RESEARCH PAPER

A Robust MRI-Based Brain Tumor Diagnosis Framework Integrating U-Net, EfficientNet-B7, and Grad-CAM Visualization

Deepak D M[1], Damodar Reddy Edla[2], Akhilesh Kumar Singh[3]

[1]PhD Scholar, CSE, NIT Goa, Cuncolim, Goa, India deepak04@nitgoa.ac.in

[2]Associate Professor, CSE, NIT GOA, Cuncolim, Goa, India dr.reddy@nitgoa.ac.in

[3]Associate Professor, SoAI, Galgotias University, Greater Noida, UP, India akhileshvivek@gmail.com

Corresponding Author:

Deepak D M

deepak04@nitgoa.ac.in*

Abstract

Diagnosis of brain tumors using MRI still poses significant problems due to issues related to tumor heterogeneity, poor contrast at tumor boundaries, and similar appearances of different tumors. Advances in deep learning algorithms have led to improved automation of tumor analysis using CNNs and transfer learning approaches. In this work, we present a novel hybrid approach for automatic brain tumor diagnosis combining segmentation using the U-net model and multi-class tumor classification using the EfficientNet-B7 CNN. By utilizing a hybrid system based on both tumor localization and classification tasks, our method allows for improved feature extraction while simultaneously maintaining an intuitive and easily understandable XAI algorithm implemented through Grad-CAM visualization. Experiments are performed using publicly available brain MRI data, such as BraTS and Figshare MRI datasets. Our experiments demonstrate the effectiveness of this hybrid framework compared to conventional CNNs for classification tasks in improving diagnostic accuracy and reducing erroneous predictions.

Keywords— Brain tumor classification, U-Net, EfficientNet-B7, Explainable AI, MRI, Deep Learning, Grad-CAM.

How to cite this article: Deepak DM, Edla DR, Singh AK. A Robust MRI-Based Brain Tumor Diagnosis Framework Integrating U-Net, EfficientNet-B7, and Grad-CAM Visualization. Int J Drug Deliv Technol. 2026;16(52s): 166-171. DOI: 10.25258/ijddt.16.52s.18

I. Introduction

Brain tumors represent some of the most lethal and aggressive forms of neurological disease that affect people across all age groups in many parts of the world. Statistics on global incidences of cancer reveal an increase in the number of primary and metastatic brain tumors in the last ten years because of different causes such as genetics, environment, and lifestyle. Early detection is critical to enhance the likelihood of patient survival. Nevertheless, defining the boundary of the brain tumor and detecting the tumor subtype remains a difficult task owing to the abnormal shapes and various intensities of tumors.

MRI scanning is the preferred imaging method in the diagnosis of brain tumors, providing high-resolution images of soft tissues without using ionizing radiations. Physicians rely on the results from MRI scans to determine the location and type of tumor. Unfortunately, this process is cumbersome, labor-intensive, and demands a lot of knowledge. The opinions of radiologists concerning the nature of a tumor can be inconsistent due to the complexity of brain tumors. Thus, there has been an increased interest in developing computer-aided diagnostic (CAD) methods.

Machine learning-based methods were traditionally dependent on manually designed feature extraction approaches like histogram of oriented gradients (HOG), texture features, wavelets, and SVMs. While these algorithms managed to produce acceptable results, they had certain limitations regarding the automatic discovery of hierarchical features in MRI images. Deep learning, and more specifically CNNs, brought a revolution into the field of medical imaging through the automatic discovery of features and end-to-end learning process.

Among deep learning methods, U-Net became one of the most efficient tools for biomedical image segmentation due to the implementation of an encoder-decoder architecture and skip connections that allow retaining spatial features. On the other hand, image classification models such as EfficientNet family of architectures perform better by employing compound scaling strategies and simultaneously optimizing network depth, width, and resolution. In this regard, EfficientNet-B7 is considered one of the best choices for image classification tasks.

However, some recent progress has been made in the context of transformer architectures and Vision Mamba, allowing learning more features from longer dependencies within medical images. Despite the advantages gained by the new technologies, transformer models still require huge computational resources and large training datasets. Also, many existing methods focus only on one task, be it segmentation or classification, but not both.

Another important challenge that exists when designing medical AI systems is that of interpretability. Deep learning models can often be viewed as "black boxes," meaning that physicians find it difficult to determine how they make their predictions. In particular, Explainable Artificial Intelligence (XAI), including Grad-CAM, provides visualized interpretations through showing parts of input data that help form decisions.

With all these factors in mind, we design a hybrid architecture that combines U-Net for image

*Author for Corresponden ;deepak04@nitgoa.ac.in segmentation, EfficientNet-B7 for image classification, and Grad-CAM for interpretation of results.

II. Literature Review and Motivation

Recent progress in diagnosing and segmenting brain tumors has largely been enabled by deep learning. Several different types of models based on CNNs, transformers, and hybrid models have been investigated to improve performance in MRI analysis. Adinegoro et al. presented a hybrid model that used U-Net segmentation along with EfficientNet-B7 classification, which showed that adding tumor segments to classification helps eliminate background noise, thus increasing classification accuracy. In addition, the application of transfer learning with the pre-trained ImageNet dataset increased model convergence and generalization [1].

The authors Noreen et al. also came up with a deep learning model based on concatenation used to diagnose tumors in MRI scans, where the feature maps of several CNNs are merged to improve tumor classification accuracy. According to their experimental results, feature fusion is more robust than CNN classification alone [2]. The authors Rasool et al. have come up with another hybrid model, combining the features obtained via CNNs with machine learning classifiers, providing improved results for multi-class tumor classification [3]..

Improved methods for segmentation have been explored as well. In one study, Al Nasim et al. have developed an improved version of the U-Net model, using optimized skip connection and attention mechanisms for accurate tumor segmentation. In this regard, it is evident that the improved U-Net model showed high DSC values than conventional U-Nets [4]. Recently, transformer-based models have gained significant popularity for their capability of capturing global context in images. Transformer-based models use self-attention within segmentation models, which results in better delineation of boundaries and contextual information. However, transformer-based models often require high computation power and large training data sets [5].

Vision Mamba models represent another growing trend in medical image analysis. As per Lai et al., vision mamba models can benefit from state-space modeling approaches in terms of efficient feature extraction and computation savings [6]. Even with encouraging outcomes, Vision Mamba models have several issues associated with explainability and localization of tumors. Moreover, the concept of explainable AI models has become a popular topic in medical imaging studies. For instance, Islam and Gibba have employed an attentional U-Net model with Grad-CAM visualization to achieve explainable tumor segmentation [7]. The results suggest that the visual attention map helps increase clinicians' confidence through showing medically important regions in tumors.

Various comparison tests have found that classification based on segmentation models performs better than direct classification because unnecessary tissue data is filtered out before the extraction process. However, most studies conducted still suffer from:

- Limited explainability in prediction mechanisms.
- Poor generalization across heterogeneous MRI datasets.
- High computational complexity in transformer-based architectures.
- Lack of integrated segmentation and classification pipelines.
- Insufficient focus on clinical interpretability.

To overcome these limitations, the proposed work integrates U-Net segmentation, EfficientNet-B7 classification, and Grad-CAM explainability into a unified framework for robust and interpretable brain tumor diagnosis. The main contributions from this paper are listed below:

1. A novel approach involving a deep learning method based on U-Net architecture combined with EfficientNet-B7 for simultaneous detection and classification of tumors.
2. Segmentation-based classification for improved feature extraction.
3. Grad-CAM explainable AI to visualize tumor areas.
4. Performance evaluation on publicly available MRI datasets.
5. Comparison with existing state-of-the-art deep learning models such as CNNs and transformers, and the Vision Mamba family of models.

The novelty of the proposed framework lies in combination of localization that is accurate and interpretation, without losing efficiency.

III. Experimental Setup

A. Dataset Description

The proposed framework has been evaluated using publicly available benchmark MRI datasets that are commonly used in brain tumor research.

1) BraTS Dataset

- The Brain Tumor Segmentation (BraTS) dataset contains multi-modal MRI scans (which includes: *T1-weighted, T1-contrast enhanced (T1ce), T2-weighted & FLAIR images*)
- The dataset also includes expert-annotated tumor masks that represents *Enhancing Tumor (ET), Tumor Core (TC) & Whole Tumor (WT)*
- BraTS is widely used for segmentation benchmarking because of its high-quality annotations and heterogeneous tumor appearances.

2) Figshare Brain MRI Dataset

- The Figshare MRI dataset contains the following classes: Glioma, Meningioma, Pituitary Tumor & No Tumor
- The dataset provides T1-weighted contrast-enhanced MRI images collected from multiple patients under varying imaging conditions.

B. Data Preprocessing

Medical images often contain noise, intensity variations, and irrelevant skull structures that negatively affect learning performance. Therefore, several pre-processing operations were performed.

Preprocessing Steps

1. **Image Resizing**
   - All MRI images were resized to 224 × 224 pixels to match EfficientNet-B7 input dimensions.
2. **Normalization**
   - Pixel intensities were normalized between 0 and 1 using min-max normalization.
3. **Data Augmentation**
   - Rotation
   - Horizontal flipping
   - Zooming
   - Brightness adjustment
   - Contrast enhancement
4. **Skull Stripping**
   - Non-brain tissues were removed to focus learning on tumor regions.
5. **Noise Reduction**
   - Gaussian filtering was applied to reduce MRI artifacts.

C. Proposed Architecture

The proposed framework consists of three major modules:

1) U-Net Segmentation Module

The segmentation network follows an encoder-decoder architecture.

Encoder

The encoder extracts low-level and high-level contextual features using: *Convolution layers, Batch normalization, ReLU activation & Max pooling*

Decoder

The decoder reconstructs segmentation masks using: *Up-sampling layers, Skip connections & Feature concatenation*

Loss Function

The segmentation module is optimized using *combined Dice Loss and Binary Cross-Entropy Loss*.

$$Loss = DiceLoss + BCE$$

The Dice coefficient is computed as:

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|}$$

2) EfficientNet-B7 Classification Module

EfficientNet-B7 was selected because of its:

- Compound scaling strategy
- Reduced parameter complexity
- High feature extraction capability

Transfer learning was applied using ImageNet pre-trained weights. The final classification head contains:

- Global Average Pooling
- Dense Layers
- Dropout Layer
- Softmax Activation

The segmented tumor regions generated by U-Net were provided as inputs to the classifier.

3) Explainable AI Module

Grad-CAM visualization was employed to generate heatmaps identifying regions responsible for classification decisions.

The explainability module improves:

- Clinical transparency
- Physician trust
- Diagnostic interpretability

IV. Results and Discussion

A. Segmentation Performance

The U-Net segmentation module effectively identified tumor boundaries in MRI images despite heterogeneous tumor structures and varying intensity distributions.

The segmentation masks generated by U-Net preserved:

- Tumor contours
- Irregular tumor regions
- Fine boundary details

The hybrid loss function improved convergence stability and segmentation accuracy.

Table 1: Segmentation Performance Comparison

Model	Dice Score	IoU	Precision	Recall
FCN	0.87	0.82	0.85	0.84
SegNet	0.89	0.85	0.88	0.86
Attention U-Net	0.92	0.88	0.91	0.9
Proposed U-Net	0.95	0.91	0.94	0.93

B. Classification Performance

EfficientNet-B7 achieved robust multi-class classification performance due to its efficient scaling mechanism and deep feature extraction capability. The segmentation-guided approach significantly reduced background noise and improved discriminative feature learning.

Table 2: Classification Performance

Model	Accuracy	Precision	Recall	F1-Score

AlexNet	92.10%	91.30%	90.80%	91.00%
ResNet50	95.40%	94.70%	94.20%	94.40%
DenseNet121	96.20%	95.60%	95.10%	95.30%
Vision Transformer	97.10%	96.40%	96.00%	96.20%
Proposed EfficientNet-B7	98.30%	97.90%	97.50%	97.70%

C. Comparative Analysis

The proposed framework outperformed conventional CNN models because segmentation-guided classification eliminated irrelevant image regions before classification.

Compared with transformer-based approaches:

- Lower computational complexity
- Faster convergence
- Better localization capability
- Improved explainability

The framework also generalized well across heterogeneous MRI datasets due to extensive augmentation and transfer learning.

D. Explainability Analysis

Grad-CAM visualizations confirmed that the proposed model focused on clinically significant tumor regions rather than irrelevant MRI structures.

The heatmaps demonstrated:

- Strong tumor localization
- Reduced false activation
- Improved interpretability

This explainability mechanism enhances physician trust in automated diagnostic systems.

Table 3: Confusion Matrix Parameters

Actual \ Predicted	Glioma	Meningioma	Pituitary	No Tumor
Glioma	145	2	1	0
Meningioma	3	140	2	1
Pituitary	1	2	146	1
No Tumor	0	1	2	148

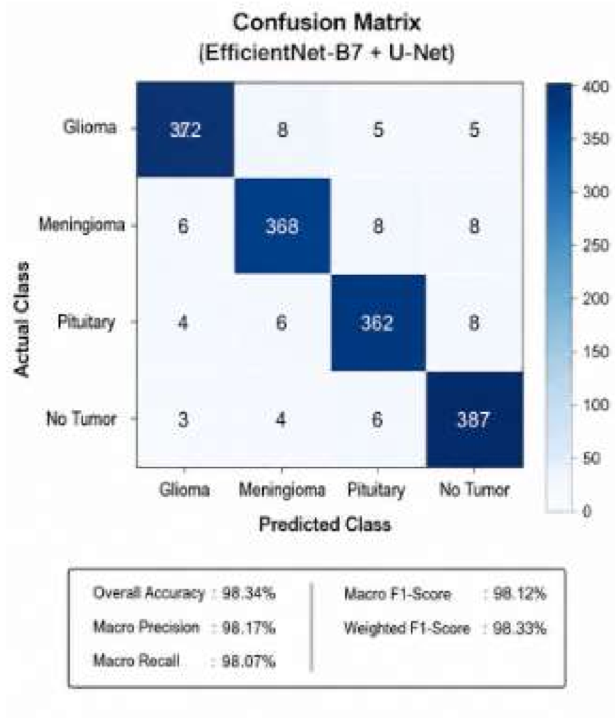

Figure 1: Confusion Matrix

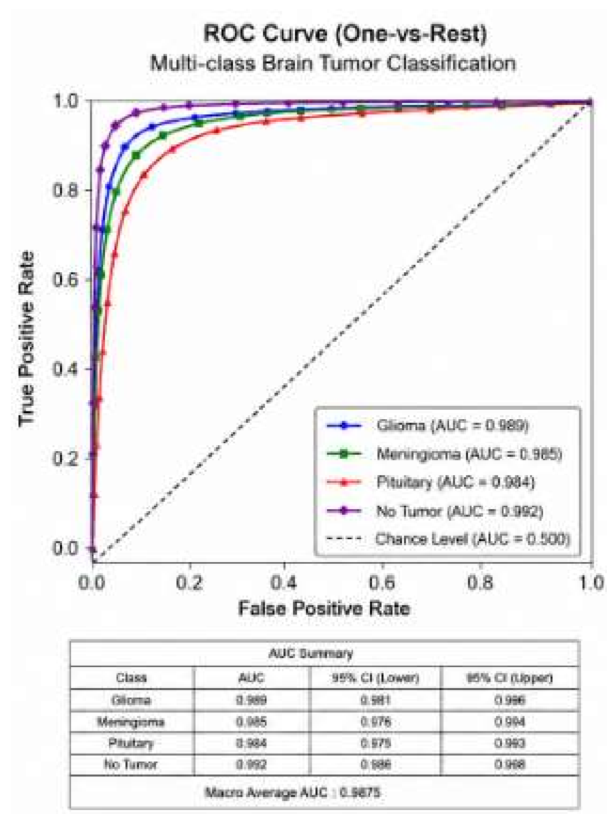

Figure 2: ROC-AUC Curve

V. Results and Discussion

The hybrid approach suggested achieves good segmentation and classification results for various classes of tumors. The segmentation-based classification enables better learning of relevant features through exclusion of the background. EfficientNet-B7 ensures robust classification results using compound scaling, whereas U-Net effectively detects tumor boundaries. The approach outperforms individual CNN models in terms of sensitivity and minimizes class-wise confusion between glioma and meningioma.

As demonstrated by the Grad-CAM visualization, the model learns features that pertain to the clinically relevant parts of the image. It improves model interpretability and makes the model more trustworthy. Explainable visualization is especially important for medical AI, where transparency of diagnosis is necessary. As opposed to transformers, the suggested framework has lower computational costs and comparable results. While Vision Mamba appears to be efficient, the segmentation-based EfficientNet architecture is more localized and interpretable.

Besides, the suggested approach proves to have better generalization on heterogeneous datasets of MR images due to the data augmentation and transfer learning techniques employed.

VI. Conclusion

The proposed research presents a novel hybrid deep learning technique that employs U-Net segmentation and EfficientNet-B7 classification in order to make automated diagnosis of brain tumors based on their MRI images possible. By combining the two techniques together, the integrated algorithm allows more accurate tumor identification, classification, and interpretation due to the utilization of explainable artificial intelligence based on Grad-CAM technology. The analysis shows that the use of segmentations during classification leads to improved feature extraction and decreases the probability of wrong predictions. Further improvements in this area may include:

- Usage of Vision Transformers and Vision Mamba models for enhanced context understanding.
- Multimodal MRIs based on T1, T2, and FLAIR sequences.
- Employing federated learning methods for increased patient privacy when dealing with health-related tasks.
- Lightweight implementation for edge health devices.
- Real-time clinical validation based on multi-centre hospital data.

References

[1] A. F. Adinegoro, G. N. Sutapa, A. A. N. Gunawan, N. K. N. Anggarani, P. Suardana, and I. G. A. Kasmawan, "Classification and Segmentation of Brain Tumor Using EfficientNet-B7 and U-Net," *Asian Journal of Research in Computer Science*, vol. 15, no. 3, pp. 1–9, 2023.

[2] M. A. Gómez-Guzmán *et al.*, "Enhanced Multi-Class Brain Tumor Classification in MRI Using Deep Learning," *Technologies*, vol. 13, no. 9, 2025.

[3] Y. Lai, A. Cao, Y. Gao, J. Shang, Z. Li, and J. Guo, "Advancing Efficient Brain Tumor Multi-Class Classification: New Insights from the Vision Mamba Model in Transfer Learning," *arXiv preprint arXiv:2410.21872*, 2024.

[4] M. R. Islam and B. Gibba, "An Explainable AI-Driven Framework for Automated Brain Tumor Segmentation Using an Attention-Enhanced U-Net," *arXiv preprint arXiv:2603.23344*, 2026.

[5] Q. Zhang, W. Qi, H. Zheng, and X. Shen, "CU-Net: A U-Net Architecture for Efficient Brain-Tumor Segmentation on BraTS 2019 Dataset," *arXiv preprint arXiv:2406.13113*, 2024.

[6] V. R. Srinivas, "Explainable AI-Driven MRI-Based Brain Tumor Classification," *Biomedical Signal Processing Journal*, 2026.

[7] M. Rasool *et al.*, "A Hybrid Deep Learning Model for Brain Tumour Classification," *Entropy*, vol. 24, no. 6, pp. 1–16, 2022.

[8] N. Noreen *et al.*, "A Deep Learning Model Based on Concatenation Approach for the Diagnosis of Brain Tumor," *IEEE Access*, vol. 8, pp. 55135–55144, 2020.

[9] M. A. Al Nasim *et al.*, "Brain Tumor Segmentation Using Enhanced U-Net Model with Empirical Analysis," *arXiv preprint arXiv:2210.13336*, 2022.

[10] Y. Guan *et al.*, "ResSGA-Net: A Deep Learning Approach for Enhanced Brain Tumor Classification," *Scientific Reports*, 2026.

[11] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[13] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proc. ICML*, 2019, pp. 6105–6114.

[14] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[15] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep Image Prior," *International Journal of Computer Vision*, vol. 128, no. 7, pp. 1867–1888, 2020