

Hybrid EfficientNet–Vision Transformer Model for Brain Tumor Detection Using MRI Images

Soumesh Acharya¹, Sumanjoy Nag¹, and *Bijoyini Bagchi²

Department of Computer Science and Engineering, Institute of Engineering
and Management, Kolkata, India

Soumeshacharya24@gmail.com, nagsumanjoy@gmail.com, *bijoyini.bagchi@iem.edu.in

Abstract:

Timely detection of deadly diseases such as brain tumor is of paramount importance. Mechanisms such as Magnetic Resonance Imaging (MRI), although considered standard and conventional involve tedious processes in interpreting and analyzing obtained results accurately. This paper, through the usage of upgraded versions of EfficientNet-B0 and Vision Transformer (ViT) models introduces an automated system of detection and segmentation of brain tumors including steps such as region of interest (ROI) extraction, preprocessing, normalization in turn vastly improving metric parameters.

Keywords:

EfficientNet-B0, Vision Transformer, Magnetic Resonance Imaging (MRI)

1 Introduction:

Convolutional Neural Networks (CNNs) specifically advanced models such as EfficientNet fail to capture broader contextual details although being fruitful in recognizing spatial image patterns, thus affecting precise detection of tumors. Introduction of Vision Transformers solved this problem however require extensive data and high computational power in order to analyze accurately. Hybrid architectures elevate optimization and utilization of resources in diagnosis by combining the respective strengths of localized extraction of features of the CNN in this case EfficientNet along with that of Vision Transformer which is global contextual modelling.

2. Implementation Framework:

2.1 Dataset Preparation:

A publicly accessible dataset from Kaggle of brain tumors containing 7,200 images. Out of this, roughly 62.2% is trained, 15.6% is validated and 22.2% is tested. Augmentation techniques include horizontal flipping, random rotation and scaling in order to elevate generalization of model.

2.2 Proposed Hybrid EfficientNet–Vision Transformer Architecture:

Normalized MRI images pass through **EfficientNet-B0**. Local spatial feature maps are computed by convolutional layers as:

$$Y_{ij} = \sigma(\sum_m \sum_n W_{mn} \cdot X_{i+m, j+n} + b)$$

The resulting feature maps are split into patches and the **Vision Transformer's self-attention** mechanism processes them, which models global contextual relationships as:

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k}) \cdot V$$

The overall representation that is extracted by combining both branches is represented as:

$$F = f(X_{\text{norm}}; \theta)$$

In the final classification stage, processing of the combined features takes place, which results in determination of predicted likelihood for each target category.

$$P(y_i) = e^{(z_i)} / \sum_j e^{(z_j)}$$

Minimization of the **categorical cross-entropy loss** leads to optimization of model parameters:

$$L = -\sum_i y_i \log(\hat{y}_i)$$

2.2 Experimental Setup :

The framework was implemented in TensorFlow/Keras and trained on Google Colab (T4 GPU). Inputs were standardized to 224 * 224 * 3 pixels. Optimization was performed using Adam (10^{-4} , batch size=32) with a 15 epoch early stopping patience.

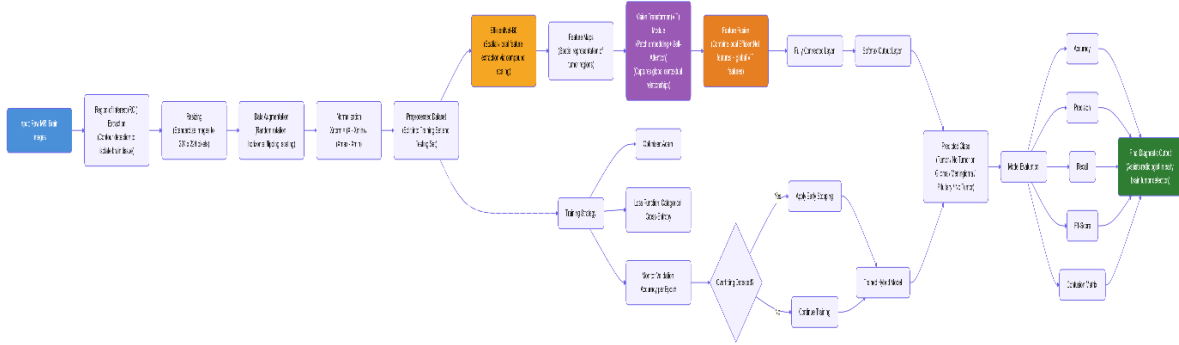


Fig. 1. Proposed workflow for Hybrid EfficientNet-Vision Transformer Architecture.

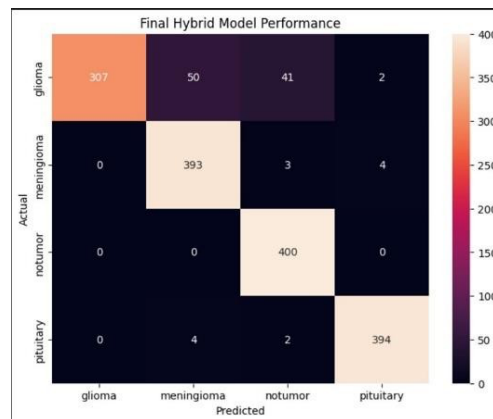


Fig. 2. Confusion Matrix of Brain Tumor Classification.

3. Results and Analysis:

The proposed architecture recorded an accuracy of 99.1%, precision of 99.3%, recall of 99%, and an F1-score of 99.1% which demonstrates high consistency, balanced performance and a robust capability of classification.

The confusion matrix is a proof of strong diagonal dominance: no tumor achieves 400/400 correct, meningioma records 393/400 correct and pituitary shows 394/400 correct with only minor classifications whereas glioma achieves only 307/400 correct making it hard to distinguish.

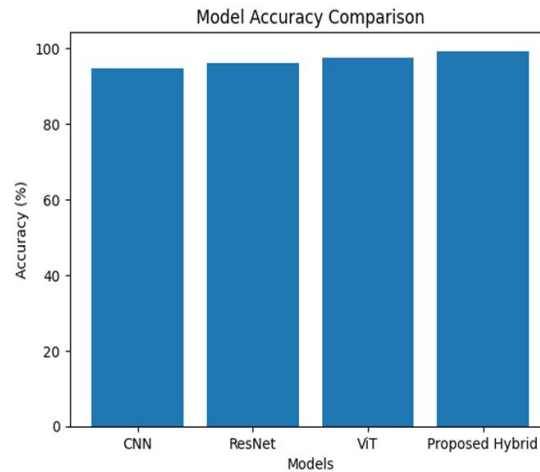


Fig. 3. Performance and Complexity Comparison

4. Conclusion:

This study showcases the efficiency of a hybridized EfficientNet-B0 and Vision Transformer (ViT) model to aid radiologists in less time consuming and more reliable tumor diagnosis having achieved 99.1% accuracy. Future work would involve reducing or diminishing the model's dependency on size of input dataset and minimizing higher computational demands in case of bulkier datasets.

References:

S. Alam and A. Sheneamer, "A fine-tuned vision transformer based enhanced multi-class brain tumor classification using MRI scan imagery," *Frontiers in Oncology*, 2024.

Biographies:

Soumesh Acharya is a B.Tech. student at the Institute of Engineering and Management, Kolkata, India, with research interests in machine learning and pattern recognition.

Sumanjoy Nag is a B.Tech. student at the Institute of Engineering and Management, Kolkata, India, interested in deep learning and computer vision.

Bijoyini Bagchi is an Assistant Professor at the Institute of Engineering and Management, Kolkata, India. Her research interests include artificial intelligence and intelligent healthcare systems.