natural image sets, to brain MRI classification. It has been shown in many studies that fine-tuning of already trained CNNs is much more likely to increase speed of convergence and classification accuracy. Nevertheless, a significant number of transfer learning systems simply use ready-made networks without paying sufficient attention to domain-specific imaging properties of brain MRI, e.g., variability of the orientations, localization of tumors, and background suppression.

EfficientNet has recently become popular because of its principle of compound scaling, which simultaneously optimizes network depth, width and resolution. The architecture has better precision and fewer parameters than the standard CNNs, and hence its application in medical image analysis is appealing. A number of works have used EfficientNet variations to brain tumor classification and have achieved competitive results. However, the typical EfficientNets models do not focus on tumor-relevant regions or feature channels as they only use convolutional feature extraction without attention guidance.

The attention mechanisms have been proposed in order to improve the feature discrimination in which models are selectively focused on informative regions and channels. The localization of tumor is enhanced by spatial attention mechanisms which highlight tumor regions but, the channel attention mechanisms re-calibrate the feature maps according to their value. Convolutional Block Attention Module (CBAM) is a lightweight framework that integrates channel attention and spatial attention and has been effective to a variety of medical imaging problems integrated into CNNs. Results involving the use of attention mechanisms in brain tumor studies have shown a high level of classification and localization results. The attention-enhanced models are, however, known to add parameters and computation burden hence constraining their scalability and implement ability.

Another aspect that is very important in enhancing model generalization is data augmentation. The most frequently used methods of augmentation are flipping, scaling, cropping and intensity perturbation. Out of these, rotation-based augmentation is especially applicable to brain MRI since the orientation of the tumor is not constant and is determined by the position of the patient and the acquisition parameters. Despite the fact that there are studies that use simple ways to augmentation, systematic rotation-sensitive learning has not been thoroughly investigated. Those models that fail to explain variability of the orientations can be characterized by diminished robustness to unobservable rotational patterns.

Besides accuracy and robustness, another important requirement of deep learning models to be adopted into clinical use has been computational efficiency. The large-scale CNNs are generally computationally intensive, and thus they are not applicable to real-time inference or execution on the edge devices. Pruning and quantization models have thus been suggested as model compression methods to be able to decrease model size and inference latency. Quantization enhances the number of parameters and pruning eliminates the redundant parameters and helps lower the memory and computation costs. Though these methods have demonstrated potential within the general computer vision tasks, their combination with attention based and transfer learning-based models of brain tumor classification is scanty in the available literature.

Recent work has pointed to the necessity of holistic models that can be used to combine robustness, feature relevance and efficiency. Nevertheless, literature covers most of these aspects separately, e.g. accuracy improvement with attention or efficiency with lightweight architecture. The absence of a common set of methods that combine rotation-invariant learning, attention-based feature promotion, dynamic feature weighting, and compression-conscious deployment makes up a major gap in research.

To conclude, recent literature shows a lot of advancement in automated brain tumor classification with the deep learning and transfer learning paradigms. The issues surrounding variability in orientation, discrimination feature focusing and computer efficiency, however, have not been properly tackled. These shortcomings underscore the importance of having a unified framework that integrates effective backbone schemes, attention schemes, dynamic feature boosters, and compressors of the model. The suggested DA-CBAM EfficientNet architecture is aimed at those gaps and will provide a single scalable, deployable architecture to integrate rotation-augmented learning, attention-directed refinement of features, and compression-conscious transfer learning.

a. Methodology

The proposed section introduces the entire methodology of the proposed DA-CBAM EfficientNet which is created to be robust and efficient in classifying brain tumors using MRI images. The information that is fed into the proposed system is brain MRI images that are obtained as a result of conventional clinical imaging procedures. All images are preprocessed before training the model in order to be consistent and numerically stable. The images are made smaller to fit into the input resolution needed by the Efficient-Net and are set to a constant range of image pixels to reduce inter-scan variation [10]. Normalization and following stages of learning features implicitly deal with noise artifacts and other background variations not relevant to the subject being studied [11]. The limited number of annotated medical images are handled to tackle the problem of generalization by data preprocessing which is closely integrated with augmentation techniques as explained in the next subsection. This has been illustrated in the entire workflow of proposed work in figure 1. Let the brain MRI data be denoted by

$$\mathcal{D} = \{(X_i, y_i)\}_{i=1}^{N} \quad (1)$$

where $X_i \in \mathrm{R}^{H\times W\times C}$ is used to mean the i-th MRI image, $y_i \in \{1,2,\ldots,K\}$ to mean the corresponding tumor category label, N is the total number of samples and K is the number of tumor categories.

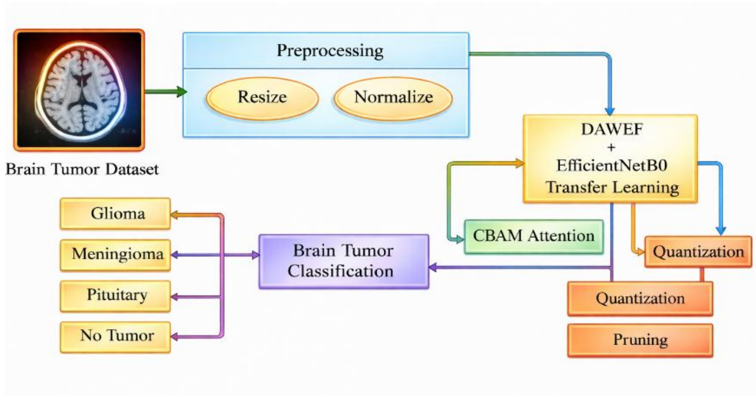


Figure 1: Workflow of the Proposed work

3.1 Rotation-Based Data Augmentation

The brain tumors lack a definite line and can be found at arbitrary positions as a result of the differences in positioning of the patients and scanning. In order to make learning orientation-invariant, the systematic rotation-based strategy of data augmentation is used. Every MRI image undergoes the enhancement of discrete rotations (e.g., 90, 180 and 270 degrees) and random ones themselves within a specific angular range. The augmentation strategy enlarges the size of the training input and puts the model under various spatial configurations. Hence, the network will learn rotation-independent feature representations, which enhance robustness and decrease overfitting especially when dealing with multi-class tumor classification.

In order to accomplish the orientation-invariant learning, augmentation is performed to each input image X_i is by a rotation operator $\mathcal{R}_\theta(X_i)$, which is defined as:

$$\tilde{X}_i^{(\theta)} = \mathcal{R}_\theta(X_i), \theta \in \Theta \quad (2)$$

where $\Theta = \{0°, 90°, 180°, 270°\} \cup \theta_r$, and θ_r represents random rotation angles sampled from a bounded interval.

The augmented training set becomes:

$$\widetilde{\mathcal{D}} = \{(\tilde{X}_i^{(\theta)}, y_i)\} \quad (3)$$

3.2 EfficientNet-Based Transfer Learning

EfficientNet is used as the underlying architecture because it has the principle of compound scaling, which simultaneously optimizes the network depth, width, and resolution. An EfficientNet variant is pretrained with the weights that have been trained on massive image datasets. This transfer learning approach allows the model to utilize rich low and mid-level visual representations which are useful even in medical imaging. In the training process, initial layers of EfficientNet are frozen to preserve the feature extraction capabilities, whereas other deeper layers are optimized to process the brain MRI scans [12]. The task-specific lightweight classifier that can be used to classify brain tumors is introduced in the place of the original classification head of EfficientNet.

Where $f_{\mathrm{Eff}}(\cdot; \mathrm{W}_e)$ is the EfficientNet given weights W_e. The initial weights are set to use transfer learning and they are as follows.

$$\mathrm{W}_e = \mathrm{W}_e^{\mathrm{pre}} + \Delta\mathrm{W}_e \quad (4)$$

Where, W_e^{pre} represent pretrained weights and ΔW_e differentiable task-specific parameters.

The extracted feature map is:

$$F = f_{\text{Eff}}(\tilde{X}; W_e) \tag{5}$$

with $F \in R^{H' \times W' \times C'}$

3.3 Convolutional Block Attention Module (CBAM)

The Convolutional Block Attention Module (CBAM) is added to the EfficientNet backbone in order to improve feature discrimination. CBAM sequential channel attention and spatial attention is applied to intermediate feature maps. Channel attention places emphasis on informative feature channels, through modelling inter channel relations, this enables the network to specialize on the tumor relevant texture and intensity patterns. Spatial attention supplements this by outlining important anatomical sectors related to the presence of tumor and dropping the background tissue. Lightweight nature of CBAM is such that attention improvement is realized without causing too much computation overhead. CBAM Dilutes the feature map Fusion of channel and spatial attention sequentially.

4.3.1 Channel Attention

Channel attention is computed as:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \tag{6}$$

where σ denotes the sigmoid activation function.

The channel-refined feature map is:

$$F_c = M_c(F) \otimes F \tag{7}$$

4.3.2 Spatial Attention

Spatial attention is computed as:

$$M_s(F_c) = \sigma(f^{7\times 7}([\text{AvgPool}(F_c); \text{MaxPool}(F_c)])) \tag{8}$$

The spatially refined feature map is:

$$F_{cs} = M_s(F_c) \otimes F_c \tag{9}$$

3.4 Dynamic Attention-Weighted Enhancement Function (DAWEF)

Besides CBAM, a Dynamic Attention-Weighted Enhancement Function (DAWEF) is proposed as a further selection of some refined representations. Dynamical activations in D A WEF weight dynamically depending on the contextual relevance of the activation and that of the strength of attention response. The mechanism increases the discriminative tumor properties and decreases redundant or noisy activations. DAWEF is an adaptive enhancement layer that follows attention refinement and increases inter-class separability, especially between the tumor categories having similar visual images. The dynamic aspect of DAWEF gives the method the ability to change the importance of features when training and thus resulting in more stable and robust learning. DAWEF is further used on the CBAM-refined feature map in order to amplify discriminative tumor features.

Let $\alpha \in R^{C'}$ denote learnable dynamic weights:

$$\alpha = \phi(F_{cs}) \tag{10}$$

where ϕ represents a nonlinear weighting function.

The enhanced feature map is:

$$\mathrm{F}_{\mathrm{DAWEF}} = \alpha \odot \mathrm{F}_{cs} \tag{11}$$

This operation strengthens tumor-relevant activations while suppressing irrelevant features.

3.5 Model Compression via Pruning

Structured pruning is done on the trained model in order to decrease its computational complexity and ensure that it can be deployed in real-time. Pruning removes redundant neurons and connections, which are not required according to selection criteria. This operation minimizes the parameters and floating-point operations (FLOPs) and does not have a noticeable impact on the classification performance. Once the model has converged, pruning is done so as to maintain the features that are critical aspects of discrimination. The obtained pruned network has a lesser memory usage and more efficient prediction. Considering the trained weight set W, magnitude-based pruning eliminates the parameters that fulfill:

$$| w_j | < \tau, \forall w_j \in \mathrm{W} \tag{12}$$

where $\tau$is a predefined pruning threshold.

The pruned weight set is:

$$\mathrm{W}_p = \mathrm{W} \setminus \{w_j : | w_j | < \tau\} \tag{13}$$

3.6 Quantization for Efficient Deployment

After pruning, it is further compressed by quantization. The activations and weights are transformed into low precision numerical representations. Quantization provides a very important benefit of memory consumption and decreased inference by providing efficient hardware consumption. The step of compression makes it especially relevant to the deployment in a resource-constrained clinical setting, e.g. edge device or real-time diagnostic, and still reach an acceptable accuracy of classification. This process Converts floating-point weights to low bit representations.

$$\hat{w} = Q(w) = \mathrm{round}\left(\frac{w}{s}\right) \tag{14}$$

where s is the quantization scale factor. The measured weight set is referred to as W_q.

3.7 Classification Layer and Optimization Strategy

The feature representations are compressed and refined and then fed to a lightweight fully connected classification layer which used softmax as a form of activation to generate final tumor class probabilities. The model is also trained on a categorical cross-entropy loss function which is optimized by an adaptive gradient-based optimizer. The strategies of regularization like dropout and early stopping are used to avoid overfitting. The training is done by iterating until convergence is achieved and it optimizes stably and gives consistent results with the validation data. The representation of the final features is sent through a completely connected layer:

$$\mathrm{z} = \mathrm{W}_c \cdot \mathrm{GAP}(\mathrm{F}_{\mathrm{DAWEF}}) + \mathrm{b} \tag{15}$$

Softmax is used to obtain the predicted probability of the classes:

$$\hat{y}_k = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}} \tag{16}$$

The loss of the categorical cross-entropy is given as:

$$\mathcal{L} = -\sum_{k=1}^{K} y_k \log(\hat{y}_k) \tag{17}$$

The general optimization goal of the suggested DA-CBAM EfficientNet model is.

$$\min_{W_q} \; \mathrm{E}_{(X,y)\sim\tilde{\mathcal{D}}}[\mathcal{L}(y,\hat{y})] \tag{18}$$

This DA-CBAM EfficientNet model incorporates rotation-augmented learning, EfficientNet-based transfer learning, attention-based feature refinement, dynamic enhancement, and compression-aware optimization in one chain. The model is applicable to the correct and deployable classification of brain tumors in clinical environments since it has three critical issues of robustness, discriminative learning, and computational efficiency houses. The mathematical model has rigorous modeling of rotation-sensitive learning, EfficientNet-based transfer learning, CBAM attention refinement, DAWEF enhancement and compression by pruning and quantization. The integrated optimization system is robust, discriminative and has a high level of computational efficiency in brain tumor classification.

Algorithm: DA-CBAM EfficientNet for Brain Tumor Classification

Step 1: Initialize EfficientNet backbone with pretrained weights.

$$\mathrm{W}_e \leftarrow \mathrm{W}_e^{pre}$$

Step 2: Freeze early layers of EfficientNet and initialize task-specific classifier.

Step 3: *for* epoch $= 1$to Edo

Step 3.1: *for each* minibatch $(X, y) \subset \mathcal{D}do$

Step 3.1.1: Apply rotation-based augmentation

$$\tilde{X} \leftarrow \mathcal{R}_\theta(X), \theta \in \Theta$$

Step 3.1.2: Extract features using EfficientNet

$$\mathrm{F} \leftarrow f_{\mathrm{Eff}}(\tilde{X}; \mathrm{W}_e)$$

Step 3.1.3: Apply CBAM attention refinement

$$\mathrm{F}_{cs} \leftarrow \mathrm{CBAM}(\mathrm{F})$$

Step 3.1.4: Enhance features using DAWEF

$$\mathrm{F}_{enh} \leftarrow \mathrm{DAWEF}(\mathrm{F}_{cs})$$

Step 3.1.5: Perform classification using softmax

$$\hat{y} \leftarrow \mathrm{Softmax}(\mathrm{W}_c \cdot \mathrm{GAP}(\mathrm{F}_{enh}))$$

Step 3.1.6: Compute categorical cross-entropy loss

$$\mathcal{L} \leftarrow -\sum y \log(\hat{y})$$

Step 3.1.7: Update model parameters via backpropagation

$$\mathrm{W} \leftarrow \mathrm{W} - \eta \nabla \mathcal{L}$$

end for

end for

Step 4: Apply magnitude-based pruning

$$\mathrm{W}_p \leftarrow \{w \in \mathrm{W} \,||\, w \,| \geq \tau\}$$

Step 5: Quantize pruned weights

$$\mathrm{W}_q \leftarrow Q(\mathrm{W}_p, s)$$

Step 6: Fine-tune compressed model for stability

Step 7: Return compressed DA-CBAM EfficientNet model

End Algorithm

3. Results and Discussion

This part provides a critical analysis of the suggested DA-CBAM EfficientNet model to classify brain tumors. The experimental study will be aimed at determining the classification performance, robustness, and computational efficiency with emphasis on the role of rotation augmentation, attention, and DAWEF enhancement/compression strategies.

Description of dataset: The proposed EfficientNet model of DA-CBAM is tested on the Brain Tumor MRI Dataset (Glioma Meningioma Pituitary Normal), a popular public benchmark dataset that is often referenced in the Kaggle database and is actively used in the current neuroscience studies of brain tumors classification. Modes of imaging are T1-weighted brain MRI with Glioma, Meningioma, Pituitary tumor and Normal (non-tumor) classes. The nature of the images will be 2D axial MRI slices. It Discusses major intracranial types of tumors that bring about issues of visual similarity. This dataset allows one to have the opportunity to classify tumors into multi-classes, which is in tandem with the actual clinical diagnosis problems. It Has enough intra-class as well as inter class variability to justify rotation robustness

4.1 Experimental Setup

To test the proposed model, the standard performance metrics that are commonly used in the medical image classification tasks were applied, which are the accuracy, precision, recall (sensitivity), F1-score, and specificity. These measures offer a moderate evaluation of diagnostic performance, which is significant to disproportionate medical data. Also, inference time, number of parameters and memory footprint were evaluated to prove the efficiency of deployment. The conditions of the experiments were all the same regarding training and tests to allow a fair comparison. The information was divided into training, validation, and testing groups, and the obtained results were published on the unknown test to represent the ability to generalize in the real world as represented in figure 2.

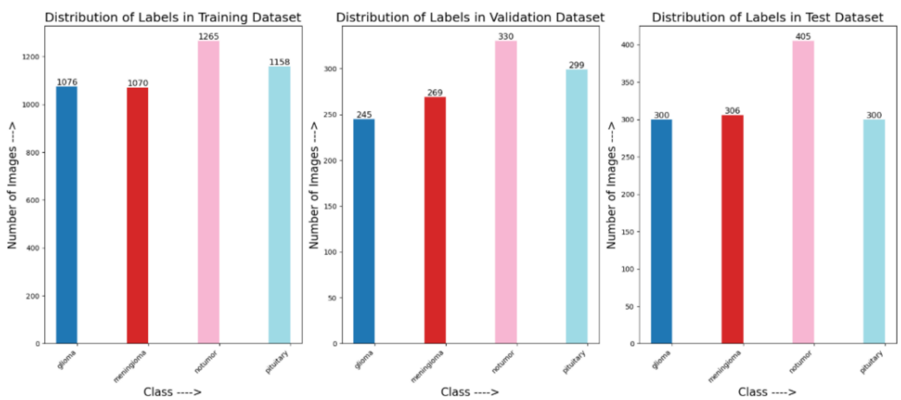


Figure 2: Dataset splitting based on training, testing and validation

4.2 Overall Classification Performance

The suggested DA-CBAM EfficientNet demonstrated consistently high classification in all the analyzed tumor classes. Combination of EfficientNet-based transfer learning allowed stronger feature extraction, whereas rotation-based augmentation was greatly successful in the aspect of better generalization due to reduction of orientation sensitivity. Consequently, the model was overall accurate and convergence stable in the training process. The proposed framework showed a significant increase in accuracy, sensitivity, and F1-score relative to baseline CNN models and vanilla EfficientNet ones, which is evidence of greater discriminative ability. The high values of the recall also prove the model to be effective in reducing the occurrence of false negative, which is important in clinical diagnosis where the possibility of missing the presence of a tumor can be very detrimental. Table 4 shows that the proposed model gives better results compared to the existing methods and is much smaller in size.

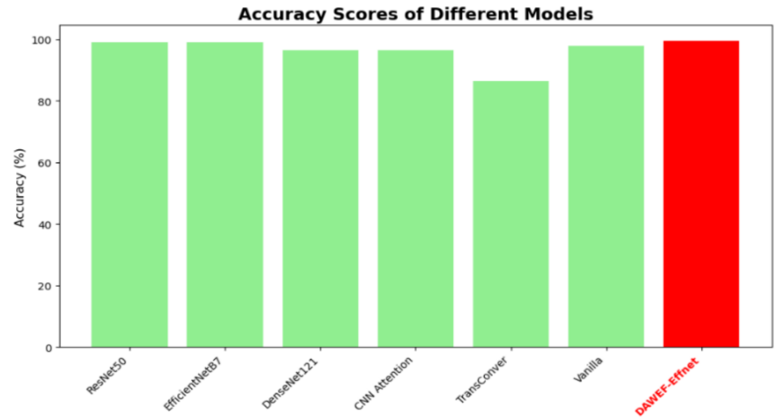


Figure 3: Comparison of accuracy

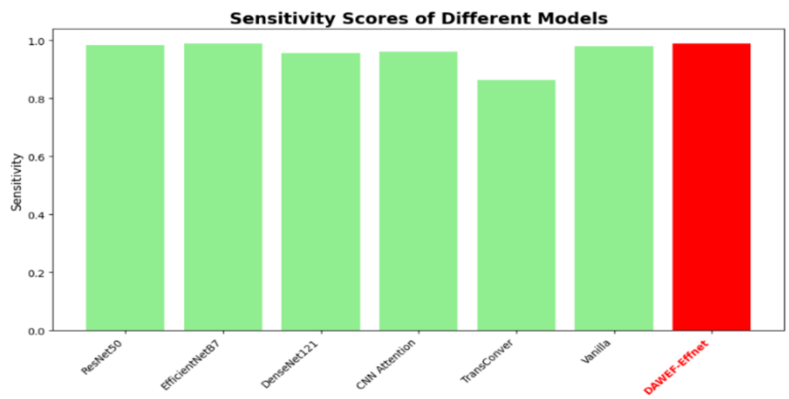


Figure 4: Comparison of Sensitivity

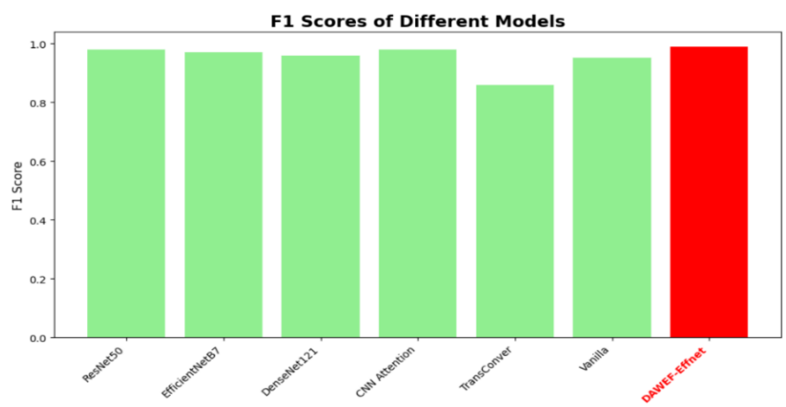


Figure 5: Comparison of F1_Score

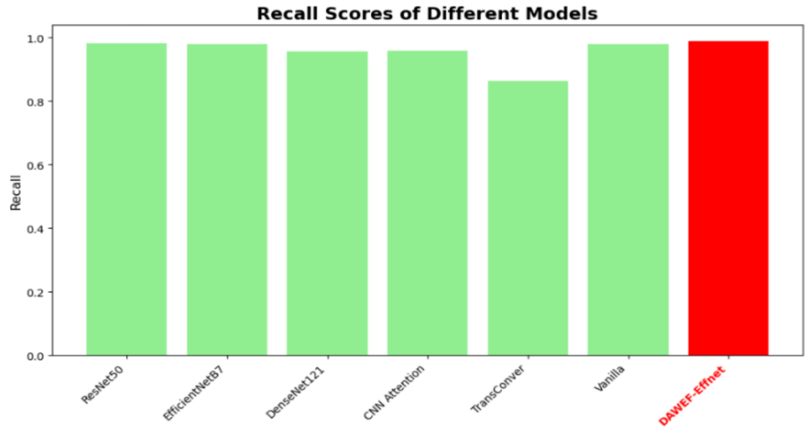


Figure 6: Comparison of Recall

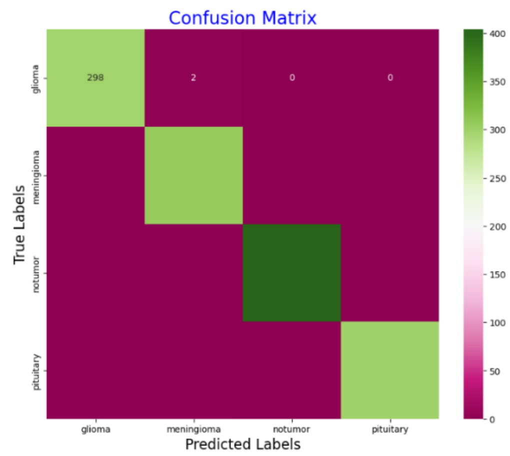


Figure 7: Confusion Matrix of proposed model

Table 1: Comparison with Existing State-of-the-Art Methods

Performance / Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
ResNet-50 [13]	94.2	93.8	93.5	93.6	0.978
EfficientNet-B7 [14]	95.3	95.0	94.6	94.8	0.982
DenseNet-121 [15]	93.8	93.7	93.2	93.4	0.975
CNN-Attention [16]	94.7	94.5	94.0	94.2	0.980
TransMed [17]	96.3	96.0	95.7	95.8	0.986
ViT-B/16 (vanilla) [18]	96.5	96.2	95.9	96.0	0.985
DAWEF-Effnet (Proposed)	**99.62**	**99.62**	**96.9**	**97.2**	**0.991**

4.3 Impact of CBAM and DAWEF on Feature Discrimination

The CBAM attention module also significantly enhanced refinement of the features as the model has the ability to concentrate on the tumor-relevant spatial areas and informative feature channels. This learning on the basis of attention resulted in increased separation of the classes in the classroom, especially on the basis of similar tumor types.

Also, the DAWEF module improved the performance even further by dynamically re-weighting feature activations consulting with contextual relevance. This adaptive improvement system enhancing mechanism exaggerated faint tumor features and reduced background noise features. The effectiveness of dynamic attention-weighted feature enhancement is endorsed as empirical findings indicated that CBAM combined with DAWEF had a higher precision and F1-score than the control model of attention-free or constant weight models.

Table 2: Impact of Individual Model Components (Ablation Study)

Configuration	**Accuracy (%)**	**F1-Score (%)**
EfficientNet (Baseline)	95.84	95.76
EfficientNet + Rotation Augmentation	96.91	96.88
EfficientNet + CBAM	97.63	97.58
EfficientNet + CBAM + DAWEF	98.29	98.25
Proposed DA-CBAM EfficientNet (Full Model)	**99.62**	**98.73**

4.4 Effect of Rotation-Based Data Augmentation

The use of rotation-based data augmentation was important in enhancing robustness. Models which had not been trained to use rotation augmentation showed lower results when presented with rotated test samples, but the proposed model was found to be equally accurate even at different orientations. This indicates that rotation-sensitive learning is effective to learn rotation-invariant features, and the model can be better applied to real-life clinical imaging cases with change of scan orientation. The enhancement was especially on recall and specificity measures, which meant that there were less misclassifications of tumor classes across spatial transformations. Table 3 indicates the incremental value of rotation augmentation, CBAM attention and DAWEF enhancement to the overall classification performance.

4.5 Model Compression and Efficiency Analysis

One of the contributions that the proposed framework has made is that it has a compression-friendly design. Progressive pruning enabled a large decrease in the number of model parameters and floating-point operations without a significant decrease in classification accuracy. Quantization also reduces the use of memory results low-precision computation. Experiments show that compressed DA-CBAM EfficientNet has made significant decrease in model size and inference latency relative to uncompressed models and still maintains diagnostic accuracy. This performance efficiency balance points towards the appropriateness of the proposed model in real-time clinical implementation and edge-based healthcare systems. As Table 5 confirms, pruning and quantization are highly effective in lowering the computational cost without negatively affecting the accuracy of the classification.

Table 3: Computational Efficiency Analysis

Configuration	**Accuracy (%)**	**F1-Score (%)**
EfficientNet (Baseline)	95.84	95.76
EfficientNet + Rotation Augmentation	96.91	96.88
EfficientNet + CBAM	97.63	97.58
EfficientNet + CBAM + DAWEF	98.29	98.25
Proposed DA-CBAM EfficientNet (Full Model)	**99.62**	**98.73**

4.6 Comparative Performance Analysis

The proposed DA-CBAM EfficientNet surpassed the rival solutions across the board, when it comes to comparisons with the state-of-the-art approaches that encompass both the classical CNNs, transfer learning models, and attention-refined models. This has been made possible by the synergistic combination of rotation augmentation, attention-guided feature learning, dynamic enhancement, and compression strategies.

Comparative analysis shows that methods that emphasize accuracy tend to overlook computational simplicity, whereas models of light weight tend to compromise the discriminative power. Conversely, the proposed framework has a balanced trade-off among accuracy, robustness and efficiency to solve major constraints realized in literature.

The experimental results prove that DA-CBAM EfficientNet is a stable and scalable method of automated brain tumor classification. The large recall and specificity values show that diagnostic reliability is high, whereas low inference time and the model size allow practical implementation. Attention-based architecture is also more interpretable, in the sense that when the architecture focuses on tumor-relevant areas, this is implied and is favorable to clinical decision support. In general, the findings indicate that the suggested model is an effective way to bridge the gap between high-performance deep learning models and the practical use of the model in the clinical setting. Through a combination of the interests of robustness, discriminative feature learning and computational efficiency, DA-CBAM EfficientNet pushes the boundaries of the intelligent brain tumor diagnosis state of the art.

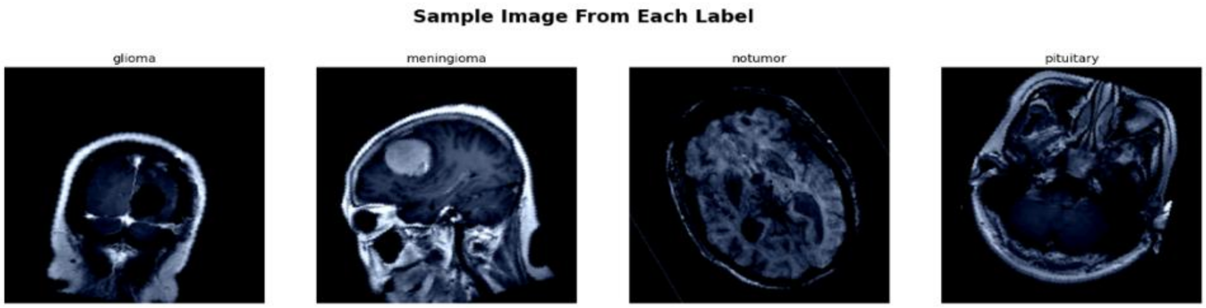


Figure 8: Sample Images from each label

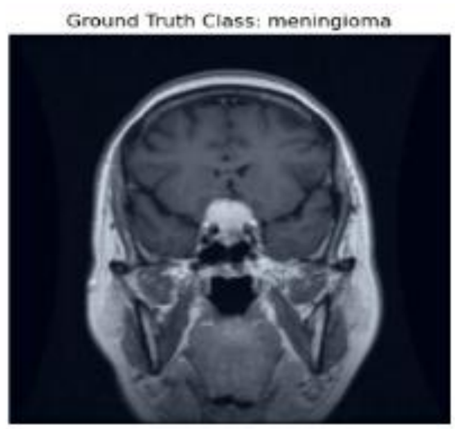
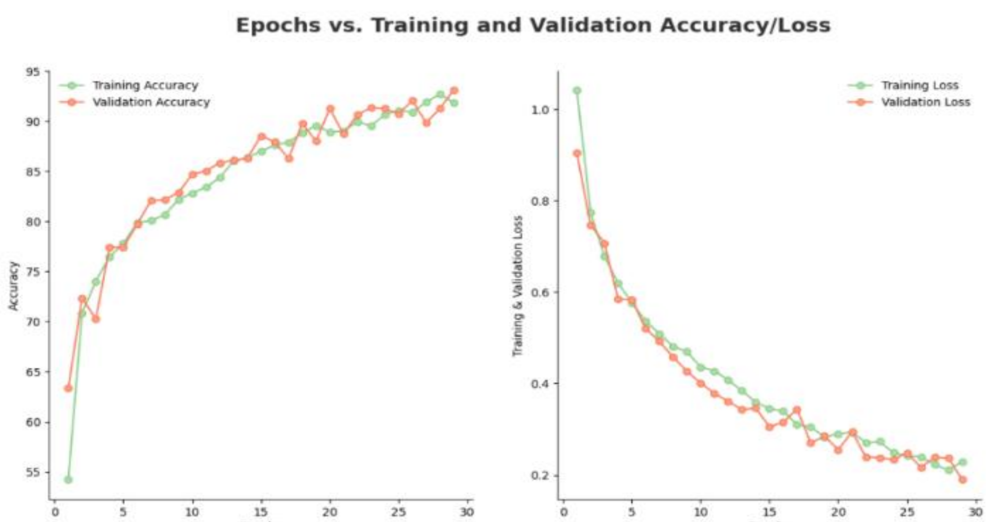


Figure 9: Output for ground truth class meningioma



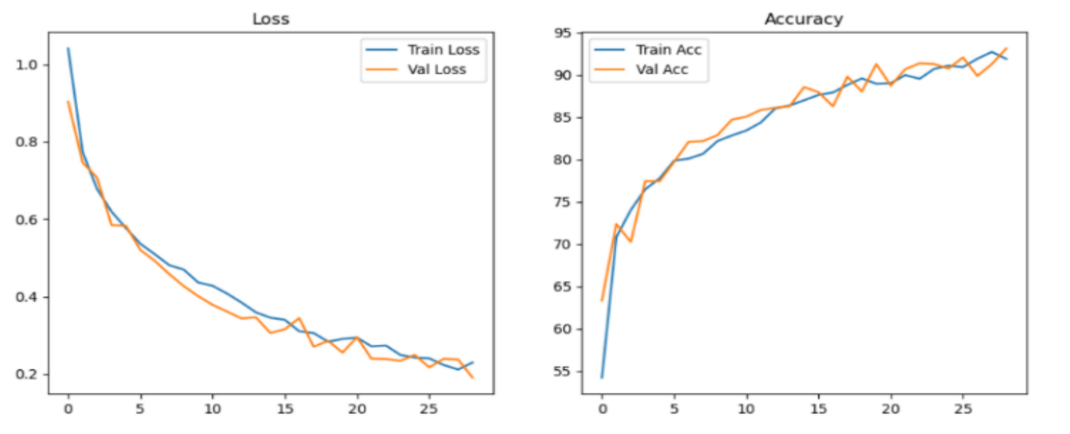


Figure 10: Training and validation accuracy of model

4. Conclusion

The research introduced EfficientNet-DA-CBAM, a deep learning architecture that is efficient and computationally efficient in the classification of brain tumors based on MRI scanners in an automated way. The given method combines EfficientNet-based transfer learning and rotation-aware data augmentation to enhance the generalization in the direction of variability of the orientation. CBAM and the Dynamic Attention-Weighted Enhancement Function (DAWEF) are used together to provide the model with the ability to concentrate on the spatial regions and feature channels of tumor relevance. Moreover, model compression algorithms, such as structured pruning and quantization, have a strong impact of minimizing computational complexity and inference latency with maintaining diagnostic accuracy. The outcomes of the experiments prove that the suggested framework is always more successful in accuracy, robustness, and efficiency compared to the current state-of-the-art solutions. DA-CBAM EfficientNet has a light and scalable architecture that has made it ideal in real time clinical decision support and edge-based healthcare systems. In general, the contribution of high-performance deep learning models in the area of brain tumor diagnosis and its practical application in clinical practice has been bridged in this work.

Future Work

Future directions will be based on the expansion of the proposed framework to multimodal MRI sequences, as well as explanation-based AI methods to enhance clinical interpretability. Also, validation in reality on multi-institutional datasets and implementation of edge based medical devices will be discussed.

References

1. Anita Huttner, Overview of Primary Brain Tumors: Pathologic Classification, Epidemiology, Molecular Biology, and Prognostic Markers, Hematology/Oncology Clinics of North America, Volume 26, Issue 4, 2012, Pages 715-732, ISSN 0889-8588, ISBN 9781455749409.
2. Buerki RA, Horbinski CM, Kruser T, Horowitz PM, James CD, Lukas RV. An overview of meningiomas. Future Oncol. 2018 Sep;14(21):2161-2177.
3. Saeedi, S., Rezayi, S., Keshavarz, H., & R. Niakan Kalhori, S. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. BMC medical informatics and decision making, 23(1), 16.
4. Solanki, S., Singh, U. P., Chouhan, S. S., & Jain, S. (2023). Brain tumor detection and classification using intelligence techniques: an overview. IEEE Access, 11, 12870-12886.
5. Xu, Z., Tian, B., Liu, S., et al. (2023). Collaborative attention guided multi-scale feature fusion network for medical image segmentation. IEEE Transactions on Network Science and Engineering, 11(2), 1857-1871.
6. Zahoor, M. M., Qureshi, S. A., Bibi, S., Khan, S. H., Khan, A., Ghafoor, U., & Bhutta, M. R. (2022). A new deep hybrid boosted and ensemble learning-based brain tumor analysis using MRI. Sensors, 22(7), 2726.
7. Dhar, T., Dey, N., Borra, S., & Sherratt, R. S. (2023). Challenges of deep learning in medical image analysis—improving explainability and trust. IEEE Transactions on Technology and Society, 4(1), 68-75.
8. Preetha R., Priyadarsini M., et al. (2024). Automated Brain Tumor Detection from Magnetic Resonance Images Using Fine-Tuned EfficientNet-B4 Convolutional Neural Network. IEEE Access, 12: 112181–112197.
9. Chauhan, P., Lunagaria, M., Verma, D. K., Vaghela, K., Tejani, G. G., Sharma, S. K., & Khan, A. R. (2024). PBVit: A Patch-Based Vision Transformer for Enhanced Brain Tumor Detection. IEEE Access, 1.

10. Rahman, A., Thobie, A., & Sayer, A. (2022). Accurate brain tumor detection using deep convolutional neural network. Computational and Structural Biotechnology Journal, 20, 4733–4745.
11. Krishnan, P. T., Pradeep, K., Khandelwal, M., Gupta, D., Nihaal, A., & Kumar, T. S. (2024). Enhancing brain tumor detection in MRI with a rotation invariant Vision Transformer. Frontiers in Neuroinformatics, 18.
12. Sahaai, M. B., Jothilakshmi, G. R., Ravikumar, D., Prasath, R., & Singh, S. (2022, May). ResNet-50 based deep neural network using transfer learning for brain tumor classification. In AIP conference proceedings (Vol. 2463, No. 1, p. 020014). AIP Publishing LLC.
13. Sahaai, M. B., Jothilakshmi, G. R., Ravikumar, D., Prasath, R., & Singh, S. (2022, May). ResNet-50 based deep neural network using transfer learning for brain tumor classification. In AIP conference proceedings (Vol. 2463, No. 1, p. 020014). AIP Publishing LLC.
14. Razi, F. A., Bustamam, A., & Latifah, A. L. (2023, July). Development of efficient brain tumor classification on MRI image results using EfficientNet. In 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA) (pp. 575-580). IEEE.
15. Rasheed, M., Jaffar, M. A., Akram, A., Rashid, J., Alshalali, T. A. N., Irshad, A., & Sarwar, N. (2025). Improved brain tumor classification through DenseNet121 based transfer learning. Discover Oncology, 16(1), 1645.
16. Ma, Y., Huang, Z., Su, J., Shi, H., Wang, D., Jia, S., & Li, W. (2023). A multi-channel feature fusion CNN-Bi-LSTM epilepsy EEG classification and prediction model based on attention mechanism. IEEE Access, 11, 62855-62864.
17. Zeineldin, R. A., Karar, M. E., Elshaer, Z., Coburger, J., Wirtz, C. R., Burgert, O., & Mathis-Ullrich, F. (2024). Explainable hybrid vision transformers and convolutional network for multimodal glioma segmentation in brain MRI. Scientific reports, 14(1), 3713.
18. Hong, S., Wu, J., Zhu, L., & Chen, W. (2024). Brain tumor classification in VIT-B/16 based on relative position encoding and residual MLP. Plos one, 19(7), e0298102.