RESEARCH PAPER

An Explainable EfficientNetV2-S and CBAM-Based Deep Learning Framework for Multiclass Brain Tumor Classification Using MRI

Pooja N. Bhandari1, Dr. Atul Nandwal2
1Research Scholar, Department of Computer science and Engineering, SAGE University, Indore, MP 452020, India.
Email:bhandaripooja453@gmail.com
2Associate Professor, Department of Computer science and Engineering, SAGE University, Indore, MP 452020, India.
Email: atulnandwal@gmail.com

Abstract

Reliable discrimination between several types of brain tumors via magnetic resonance imaging (MRI) is a difficult diagnostic problem because of the significant variability in tumor morphology and the often observed overlap in the visual properties of the tumors in the different pathological classes. To address those inherent challenges this work suggests a comprehensive deep learning architecture based on the EfficientNetV2-S as a transfer learning foundation supplemented by a Convolutional Block Attention Module (CBAM) to improve the ability to discriminate the features. MixUp data augmentation, Focal Loss for focusing on difficult instances, AdamW optimization and Cosine Annealing scheduler with warm restarts further enhance the training pipeline. To create an interpretable model, Score-CAM is used to produce qualitative heatmaps that reveal areas that are important to the model decisions. The system was evaluated on a dataset containing 8334 MRI images of tumor type (Glioma, Meningioma, Pituitary tumor, No Tumor) from which the images were divided into three sets: training set (5712 images), validation set (1311 images) and test set (1311 images). The proposed model obtained a testing accuracy of 99.47%, and only seven samples were misclassified in the test set of 1311 samples. The results suggest good results in multibox categorization and better transparency in decision-making but their performance needs to be further tested on independent external datasets to ensure that the model generalizes well to different clinical environments.

Keywords— Brain Tumor Classification, Magnetic Resonance Imaging, EfficientNetV2-S, Convolutional Block Attention Module (CBAM), Transfer Learning, Score-CAM, Explainable Artificial Intelligence

How to cite this article: Pooja N. Bhandari1 Dr. Atul Nandwal2. An Explainable EfficientNetV2-S and CBAM-Based Deep Learning Framework for Multiclass Brain Tumor Classification Using MRI. Int J Drug Deliv Technol. 2026;16(90s):390-399. DOI: 10.25258/ijddt.16.90s.39

Introduction

Brain tumors represent a significant category of neurological disorders and their early detection is critical for effective treatment planning. Among the various imaging methods available MRI stands out as a non-invasive tool that offers detailed, high-resolution images of brain structures. Although the technique is useful there is still a great degree of variability in both the morphology and the signal intensity size, shape and distribution of the tumors in their space on MRI scan. The similarities in the visual appearance between different tumor types make it difficult to differentiate one tumor type from another, highlighting the importance of a strong and automatic decision support system that can assist radiologists in making consistent and accurate tumor typing. Over the past few years computational methods have been more and more studied to automate brain MRI analysis. Most machine learning systems are based on manually designed features or representations based on different feature extraction modules which restricts their ability to capture the complex high-order spatial structure found in medical images. However, with the advent of deep learning and especially convolutional neural networks (CNNs) this has changed by allowing for end-to-end learning of hierarchical structure of

features directly from raw image data. In addition transfer learning has been shown to be effective for tailoring large-scale pre-trained models to particular medical tasks, reliving the restrictions of training on a restricted number of labeled data. These models face several challenges including data variability, class imbalance, overfitting and the lack of transparency in their decision-making which can hinder their use in critical healthcare environments. From a review of the literature it is also evident that there has been significant advancements in various methodological aspects such as machine learning, deep CNNs, hybrid architectures, transfer learning, segmentation based pipeline, multimodal integration and XAI. The reported classification accuracy rate in this field can be as high as 99.3% and as low as 91% but there are still several issues that have not been addressed. For example, many approaches either divide feature extraction from the classification task or focus only on binary tumor detection or rely on segmentation annotations that are expensive to acquire. Solutions that are multimodal are often costly in terms of computation. Cutting-edge performance does not automatically imply the transparency of the model.

In order to address these challenges a comprehensive deep learning framework for multiclass classification of brain MRI images into four classes: Glioma, Meningioma, Pituitary tumor and No Tumor is introduced in the present work. The key goal is to build an accurate and meaningful system that in addition to high classification accuracy, provides visual explanations of why a prediction was made. This is done through the synergic integration of several complementary components: a transfer learning backbone, attention module for feature refinement, targeted data augmentation, a loss function that emphasizes difficult instances and an optimization strategy designed for stable convergence.

The proposed architecture uses EfficientNetV2-S as the main feature extractor with a pretrained version of the network to recognize relevant visual patterns. The CBAM is introduced to further strengthen the discriminative capacity of the learned features with both channel-wise and spatial attention. MixUp augmentation is used to spread the distribution of the training data and Focal Loss to shift the model attention towards those more difficult to classify. To support adaptive learning rate scheduling and convergence behavior, the implementation of the combination of AdamW and Cosine Annealing with Warm Restarts is used for optimization. The classification head consists of Batch Normalization, GELU activation, Dropout and a final Softmax layer for the prediction of 4 classes. Score-CAM is incorporated in the framework to create class-specific saliency maps that show the influential image regions and the heatmap refinement and contour extraction are done just for further visualization without changing the Score-CAM mechanism.

The Contribution of this research as follows:

- A common classification pipeline combining EfficientNetV2-S transfer learning with attention using CBAM to provide strong feature representation for brain MRI data.
- The introduction of MixUp augmentation and Focal Loss to enhance the robustness of learning and to boost the difficulty of easy samples.
- Effective optimization and adaptive learning rate management by using AdamW along with Cosine Annealing with Warm Restarts.
- Post-hoc explainability: Integration of Score-CAM for qualitative visualization of areas of the image relevant to the decision.

Results from the experiments show the effectiveness of the proposed system which yielded an overall accuracy of 99.47% with 1304 correct classifications from 1311 test instances. Transfer learning, attention, regularization, optimization and explainability are all shown to be useful techniques in this study and when combined in a unified framework they are a promising approach to improving automated brain tumor classification. The results are subject to the design of the study. There was no testing or ablation experiments conducted with the model using an external dataset and the localization fidelity of Score-CAM was not quantitatively compared with ground-truth annotations. So, future studies of independent validation cohorts and stringent interpretability metrics are needed before making claims about clinical generalizability.

Literature Survey

Bhagyalaxmi and Dwarakanath were trying to create a hybrid detection system which used multiple classifiers, such as logistic regression, support vector machine, naïve bayes, XGBoost, AdaBoost, and Random Forest, to detect the pattern and VGG16 for feature extraction. Out of these combinations, CNN Logistic Regression achieved the best precision, recall, F1 Score, and accuracy at 94%, 94%, 94%, and 91% respectively [1]. To identify tumors in the MRI images, Khairandish M. O. et al. proposed the hybrid approach of CNN-SVM. They followed a methodology of threshold based segmentation for tumor localization, CNN for features extraction and SVM for final classification. The system showed 98.50% accuracy which is higher than the classical machine learning and deep learning methods [2]. Sreenu K. et al. proposed a single deep learning architecture that includes CNN classification and U-Net segmentation. The classification part was trained based on the Kaggle Brain MRI database and segmentation was performed using the BraTS benchmark data and expert annotations. The integrated framework achieved a classification accuracy of 97.0% and a Dice coefficient of 0.88, indicating its effectiveness in tumor classification and localization in real-time use [3]. Mehan V. presented a hybrid approach for binary classification of brain tumour from MRI image using inceptionV3 for feature extraction. Logistic Regression, Random Forest, and AdaBoost were evaluated using 50 random splits and 10-fold cross-validation. Logistic Regression showed the highest results of accuracy (98.2%) and the highest AUC (0.999), indicating its potential for clinical decision support [4]. In this extensive review, Aamir M. et al. examined a wide range of ML and DL methods for classification of brain tumors based on various public available datasets. They analysed traditional ML algorithms, CNNs, RNNs, vision transformers, preprocessing techniques, and feature extraction techniques. While most of the advanced DL models achieved accuracy exceeding 95%, the review pointed out the current difficulties involved in the models such as interpretability limitations, class imbalance, cross-dataset generalization limitations, and clinical deployment barriers [5]. Saeedi S. et al. have developed a DL system that was trained using 3,264 MRIs containing glioma, meningioma, pituitary tumors, and healthy cases. They used a pipeline consisting of pre-processing, augmentation, convolutional autoencoder, a custom 2D CNN, and six ML classifiers for comparison. The proposed CNN showed the best performance among all the ML techniques, with 96.47% training accuracy, 95% average recall, and 0.99 AUC [6]. Bansal S. et al. proposed a lightweight hybrid CNN-SVM network trained using Dataset_MC and Dataset_BC. To test the robustness of the framework, CNN-extracted features were classified by SVM with 98% and 99% accuracy for the two datasets respectively [7].

Kumar S. et al. used transfer learning to an improved ResNet50 network to distinguish between benign and malignant tumors. The proposed model was found to be superior when compared with the ResNet, AlexNet, U-Net and VGG16 with performance of 99.30% in accuracy for benign and 98.40% in accuracy for malignant cases with high precision, recall and F1 score [9]. The authors in Yadav P. et al. presented a comprehensive review of CNN, ResNet, CRNN, Vision Transformers, transfer learning architectures, hybrid feature fusion and XAI methods. Most of the contemporary models obtained a 95 to 99% accuracy but there are issues like data imbalance, overfitting, interpretability, clinical translation etc. [10]. Rita Appiah et al. suggested Proper Orthogonal Decomposition integrated CNN (POD-CNN) to reduce the computational complexity. The pre-processing of dimensionality reduction before classification was able to significantly reduce the processing time while retaining the classification performance and with the classification accuracy achieving 95.88% and the computational requirement reduced to about a third of that needed by conventional CNN [11]. Llani M. A. et al. proposed a hybrid framework with a combination of T1-weighted MRI images: U-Net was applied as the segmentation model and EfficientNetB4 with VGG19 was used as the classification model. The U-net-based approach achieved accuracy of 98.56%, F score of 99%, AUC of 99.8% and a cross-dataset validation accuracy of 96.01% reflecting a good generalization across clinical datasets [12]. Gottipati S. B. and Thumbur G. presented a multimodal fusion deep transfer learning (MMFDTL) system that fused the contour-based and original MRI modalities. The feature extraction was done using VGG16, InceptionV3 and ResNet50 and the fusion was performed at the decision level. When multimodal integration was applied the accuracy was 98.80%, sensitivity 92.96%, specificity 98.54%, precision 93.60%, F1-score 93.26% and Cohen's kappa coefficient was

91.86% showing the significance of multimodal integration [13]. Shamshad N. et al. have conducted a comprehensive comparison of the transfer learning models of VGG16, VGG19, InceptionV3, ResNet50, DenseNet, and MobileNet using confusion matrix, ROC curve and AUC value. The highest classification accuracy of up to 97% with a reduction in execution time of about 22% from the baseline was obtained with VGG16 which demonstrates its computational efficiency and diagnostic capability [14]. In an attempt to give a better understanding of the image Nabila T. and Salam A. designed a hybrid CNN-SVM architecture to classify binary MR images. The framework consisted of a preprocessing CNN feature extraction and classification by SVM optimized with a learning rate of 0.001 and the Adam optimizer. The experimental results showed high accuracy of 98.92%, precision of 98.92%, recall of 98.92% and F1-score of 98.92% indicating the effectiveness of CNN-SVM integration [15].

Methods and Materials

In this study a novel deep learning based approach is presented for multiclass classification of Brains Tumors with Magnetic Resonance Imaging (MRI). The experimental setting was built and tested with a benchmark corpus available in the Kaggle data repository which is publicly accessible. The dataset consists of 8,334 labelled MRI volumes representing a wide range of pathological and healthy brain morphologies and providing a balanced and representative subset of all categories of diagnostic interest. The scope and inherent variability of this imaging collection is suitable for intensive model training and cross validation and allows for computer-aided diagnostic system performance benchmarking making it a strong reference standard for insuring reliability and generalizability of computer-aided diagnostic systems in the field of neuro-oncology.

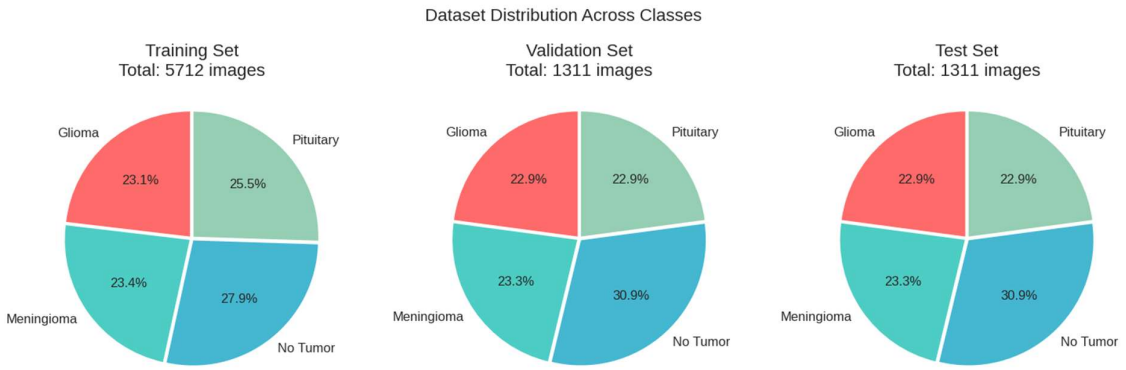

Figure 1: Dataset Distribution

The entire set of 8334 MRIs was divided in a systematic manner into three disjoint subsets for model training, hyperparameter tuning and final evaluation as shown in Figure 1. The ratio of the allocation is approximately 70:15:15 and there are 5,712 (68.5%) samples in the training partition, 1,311 (15.7%) samples in the validation partition, and 1,311 (15.7%) samples in the test partition. A stratified sampling procedure was used when splitting the generation to ensure that the original class prevalence for the entire corpus is maintained. This design allows the relative frequency for each pathological category to be very close to equal for all three subsets. The ratio of images per class in the training partition is as follows: Glioma of 23.1%, Meningioma of 23.4%, No Tumor of 27.9% and Pituitary of 25.5%. The validation and testing sets are both similarly distributed with Glioma of 22.9 %, Meningioma of 23.3 %, No Tumor of 30.9 % and Pituitary of 22.9 %. The fact that this is a well-balanced class across data splits means that there is less potential for selection bias and ensures that all types of tumors are exposed during the learning process. The proposed experimental design maintains the same class ratios helps limit over-fitting potential and enables reliable measure of model predictive performance on unseen MRI volumes thereby ensuring training stability.

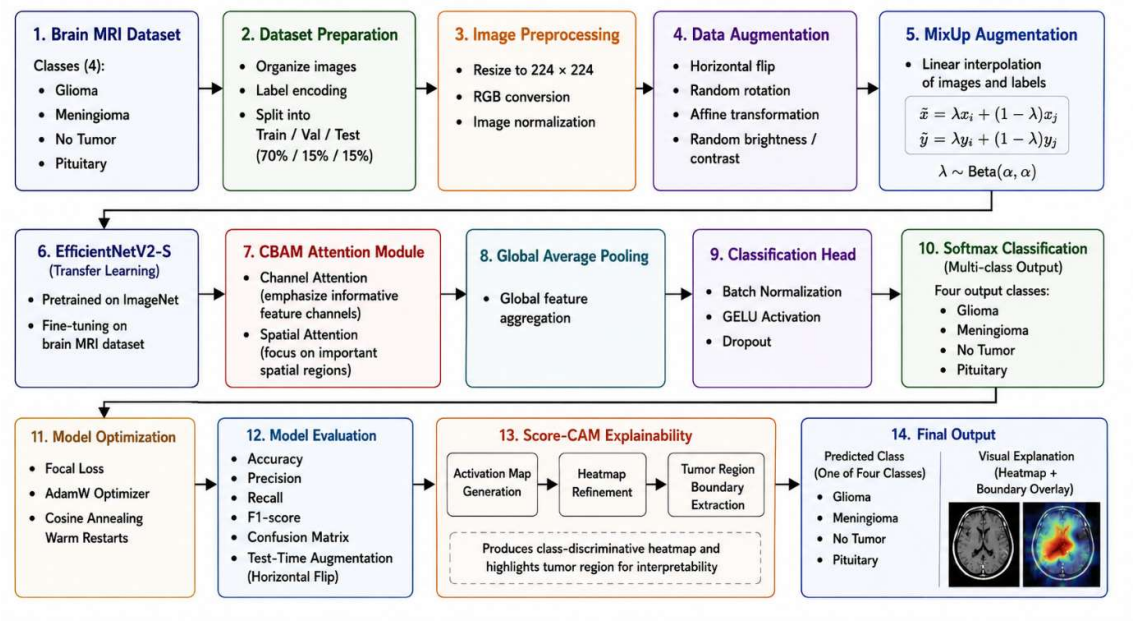

Figure 1: Proposed Methodology

The proposed methodology proposes an extensive framework of deep learning for tumor diagnosis based on the magnetic resonance imaging (MRI) scans categorized into multiclasses. Three major components of the system include EfficientNetV2-S for extracting features by transfer learning, Convolutional Block Attention Module (CBAM) for feature adaptive processing and Score-CAM for obtaining visual explanations and interpretability of the model. As shown in Figure 1 the workflow consists of multiple sequential steps starting from data acquisition, then data preprocessing, augmentation, data synthesis using MixUp, deep feature encoding, attention-guided refinement, classification, hyperparameter tuning, performance evaluation and interpretability evaluation. The intermediate steps result in outputs which are fed forward to the next step and ultimately form a categorical prediction of the tumor type and a corresponding heatmap which shows the regions of the image that are most relevant in the process of classification. The first part of the proposed pipeline targets curating the brain MRI data set that is to be classified into multiple classes. The data set is divided into four groups: Glioma, Meningioma, Pituitary and No Tumor. The first three groups represent pathological entities, the fourth serves as a healthy control group. Multiple tumor types define the problem as a multi-class discrimination task instead of a binary problem of tumor or non-tumorous cases. The next step is to make sure that all MRI samples are laid out systematically based on their class memberships. Categorical encoding labels each instance with a numeric label enabling the model to understand the target variable in supervised learning. The entire data set is then randomly split into three mutually exclusive sets: 70% for learning the model, 15% for validation and 15% as a test set that was not used to train or validate the model as shown in Figure 1. The training set will be used to estimate the weights for the network the validation set will be used for iterative tuning of the hyper parameters and early stopping while the test set will only be used for final unbiased evaluation. This hierarchical split allows clear separation of data roles to reduce the danger of data leakage and to establish a solid basis for evaluating generalizability.

Prior to feature extraction all MRI scans are standardized through a fixed pre-processing routine. To ensure the backbone was compatible with the pretrained image sizes and to normalize the data to stabilize the learning process by gradient-based methods each image is uniformly resized to 224×224 pixels converted to RGB and normalized. These steps help to ensure that the inputs are uniform prior to the augmentation phase. The framework uses a variety of geometric and photometric transformations to improve training diversity and prevent overfitting such as horizontal flips, random rotations, affine warps and brightness-contrast adjustments. These operations add noise that pushes

the model toward learning more general features than dataset-specific artifacts. The retransformed samples are then sent to the MixUp module.

MixUp augmentation is also introduced to the framework to create synthetic training examples through linear combination of pairs of input images and their labels. The mixed image and label can be written as shown in Figure 1:

$$X' = \lambda x_i + (1-\lambda)\, x_j$$

And

$$y' = \lambda y_i + (1-\lambda)\, y_j$$

Where xiand xj represent two input images, yi and yj denote their corresponding labels, and λ is the mixing coefficient sampled from a Beta distribution. The generated sample represents a weighted combination of two training instances rather than an exact copy of either image. This strategy encourages the classifier to learn smoother decision boundaries and reduces excessive dependence on individual training examples. The resulting augmented samples are then provided to the deep feature extraction network.

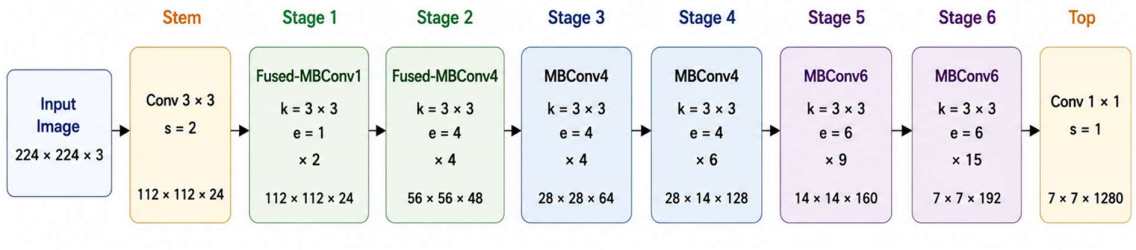

Figure 2:EfficientNetV2-S architecture

In the above Figure 2 the proposed framework is based on EfficientNetV2-S model which is used as a backbone network and is initialized with pre-trained weights on ImageNet data and then fine-tuned on brain MRI data. The model takes in standard size input images of dimensions 224 × 224 × 3 followed by a 3 × 3 convolutional layer with stride 2 that extracts basic visual features while reducing the spatial resolution. After this initial operation the resulting feature maps are hierarchically refined at successive stages called Fused-MBConv1, Fused-MBConv4, MBConv4 and MBConv6 with different expansion ratios and number of repetitions of layers used in each stage. These blocks in the earlier layers of the network namely the Fused-MBConv layers are effective at learning basic spatial structures while the deeper MBConv layers progressively acquire increasingly meaningful and complex representations that are especially sensitive to morphologically salient tumor-specific cues. The depth of the network gets proportionately deeper and its spatial dimensions get progressively smaller while the number of channels increases and the representational capacity of the network is systematically expanded. The multi-channel information is reduced to high level feature maps by a terminal 1 × 1 convolution which are then fed to later processing stages.

Table 1: EfficientNetV2-S Architecture Details

Component	Details
Backbone	EfficientNetV2-S
Attention Mechanism	Convolutional Block Attention Module
Regularization	MixUp (alpha = 0.2)
Loss Function	Focal Loss (gamma = 2.0)
Optimizer	AdamW
Learning Rate Schedule	Cosine Annealing warm Restarts
Explainability	Score-Cam

The numbers in the figure 2 indicate the architectural parameters these are kernel size (k), expansion ratio (e), stride (s) and number of repeated blocks (×n). The goal of transfer learning is achieved by utilizing the general visual knowledge learned from the ImageNet weights and fine-tuning on the MRI corpus. The backbone-generated refined feature maps are then fed into a convolutional block attention module (CBAM) which performs channel-wise and spatial attention to further improve task-relevant features before classification.

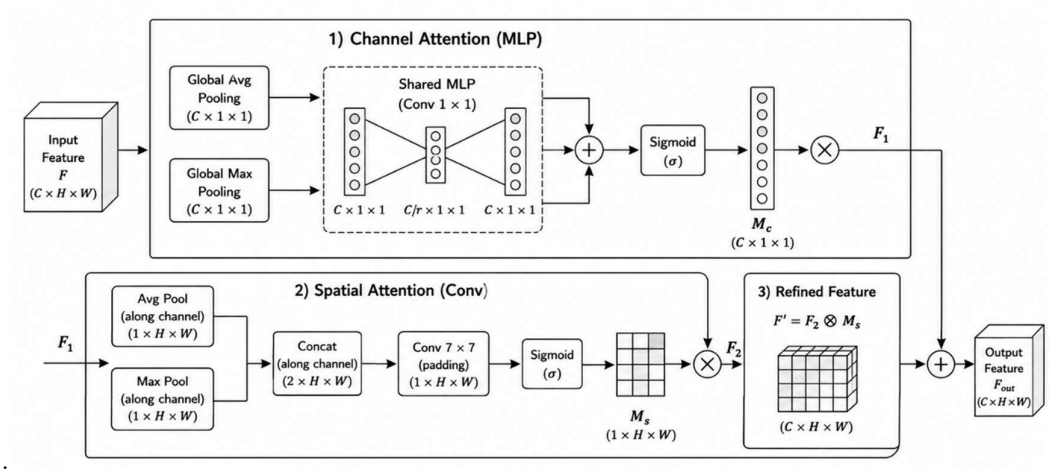

Figure 3: Convolution Block Attention Module

The Convolutional Block Attention Module (CBAM) as illustrated in above figure 3 enhances feature maps by processing them first through channel-wise attention and then through spatial attention. The module takes in an input feature map F of shape C×H×W and first produces two descriptors for each channel using the global average pooling and global max pooling respectively of size C×1×1. They are input to a shared multilayer perceptron (MLP) that uses 1×1 convolutional layers with a reduction ratio r that reduce the channel dimension and then expand it back to the original size. Finally the resulting feature vectors are summed element-wise and fed into a sigmoid activation function to get the channel attention map M_c. This map maps the input feature map's channels adaptively to rescale the output feature map F_1. Average and max pooling operations are applied over the channel axis to produce two different spatially refined feature descriptors 1×H×W. These descriptors are simply concatenated parallel to the channel dimension and passed through a 7×7 kernel which is followed by a sigmoid function to obtain the spatial attention map M_s. This map is used to multiply with F_1 for the final refined feature representation. CBAM can effectively focus the network's attention on relevant channel-wise information and spatially important regions in a two-stage attention process while reducing the impact of irrelevant or noisy features. This adaptive recalibration is used to enhance the discriminative power of the extracted features and hence the performance of the brain tumor classification task.

This is done by the classification head which processes the pooled feature vectors and outputs a feature vector that is suitable for multi-class prediction using the following operations: Batch Normalization, GELU activation and Dropout. To stabilize the distribution of activations throughout the training process, Batch Normalization is used. The non-linear activation function is GELU. To avoid overfitting the technique of dropout is used to randomly turn off a portion of the network during each forward pass. The processed features are then passed to a Softmax layer that generates a probability distribution across the four target classes: Glioma, Meningioma, No Tumor and Pituitary. The last predicted label is the class with the highest probability score.

The training pipeline for model optimization also uses Focal Loss as an objective function which gives more importance to difficult samples thus solving the class imbalance problem. The AdamW optimizer

is used for parameter updating which is the benefit of a decoupled weight decay for better generalisation. To control the learning rate schedule Cosine Annealing with Warm Restarts is followed to periodically decrease and reset the learning rate which helps to ensure convergence stability and avoid local minima. The performance is evaluated by implementing the standard metrics like Accuracy, Precision, Recall, F1-score and a Confusion Matrix to get both overall and perclass performance insights. In cases where Horizontal Flip Test-Time Augmentation (TTA) is applied during inference it is explicitly included in the evaluation pipeline otherwise it is omitted from the methodological framework. For improving the interpretability of the model, Score-CAM is used to obtain class discriminative maps that show the regions of an input image that most affect the model's prediction. This work introduces extra post-processing procedures including heatmap refinement and tumor boundary extraction to the raw Score CAM outputs for more precise and clinically meaning visualizations. These improvements are not part of the standard implementation of Score-CAM but are added to the original Score-CAM algorithm.

Result and Discussion

This chapter will include the outcomes of our experiments and make a comparison with the previous experiments. We take a transfer learning approach with deep neural networks to automatically learn meaningful patterns from the raw data provided. The features extracted from these are then fed into a classifier that predicts the results accurately. The whole system was developed using Python and all experiments were executed with the Google Colab and with the support of an NVIDIA T4 GPU. This facilitated faster training and testing process and made the assessment process efficient and practical.

This chapter summarizes the performance of the model proposed in this work using the backbone feature extractor EfficientNetV2-S. We did not use the pre-trained network as it is but we fine-tuned the final three blocks to learn the patterns more accurately in brain MRI scans related to the disease. To enhance the features representation we added a CBAM attention module which enables the model to attend on the most relevant regions of every image. We utilized MixUp data augmentation to enhance data diversity and prevent overfitting during training. To overcome the class imbalance problem utilized Focal Loss for optimization and used cosine annealing with warm restarts for the learning rate. In order to understand the decision making of the model employed Score-CAM a gradient-less visualization method that generates interpretable heatmap images that show the areas of influence in each prediction.

The proposed model was developed and evaluated using a publicly accessible brain MRI dataset comprising 8,334 images across four categories: glioma, meningioma, pituitary tumor and non-tumor cases. The dataset was partitioned into training, validation and test subsets with 5,712 images allocated for training, 1,311 for validation and 1,311 for testing. To evaluate model performance standard classification metrics including accuracy, precision, recall and F1-score were employed. Additionally, confusion matrices were generated to examine classification errors, and Score-CAM-based saliency maps were produced to provide visual confirmation that the model focused on pathologically relevant regions during decision-making.

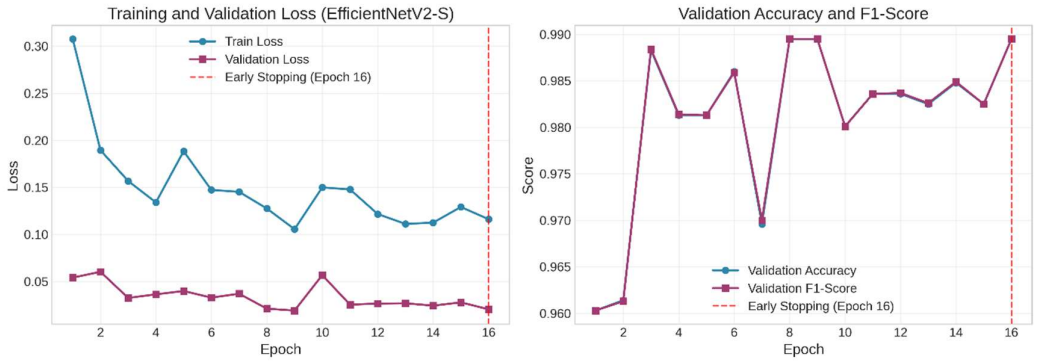

Figure 4: Performance Evaluation

The performance of the model was plotted throughout the training and validation process for 16 epochs in total as seen in Figure 4. The training loss began at 0.3078 and continued to decrease to 0.1162 the lowest at Epoch 9 being 0.1056. This gradual decrease indicates that the model was able to acquire significant patterns in MRI data. The validation loss also began to decline and dropped from 0.0541 to 0.0191 by epoch 9 and remained fairly steady after that before settling at 0.0204 in epoch 11. The validation accuracy and F1-score also significantly increased with epochs reaching approximately 96.03% in the first epoch and 98.95% in Epochs 8, 9 and 16. Note that the minor fluctuations in loss and validation scores are more likely attributed to the use of MixUp augmentation, Focal Loss and the cosine annealing learning rate schedule instead of overfitting. The results in terms of accuracy and F1-score closely matched throughout the validation meaning that the accuracy of the model was consistent across the different measurements used to evaluate it.

To avoid expending unnecessary computation training was stopped automatically at epoch 16 due to early stopping. To avoid such pointless computation training was stopped automatically at epoch 16 using early stopping. We applied test-time augmentation (TTA) to the held-out test set with the model getting an accuracy and F1 score of 99.47% demonstrating its generalizing power to new unseen data.

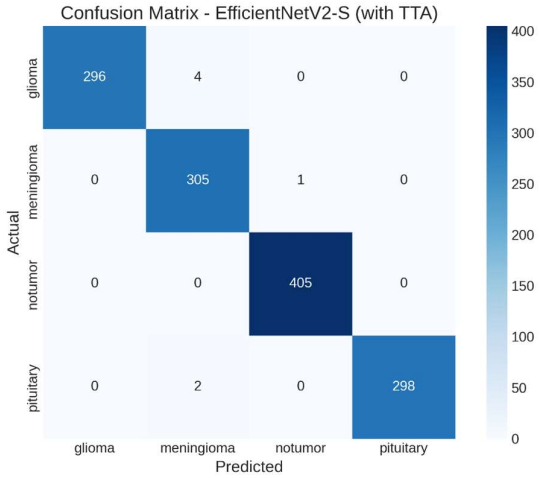

Figure 5: Confusion Matrix

The confusion matrix of our EfficientNetV2-S model trained with CBAM attention on a test-time augmentation set of 1,311 brain MRI scans is shown in figure 5. The model could accurately classify 296

out of 300 Glioma cases as Glioma and classify the remaining 4 cases as Meningioma. In the Meningioma category 305 samples were correctly classified and 1 was incorrectly predicted as No Tumor. With the 405 normal control images the model classified them perfectly. In the Pituitary category 298 of the 300 samples are correctly identified with 2 samples misclassified as Meningioma. The model achieved an accuracy of 99.47% on 1311 test samples with 1304 of them being correctly predicted. There were only 7 misclassifications in all categories. Most of these errors were limited to the Meningioma class and sometimes the model misidentified it as Glioma or Pituitary tumors. This indicates that there are some commonalities between the signals of the various tumor types but the overall error rate is very small. The results indicate that the model reliably discriminates among the four diagnostic categories.

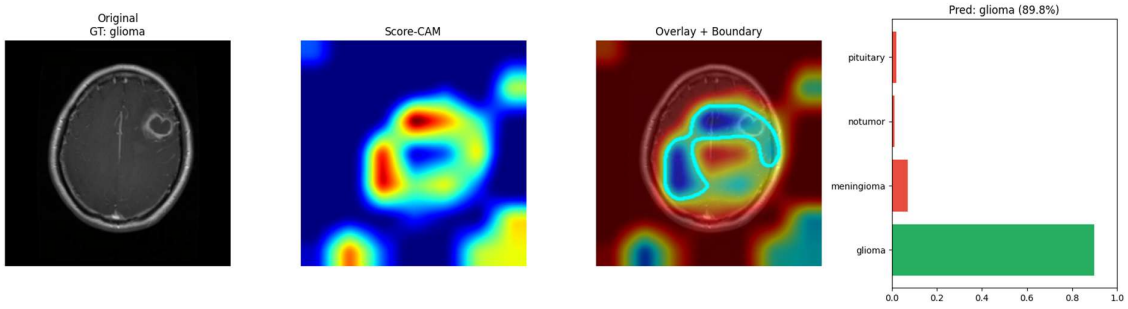

Figure 6: Grad CAM Visualization

Figure 6 demonstrates the interpretability analysis using Score-CAM for the correctly classified case of an MRI of a Glioma tumor. The input image is annotated with the ground truth class Glioma and the output image is labeled with the same class with 89.8% confidence. The Score-CAM heatmap highlights prominent regions in the suspected tumor region indicating that the spatial patterns have a significant impact on the network prediction. The overlay applied in a superimposed manner further illustrates the visual similarity of the activation map and underlying lesion site and the extracted contour provides an additional demarcation of the focus of lesion interest. These visual outputs provide qualitative evidence that the model decision is based on diagnostically salient features rather than global features of the image. Please note that boundary delineation and heatmap enhancement make up additional post-processing operations on the original Score-CAM results which are not included in the core algorithm. However, the contours shown should not be interpreted as exact tumour boundaries, or clinically validated contours without the aid of ground-truth segmentation masks or quantitative localization metrics.

Table 2: Hyperparameter

Hyperparameter	**Value**
Backbone Network	EfficientNetV2-S
Input Image Size	224 × 224 × 3
Batch Size	32
Maximum Epoches	25
Optimizer	AdamW
Initial Learning Rate	3×10^{-4}
Loss function	Weighted Cross Entropy Loss
TTA	Horizontal Flip
XAI	Score CAM
LR Scheduler	Cosine Annealing with warm Restarts

Table 2: Comparative Analysis

Technology Category	Methodology	Best Reported Performance	Major Limitation

Machine Learning [1] [4]	CNN/transfer-learning features combined with Logistic Regression, SVM, RF, AdaBoost, XGBoost, etc.	Up to 98.2% accuracy [4]	Separate feature-extraction and classifier stages and limited interpretability.
Deep Learning [3],[6],[11]	CNN, convolutional autoencoder and POD-CNN-based architectures.	Up to 97.0% accuracy [3]	Architectural complexity and limited integrated explainability
Hybrid DL Models [2], [7], [15]	CNN feature extraction combined with SVM classification	Up to **99% accuracy** [7]	Additional classifier stage and in some studies binary classification.
Transfer Learning [8],[9],[12],[14]	Pretrained VGG, ResNet, EfficientNet, Inception, DenseNet, and MobileNet models fine-tuned for MRI classification.	Up to 99.30% accuracy [9]	Different datasets/tasks make direct comparison difficult. Attention and explainability are not consistently integrated.
Segmentation + Classification [3], [12]	U-Net-based tumor segmentation combined with CNN/transfer-learning classification.	98.56% accuracy and 99% F-score [12]	Requires segmentation annotations and an additional processing stage.
Multimodal DL [13]	Fusion of original and contour-based MRI representations using multiple pretrained networks.	98.80% accuracy [13]	Requires multiple modalities and greater computational complexity.
Explainable AI [5],[10]	XAI techniques reviewed to improve interpretation of DL decisions.	Most reviewed models achieved 95 to 99% accuracy [5], [10]	Reliable localization, generalization, and clinical translation remain challenging
Proposed Integrated Framework	EfficientNetV2-S + CBAM + MixUp + Focal Loss + AdamW + Cosine Annealing Warm Restarts + Score-CAM.	99.47% accuracy, precision, recall, and F1-score. 1,304/1,311 correct predictions.	Single-dataset evaluation, external validation.

The current literature on brain tumor classification shows that there are a wide range of techniques such as machine learning algorithms, deep neural networks, hybrid CNN–SVM, transfer learning, segmentation-based, multimodal data integration and explainable AI (XAI). The accuracy of these studies in such reports are generally in the range of 91 to 99.30% [1–15]. Despite these progresses most proposed solutions are still limited to either independent classifiers, different preprocessing segmentation units, multiple feature extraction streams or lack of model transparency. To overcome these drawbacks the present study proposes a unified multiclass classification framework that integrates EfficientNetV2-S with transfer learning, convolutional block attention modules (CBAM), MixUp data augmentation, Focal Loss optimization, AdamW weight decay, cosine annealing scheduling with warm restarts and Score-CAM for visual explainability. The proposed system was able to obtain classification accuracy of 99.47%, precision of 99.47% and recall of 99.47% with the F1 score of 99.47% from the test set of 1311 test samples out of which 1304 samples were correctly classified. It is worth pointing out though, that the model was only validated using one dataset an analysis of ablation was not performed and an external validation on another independent dataset or a quantitative evaluation of XAI localization fidelity was not performed within this work.

Conclusion

This paper addresses the problem of multiclass brain tumor classification using MRI images by proposing a deep learning architecture, EfficientNetV2-S and CBAM which improves feature representation. MixUp augmentation is used to make the data more varied, Focal Loss is used to focus learning on harder examples and AdamW optimizer is used with Cosine Annealing which uses warm restarts to ensure stable convergence. To make the model more transparent, Score-CAM is used to generate visual

maps of which regions of the image are most important for the network's prediction. The proposed system was evaluated with a set of 8,334 MRI volumes belonging to four classes: Glioma, Meningioma, Pituitary tumor and No Tumor that achieved a classification accuracy of 99.47% on 1,304 out of 1,311 test instances. Cross-dataset performance evaluation and quantitative interpretability analysis should be studied first before real-world clinical deployment in the future.

References

1. Bhagyalaxmi K., Dwarakanath B. Hybrid model for detection of brain tumor using convolution neural networks. *Computer Science and Information Technologies, 5*(1) https://doi.org/10.11591/csit.v5i1.p84-90
2. Khairandish M. O., Sharma M., Jain V., Chatterjee J. M., Jhanjhi N. Z. (2022) A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images**.** *IRBM, 43*(4), 290–299. https://doi.org/10.1016/j.irbm.2021.06.003
3. Sreenu K., Satya N. S. K., Gowtami V. P., Suravarapu S. and Teja S. (2026). Hybrid deep learning framework for brain tumor classification and segmentation in MRI images. In *2026 International Conference on Electrical and Electronics for Sustainable Innovations (ICEESI).* https://doi.org/10.1109/ICEESI2026
4. Mehan V. (2025). Deep Inception-based hybrid machine learning framework for binary classification of brain tumor MRI scans. *Complex Systems Informatics and Modeling Quarterly, 44*, 1–16. https://doi.org/10.7250/csimq.2025-44.01
5. Aamir M., Rahman Z., Bhatti U. A., Bhutto J. A., Choudhry N., & Abro W. A. (2025) Brain tumor classification based on machine learning and deep learning technologies: A comprehensive review. *2025 IEEE 3rd International Conference on Computer, Vision and Intelligent Technology (ICCVIT)*. IEEE. https://doi.org/10.1109/ICCVIT67848.2025.11391405
6. Saeedi S., Rezayi S., Keshavarz H. & Niakan Kalhori, Sharareh R. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques**.** *BMC Medical Informatics and Decision Making, 23*, Article 16. https://doi.org/10.1186/s12911-023-02114-6
7. Bansal S., Jadon Rakesh S. & Gupta Sanjay K. (2024). A robust hybrid convolutional network for tumor classification using brain MRI image datasets. International Journal of Advanced Computer Science and Applications (IJACSA)*, 15*(4). DOI: https://doi.org/10.14569/IJACSA.2024.0150459
8. T. Lakshmi Prasanthi, N. Neelima, "Improvement of Brain Tumor Categorization using Deep Learning: A Comprehensive Investigation and Comparative Analysis" Procedia Computer Science Volume 233, 2024, Pages 703-712. https://doi.org/10.1016/j.procs.2024.03.259
9. Kumar S., Choudhary S., Jain A., Singh K., Ahmadian A.and Bajuri M. Y. (2023). Brain tumor classification using deep neural network and transfer learning. *Brain Topography, 36*, 305–318. https://doi.org/10.1007/s10548-023-00968-3
10. Yadav P., Shukla S., Pandey O. and Nikita. (2026). Brain tumor detection and classification from MRI scans using deep learning. *Fourth International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*. IEEE. https://doi.org/10.1109/ICAISS68683.2026.11526152
11. Rita Appiah, Venkatesh Pulletikurthi, Helber Antonio Esquivel-Puentes, Cristiano Cabrera, Nahian I. Hasan , Suranga Dharmarathne, Luis J. Gomez , Luciano Castillo, "Brain tumor detection using proper orthogonal decomposition integrated with deep learning networks" Computer Methods and Programs in Biomedicine 250 (2024) 108167 doi: 10.1016/j.cmpb.2024.108167
12. Llani M. A., Shi Dingjing and Banad Y. M. (2025) T1-weighted MRI-based brain tumor classification using hybrid deep learning models. Scientific Reports*, 15*, Article 7010. https://doi.org/10.1038/s41598-025-91398-0
13. Gottipati S. B. and Thumbur G. (2024). Multi-modal fusion deep transfer learning for accurate brain tumor classification using magnetic resonance imaging images. Indonesian Journal of Electrical Engineering and Computer Science*, 34*(2). https://doi.org/10.11591/ijeecs.v34.i2.pp1134-1143
14. Shamshad N., Sarwr D., Almogren A., Saleem K., Munawar A. and Ur Rehman A. (2025). Enhancing brain tumor

classification by a comprehensive study on transfer learning techniques and model efficiency using MRI datasets. *IEEE Access.* https://doi.org/10.1109/ACCESS.2025.3565380

15. Nabila T. and Salam A. (2024). Classification of brain tumors by using a hybrid CNN-SVM model. Journal of Applied Informatics and Computing
16. Sandeep Kumar Mathivanan , Sridevi Sonaimuthu , Sankar Murugesan and other "Employing DLand transfer learning for accurate brain tumor detection" Nature Scientific report, 2024:14-7232
17. K. Bhagyalaxmi, B. Dwarakanath, P. Vijaya Pal Reddy "Deep learning for multi grade brain tumor detection and classifcation: a prospective survey" Springer Multimedia Tools and Applications, 2024
18. Praveen Kumar Ramtekkar, Anjana Pandey, Mahesh Kumar Pawar "Accurate detection of brain tumor using optimized feature selection based on deep learning techniques" Springer Multimedia Tools and Applications (2023) 82:44623–44653.
19. Kamini Lamba , Shalli Rani, Monika Anand, Lakshmana Phaneendra Maguluri, "An integrated deep learning and supervised learning approach for early detection of brain tumor using magnetic resonance imaging" , Healthcare Analytics 5 (2024) 100336