A Global-Local Interaction Modeling Network with Adaptive Feature Optimization for Brain Tumor Classification Using MRI

Subia Salma
Department of Computer Science and Engineering, SVCE, Bengaluru, India
subiasalma89@gmail.com (corresponding author)

S. C. Lingareddy
Technology Transfer Department, SVCE, Bengaluru, India
sclingareddy@gmail.com

S. Shilpashree
Department of Computer Science and Engineering, NCET, Bengaluru, India
shilpatani@gmail.com

Vineet Kumar
VVIT, Bengaluru, India
vineetkumarvk97@gmail.com

Received: 20 March 2026 | Revised: 20 April 2026 | Accepted: 1 May 2026

 DOI: https://doi.org/10.48084/etasr.18843

ABSTRACT

Reliable classification of brain tumors remains challenging for computer-aided diagnosis, since each tumor type can look very different, only sparse data exist, and images and genomic profiles contain noise. This study presents an efficient deep learning framework that mixes three components: a dimensional transformation, an attention step, and regularized learning. First, a method turns high-dimensional sparse imaging feature intensity variation pattern records into structured 2D RGB images. This step allows ordinary convolutional and attention networks to process the data, while intensity variation patterns that matter biologically stay intact. To reduce noise and sparseness, an Adaptive Feature Optimization Algorithm (AFOA) applies clustering-based imaging feature selection, keeps informative candidates, and drives tumor-aware multi-channel MRI features, raising feature quality and stabilizing training. On the cleaned data, a global-local interaction modeling network uses a Tumor-Aware Feature Refinement Block (TAFRB) and multi-head self-attention to learn local intensity pattern variation details or global context with reduced computation. Label smoothing regularization addresses overfitting when data are scarce and widens gaps between classes. Tests on public brain tumor datasets show that the proposed framework outperforms current CNNs with transformer models in accuracy, precision, recall, and F1-score, and generalizes well across dataset sizes. The results show that the proposed approach offers a robust, scalable, and fast route for brain tumor classification and can support clinical decisions.

Keywords-brain tumor classification; intensity pattern variation; Adaptive Feature Optimization Algorithm (AFOA); global-local interaction modelling network; multi-head self-attention; label smoothing regularisation; computer-aided diagnosis

I. INTRODUCTION

Brain tumors are a challenging and severe kind of cancer. Deep Learning (DL) software that can automatically classify tumors can help staff and greatly shorten the task [1]. The National Brain Tumor Society estimates that more than 94,000 people are diagnosed with a primary brain tumor yearly [2]. Around one million people in the United States already live with a brain tumor. Tumors that grow slowly but also stay in the brain are categorized as benign and grade I or II. Malignant tumors vary widely in shape, contrast, and growth behavior, which complicates automated classification. Compact Convolutional Neural Networks (CNNs) can identify metastatic tissue with high reliability, and learned hierarchical features clearly outperform hand-crafted descriptors when tumor appearance is heterogeneous [3]. This finding directly

supports the deep feature-learning strategy for MRI-based brain tumor classification adopted in this work, where inter-class visual heterogeneity is equally pronounced. Brain tumors form a major group of neurological diseases that evade diagnosis because each mass differs in form or biological detail. Surveys show a rise in new cases and, among adults, gliomas, meningiomas, and pituitary tumors appear more frequently. Doctors must quickly determine the exact type of tumor to choose the right therapy, stop progression, and improve the odds of survival. However, since tumor subtypes look similar in images, their brightness patterns overlap, and their size and site differ, standard clinical methods struggle in labeling [4].

Magnetic Resonance Imaging (MRI) is the routine method for imaging the brain, as it shows soft tissue better than other tests. Manual diagnosis using these scans depends on the radiologist's skill and varies between observers, especially in hospitals. As the volume of neuroimages increases, there is a greater need for automated computer-aided systems to diagnose tumors accurately and with low delay, helping physicians make better and more reliable decisions [5]. In recent years, DL algorithms have shown strong results in tumor diagnosis. CNNs and their variants surpass traditional ML techniques, but remain inefficient in daily use because they ignore wide context, react to noise, and demand heavy computation [6]. Transformer-based or hybrid designs can overcome some limits, capturing long-range links, but increase model size and complicate use in clinics with low hardware resources. Generalization across datasets, tolerance of sparse data, and steady performance remain open-ended problems [7].

Meningiomas and gliomas dominate among adults; gliomas include astrocytomas, oligodendrogliomas, and ependymomas. The fifth edition of the WHO central nervous system guide lists 15 subtypes of meningioma, most of which are harmless grade I lesions. Pituitary adenomas are usually benign, slow-growing masses that arise within the gland. Pituitary tumors are located near the gland, and meningiomas attach to the skull vault. Early detection is important, but difficult. Specialists rely on physical checks and neurologic exams, but the lesions' complexity hampers certainty [1].

AI-based non-invasive tools process MRI or CT data to suggest a diagnosis. Biopsy, as well as pathology, can be used to confirm the result. MRI avoids radiation and surgery while giving fine anatomic details. Reading these images is demanding, and classification is difficult or slow when done by eye. Hospitals now adopt semi-automatic software because radiologists need relief from labor-intensive labeling. DL for brain tumor classification has advanced quickly; early work used Support Vector Machines along with Random Forests, but handcrafted features limited accuracy. CNNs and their descendants, ResNet, DenseNet, Inception, and others, learn hierarchical features directly from slices, improving performance. VGG16 and VGG19 locate details well, but at a high computational cost. Later approaches added residual and inception blocks to address vanishing gradients. Recent hybrids blend vision transformers with CNNs, mixing global context and local detail. Variability between patients and complex tissue texture limit final precision [8, 9].

This study addresses three outstanding gaps in tumor classification: first, extending current models to full 3D MRI stacks and to high-resolution data; second, settling on uniform benchmarks that include speed as well as cross-protocol generalization relevant to real-time clinics; third, proving robustness against domain shift and giving clinicians interpretable evidence. To address these gaps, this paper proposes an end-to-end framework whose contributions can be summarized as:

- Novel Intensity Pattern Variation (IPV)-to-image dimensional transformation framework: This work introduces a structured dimensional data transformation strategy that converts high-dimensional and sparse gene IPV profiles into compact RGB-based 2D representations, enabling effective exploitation of DL models while preserving biologically relevant IPV patterns across tumor cohorts.
- Adaptive Feature Optimization Algorithm (AFOA) for noise and sparsity reduction: A clustering-driven AFOA module is proposed to eliminate noisy and low-informative passenger tumor-aware multi-channel MRI features, selectively retaining candidate and potential driver ones. This strategy significantly reduces data sparsity and redundancy, enhances feature relevance, and improves overall training stability and classification performance.
- Global-local interaction modeling network-based DL architecture: The proposed network integrates a Tumour-Aware Feature Refinement Block (TAFRB) with a multi-head self-attention mechanism to jointly capture local IPV patterns and global contextual dependencies, achieving improved class separability with reduced computational complexity compared to conventional transformer-heavy models.
- Regularization-enhanced learning with label smoothing: A label-smoothing-based regularization scheme is incorporated to mitigate overfitting under limited-data conditions, improve inter-class margin separation, and enhance generalization capability, leading to more robust and consistent classification performance across different tumor cohorts.

Recent studies have advanced MRI-based brain tumor analysis in several complementary directions. In [10], a decentralized attention-embedded CNN combined inception, attention, and residual links, reporting 98.70 % accuracy across four MR classes. TTG-U-Net [11] is a cross-modal transformer that balances accuracy, interpretability, and speed through a low-rank tensor module and a modality-adaptive gate. In [12], an anatomy-induced region decoupler learned each tumor sub-region independently across MRI images. In [13], residual U-Net blocks, attention layers, and a feature pyramid with a weighted focal loss were combined to address class imbalance. In [14], a hybrid explainable-AI pipeline was based on an improved rolling guidance filter, a modified dilated U-Net, a swin patch transformer, an oppositional optimizer, and a capsule-routing autoencoder. In [14], a hybrid explainable-AI pipeline was based on an Improved Rolling Guidance Filter, a Modified Dilated U-Net, a Swin patch transformer, an

oppositional optimizer, and a capsule-routing autoencoder. In [15], a frequency-domain encoder with Fourier-based edge enhancement was introduced. Multimodal approaches, such as [16], fuse diffusion-weighted and T2-weighted MRI through inter-region attention with a trainable high-frequency retention filter. HSViT [17] is a lightweight hierarchical transformer that couples a pre-encoder, cross-layer attention refinement, and a progressive token filter with a tumor context module.

In contrast to transformer-heavy encoder-decoder stacks [11-14], the proposed GLIM-Net uses a light-weight inverted-residual TAFRB backbone combined with a single multi-head self-attention block that reduces parameter count without sacrificing accuracy. Unlike [16, 17], the AFOA step explicitly removes noisy, low-informative features before training, stabilizing learning in small cohorts. Unlike [10], the proposed framework is designed to work with limited data through label smoothing and does not require a decentralized training protocol. These design choices yield a consistent gain over [10, 16, 17] on three public datasets.

II. PROPOSED METHODOLOGY

This study presents an effective and robust framework for brain tumor classification by transforming high-dimensional and sparse IPV data into structured RGB-based 2D representations suitable for DL. An AFOA is employed to eliminate noise and retain informative driver tumor-aware multi-channel MRI features, reducing sparsity and improving feature quality. The filtered representations are then processed using a global-local interaction modeling network that combines tumor-aware feature refinement learning. The system uses multi-head self-attention to record local and long-range links in the data. Label smoothing regularization lowers the risk of overfitting and raises the ability to generalize, giving a classifier that runs in an even way with low cost. Figure 1 presents the proposed model architecture, while Figure 2 presents a simplified pipeline flowchart.

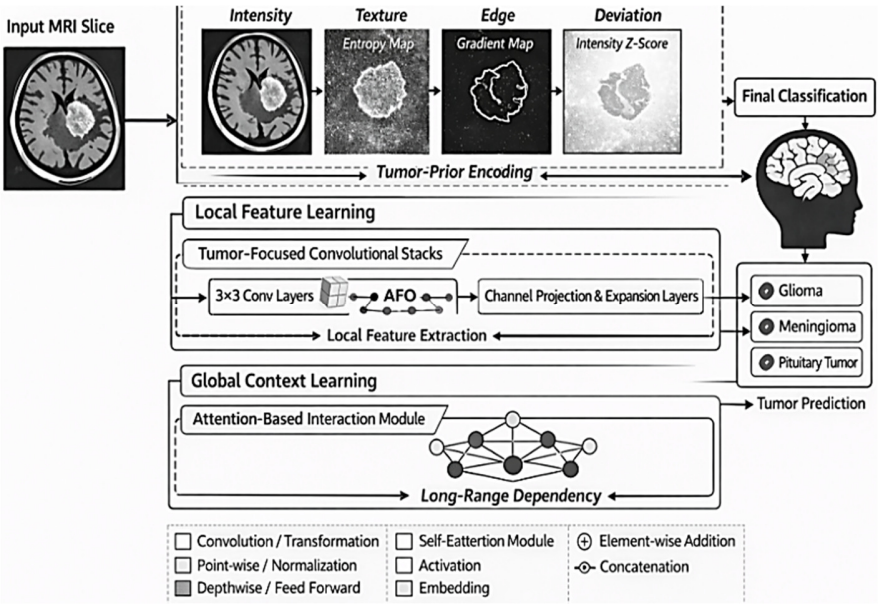


Fig. 1. Proposed GLIM-net framework with tumor-aware feature maps (Intensity, Texture, Edge, and Deviation) feeding parallel local and global branches for multi-class brain tumour classification.

A. Dimensional Data Transformation

The dataset is different from standard 1-D data because it is random and not very common. The features are randomly spread out in the long, redundant sequence, which makes it hard for the current deep learning model to choose from them. The long data length also means that the model needs to be very complex to perform well, and it is hard to use traditional methods to extract features for classification tasks. Following any change, the data are re-cast to the same three-channel layout that an RGB image uses, training a DL network from a familiar three-layer array.

The dataset is processed by counting each cohort and sorting them according to the positions of the features. The i-th feature is arranged in the form of a matrix U_{kl}, where $0 \leq k \leq 30$ and $0 \leq l \leq 18$. The other characteristics of various tumors inside the k-th MRI modality are N_k. By presuming a graph in the form of $P * P$ within this point, the k-th feature lodges o_k. The graph is as shown in (1). All features accommodate O rows, where O is defined in (2).

$$o_k = \begin{Bmatrix} {N_k}/{P} + 1 & N_k \; mod \; P \text{ not equal to } 0 \\ {N_k}/{P} & N_k \; mod \; P \text{ equal to } 0 \end{Bmatrix} \quad (1)$$

$$O = \sum_{k=0}^{23} o_k = \sum_{k=0}^{23} \begin{matrix} N_k \; mod \; P \text{ not equal to } 0 \\ N_k \; mod \; P \text{ equal to } 0 \end{matrix} \quad (2)$$

where $O \leq P$.

Considering this definition, a tumor-aware multi-channel map of size $P * P$ includes the features of 19 tumor cohorts; here, the features considered are inserted, deleted, and associated with single centric nucleotide variations in correspondence within the RGB channels of the image. The IPV information linked to the image is embedded, and the method picks the first value with the highest IPV. Max_{mut} is then using one of the indicator variables upon mapping the IPV quantity to the grayscale values. Assuming that the single-nucleotide tumor-associated feature channel I makes c replacements within a single sample feature channel fragment, this value in the red channel is considered the relationship between c and d in:

$$\frac{d}{255} = \frac{c}{Max_{mut}},\ d \leq 255, \mathrm{c} \leq Max_{mut} \quad (3)$$

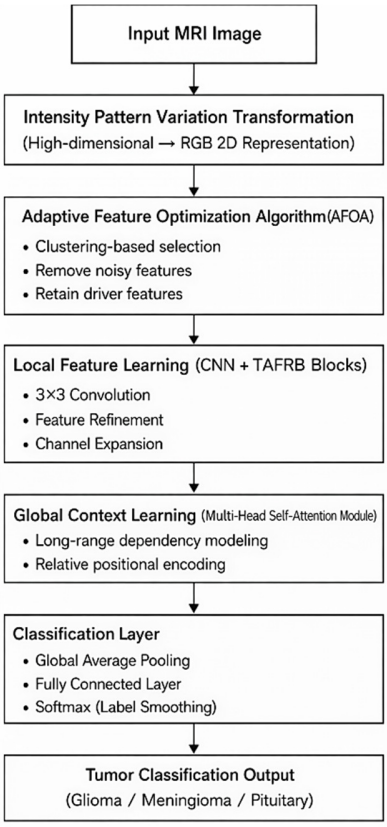


Fig. 2. Simplified pipeline flowchart of the proposed framework.

For every sample, we count how many times each IPV type occurs. Those counts are turned into grayscale values and placed in one image channel. One channel is built for every IPV type—each sample ends up with a separate single-channel image per type. All single-channel images that belong to the same sample are taken and overlaid. The result is a 2D image that shows the combined IPV pattern for that sample. Similarly, substitution, insertion, and deletion are handled within any feature channel that carries a single nucleotide change. Each of those three IPV types is assigned to one RGB channel. The counts for substitutions go to the red channel, insertions to green, and deletions to blue. After the values are scaled to the 0-255 range, the three channels are merged into one color image.

B. Adaptive Feature Optimization Algorithm (AFOA)

The noise and sparsity issues within the tumor-associated profile affect the extraction of relevant feature data. Upon learning, noise enhances the complexity of the model, resulting in overfitting. The intervention caused by noise and sparsity in data is avoided by narrowing the feature selection, considering the previous techniques. Clustering algorithms are employed to filter the candidate driver tumor-aware multi-channel MRI features for the AFOA. The passenger tumor-aware multi-channel MRI features do not cot consist of any significant tumor-related information and are discarded. This helps improve the model's overall training effectiveness and classification performance. Clusters with fewer than two IPVs are removed. The primary removal of noise pixels represents feature channel fragments consisting of only single IPV points. These tumor-aware multi-channel MRI features have more irrelevant IPVs compared to other passenger ones. Among the candidate driver tumor-aware multi-channel MRI features chosen through various other clustering algorithms, a large number consist of true ones, mainly localized in low-frequency IPV regions that are difficult to capture in the cohort. The proposed clustering algorithm captures the driver tumor-aware multi-channel MRI features within low-frequency IPV backgrounds, which other methods lack.

However, there are still many tumor-aware multi-channel MRI features within low-frequency IPV background that are not included in the candidate driver features. To capture more IPV addresses and information and address the limitations of the existing IPV-based profiling methods, this method not only focuses on candidate driver tumor-aware multi-channel MRI features with s-values less than 0.25 as benchmarks for AFOA, but also consists of some potential driver tumor-aware multi-channel MRI features for reference purposes.

The tumor-aware multi-channel input after AFOA is defined as $P' * P'$, and c' lists the feature tumor-aware multi-channel MRI features selected for each group. Since this follows the rules of a tumor-aware multi-channel, a clear connection must link c' and P', denoted as

$$24P' + 19c' - 24 \leq P^{v^2}$$

This method changes the size of the data—each line of the tumor-aware multi-channel map stores the IPV data for one MRI modality. The tumor-aware multi-channel MRI features are not spread evenly across the MRI modalities. If one MRI Modality holds $P' + 1$ tumor-aware multi-channel MRI features, it occupies two lines in the map, and the model learns better. This study keeps P' fixed at 56. The highest possible whole number c' is 95, but the chance of reaching this value is close to zero. c' is set to 100, which means that for every tumor group, the 100 tumor-aware multi-channel MRI features with the highest cluster scores are chosen. If fewer than 100, more potential drivers are added until the list reaches 100. After the AFOA step, the feature channel IPV matrix is reduced to 56×56 to lower noise and sparsity that troubles the earlier size change approach.

C. Global-Local Interaction Modeling Network

The network starts from a tumor-aware multi-channel input, where a 3×3 convolution layer with a stride of 2 reads the map and produces shallow features. After that, the stacked TAFRB raises the dimension of the representation. Every block follows one of two layouts, called type A or type B. Close to the end, a set of multi-head self-attention blocks captures long-range patterns. Each of those blocks contains two 1×1 convolutions and one multi-head self-attention layer. A 7×7 global average pooling layer produces the final output. Figure 3 presents the detailed GLIM-Net architecture. The model integrates TAFRBs for robust low-level feature extraction, CNN-based local feature learning, and multi-head self-attention for global context modeling. Together, these components capture both fine-grained tumor textures and long-range spatial dependencies, enabling accurate multi-class tumor classification.

D. Tumor-Aware Feature Refinement Block (TAFRB)

The residual structure first shrinks the data with a 1×1 convolution, then restores it with another 1×1 convolution, and finally enlarges it once more through a third 1×1 convolution. In this module, the 1×1 convolution handles the expansion, with the last expansion step also relying on a 1×1 convolution. After that, a 3×3 convolution reduces the size, and a final 1×1 convolution sets the ultimate dimension. At every spatial position, the method combines information from every channel. Standard residual blocks apply ReLU after adding the shortcut—the inverted version instead places ReLU inside the expansion path:

$$a = ReLU6(z) = \min(ReLU(z), 6) = \min(\max(0, z), 6) \quad (4)$$

As a result of information loss, the output value of the $ReLU$ function is in the range $0\ to\ \infty$. A linear activation function performs dimension reduction and keeps all information intact.

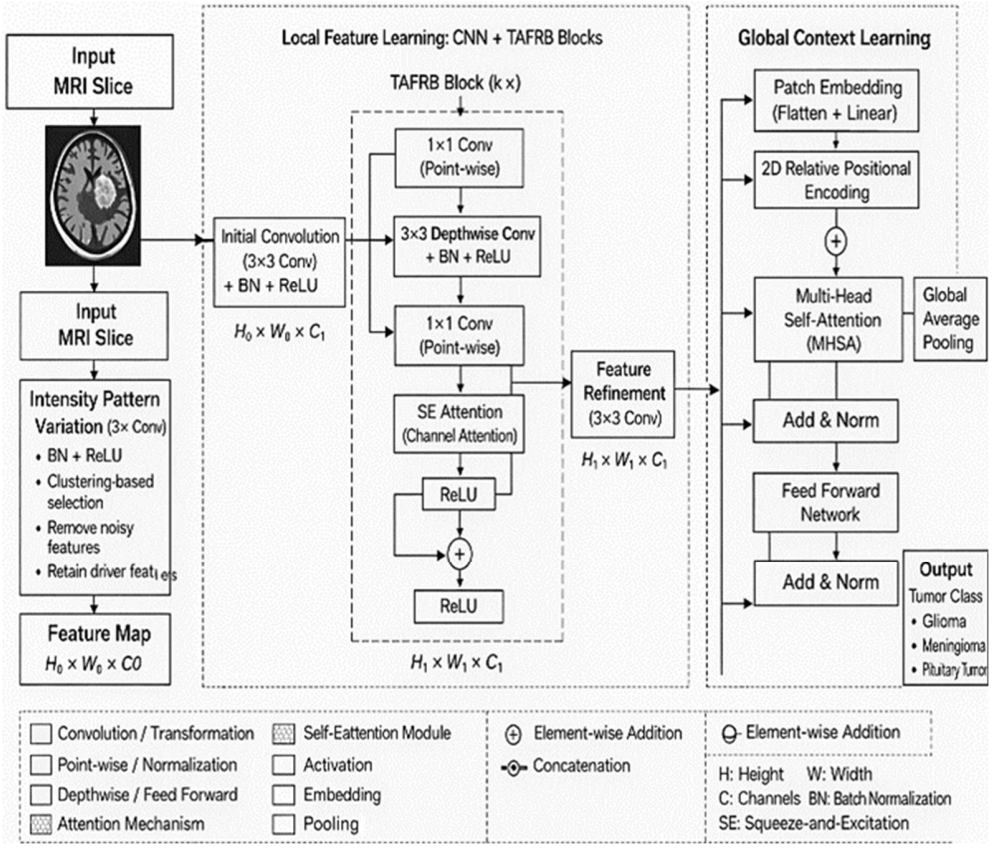


Fig. 3. Detailed GLIM-Net architecture showing TAFRB blocks, tokenization, local feature learning (CNN), global context learning (multi-head self-attention), and the classification head.

E. Global Context Interaction Module (GCIM)

The core part of the transformer network is the GCIM, a specific multi-head attention mechanism that receives inputs M, X, and S. The hidden state matrices are $J = T^{f*N}$ for a given input sequence, where f stands for the hidden state dimension and N is the sequence length. The transformer adds positional encodings so that the attention blocks know where each token sits. A distance-aware variant works well for vision, as it lets attention weigh features by how far apart they are, and objects in different places remain linked.

In the 2D relative positional encoding attention module, every attention mechanism operates on a 2D feature map. The relative distance positional encodings are like that, as T_j and T_y shows how tall and wide the feature map is. The attention log is represented by $S^{m^t} + S^{t^t}$, where s, m, and t denote the query, key, and positional encoding. 1×1 represents point-wise convolution. The transformer network is then normalized through layer normalization. Finally, the Self-Adaptive Attention Module (SAAM) transformer ends with an output projection layer. Inside its feed-forward network, the block applies the same nonlinear activation function three times in sequence.

F. Label Smoothing with Regularization

Classification algorithms usually assign data to categories through hard labels. Each label appears as a one-hot vector, where every position in the vector is set to zero except for one:

$$a_k = \begin{cases} 0, & k \text{ equal to target} \\ 1, & k \text{ not equal to target} \end{cases} \quad (5)$$

In the label vector, the probability of the target class is 1, and the probability of all non-target classes is 0. Cross-entropy is given by:

$$J = -\sum_k a_k \log r_k \quad (6)$$

where a_k and r_k are the real and predicted probabilities, respectively. The goal is to lower the loss during model training. The predicted probability r_k is obtained by applying the $softmax$ function to the logits vector b_k, which is what the last layer of the model gave:

$$r_k = softmax\,(a_k) = \frac{\exp\,(a_k)}{\sum_l a_l} \quad (7)$$

To reduce the loss through one-hot encoded label data, in an ideal situation, the model's prediction should approach a probability of 1 for the target class and a probability close to 0 for non-target classes. The training set is too small to reveal the real spread of the data—the network easily fits the limited samples and loses its ability to work on new cases. To lower this risk, the one-hot hard label vector a_k is replaced with a new vector that mixes the original label with a uniform distribution, injecting controlled noise into the target and steering the model away from extreme confidence:

$$a_k = \begin{cases} 1-\beta & k = target \\ \frac{\beta}{M-1} & k \text{ not equal to target} \end{cases} \quad (8)$$

Here β denotes the hyperparameter (set to 0.1) and M represents the count of classes. This technique stops the network from memorizing the exact pattern of the training set. Noise injected into the examples forces the classifier to stay within a bounded region and reduces overfitting. The same smoothing rule also pushes apart the class centroids, widening the gap between them. This improves performance on unseen data.

III. PERFORMANCE EVALUATION

The proposed model was tested on open MRI brain tumor datasets. The results were obtained in terms of accuracy, precision, recall, and F1-score for each tumor type, giving a clear view of how well the model labels each class. The proposed method was compared with leading CNN and transformer models to evaluate its performance.

A. Dataset Details

1) FigShare Brain Tumor Dataset

The open-access FigShare brain tumor collection [18] consists of 3,064 axial T1-weighted contrast-enhanced MRI slices with one of three class tags, Glioma, Meningioma, or Pituitary, together with the related segmentation mask. The 2D slices show realistic spread in brightness, shape, and position, and can be used for tumor type classification and for localizing each mass. One class is slightly larger than the others, but every class supplies enough examples for supervised training. This dataset was chosen for its trustworthy labels and for treating the three main tumor kinds evenly. Each image comes with an aligned axial slice but also a fixed pixel size, which simplifies preprocessing, limits memory use, and fits compact transformer models, even though it omits the full 3D contexts.

2) 3K-DS and 7K-DS Datasets

The 3K-DS dataset [19] consists of 3,264 MRI scans that are classified into four groups. The first has 901 scans showing a pituitary tumor, the second has 926 scans with glioma, the third has 937 scans with meningioma, and the last group has 500 scans that show no tumor. For building and testing models, 2,937 scans were used for the training set, and 327 for the test set. Much data is needed to teach stable patterns, while the test subset gives a fair measure of performance. With this size, researchers can test how the model behaves when data are scarce and judge how the features channelize in small medical imaging conditions.

The 7K-DS dataset [20] consists of 7,023 MRI scans, with 1,757 pituitary adenomas, 1,621 gliomas, 1,645 meningiomas, and 2,000 scans free of tumor. This spread raises the share of normal cases and balances the diagnostic classes. 6320 cases form the training pool, 703 remain for testing. This split lets algorithms learn from a wide variation, but still gives enough unseen data for sound measurement.

TABLE I. DS DATASETS' DESCRIPTION

Category	3K-DS	7K-DS
Pituitary	901	1757
Glioma	926	1621
Meningioma	937	1645
No tumour	500	2000
Total Images	3264	7023
Training set	2937	6320
Testing set	327	703

B. Results on FigShare

Figure 4 shows a comparative performance analysis for glioma classification using two approaches, namely HSViT [17] and the proposed GLIM-Net. Besides the F1-score, each method reached high precision, which indicates that it correctly labeled glioma cases and rarely yielded false positives. GLIM-Net scored slightly higher than HSViT [17]. A more noticeable performance gain is observed in recall, demonstrating a

superior capability to correctly detect a higher proportion of true glioma cases and reduce false negatives. This improvement in recall translates into a higher F1-score for GLIM-Net, reflecting a better balance between precision and recall. The results show that while both models are effective for glioma detection, the GLIM-Net approach provides more robust and reliable performance, particularly in capturing glioma cases.

TABLE II. RESULTS COMPARISON (PER-CLASS) BETWEEN HSVIT [17] AND GLIM-NET ON THE FIGSHARE DATASET

Tumor Type	Precision (%)	Recall (%)	F1-score (%)
Glioma HSViT [17]	99.32	93.55	96.35
Glioma GLIM-Net	99.67	99.10	99.38
Meningioma HSViT [17]	98.01	96.73	97.37
Meningioma GLIM-Net	99.15	99.05	99.10
Pituitary HSViT [17]	99.34	97.4	98.36
Pituitary GLIM-Net	99.68	99.16	99.42
Overall HSViT [17]	99.11	98.66	98.88
Overall GLIM-Net	**99.5**	**99.10**	99.30

Figure 5 presents a comparative evaluation of meningioma classification performance between the baseline HSViT [17] and the proposed GLIM-Net. Both methods demonstrate high precision, indicating reliable identification of meningioma cases; however, GLIM-Net consistently achieves higher precision, reflecting fewer false-positive predictions. A more pronounced improvement is observed in recall, where GLIM-Net substantially outperforms HSViT [17], showing its enhanced ability to correctly detect meningioma cases and minimize false negatives. This improvement in recall leads to a superior F1-score, highlighting a more balanced and robust classification performance. The results confirm that the GLIM-Net model offers improved diagnostic reliability and consistency for meningioma detection compared to the HSViT-based approach. Figure 6 compares how well the baseline HSViT [17] and the proposed GLIM-Net classify pituitary tumours, using precision, recall, and F1-score. Each method reaches almost perfect scores on all three measures, showing that the classification is highly reliable. GLIM-Net gains a slight but steady edge over HSViT [17], mainly in recall or F1-score, meaning that it finds a few more true pituitary cases while still keeping precision high. In short, both models label pituitary tumors very accurately, but GLIM-Net delivers a small but noticeable improvement that confirms its strength in spotting those tumors. Figure 7 presents an overall performance comparison between the HSViT [17] and the proposed GLIM-Net using precision, recall, and F1-score. Both methods demonstrate exceptionally high performance, with values consistently close to 100%, indicating strong classification reliability across all classes. However, GLIM-Net achieves uniformly higher scores across all three metrics, with the most notable improvement observed in recall, reflecting its enhanced ability to correctly identify true positive cases and reduce false negatives. The corresponding increase in F1-score highlights a better balance between precision and recall for GLIM-Net. Overall, these results confirm that GLIM-Net delivers superior and more stable performance compared to HSViT [17], reinforcing its effectiveness as a robust and generalizable classification framework.

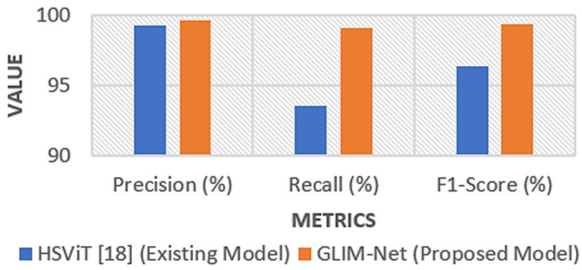


Fig. 4. Comparative performance analysis of glioma classification on the FigShare dataset.

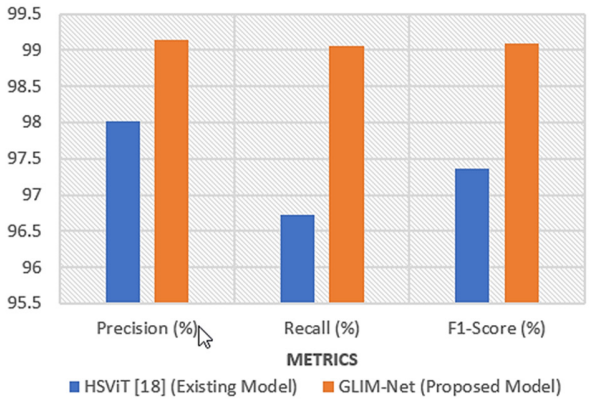


Fig. 5. Comparative performance analysis of meningioma classification on the FigShare dataset.

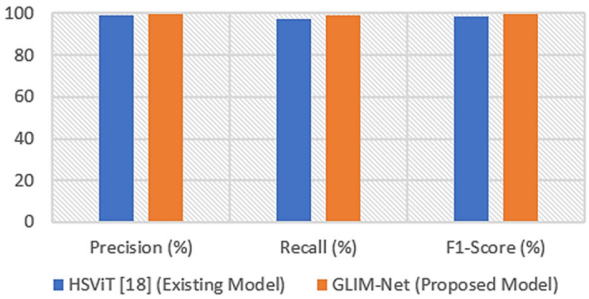


Fig. 6. Comparative performance analysis of pituitary classification on the FigShare dataset.

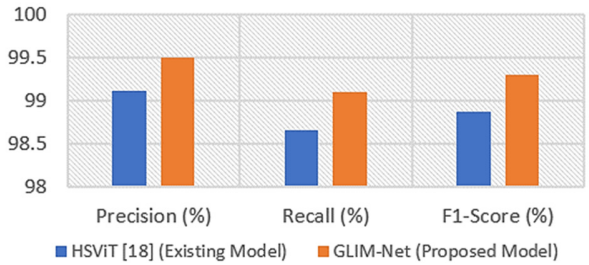


Fig. 7. Overall performance analysis of classification on the FigShare dataset.

Figure 8 compares multiple models. GLIM-Net earns the top mark in every measure, which shows that it classifies better and more reliably than the others. HSViT [17] also performs strongly, closely following GLIM-Net, indicating its effectiveness among existing methods. Models such as RVIT2

[21] and Hybrid ViT+GRU [22] exhibit competitive precision and recall values but show comparatively lower accuracy, suggesting limitations in overall prediction consistency. In contrast, traditional architectures such as ResNet-50+Multi-CNN [23] and VGG16 [23] achieve moderate performance across metrics, highlighting the advantages of transformer- and hybrid-based models. Overall, the comparison confirms that the GLIM-Net approach provides the most balanced and robust performance among all evaluated methods, effectively outperforming both conventional CNNs and recent hybrid architectures.

TABLE III. COMPARISON ON THE FIGSHARE DATASET

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
RViT2 [21]	97.3	98.1	95.6	96.8
Hybrid ViT+GRU [22]	96.08	97	96	96.5
VGG16 [23]	97.8	96.5	96.4	96.4
ResNet-50+Multi-CNN [23]	95	94	94	94
OS-DET [24]	89.57	95	94.2	94.6
HSViT [17]	97.83	99.11	98.66	98.88
GLIM-Net	99.68	99.5	99.10	99.30

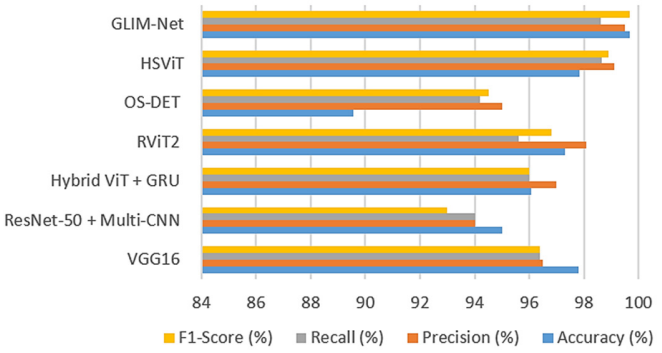


Fig. 8. Comparative evaluation of multiple classification models on the FigShare dataset.

C. Results on 3K-DS and 7K-DS Comparison

Table IV illustrates a sample-level prediction comparison on the 3K-DS dataset [19], highlighting the agreement between ground-truth labels and model predictions. Most instances show correct classification, where samples belonging to class 0 and class 1 are accurately predicted, indicating that the model effectively learns and distinguishes the dominant class patterns present in the dataset. However, one misclassification is observed, where a sample from class 2 is incorrectly predicted as class 0, revealing a class confusion that may arise due to limited inter-class separability or class imbalance in the relatively smaller 3K-DS dataset. Overall, the predominance of correct predictions demonstrates strong model reliability on 3K-DS, while the isolated error suggests that performance for less-represented or visually similar classes could be further improved through enhanced feature representation or data augmentation.

TABLE IV. SAMPLE RESULTS ON THE 3K-DS DATASET

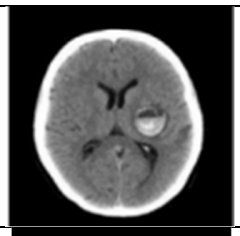
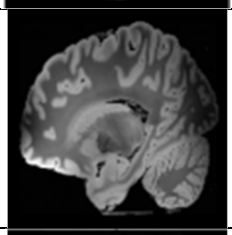
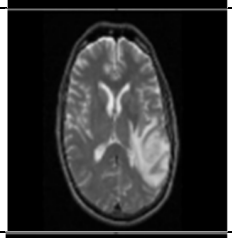
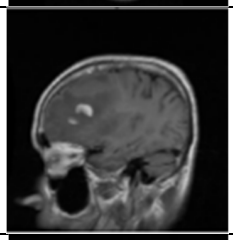
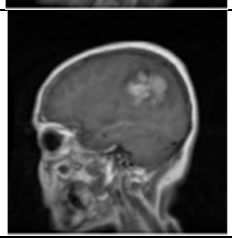
True=0, Pred=0	
True=1, pred=1	
True=0, pred=0	
True=0, pred=0	
True=2, pred=0	

Table V presents sample-level prediction results on the 7K-DS dataset [20], reflecting both correct classifications and residual class confusions. The model correctly predicts multiple instances belonging to class 0 and class 1, as well as accurately identifies a class 3 sample, demonstrating its strong discriminative capacity across several classes when trained on the larger and more diverse 7K-DS dataset. However, a misclassification is observed, where a sample from class 0 is incorrectly predicted as class 2, indicating a minor confusion between these two classes. Despite this isolated error, the overall results suggest improved class separability and robustness compared to smaller datasets, as the increased data volume in 7K-DS enables more stable learning and better IPV. The remaining misclassification highlights potential visual or feature-level similarities between certain classes, which could be further mitigated through refined feature extraction or class-specific optimization strategies.

TABLE V. SAMPLE RESULTS ON THE 7K-DS DATASET

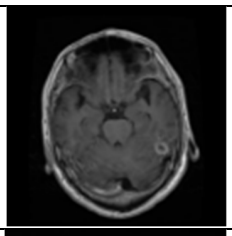
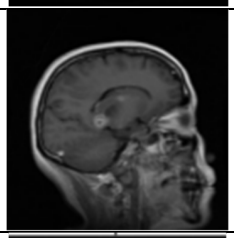
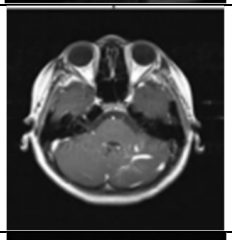
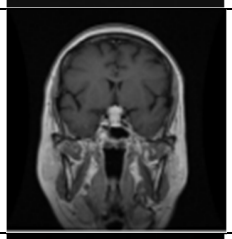
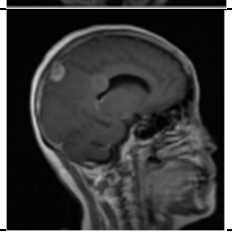
True=0, pred=0	
True=0, pred=0	
True=1, pred=1	
True=3, pred=3	
True=0, pred=2	

Table VI compares standard learners, such as XGBoost [27], and ensemble voters, which achieve only moderate scores because they miss many fine-grained tumor details. As the models progress toward deep learning architectures, including block-wise CNNs [25], encoder-decoder GANs [32], and multi-level attention networks, a steady improvement in accuracy is observed, demonstrating the effectiveness of hierarchical and attention-driven feature extraction. Hybrid CNN-based models [26] further enhance performance by combining complementary learning strategies. The GLIM-Net encodes features well, applies attention sensibly, and generalizes the unseen data, achieving the highest accuracy, showing its superior ability to capture discriminative features and its strong generalization capability compared to existing state-of-the-art methods.

TABLE VI. ACCURACY COMPARISON

Model	**Dataset**	**Accuracy (%)**
Block-wise fine-tuning CNN [25]		94.82
Hybrid CNN [26]		98.4
XGBoost [27]		88.51
Ensemble of deep features with SVM [28]		93.72
Regularised Extreme Learning Machine [29]		94.33
Multi-level attention network [30]		96.51
CGAN [31]		97
Deep-GAN [32]		96
Deep-EFNet [33]	3K-DS	95.4
GLIM-Net (Proposed)	3K-DS	99.10
Deep-EFNet [33]	7K-DS	98.2
GLIM-Net (Proposed)	7K-DS	99.3

IV. CONCLUSION

This study presented a robust and efficient deep learning framework for brain tumor classification, addressing key challenges related to data sparsity, noise, limited generalization, and computational complexity. The proposed approach integrates a novel IPV-to-image dimensional transformation strategy, a clustering-driven AFOA, and a global-local interaction modeling network-based architecture to effectively capture both local and global discriminative patterns. By transforming high-dimensional IPV profiles into structured RGB-based 2D representations, the framework enables effective feature learning while preserving biologically meaningful information. The AFOA further enhances model reliability by eliminating noisy and low-informative passenger tumor-aware multi-channel MRI features, thereby improving training stability and classification performance. The SAAM-based network, which combines tumor-aware feature refinement learning with multi-head self-attention, demonstrates strong capability in modeling complex dependencies with reduced computational overhead compared to conventional transformer-heavy architectures. In addition, the incorporation of label-smoothing-based regularization mitigates overfitting under limited-data conditions and improves inter-class separability, leading to more consistent and generalizable predictions. Extensive experimental evaluations conducted on multiple benchmark datasets confirm that the proposed model consistently outperforms state-of-the-art CNN and transformer-based methods across key performance metrics.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the results reported in this paper.

ACKNOWLEDGMENT

No external funding was received for this work.

DATA AVAILABILITY

All datasets used in this study are publicly available. The FigShare brain tumour dataset is available at [18], the 3K-DS is available at [19], and the 7K-DS dataset is available at [20]. No new datasets were generated during this study.

REFERENCES

[1] D. N. Louis *et al.*, "The 2021 WHO Classification of Tumors of the Central Nervous System: a summary," *Neuro-Oncology*, vol. 23, no. 8, pp. 1231–1251, Aug. 2021, https://doi.org/10.1093/neuonc/noab106.

[2] "Brain Tumor Facts," *National Brain Tumor Society*. https://braintumor.org/brain-tumors/about-brain-tumors/brain-tumor-facts/.

[3] O. Al-Omari, O. Alkhatib, and T. Al-Omari, "CNN-Based Automated Detection of Metastatic Cancer in Histopathology Images," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 24478–24485, Aug. 2025, https://doi.org/10.48084/etasr.10888.

[4] M. Havaei *et al.*, "Brain tumor segmentation with Deep Neural Networks," *Medical Image Analysis*, vol. 35, pp. 18–31, Jan. 2017, https://doi.org/10.1016/j.media.2016.05.004.

[5] 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, Feb. 2016, https://doi.org/10.1109/TMI.2016.2538465.

[6] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, https://doi.org/10.1016/j.media.2017.07.005.

[7] M. Rasool, A. Noorwali, H. Ghandorh, N. A. Ismail, and W. M. S. Yafooz, "Brain Tumor Classification using Deep Learning: A State-of-the-Art Review," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16586–16594, Oct. 2024, https://doi.org/10.48084/etasr.8298.

[8] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv, June 03, 2021, https://doi.org/10.48550/arXiv.2010.11929.

[9] S. Salma, S. C. Lingareddy, Shilpashree. S, and V. Kumar, "STRIVE-Net: A Dual Stream Semantic Transformer Framework with Reference Embedding Refinement for 3D brain Tumor Segmentation in MRI," in *2025 2nd International Conference on New Frontiers in Communication, Automation, Management and Security (ICCAMS)*, July 2025, pp. 1–7, https://doi.org/10.1109/ICCAMS65118.2025.11234183.

[10] A. B. Subba, A. K. Sunaniya, and A. Mukherjee, "Multiclass Brain Tumor Detection with Attention-Embedded CNN Framework: Advancing Toward Decentralized Deep Learning-Based Health Monitoring," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–14, 2025, https://doi.org/10.1109/JBHI.2025.3638154.

[11] X. Zhang, A. Bahri, C. Desrosiers, H. Liu, and F. Bao, "M^2SegMamba: Mamba-Based Incomplete Multimodal Learning for Brain Tumor Segmentation With Few Samples," *IEEE Journal of Biomedical and Health Informatics*, vol. 30, no. 3, pp. 2455–2465, Mar. 2026, https://doi.org/10.1109/JBHI.2025.3600652.

[12] F. Xiao, C. Ji, Z. Zhang, and R. Wang, "Decouple-and-Couple Learning in Multi-Modal Brain Tumor Segmentation," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–11, 2025, https://doi.org/10.1109/JBHI.2025.3542394.

[13] M. R. Shoaib, D. Niyato, H. M. Emara, and J. Zhao, "Automated Segmentation and Classification of Brain Magnetic Resonance Images for Brain Tumor Diagnosis Based on Deep Learning," *IEEE Internet of Things Journal*, vol. 13, no. 5, pp. 9872–9889, Mar. 2026, https://doi.org/10.1109/JIOT.2025.3647773.

[14] V. D. R. Kumar and P. Singh, "HCAC: Trustworthy Hybrid Attentive Framework Tackling Brain Tumor Identification With XAI," *IEEE Access*, vol. 14, pp. 16158–16175, 2026, https://doi.org/10.1109/ACCESS.2026.3656212.

[15] C. Li, Q. Liu, and J. Teng, "Edge Feature Extraction Medical Image Segmentation Network Based on Frequency Domain Data Enhancement," *IEEE Access*, vol. 13, pp. 218192–218206, 2025, https://doi.org/10.1109/ACCESS.2025.3648505.

[16] I. Abdelhaliem, J. Dixon, A. Abdelhamid, G. A. Saleh, and F. Khalifa, "A Multimodal Adaptive Inter-Region Attention-Guided Network for Brain Tumor Classification," *IEEE Access*, vol. 13, pp. 187964–187975, 2025, https://doi.org/10.1109/ACCESS.2025.3627777.

[17] P. Chauhan, M. Lunagaria, and D. K. Verma, "Hierarchically Scaled Vision Transformer (HSViT) With Digital Twinned Non-Invasive Neurotherapy to Detect Brain Tumor," *IEEE Transactions on Consumer Electronics*, vol. 71, no. 4, pp. 11026–11033, Nov. 2025, https://doi.org/10.1109/TCE.2025.3624311.

[18] J. Cheng, "brain tumor dataset." figshare, 2017, https://doi.org/10.6084/M9.FIGSHARE.1512427.V5.

[19] "Brain Tumor Classification (MRI)." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri.

[20] "Brain Tumor MRI Dataset." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset.

[21] P. T. Krishnan, P. Krishnadoss, M. Khandelwal, D. Gupta, A. Nihaal, and T. S. Kumar, "Enhancing brain tumor detection in MRI with a rotation invariant Vision Transformer," *Frontiers in Neuroinformatics*, vol. 18, June 2024, https://doi.org/10.3389/fninf.2024.1414925.

[22] M. M. Ahmed *et al.*, "Brain tumor detection and classification in MRI using hybrid ViT and GRU model with explainable AI in Southern Bangladesh," *Scientific Reports*, vol. 14, no. 1, Oct. 2024, Art. no. 22797, https://doi.org/10.1038/s41598-024-71893-3.

[23] A. A. Asiri *et al.*, "A Novel Hybrid Machine Learning Approach for Classification of Brain Tumor Images," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 641–655, 2022, https://doi.org/10.32604/cmc.2022.029000.

[24] K. Deng *et al.*, "OS-DETR: End-to-end brain tumor detection framework based on orthogonal channel shuffle networks," *PLOS ONE*, vol. 20, no. 5, 2025, Art. no. e0320757, https://doi.org/10.1371/journal.pone.0320757.

[25] Z. N. K. Swati *et al.*, "Brain tumor classification for MR images using transfer learning and fine-tuning," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 34–46, July 2019, https://doi.org/10.1016/j.compmedimag.2019.05.001.

[26] H. Ayaz, M. Ahmad, D. Tormey, I. McLoughlin, and S. Unnikrishnan, "A Hybrid Deep Model for Brain Tumor Classification," in *Proceedings of 2021 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2021)*, 2022, pp. 282–291, https://doi.org/10.1007/978-981-16-3880-0_29.

[27] D. L. De Souza, A. M. Coelho, J. C. Felipe, and M. De Freitas Oliveira Baffa, "Brain Tumor Classification on Multi-Modality MRI Using Radiomic Features," in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2021, pp. 1799–1804, https://doi.org/10.1109/BIBM52615.2021.9669791.

[28] J. Kang, Z. Ullah, and J. Gwak, "MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers," *Sensors*, vol. 21, no. 6, Mar. 2021, https://doi.org/10.3390/s21062222.

[29] A. Gumaei, M. M. Hassan, M. R. Hassan, A. Alelaiwi, and G. Fortino, "A Hybrid Feature Extraction Method With Regularized Extreme Learning Machine for Brain Tumor Classification," *IEEE Access*, vol. 7, pp. 36266–36273, 2019, https://doi.org/10.1109/ACCESS.2019.2904145.

[30] N. S. Shaik and T. K. Cherukuri, "Multi-level attention network: application to brain tumor classification," *Signal, Image and Video Processing*, vol. 16, no. 3, pp. 817–824, Apr. 2022, https://doi.org/10.1007/s11760-021-02022-0.

[31] A. A. Asiri *et al.*, "Next-Gen brain tumor classification: pioneering with deep learning and fine-tuned conditional generative adversarial networks," *PeerJ Computer Science*, vol. 9, Nov. 2023, Art. no. e1667, https://doi.org/10.7717/peerj-cs.1667.

[32] A. A. Asiri *et al.*, "Multi-Level Deep Generative Adversarial Networks for Brain Tumor Classification on Magnetic Resonance Images," *Intelligent Automation & Soft Computing*, vol. 36, no. 1, pp. 127–143, 2023, https://doi.org/10.32604/iasc.2023.032391.

[33] H. Alshaari and S. Alqahtani, "Deep-EFNet: An Optimized EfficientNetB0 Architecture With Dual Regularization for Scalable Multi-Class Brain Tumor Classification in MRI," *IEEE Access*, vol. 13, pp. 85682–85697, 2025, https://doi.org/10.1109/ACCESS.2025.3567919.