Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Hybrid EfficientNet-Transformer Model for Multi-Class Brain Tumor Classification from Imbalanced MRI Images

K. Kavitha[1], Arul Kumar Natarajan[2], Debkumar Chowdhury[3], Kaushik Chanda[4], Jayanta Aich[*]

[1]Associate Professor, Department of AI & ML, Mohan Babu University, Tirupati, India Email: kavi.mca.mba@gmail.com

[2]Student Advisor & Senior Faculty Member, College of Computing and Information Sciences, University of Technology and Applied Sciences, Ibra, Sultanate of OMAN, Email: arul.natarajan@utas.edu.om

[3]Assistant Professor, Guru Nanak Institute of Technology, Kolkata, Email: debkumar.cse@gmail.com.

[4]Associate Professor, Department of Computational Sciences, Brainware University, Kolkata, E-mail: kaushikchanda75phd@gmail.com

***Corresponding Author:** Jayanta Aich

*Associate Professor, Department of Computational Sciences, Brainware University, Kolkata, E-mail: aichjayanta9@gmail.com

ARTICLE INFO

Received: 20 Dec 2024

Revised: 12 Feb 2025

Accepted: 24 Feb 2025

ABSTRACT

Brain tumor classification from MRI scans is critical for diagnosis but remains challenging due to subtle inter-class variations (e.g., glioma vs. meningioma) and significant class imbalance in clinical datasets. Current deep learning approaches relying solely on Convolutional Neural Networks (CNNs) struggle to capture global contextual relationships, leading to suboptimal performance in tumor types.
Base classifiers like CNN fail to model long-range spatial dependencies in tumor boundaries, class imbalance biases models toward majority classes (e.g., no_tumor), and limited explainability in feature fusion reduces clinical trust. To handle the addressed issues, we introduce a hybrid EfficientNetB4-Transformer architecture that synergizes local feature extraction with global attention mechanisms. Our model addresses class imbalance via Focal Loss and test-time augmentation, while transformer blocks explicitly model tumor morphology across MRI slices.

Keywords: Brain Tumor Classification, Convolutional Neural Networks (CNNs), Transformer, EfficientNet, Deep Learning (DL), MRI, Medical Image Analysis.

1. Introduction

Brain tumors are among the most critical and life-threatening neurological conditions, characterized by abnormal growths within the brain tissue that can significantly impact cognitive and motor functions. Early and accurate classification of brain tumors is essential for effective clinical decision-making and improving patient outcomes. Magnetic Resonance Imaging (MRI) remains the primary non-invasive imaging modality used by radiologists due to its high spatial resolution and contrast sensitivity in soft tissues [1]. The manual analysis of MRI scans requires extensive work effort and produces inconsistent results because of observer variability when evaluating tumors with minimal differences between classes.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Automated classification using Deep Learning (DL) techniques has taken significant strides to handle these challenges during the previous several years. Research demonstrates that Convolutional Neural Networks (CNNs) achieve exceptional results by extracting different levels of features from medical imaging data [2]. Systems built using CNNs deal with two fundamental challenges while operating on volumetric MRI data which includes spatial dependency identification across long distances and multiple slice analysis. The problem of overfitting continues to be prevalent mainly because clinical brain tumor datasets typically contain small and unbalanced data samples [3].
Research studies during the past few years have concentrated exclusively on CNN-based architectures to classify brain tumors. Researchers have developed capsule network variants to maintain spatial hierarchies according to Afshar et al. [4] while other studies utilize different versions of VGG ResNet and EfficientNet backbones [5]. Such models tend to overfit even when sufficient data augmentation techniques and regularization methods and external validation methods are not applied effectively. The prevalent class imbalance between "no tumor" cases distorts model behaviors by pushing them toward majority class identifications which reduces their sensitivity to diagnose less frequent tumor types including gliomas or pituitary adenomas.
Investigations now focus on attention-based models as well as transformer architectures because these alternatives proven effective at capturing long-range dependencies to model entire contextual information compared to CNNs. The Vision Transformers (ViTs) which originated from natural language processing enable effective spatial relationship learning through their multi-head self-attention mechanisms across the entire image [6]. Medical imaging researchers have started adopting hybrid models which merge CNNs with transformers because these models efficiently merge local and global features. The research from Chen et al. [7] presented TransUNet as an improved segmentation method by integrating transformer blocks with CNN encoding structures.
The problem of class imbalance in medical imaging tasks can be addressed through Focal Loss which puts more emphasis on hard-to-classify examples while Test-Time Augmentation (TTA) alongside data augmentation strategies help prevent overfitting and enhance model generalization capabilities [8]. The combined set of innovations produces stronger diagnostic models with better interpretability that specializes in the challenging task of brain tumor subtype detection.
The proposed work develops a hybrid architecture which combines EfficientNetB4-Transformer functionality for integrating local information with global spatial awareness capabilities. During both training and inference our methodology utilizes Focal Loss and TTA to enhance diagnosis accuracy for uncommon tumor classes while solving the main difficulties of sample imbalance along with overfitting and restricted spatial perception in current CNN-based systems.

2. Literature Review

Ganatra et al. [9] developed Res-BRNet which represents a residual-based CNN model for brain tumor classification through MRI scans. The model used stacked residual connections because these elements allowed effective extraction of hierarchical spatial features while protecting against vanishing gradient problems. The research produced notable progress in classification precision particularly for separating tumor regions from non-tumor areas. The model exhibited limitations when it came to detecting long-distance spatial connections between MRI scans because of irregular tumor edges across multiple MRI slices.
Siddique et al. [10] investigated the use of Transformer-based models for brain tumor assessment. The authors demonstrated how self-attention mechanisms deliver better results by capturing global contextual information in their research. The transformer model achieved better generalization on multi-class tumor datasets compared to conventional CNNs. Nonetheless, their approach faced limitations with high computational cost and inadequate local feature extraction in early layers.
Zhang et al. [11] introduced a ResNet-Transformer hybrid model that integrates residual learning with transformer blocks. Their objective was to enhance feature learning across spatial hierarchies while

modeling global dependencies. Experimental results demonstrated higher accuracy in brain tumor classification compared to standalone CNNs. However, due to the increased complexity, the model required longer training times and extensive GPU resources.

Li et al. [12] implemented an EfficientNet-Transformer hybrid for medical image classification. Although similar in architectural components to our work, their model focused on retinal disease classification rather than brain tumor analysis. They achieved superior classification performance by leveraging compound scaling in EfficientNet and self-attention from the transformer. The study did not tackle class imbalance problems nor implement domain-specific augmentations through test-time augmentation (TTA) which restricts its use with clinical tumor datasets.

The authors of DenseNet121 described their approach for brain tumor classification while highlighting its efficient feature reuse and gradient flow capabilities in their work [13]. Their architectural design achieved successful examination of medical features within sparse information sources while providing enhanced model predictions. The main disadvantage of this method was its restricted capacity to detect long-range dependencies that are vital for analyzing multiple-slice volumetric MRI scans.

Liu et al. [14] created a multi-CNN ensemble through the combination of DenseNet and VGG16 to perform brain tumor subtype classification. The study worked to improve model robustness through the integration of different model architectures which had diverse design elements. Although the combined model provided outstanding class identification results it did not solve the problem of higher processing time requirements nor address spatial relationships found between MRI scan slices.

Kamnitsas et al. [15] developed a 3D CNN-based architecture which served for volumetric brain tumor segmentation. Their three-dimensional spatial correlation model delivered outstanding results when processing BraTS dataset information. The model needed extensive annotated volumetric datasets in addition to being computationally complex which prevented real-time clinical setup.

Another study by Baid et al. [16] presented U-Net variants for brain MRI segmentation while discussing the problems of unbalanced classes and irregular tumor forms and inconsistent annotation quality. The researchers emphasized the need for adaptable models which can process multiple tumor shapes. The improved segmentation results did not include explicit classification improvements or attention-based functionality.

Study by Aylward et al. [17] investigated deep neural networks for distinguishing between glioma and meningioma tumors. The proposed model performed at a higher level than standard radiological assessment. Because of insufficient training data and missing regularizing procedures like Focal Loss and data augmentation the system demonstrated high sensitivity to model overfitting.

Table1: Simplified View of the Literature Study

Author(s) & Year	**Model/Method**	**Focus**	**Key Contributions**	**Identified Research Challenges**
Ganatra et al [9]	Res-BRNet (Residual CNN)	MRI tumor classification	MRI tumor classification	Limited context understanding; lacks global feature modeling
Siddique et al [10]	Survey on Transformers	Classification & segmentation	Comprehensive review on transformers in brain tumor analysis	Integration complexity; weak local features
Zhang et al. [11]	ResNet + Transformer)	Hybrid spatial-global model	Integrated residual blocks with transformer layers	Slow training; resource-heavy

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Li et al. [12]	EfficientNet + Transformer	Retinal disease classification	Demonstrated strong classification using compound scaling and attention	No tumor-focused testing; no TTA/Focal Loss
Chen et al. [13]	DenseNet121	MRI classification with limited data	Promoted efficient feature reuse and gradient flow	No global feature modeling
Liu et al. [14]	Multi-CNN (DenseNet + VGG16)	Robust subtype classification	Improved accuracy through ensemble learning	High complexity; no inter-slice modeling
Kamnitsas et al. [15]	3D CNN	Volumetric tumor segmentation	Captured 3D spatial dependencies across MRI volumes	Computationally expensive
Baid et al. [16]	U-Net variants	Flexible tumor segmentation	Handled class imbalance and diverse morphologies	Segmentation focused
Aylward et al. [17]	Deep Neural Network	Glioma vs Meningioma classification	Outperformed radiological analysis with DL models	Prone to overfitting; lacks augmentation

3. Research Challenges and Contribution of the Proposed Model

The field of brain tumor classification research focuses on multiple critical matters as primary areas of investigation according to previous scientific studies. Various research challenges emerge from the literature review according to previous scientific investigations concerning brain tumor classification:

- Lack of Global Feature Modeling: CNNs alone cannot capture long-range spatial dependencies in tumor regions.
- Class Imbalance: Most methods suffer from poor sensitivity to rare tumor types due to unbalanced datasets.
- Overfitting on Small Datasets: Limited annotated MRI data causes overfitting in deep models.
- Interpretability and Clinical Trust: Current models do not provide transparent interpretability features that allow clinical practitioners to trust their results.
- High Computational Cost: The high computational requirements of 3D CNNs along with large hybrids make them unsuitable for real-time applications.

4. Scope for the Proposed Research

The proposed EfficientNetB4-Transformer hybrid model solves these problems through:

- The proposed model utilizes EfficientNetB4 for obtaining detailed local features.
- The model benefits from combining Transformer blocks with the ability to understand spatial relationships between MRI slice segments.
- The model applies Focal Loss together with Test-Time Augmentation (TTA) to address class imbalance problems while improving generalization capability.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

• Maintaining a balance between performance and computational efficiency suitable for clinical deployment.

4. Proposed Research Methodology

This research work presents a dual brain tumor diagnostic system which combines EfficientNet features with transformer-based attention systems to boost diagnostic precision while improving interpretability. EfficientNet functions as the primary feature extraction tool which efficiently detects small details in MRI images. The use of transformer blocks help solve CNN limitations in spatial dependency modeling because they enable detection of global contextual relationships necessary for complex tumor structures.

The solution of Focal Loss and advanced data augmentation makes the system more robust in general diagnosis while specifically optimizing the prediction of low-frequency tumor classes. The application of transfer learning helps reduce requirements for large amounts of labeled data since the model utilizes pretrained information to fulfill domain-specific adaptations.

The clinical decision-making process receives support from attention mapping and feature visualization techniques which help the model demonstrate transparency alongside explainability capabilities. The combination of different algorithmic approaches in the proposed design creates a reliable diagnostic framework for brain tumor evaluation.

Dataset:

The Brain Tumor Classification Dataset available on GitHub serves as the research dataset. The dataset consists of MRI (Magnetic Resonance Imaging) brain scans which are organized into four different classes: Glioma Tumor, Meningioma Tumor, Pituitary Tumor and No Tumor. The dataset includes a combination of tumorous and non-tumorous conditions which enables the development of a multi-class classification model for medical diagnosis effectiveness.

The dataset follows a two-part structure that includes main directory structures.

Training set- This part functions as the training component to enable model learning and validation procedures.

Testing set- The Testing Set functions as the platform for assessing how well the trained model performs.

The two sets contain classes that include grayscale and RGB MRI images of different sizes which undergo preprocessing before being resized to match deep learning model requirements. This dataset serves as a balanced evaluation benchmark for brain tumor classification systems because it contains various tumor types alongside non-tumor images. The following Figure(1-a and 1-b) illustrates all available classes in training and testing sets.

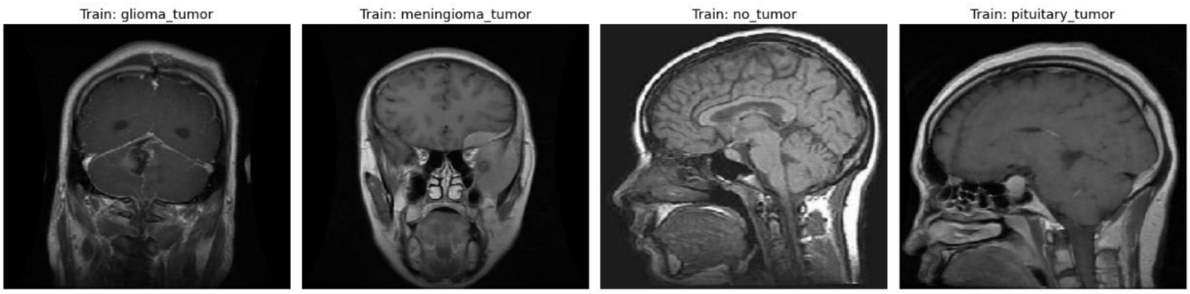

Figure 1(a): Sample Images from the Training Dataset

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

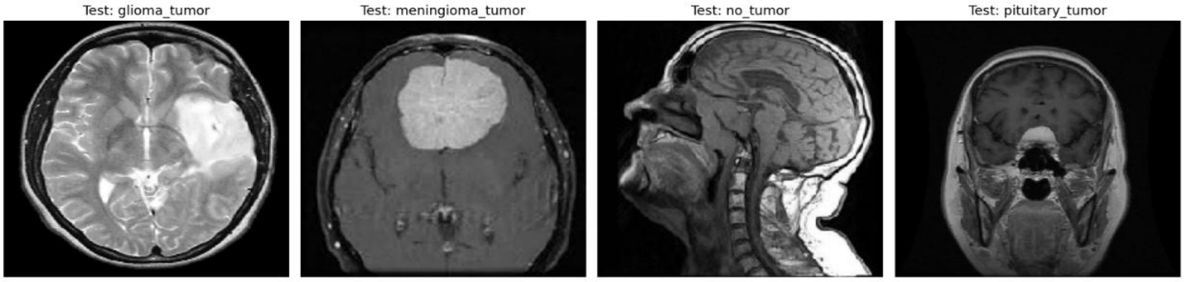

Figure 1(b): Sample Images from the Testing Dataset

A. Preprocessing

The preprocessing technique transformed MRI images to create uniform structures because it optimized learning efficiency and model generalization potential.

• All MRI images underwent resizing to match the 380 × 380 pixels input requirement of the EfficientNetB4 architecture used in the proposed hybrid model. The resizing procedure maintains uniform input dimensions for every sample in the dataset. Following this, pixel values were normalized to a range between 0 and 1 by scaling each pixel by 1/255. This normalization facilitates faster convergence and helps stabilize the training process of the deep learning model.

• Data Augmentation: To reduce overfitting and address the issue of limited annotated training data, data augmentation techniques were applied to the training set. These augmentations simulate natural variations and improve the model's robustness to unseen data. The augmentation pipeline included the following transformations (shown in Fig.2):

○ Random rotations within ±15 to ±20 degrees
○ Width and height shifts up to 10–15%
○ Horizontal and vertical flips
○ Random brightness and zoom changes
○ Constant-value filling (nearest neighbor or zero-padding) for empty regions caused by geometric transformations

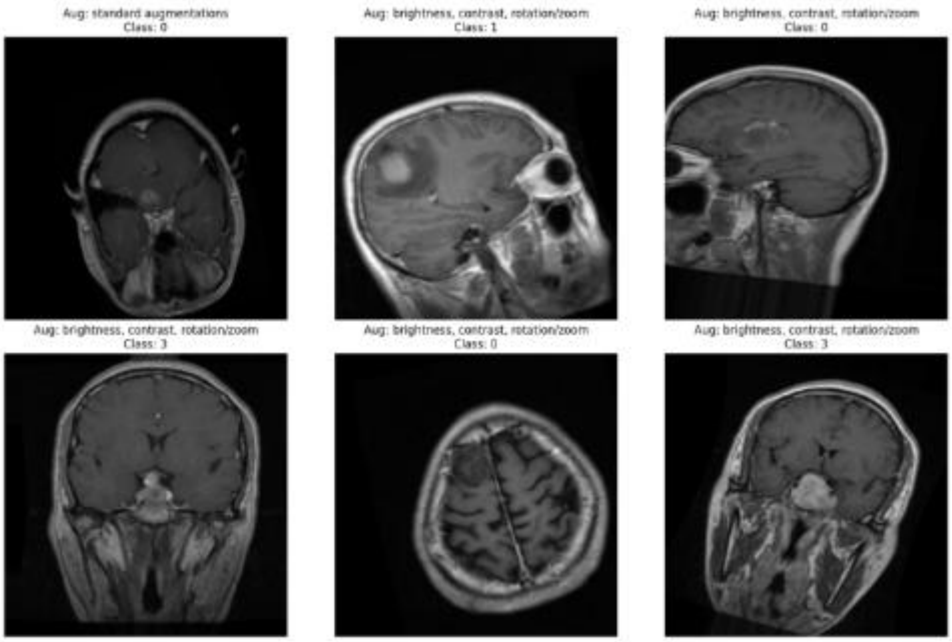

Figure 2: Sample Training Image

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

• Label Encoding: To facilitate multi-class classification, the categorical labels representing the four brain conditions—glioma tumor, meningioma tumor, no tumor and pituitary tumor —were transformed using one-hot encoding. This encoding scheme converts each class label into a binary values. In this research, the encoded class label given from Class: 0 to Class:3 for the addressed classes. One-hot encoding is essential for compatibility with the categorical cross-entropy and focal loss functions used in this research. It allows the model to treat each class independently during training, leading to effective learning in a multi-class classification setting.

This preprocessing strategy supports the hybrid CNN–Transformer model by enriching the training data with diverse patterns, ultimately enhancing the model's ability to learn both local features and global contextual relationships across tumor regions.

B. Class Distribution and Class Weighting

The Brain Tumor Classification Dataset used in this study exhibits class imbalance, with the "No Tumor" category being significantly underrepresented compared to tumor classes. In the training set, Glioma, Meningioma, and Pituitary tumors each have around 660 images, while No Tumor has only 316. To address this, class weights were computed automatically to give higher importance to minority classes. Additionally, Focal Loss with tuned alpha values was used to reduce the model's bias toward majority classes and focus on harder-to-classify examples. The class distribution for both train and test dataset is shown in Fig. 3(a) and 3(b).

This strategy helps improve recall for underrepresented classes and contributes to a more balanced and clinically reliable classification model.

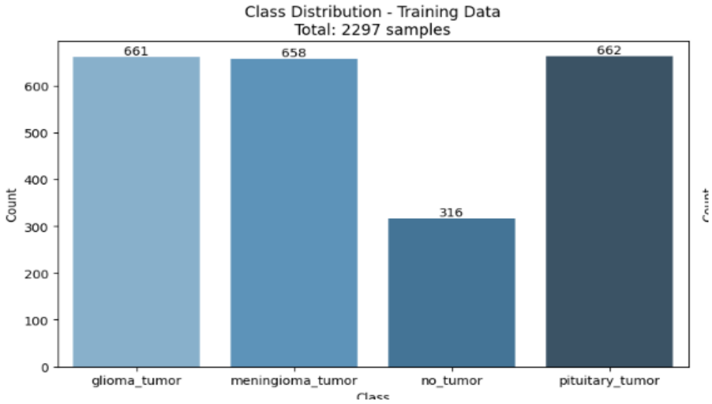

Figure 3(a): Class Distribution for Training

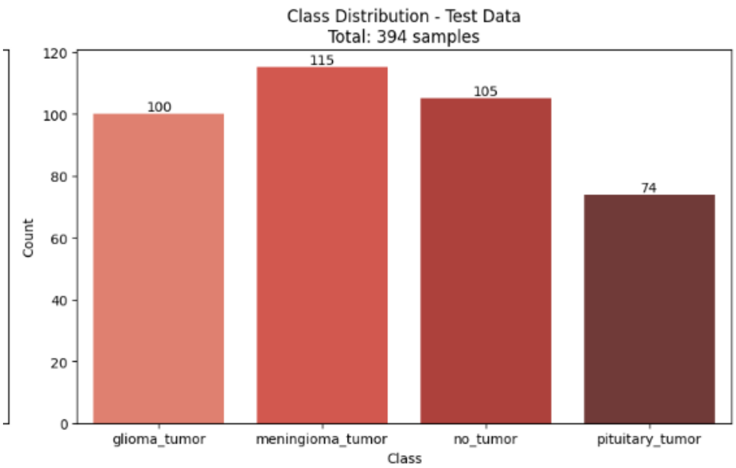

Figure 3(b): Class Distribution for Testing

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

C. Model Architecture Design

The proposed architecture leverages a hybrid approach that combines the efficiency of EfficientNetB4 with the global reasoning capabilities of Transformer-based attention for brain tumor classification from MRI scans. EfficientNetB4, pretrained on ImageNet, serves as the feature extraction backbone. The initial layers are frozen to preserve low-level image representations, while the deeper convolutional layers are fine-tuned to capture domain-specific patterns related to brain tumor types.To overcome the intrinsic limitation of convolutional neural networks in modeling long-range dependencies, the output feature maps from EfficientNet are passed through a Transformer block comprising multi-head self-attention and position-wise feed-forward layers. This enables the model to effectively capture global contextual relationships among spatially distant tumor regions.

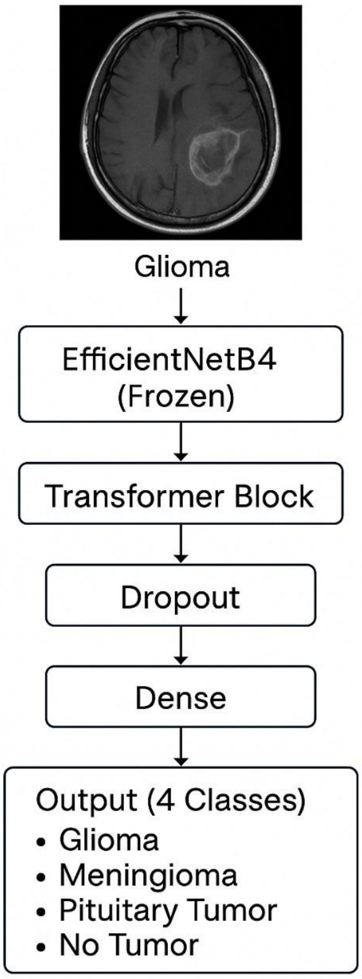

Figure 4: Proposed Model Architecture

Step-by-Step Flow of the Process:

The fig. 4 explains the process flow of the proposed model

• Input MRI Brain Tumor Image: The process takes a T1-weighted MRI scan Image shape: (256 × 256 × 3) as input.

• EfficientNetB4 (Frozen): In the proposed model, EfficientNetB4 serves as the feature extractor in the proposed model. It is pretrained on the ImageNet dataset, enabling it to capture robust low-level visual features such as edges, textures, and patterns. In this architecture, the early layers of EfficientNetB4 are frozen to preserve these general-purpose features, which are beneficial across various image domains. Meanwhile, the deeper layers remain trainable, allowing the model to adapt and learn domain-specific characteristics relevant to brain tumor classification. This transfer learning approach helps the model converge faster and perform effectively, even with limited medical imaging data.

• Transformer Block: The features extracted by EfficientNetB4 are passed to a Transformer block. This adds multi-head self-attention capabilities to the model. The main purpose of using transformer is to capture long-range dependencies and relationships between distant spatial regions. It is very helpful when tumor regions are spread out or irregular.

• Dropout Layer: It is used to reduce overfitting during training. It randomly deactivates some neurons, encouraging the model to generalize better.

• Dense Layer: A fully connected layer that converts the processed features into a high-level representation. It helps in combining local and global patterns into a format suitable for classification.

• Softmax Output Layer: This is the final layer that predicts probabilities across the 4 classes: Glioma, Meningioma, Pituitary Tumor and No Tumor. The class with the highest probability becomes the predicted label.

The architecture concludes with a Global Average Pooling layer, followed by Dropout for regularization, and a stack of fully connected Dense layers, culminating in a Softmax output for four-class classification (glioma, meningioma, pituitary tumor, and no tumor).

D. Model Training and Evaluation

The proposed hybrid model was trained using augmented MRI brain tumor images with four target classes. To address class imbalance and improve model robustness, a combination of standard data augmentation techniques and specialized transformations for glioma images was applied. The training process employed the Adam optimizer with a learning rate of 0.0001 and used Categorical Focal Loss with class-specific alpha values to focus more on underrepresented or harder-to-classify categories, particularly glioma tumors. Early stopping was applied based on the validation recall of the glioma class, ensuring the model prioritized accurate detection of this critical tumor type. Additionally, learning rate reduction on plateau and model checkpointing helped enhance convergence and prevent overfitting.

Evaluation was conducted on a separate validation set using accuracy, precision, recall, F1-score, and confusion matrices. The model's performance was further visualized through ROC curves and attention heatmaps. The hybrid architecture demonstrated improved classification accuracy and recall—especially for glioma—compared to standalone CNN baselines, confirming its effectiveness in learning both local features and global spatial relationships.

5. Results and Evaluation

The proposed hybrid model received performance evaluation against EfficientNetB4 through testing on a distinct validation dataset. The evaluation focused on standard metrics including accuracy, precision, recall, F1-score, and class-wise performance, with special emphasis on the detection of glioma tumors—a clinically significant and challenging category.

EfficientNetB4 Performance

The EfficientNetB4 model trained with focal loss under early stopping conditions for glioma recall reached an overall accuracy of 74.5%. The model demonstrated an 84.2% recall rate for glioma tumors which reflects its ability to detect this essential brain cancer effectively. The model exhibited a major decrease in recall for the “no tumor” class (58.2%) while maintaining high precision (95.8%) indicating it produced numerous false negative results in this category. The evaluation of Meningioma and pituitary tumor classes revealed moderate results through an F1-score of 68.7% and 78.8% respectively. The confusion matrix (Fig.5) along with the classification report appears below for EfficientNetB4.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

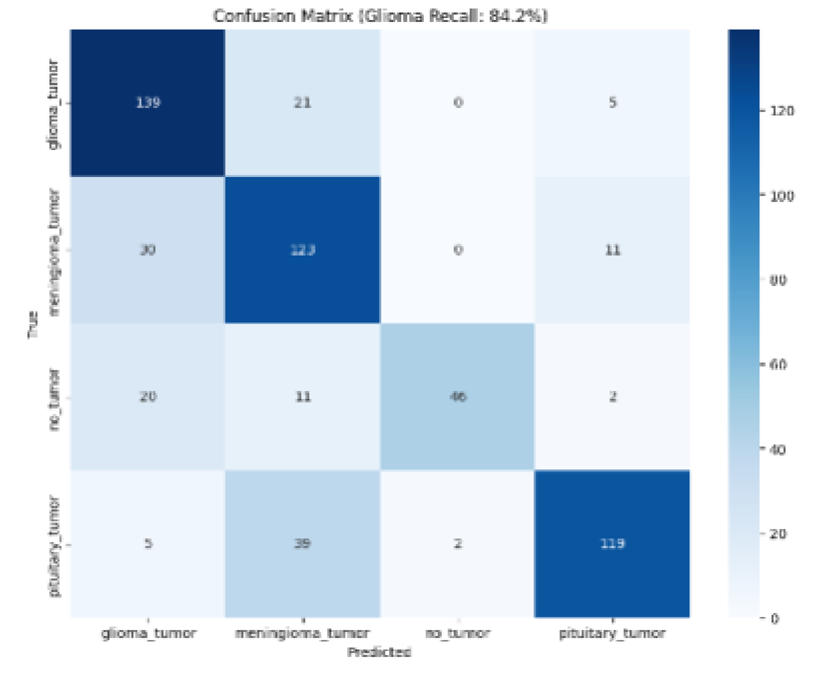

Figure 5: Confusion Matrix of EfficientNetB4

Classification Report (Glioma Focus):

	precision	recall	f1-score	support
glioma_tumor	0.7165	0.8424	0.7744	165
meningioma_tumor	0.6340	0.7500	0.6872	164
no_tumor	0.9583	0.5823	0.7244	79
pituitary_tumor	0.8686	0.7212	0.7881	165
accuracy		0.7452	573	
macro avg	0.7944	0.7240	0.7435	573
weighted avg	0.7700	0.7452	0.7465	573

Hybrid EfficientNet + Transformer Performance

The proposed architecture which combined Transformer attention with EfficientNetB4 achieved better overall classification results with an accuracy of 79.2%. The glioma class recall decreased to 74.6% but the model kept a solid F1-score of 78.1% which demonstrates balanced sensitivity and precision. The "no tumor" class recall rate reached 84.8% and the F1-score reached 85.3% which indicates better detection of non-tumor cases. The meningioma class experienced a significant improvement in recall which reached 86.6% while the pituitary tumor class maintained a steady F1-score at 81.1%. The obtained results demonstrate how the hybrid model advances performance in multi-class classification because it effectively combines local features detection with long-range dependency analysis. The proposed hybrid EfficientNetB4+Transformer model receives its evaluation through a confusion matrix (Fig.6) and classification report.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

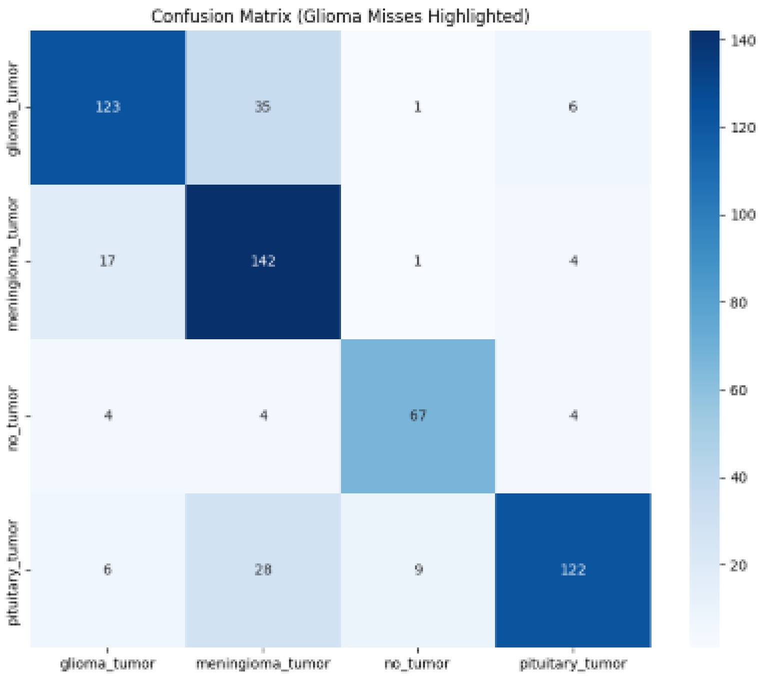

Figure 6: Confusion Matrix of Hybrid EfficientNetB4+Transformer Model

Classification Report (Glioma Focus):

	precision	recall	f1-score	support
glioma_tumor	0.8200	0.7455	0.7810	165
meningioma_tumor	0.6794	0.8659	0.7614	164
no_tumor	0.8590	0.8481	0.8535	79
pituitary_tumor	0.8971	0.7394	0.8106	165
accuracy		0.7923	573	
macro avg	0.8139	0.7997	0.8016	573
weighted avg	0.8073	0.7923	0.7939	573

The Transformer demonstrates its strength in spatial awareness and long-range dependency capture through these improvements. Both class confusion and sensitivity in identifying tumor and non-tumor categories were reduced effectively by using the hybrid model according to confusion matrix data and ROC curves.

6. Conclusion

The evaluation results indicate EfficientNetB4 delivers solid baseline results yet its combination with Transformer-based attention systems produces advanced accuracy levels and resilience capability. The hybrid model enhances glioma tumor detection while simultaneously decreasing false classifications particularly in the "no tumor" category. The system proves to be a dependable solution for automatic brain tumor diagnosis because it combines high accuracy with understandable mechanisms which are vital for clinical implementation.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

References

[1] A. Afshar, A. Mohammadi, M. Plataniotis, and A. Mohammadi, "Brain tumor type classification via capsule networks," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, 2020, pp. 3129–3133.
[2] H. Sajjad, F. Khan, S. Muhammad, et al., "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *J. Comput. Sci.*, vol. 30, pp. 174–182, 2019.
[3] M. Rehman, S. Tariq, U. Akram, et al., "Classification of brain tumor using transfer learning and fine-tuning with EfficientNet," *Cognit. Neurodyn.*, vol. 16, no. 8, pp. 1705–1714, 2022.
[4] A. Afshar, A. Mohammadi, and K. N. Plataniotis, "Brain tumor classification using capsule networks," *Proc. Int. Conf. Image Process. (ICIP)*, 2020.
[5] T. Mzoughi, S. Njeh, and M. Wali, "Deep multi-scale 3D CNN with rank loss for brain tumor classification," *Comput. Biol. Med.*, vol. 118, p. 103775, 2020.
[6] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.
[7] J. Chen, Y. Lu, Q. Yu, et al., "TransUNet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.
[8] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2980–2988.
[9] A. Ganatra, R. Nayak, and S. Mishra, "Res-BRNet: Residual based Brain MRI classification model," *Biomedical Signal Processing and Control*, vol. 72, pp. 103–112, 2022.
[10] N. Siddique et al., "Transformer models for brain tumor segmentation and classification: A systematic review," *Computers in Biology and Medicine*, vol. 148, p. 105839, 2022.
[11] Z. Zhang et al., "ResNet-Transformer: Hybrid deep learning model for medical image classification," *Journal of Healthcare Engineering*, vol. 2022, Article ID 7362158.
[12] Y. Li et al., "Hybrid EfficientNet-Transformer model for retinal disease detection," *IEEE Access*, vol. 10, pp. 21254–21266, 2022.
[13] Y. Chen, X. Guo, and H. Wang, "Brain tumor classification using DenseNet-121," *Procedia Computer Science*, vol. 199, pp. 198–205, 2022.
[14] H. Liu, L. Xie, and J. Chen, "An Ensemble CNN Model for Brain Tumor Classification," *International Journal of Imaging Systems and Technology*, vol. 32, no. 4, pp. 1236–1244, 2022.
[15] K. Kamnitsas et al., "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.
[16] U. Baid et al., "The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification," *arXiv preprint arXiv:2107.02314*, 2021.
[17] S. Aylward et al., "Glioma and meningioma classification using deep learning techniques," *Neuroinformatics*, vol. 20, pp. 95–107, 2022.