Swin-HAFNet: A Hierarchical Multi-Task Transformer for Brain Tumor Segmentation and Classification

Amirreza Fateh[a], Yasin Rezvani[b], Sadjad Rezvani[b], Mansoor Fateh[b,*] and Vahid Abolghasemi[c,*]

[a]School of Computer Engineering, Iran University of Science and Technology (IUST), Tehran, Iran
[b]Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran
[c]School of Computer Science and Electronic Engineering, University of Essex, Colchester, United Kingdom

ARTICLE INFO

Keywords:
Brain tumor Classification
Brain tumor Segmentation
Magnetic resonance imaging
Medical imaging

ABSTRACT

Precise neuro-oncological diagnosis necessitates the accurate delineation and histological typing of brain tumors. However, current automated systems often fail to address the extreme morphological variability of lesions and lack unified frameworks capable of executing segmentation and classification simultaneously without computational redundancy. To overcome these limitations, we introduce Swin-HAFNet, a novel hierarchical multi-task transformer designed for robust, dual-task analysis. The architecture leverages a Swin Transformer backbone to extract rich multi-scale representations. Following this feature extraction, a Contextual Bottleneck Enhancer employs shifted-MLPs and gated encoding units to refine latent spatial dependencies. Furthermore, a Hierarchical Attention Fusion module integrates self-attention with deformable convolutions to adaptively merge encoder-decoder features and preserve boundary details. Complementing these components, a dedicated classification branch utilizes dimension and spatial reduction to synthesize hierarchical features for precise multi-class grading. Comprehensive validation on the BRISC and diverse Kaggle datasets demonstrates the superior generalization and robustness of the model. Swin-HAFNet achieves a weighted mean Intersection over Union of 82.4% for segmentation and an accuracy of 99.63% for classification on BRISC dataset. By seamlessly integrating pixel-level localization with image-level diagnosis, this work validates the efficacy of multi-task learning and establishes a new benchmark for unified and clinically translatable brain tumor analysis.

1. Introduction

Brain tumors represent one of the most fatal and complex categories of cancer, necessitating precise localization and accurate histological typing for optimal therapeutic intervention [1, 2]. Magnetic Resonance Imaging (MRI) remains the gold standard for non-invasive neuro-imaging, providing high-resolution visualization of brain structures and pathological lesions [3, 4]. Despite the potential of automated systems to reduce radiologist workload and inter-observer variability, the clinical deployment of these tools remains challenging [5]. The primary difficulty lies in the dual requirement of segmentation, which involves pixel-wise delineation of tumor boundaries (e.g., glioma, meningioma, or pituitary), and classification, which identifies the specific malignancy type or non-tumorous condition [6, 7]. Accurate diagnosis depends on capturing both the fine-grained texture of the lesion and its global anatomical context relative to critical brain structures.

Recent advancements in medical image analysis have transitioned from traditional machine learning to deep learning paradigms, specifically Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) [8, 9]. Standard architectures like U-Net and its variants have been widely adopted for segmentation due to their hierarchical structure; however, they often lack the long-range dependency modeling required to understand complex tumor morphologies [10, 11]. In contrast, transformer-based models have emerged as powerful alternatives, utilizing self-attention mechanisms to capture global contextual information [12, 13]. While many studies focus solely on a single task, the development of multi-task frameworks that can simultaneously handle both has become a prominent research frontier in recent years [14, 15].

*Corresponding author
amirreza_fateh@comp.iust.ac.ir (A. Fateh); yasinrezvani@shahroodut.ac.ir (Y. Rezvani); sadjadrezvani@shahroodut.ac.ir (S. Rezvani); mansoor_fateh@shahroodut.ac.ir (M. Fateh); v.abolghasemi@essex.ac.uk (V. Abolghasemi)
ORCID(S):

The integration of segmentation and classification tasks presents several persistent challenges. First, brain tumors exhibit extreme heterogeneity in size, shape, and intensity, often blending into healthy tissue or mimicking non-neoplastic lesions like abscesses or cysts [16, 17]. Second, intensity inhomogeneity and variations in MRI acquisition protocols across institutions introduce domain-shift issues that degrade model generalization [18, 19]. Most critically, existing multi-task frameworks often utilize a shared encoder with separate decoders, a design that leads to task interference and sub-optimal feature sharing. These models frequently fail to adaptively fuse multi-scale features for the opposing requirements of pixel-level segmentation and image-level classification, resulting in significant computational overhead and a loss of spatial resolution in the decoding path [20, 21]. Addressing these issues requires a model that not only extracts robust features but also adaptively aggregates them across different stages of the network.

To overcome these limitations, we propose Swin-HAFNet, a lightweight yet powerful hierarchical transformer designed for robust dual-task analysis. Our method utilizes a Swin Transformer backbone to extract multi-scale semantic representations. We introduce the Hierarchical Attention Fusion (HAF) module to integrate encoder and decoder features more effectively than standard skip connections, ensuring that spatial details are preserved during reconstruction. Furthermore, we incorporate a Contextual Bottleneck Enhancer (CBE) to refine latent representations by modeling spatial context in both horizontal and vertical directions. This dual-task framework is optimized through a compound loss function that ensures boundary precision for segmentation and high categorical accuracy for classification.

The main contributions of this work are summarized as follows:

- **A unified, lightweight multi-task architecture** where a single, shared feature hierarchy is optimally distilled for both tasks via specialized, efficient heads, thereby minimizing task interference.
- **A novel HAF module** that replaces standard skip connections with an adaptive, attention-based mechanism for precise boundary preservation.
- **CBE** for explicit long-range spatial context modeling in the latent space, crucial for irregular tumor morphology.

2. Background

Overview of Brain Tumors

Brain tumours are amongst the most fatal of all cancers [22]. Amongst paediatric solid tumours, brain tumours are most fatal and commonly occurring [22]. There is diversity in the types of brain tumours, including but not limited to gliomas which account for 45% of brain tumours with pituitary tumours and meningiomas accounting for 15% each [23]. The gold standard imaging for diagnosis of a brain tumours or brain metastases is an MRI scan with gadolinium [24]. When possible, management is initially via surgery to remove the lesion which is then sent for histological and molecular genotype identification [24]. Pre, intra and post-operative MRIs can be used to guide surgical resection and management [24, 25, 26]. Intra-operatively, functional MRIs visualise cerebrovascular activity which can be correlated with neuronal activity, aiding the surgical team. Other techniques such as cord simulation can also be used [25]. In many scenarios, regardless of skill, neurosurgical reach has to be limited for safety due to the presence of many functionally important regions within the organ [22, 24]. Further management includes the use of medical interventions for symptomatic management, radiotherapy and chemotherapy [22]. The blood brain barrier poses challenges to medical interventions and chemotherapy, this barrier filters material entering the brain via circulation, limiting medical access to the brain [22]. Localisation of the lesion and adjacent structures via MRIs can guide both surgical and radiotherapeutic planning [27]. In summary, MRI scans are used in the initial diagnosis and management planning of brain tumours, including pre-surgical use and as guidance for radiotherapy [27].

Glioma

Gliomas are primary brain tumours, they are the most common of malignant primary brain tumours in adults [24]. They arise from glial cells or stem cells which develop glial properties during neoplastic changes [28]. Glial cells designate a group of different cells which provide support for neurons, for example by the formation of axonal myelin sheaths [29]. For adults, the most aggressive form of gliomas, the glioblastoma, has a two-year survival time [22]. Gliomas can be classified according to the WHO 2016 classification of Central Nervous System (CNS) tumours [28].

Diffuse forms of gliomas can grow in irregular shapes, extensively infiltrating brain parenchyma [28], this makes neurosurgical management difficult as safe maintenance of functional brain tissue is required during resection [22, 24].

Meningioma

Meningiomas in adults are the most common, being 30% of central nervous system tumours, whilst they are rare amongst children [30]. They arise from cells on the outer layer of the arachnoid mater, a part of the meninges [30]. The meninges is a layer in the central nervous system which encompasses the brain, cerebrospinal fluid and spinal cord [31]. Though, a meningioma could arise anywhere on the meninges, 98% of meningiomas are intracranial [30]. Usually benign lesions, they can be slow growing [30]. MRI scans in conjunction with CT scan can be used for diagnosis and treatment planning [30]. Treatment generally consists of neurosurgical treatment, occasionally in adjunct with radiotherapy [30]. Benign meningiomas can grow to a large size with pressures on the brain causing symptoms [32]. Non-benign meningiomas are associated with irregular shapes and tumours heterogeneity and therefore, benign meningiomas tend to be associated with regular shapes and homogeneity [33].

Pituitary Tumors

These are tumours originating in the pituitary gland, a small structure at the base of the brain, above the sphenoid bone [34]. The pituitary gland has an essential role in growth, metabolism and reproduction [35]. Due to this, pituitary tumours can cause a wide range of symptoms including but not limited to; mood disorders, diabetes mellitus, obesity, infertility and visual disturbances [35]. However, only one third of these tumours are symptomatic [35]. The majority of pituitary tumours are benign and when treated, treatment generally includes neurosurgical resection and radiotherapy [36].

Non-Tumorous Conditions

Non-tumorous conditions include both normal brain scans (from subjects without visible abnormalities) and scans with non-neoplastic lesions that can mimic tumorous appearance but are not neoplastic in origin. These may represent inflammatory or vascular pathologies such as abscesses, cysts, haematomas, or aneurysms [37]. The inclusion of such cases provides a broader spectrum of appearances encountered in clinical neuroimaging and enables models to better distinguish tumorous from non-tumorous abnormalities. Due to the high soft-tissue sensitivity of MRI, these conditions can often be characterised radiologically, although histopathological confirmation (e.g., via biopsy) may still be required in practice [38].

Anatomical planes

The main anatomical planes are coronal, transverse and sagittal planes [39]. These may be more simply described as a ‘front to back’ vertical plane, a ‘top to bottom’ horizontal plane, a longitudinal ‘side to side’ plane [39]. For a radiologist reporting an MRI scan, the above planes are available for viewing [40].

Challenges in Brain Tumor Diagnosis

As mentioned above, some tumours do not cause symptoms until reaching a certain growth [32], this may lead to late clinical suspicion to warrant imaging. When available and possible, the best imaging modality are MRI scans [24]. Even the use of neurological imaging can lead to misdiagnosis as neoplastic and non neoplastic conditions can mimic each other. As mentioned above, there are non tumorous space occupying lesions, these can be benign, meaning surgical resection and biopsy exposes many patients unnecessarily to the risks of surgery [16]. There are also neoplastic brain lesions which do not appear as a space occupying lesion [16]. Not neglecting T1 precontrast imaging can aid avoidance of misdiagnosis [16]. Further MR imaging modalities and a thorough clinical assessment alongside some further investigations can aid in reducing errors [16]. There are certain tumours which are difficult to visualised on MRI, though MRIs provide a high level of diagnostic accuracy for most tumours [41].

3. Related Work

Deep learning has become the dominant paradigm for automated brain tumor analysis, with notable progress in both segmentation and classification tasks. Recent research has shifted from conventional convolutional architectures toward advanced models that leverage attention mechanisms, multi-scale feature learning, and transformer-based representations. This section reviews representative methods for brain tumor segmentation and classification.

3.1. Brain Tumor Segmentation

The main goal of segmentation is the precise delineation of tumor subregions, including necrotic core, edema, and enhancing tumor. Modern strategies focus on enhancing encoder-decoder architectures. For example, frameworks that

adaptively optimize encoder, bottleneck, and decoder configurations achieve a better trade-off between segmentation accuracy and computational cost [42]. Similarly, integrating recurrent residual attention mechanisms into U-Net variants enables refinement of feature maps and emphasizes salient tumor regions while suppressing irrelevant background information [43].

Recent approaches also exploit multi-scale and context-aware designs. 3D convolutional networks capture volumetric information across slices, preserving spatial continuity in tumor regions [44]. Attention modules and feature refinement blocks enhance the representation of fine-grained structures, improving boundary delineation. Parallel network paths or dilated convolutions maintain a wide receptive field to simultaneously capture local and global features, which is crucial for heterogeneous tumor tissues [45].

Another trend is the integration of transformer-based modules into segmentation architectures. ViTs or hybrid CNN-ViT networks are employed to capture long-range dependencies across the image, complementing the local receptive fields of CNNs. These methods have demonstrated improved performance in segmenting tumors with irregular shapes or diffuse boundaries [46].

Additionally, attention mechanisms such as channel-wise, spatial, or self-attention layers are frequently incorporated to refine features at multiple scales. These layers help networks focus on relevant tumor regions, enhancing segmentation accuracy, especially in challenging cases with low contrast between tumor and healthy tissue [47]. Data augmentation and multi-modal input integration further improve model robustness and generalization across different patients and imaging protocols [48].

Despite the significant progress in automated brain tumor segmentation, most existing architectures still face difficulties in accurately delineating tumor boundaries where intensity gradients are subtle. While standard skip connections in U-shaped networks attempt to recover spatial details, they often introduce semantic noise by fusing misaligned features from the encoder and decoder [49, 50]. Furthermore, traditional convolutional kernels lack the global receptive field necessary to capture the full morphological context of heterogeneous lesions [51]. These limitations highlight the need for more advanced fusion strategies and contextual enhancement mechanisms that can adaptively prioritize relevant spatial and semantic information for precise boundary reconstruction [52].

3.2. Brain Tumor Classification

Classification tasks have progressed from binary detection to multi-class grading and survival prediction, often leveraging multimodal feature integration. Attention mechanisms are widely adopted to focus on the most informative regions in high-dimensional radiomics or image features [53].

Hybrid architectures combining CNNs for local feature extraction with ViTs for global context modeling have gained prominence. These models often include explainability techniques such as SHAP or Gradient-weighted Class Activation Mapping (Grad-CAM) to highlight the spatial and semantic relevance of predictions [54]. Hyperparameter optimization remains critical; systematic tuning of learning rates, filter sizes, and dropout ratios improves diagnostic robustness [55].

Feature extraction is increasingly performed using multi-path or multi-scale designs. Dilated convolutions, parallel CNN streams, or pyramid pooling modules allow models to capture features at varying resolutions, preserving both fine-grained details and global context [45]. Transfer learning from pre-trained networks such as ResNet or DenseNet remains popular for reducing training time and improving feature representation, especially in datasets with limited labeled samples [46].

Recent models also leverage attention-based fusion of multimodal features, integrating structural MRI with functional or metabolic imaging to improve classification accuracy [44]. Such fusion frameworks can dynamically weight the contribution of different modalities, enabling the model to focus on the most predictive sources of information. For survival prediction, deep learning models estimate tumor progression metrics, including proliferation rates and diffusion characteristics, offering clinicians actionable insights for prognosis and treatment planning [47].

While deep learning models have achieved high accuracy in brain tumor grading, many current approaches rely on single-scale feature extraction or isolated task training [56, 57]. Such methods often fail to leverage the rich, multi-scale semantic information that is inherently generated during the segmentation process [58]. Moreover, the lack of unified frameworks that can simultaneously handle pixel-level localization and image-level classification often results in sub-optimal feature representations [59].

4. Proposed method

4.1. Overall Design Philosophy

To address the dual challenges of precise pixel-level localization and accurate image-level grading, we propose a unified architecture governed by three core design principles. First, rather than employing separate encoders for each task, we utilize a shared Hierarchical Swin Backbone. This backbone extracts a single, robust multi-scale feature pyramid used simultaneously by both task heads. Second, the segmentation pathway prioritizes spatial fidelity through our novel modules. The CBE, which captures anisotropic (direction-specific) dependencies often missed by standard isotropic kernels, and the HAF, which actively aligns and recalibrates semantic features between encoder and decoder rather than resorting to passive concatenation. Third, the classification pathway is designed for maximum efficiency; it leverages an intelligent Feature Fusion strategy that distills the pre-computed backbone features into a compact descriptor, allowing for high-accuracy grading without the overhead of a secondary feature extractor.

4.2. Segmentation task

4.2.1. Overview

In this part, we present a novel transformer-based architecture for accurate and efficient tumor segmentation in brain MRI scans. The overall framework adopts an encoder-decoder structure, where the encoder extracts multi-scale semantic representations from the input image, and the decoder progressively reconstructs the segmentation map using enhanced contextual features.

The encoder is built upon a hierarchical Swin Transformer backbone, which efficiently captures both local and global dependencies through shifted window-based self-attention. To further refine the extracted features, we use the CBE, which enriches feature representations using a sequence of lightweight yet effective operations, including shifted multilayer perceptrons (MLPs) and gated encoding units.

To preserve high-resolution semantic details during decoding, we design a lightweight decoder that includes Adaptive Context Aggregator blocks, which adaptively fuse local and global context from the encoder outputs. Additionally, we propose a HAF module that integrates multi-scale features from different encoder levels through a combination of Swin Transformer blocks and deformable convolutions, allowing the model to capture hierarchical dependencies and spatial variations effectively.

Together, these components enable our model to achieve robust segmentation performance while maintaining computational efficiency. The complete architecture is illustrated in Figure 1.

4.2.2. Encoder Architecture

The encoder of the proposed segmentation model is designed to extract rich hierarchical features from brain MRI scans using a multi-stage Swin Transformer-based backbone. It begins with a *patch partition* module, which splits the input image into non-overlapping patches. These patches are then flattened and mapped to a fixed-dimensional embedding space through a Linear Embedding layer.

Following this, the encoder comprises three repeated stages, each consisting of two Swin Transformer Blocks followed by a *patch merging* layer. The Swin Transformer blocks utilize a window-based self-attention mechanism with a shifted window strategy, allowing the model to effectively capture local and non-local dependencies with reduced computational cost. The patch merging operation downsamples the spatial resolution while increasing the feature dimensionality, forming a hierarchical representation.

This hierarchical structure enables the encoder to progressively capture multi-scale semantic information, which is crucial for segmenting tumors of varying sizes and shapes. The feature maps from different stages are later passed to the HAF modules, enabling multi-level feature interaction and refinement in the decoding process.

4.2.3. Hierarchical Attention Fusion (HAF) Module

Standard U-shaped architectures typically rely on simple skip connections (concatenation) to recover spatial details. However, this approach often introduces "semantic noise" because the low-level encoder features are not aligned with the high-level decoder context. The HAF module addresses this by introducing an active alignment mechanism. Unlike a standard transformer block, HAF employs a dual-branch design: it utilizes a Swin block to capture long-range semantic context and a parallel Deformable Convolution branch to adaptively align local geometric structures. This ensures that the encoder features (x_{skip}) are spatially and semantically recalibrated before being merged with the decoder stream (x_{decoder}), significantly reducing the semantic gap and improving boundary precision.

At each stage of the decoder, the HAF module takes two inputs:

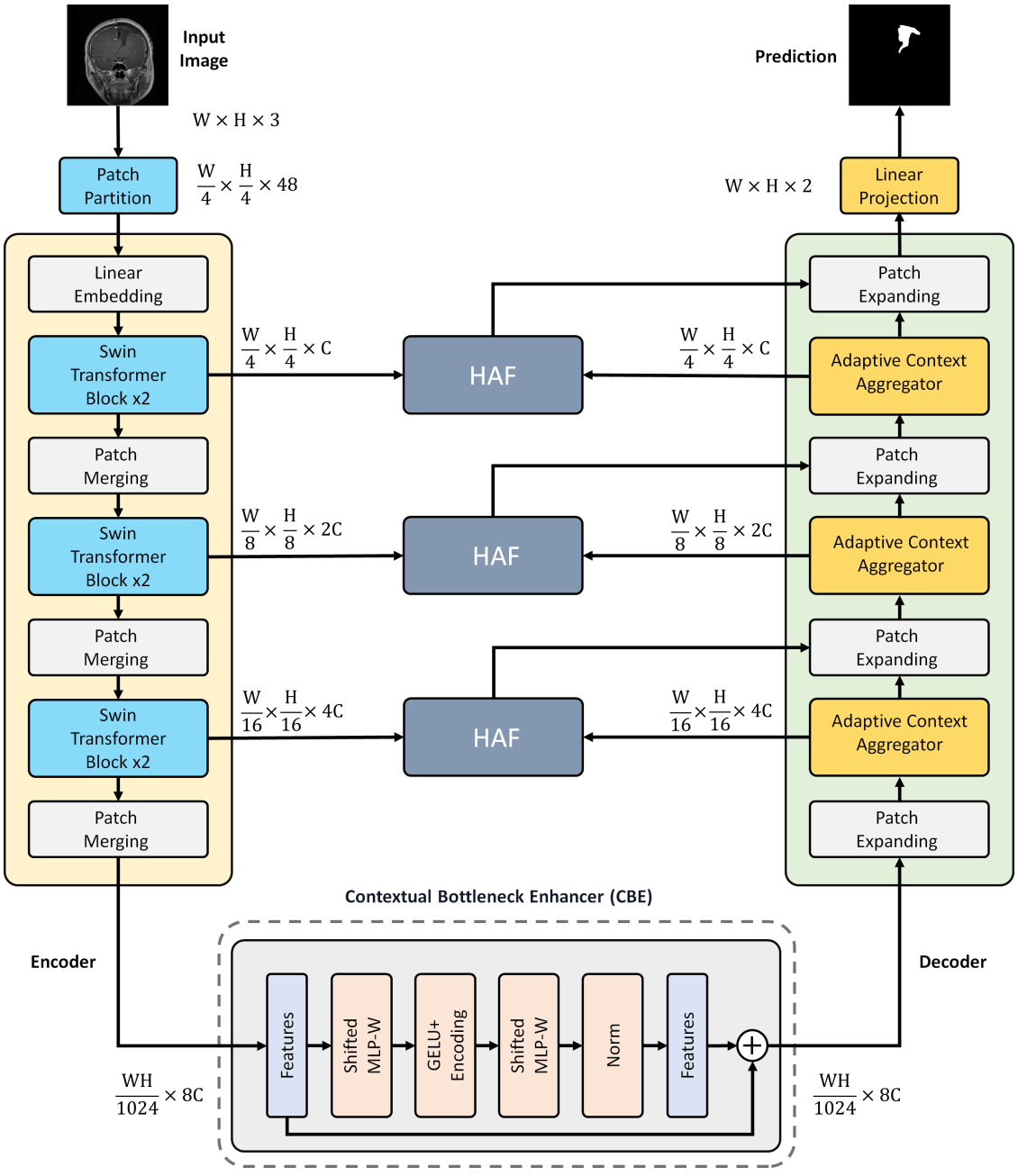

Figure 1: The overview of the Swin-HAFNet

- x_{skip}: The skip connection feature map from the encoder.
- x_{decoder}: the upsampled feature map from the previous decoder layer.

The decoder feature x_{decoder} is first passed through a *Swin Transformer Block* to refine contextual dependencies and enhance representation. As shown in Equation 1, the refined feature is then concatenated with the corresponding encoder feature x_{skip} along the channel dimension.

$$x_{\text{cat}} = \text{Concat}(x_{\text{skip}},\ \text{Swin}(x_{\text{decoder}})) \tag{1}$$

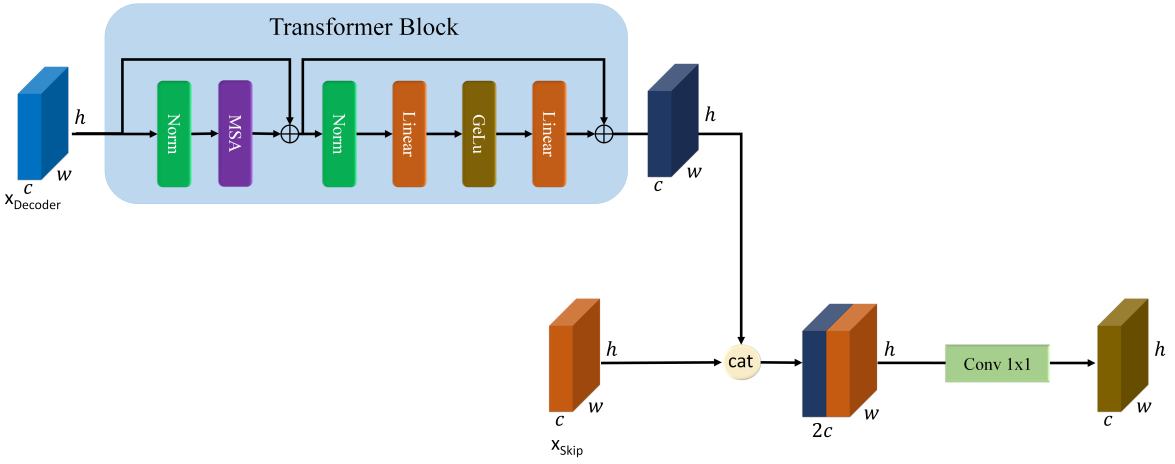

Figure 2: Hierarchical Attention Fusion module

This concatenated feature map x_{cat} is then projected through a 1×1 convolutional layer (Equation 2).

$$x_{\text{out}} = \text{Conv}_{1\times1}(x_{\text{cat}}) \tag{2}$$

The resulting output x_{out} maintains the same spatial resolution as x_{decoder} and serves as the input for the next step in the decoder. This fusion strategy allows the network to maintain fine spatial information while enriching the semantic features through transformer-based attention. The Architecture of HAF is shown in Figure 2.

4.2.4. Contextual Bottleneck Enhancer (CBE)

While standard Transformer blocks effectively model global dependencies, they are computationally intensive and often treat spatial directions isotropically. Medical images, however, frequently contain anatomical structures with distinct directional biases (e.g., elongated tumor boundaries). The CBE is designed to capture these anisotropic spatial dependencies efficiently. Instead of heavy self-attention, CBE decomposes spatial context modeling into two orthogonal steps: horizontal and vertical token mixing. By sequentially processing features along the width and height axes using shifted-MLPs, the module captures long-range dependencies with a fraction of the parameters of a full transformer block.

Given an input feature map X, we first perform a spatial shift along the width axis to encourage cross-region interaction:

$$X^{W}_{\text{shift}} = \text{Shift}_{W}(X). \tag{3}$$

The shifted features are then projected into token embeddings using a width-aware MLP, as shown in Equation (3):

$$T_{W} = \text{MLP}_{W}(X^{W}_{\text{shift}}). \tag{4}$$

Next, the tokens are refined using a depth-wise convolution followed by a GELU activation to introduce non-linearity and enhance local context:

$$Y = \text{GELU}(\text{DWConv}(T_{W})). \tag{5}$$

To further aggregate spatial information, the refined features are shifted along the height axis:

$$Y^{H}_{\text{shift}} = \text{Shift}_{H}(Y), \tag{6}$$

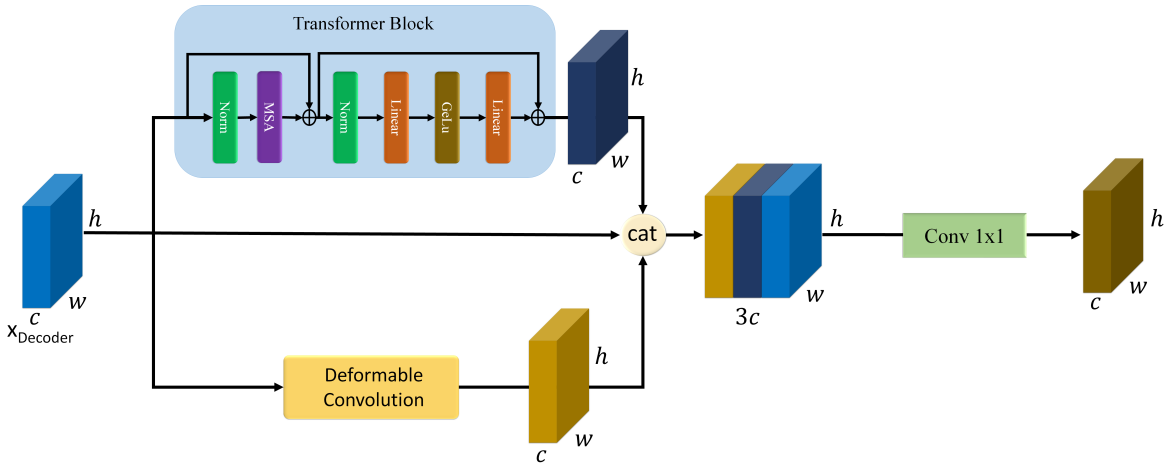

Figure 3: Structure of the Adaptive Context Aggregator module.

and subsequently processed by a second directional MLP:

$$T_H = \mathrm{MLP}_H(Y^H_{\mathrm{shift}}). \tag{7}$$

Finally, a residual connection combines the transformed tokens with the original tokenized input T, and the result is normalized to stabilize training:

$$Z = \mathrm{LN}(T_H + T). \tag{8}$$

As indicated by Equations (3)–(8), the CBE sequentially captures contextual dependencies along both spatial directions while preserving the original semantic information via residual learning.

4.2.5. Decoder Architecture

The decoder receives as input the enhanced representation produced by CBE and gradually reconstructs the segmentation map through a hierarchical upsampling process. Unlike traditional encoder-decoder architectures, our decoder integrates semantic context at multiple levels by leveraging the Adaptive Context Aggregator and HAF module.

At each stage of the decoder, the feature map undergoes a Patch Expanding operation to increase the spatial resolution. Following this, contextual features generated by the Adaptive Context Aggregator are fused with encoder features using the HAF module. The output of the HAF block serves as an enhanced skip connection, injected into the decoder pathway to guide the reconstruction process with both fine-grained and semantic details.

This structured design ensures that skip connections are not merely concatenations of encoder features, but rather semantically enriched representations aligned with the decoder's current context. After multiple stages of patch expansion and fusion, the final feature map is passed through a linear projection layer to produce the segmentation output. culminating in a final linear projection to generate the final prediction map.

4.2.6. Adaptive Context Aggregator

To effectively inject adaptive context into the decoder path, we employ the Adaptive Context Aggregator module. This block is responsible for enhancing the decoder features by integrating both local geometric and global semantic information.

As illustrated in Figure 3, the Adaptive Context Aggregator module receives the decoder feature map as input. It first processes this input through a Swin Transformer Block to capture long-range dependencies and global contextual cues. In parallel, the same input is passed through a Deformable Convolution layer to focus on important local structures and spatially variant patterns.

The outputs of both the Swin Transformer and the Deformable Convolution branches are concatenated and fused via a 1×1 convolution layer. This fusion ensures that both global and local contexts are adaptively aggregated in a computationally efficient manner. The resulting feature map serves two purposes: it is passed to the HAF module to refine the skip connection at the current decoder level, and it is also forwarded to the subsequent Patch Expanding block in the decoder pipeline. This dual role ensures both better feature fusion and more informed upsampling in the reconstruction process.

4.3. Classification task

4.3.1. Overview

In addition to tumor segmentation, we design a classification pipeline to discriminate between four diagnostic categories: *glioma*, *meningioma*, *pituitary*, and *non-tumorous*. A critical advantage of our design is its efficiency: rather than training a separate classification network, we intelligently reuse the frozen multi-scale feature representations already extracted by the segmentation backbone. The classification branch transforms these heterogeneous, high-dimensional segmentation features into a compact, discriminative descriptor through a sequence of dimension and spatial reduction steps. This allows the model to perform robust multi-class grading with negligible added computational cost, effectively turning the heavy lifting of the segmentation encoder into a "free" resource for classification. The overall structure of the classification framework is illustrated in Figure 4.

4.3.2. Backbone Architecture

The classification framework employs the Swin Transformer backbone, which naturally produces multi-scale feature representations through its hierarchical design. The backbone makes features at four progressive stages, capturing both fine spatial details in early layers and high-level semantic patterns in deeper layers. This multi-scale approach is particularly valuable for tumor classification, where diagnostic decisions depend on both localized texture features and global anatomical context.

4.3.3. Dimension Reduction

Each feature map $\mathbf{F}_i \in \mathbb{R}^{C_i \times H_i \times W_i}$ is first passed through a 1×1 convolution to reduce the channel dimension to 64. This projection ensures uniform channel size and reduces computational overhead. as shown in Equation 9 the operation is followed by a ReLU activation and Batch Normalization (BN) to improve representation stability and training convergence.

$$\mathbf{F}'_i = \mathrm{BN}\big(\sigma(\mathrm{Conv}_{1\times1}(\mathbf{F}_i))\big), \quad \mathbf{F}'_i \in \mathbb{R}^{64\times H_i \times W_i}, \tag{9}$$

where $\sigma(\cdot)$ denotes the ReLU function.

4.3.4. Spatial Reduction

To unify the spatial dimensions, we apply a 3×3 convolution with different stride settings to each feature map. Specifically, strides of $\{8, 4, 2, 1\}$ are used for the four successive stages of the Swin backbone. This design downscales all feature maps to a common resolution of 7×7, while preserving semantic richness (Equation 10).

$$\mathbf{S}_i = \mathrm{BN}\big(\sigma(\mathrm{Conv}_{3\times3,\ \mathrm{stride}=s_i}(\mathbf{F}'_i))\big), \quad \mathbf{S}_i \in \mathbb{R}^{64\times7\times7}. \tag{10}$$

4.3.5. Feature Fusion

Once channel and spatial dimensions are aligned, as shown in Equation 11 the four feature maps are concatenated along the channel axis:

$$\mathbf{F}_{cat} = \mathrm{Concat}(\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \mathbf{S}_4), \quad \mathbf{F}_{cat} \in \mathbb{R}^{256\times7\times7}. \tag{11}$$

This fusion operation aggregates complementary multi-scale cues, allowing the classifier to benefit simultaneously from low-level structural features and high-level semantic context.

4.3.6. Global Representation

To generate a fixed-dimensional representation invariant to spatial variance, a Global Average Pooling (GAP) layer is applied over $\mathbf{F}_{cat}$, producing a 256-dimensional vector:

$$\mathbf{z} = \mathrm{GAP}(\mathbf{F}_{cat}), \quad \mathbf{z} \in \mathbb{R}^{256}. \tag{12}$$

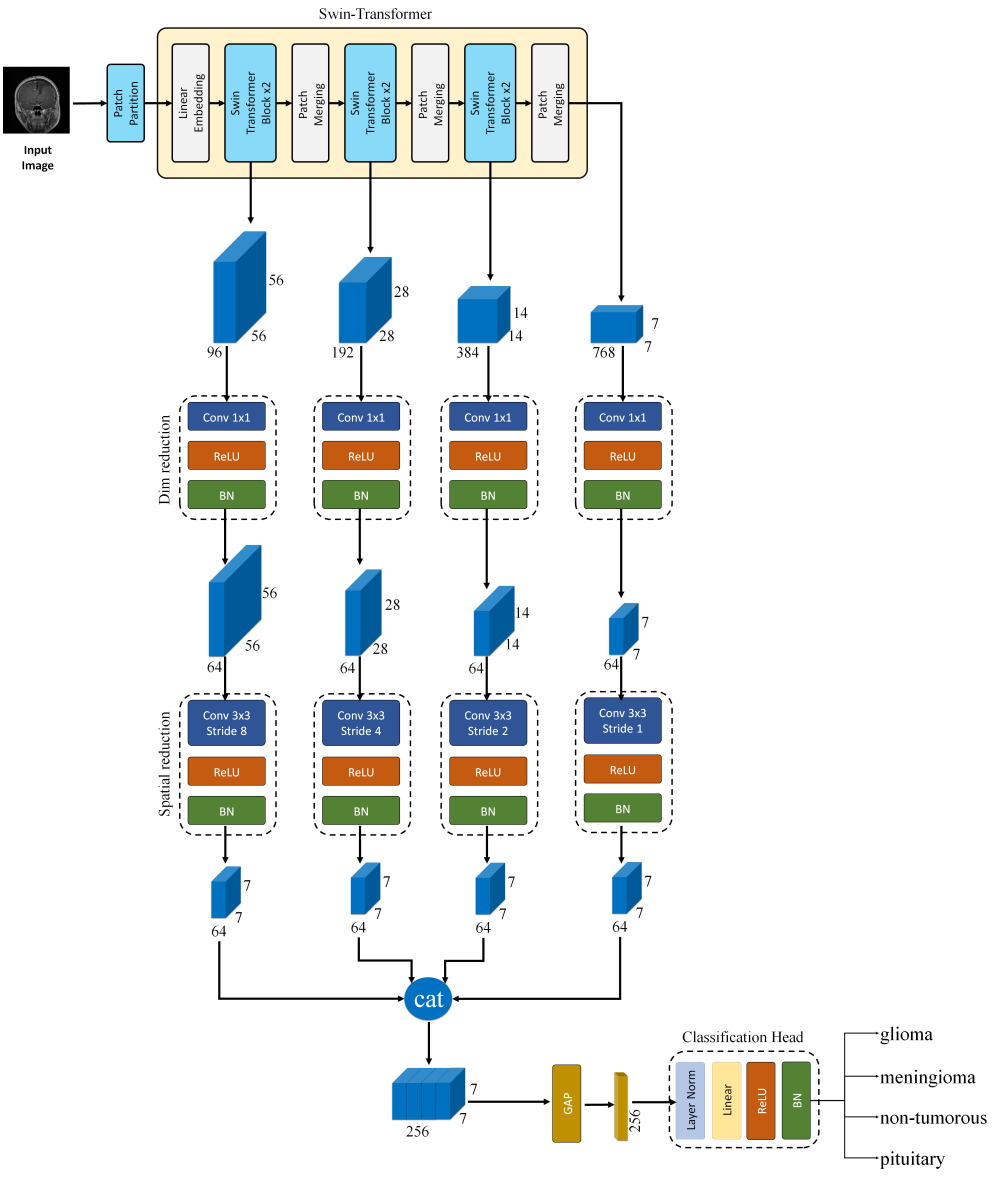

Figure 4: Overview of the proposed classification module. multi-scale feature maps from the Swin Transformer backbone are reduced, aligned, fused, and projected into class predictions.

4.3.7. Classification Head

The pooled vector $\mathbf{z}$ is passed through a lightweight classification head composed of Layer Normalization (LN), a fully connected linear projection, ReLU activation, and Batch Normalization:

$$\mathbf{y} = \mathrm{BN}\big(\sigma(\mathrm{Linear}(\mathrm{LN}(\mathbf{z})))\big), \tag{13}$$

where $\mathbf{y} \in \mathbb{R}^4$ are the class logits. A softmax function is then applied to obtain the final categorical distribution across the four tumor classes.

4.4. Loss Function

To train the proposed segmentation model, we utilize a compound loss function that combines the Binary Cross-Entropy (BCE) loss and the Dice loss. The BCE component focuses on pixel-level classification accuracy, while the

Dice loss emphasizes region-level consistency, which is particularly effective in addressing class imbalance commonly present in medical image segmentation tasks.

Given the ground truth mask $y \in \{0,1\}^N$ and the predicted probabilities $\hat{y} \in [0,1]^N$ for N pixels, the BCE loss is formulated as shown in Equation 14.

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right], \tag{14}$$

The Dice loss, which evaluates the overlap between predicted and ground truth regions, is defined as shown in Equation 15.

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \epsilon}, \tag{15}$$

where ϵ is a small constant added to avoid division by zero.

The final loss used to optimize the network combines these two components, as shown in Equation 16.

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{BCE}} + \mathcal{L}_{\text{Dice}}. \tag{16}$$

This joint formulation encourages both accurate boundary delineation and robust region-level segmentation.

For the classification task, we adopt the standard Cross-Entropy Loss (CE), which is widely used in multi-class recognition problems due to its effectiveness in penalizing incorrect predictions and encouraging confident probability distributions across classes. The CE loss measures the dissimilarity between the predicted categorical distribution and the ground truth label distribution.

Let C denote the total number of classes, $y \in \{1, \ldots, C\}$ be the ground truth label, and $\hat{p}_c$ be the predicted probability for class c after the softmax activation. The CE loss is defined as:

$$\mathcal{L}_{\text{CE}} = -\sum_{c=1}^{C} \mathbb{1}_{[y=c]} \log(\hat{p}_c), \tag{17}$$

where $\mathbb{1}_{[y=c]}$ is an indicator function that equals 1 if the true class is c and 0 otherwise. The predicted probabilities are obtained from the logits $\mathbf{y} \in \mathbb{R}^C$ via the softmax function:

$$\hat{p}_c = \frac{\exp(y_c)}{\sum_{j=1}^{C} \exp(y_j)}. \tag{18}$$

This formulation encourages the network to assign high probability to the correct class while suppressing the probabilities of incorrect classes. Unlike segmentation tasks where pixel-level overlap is a major concern, classification operates at the image level, making the Cross-Entropy loss a natural and effective choice.

Therefore, the total loss function for the classification branch is expressed as:

$$\mathcal{L}_{\text{classification}} = \mathcal{L}_{\text{CE}}. \tag{19}$$

5. Experimental Results

In this section, we present the experimental results of our proposed models on brain tumor MRI analysis, focusing on both segmentation and classification tasks. The experiments are designed to rigorously evaluate the performance and generalizability of the models, using standard metrics to provide a clear comparison across different settings. The results highlight the effectiveness of the proposed methods and offer insights into their potential applications in medical imaging, demonstrating how they can contribute to accurate and efficient diagnosis.

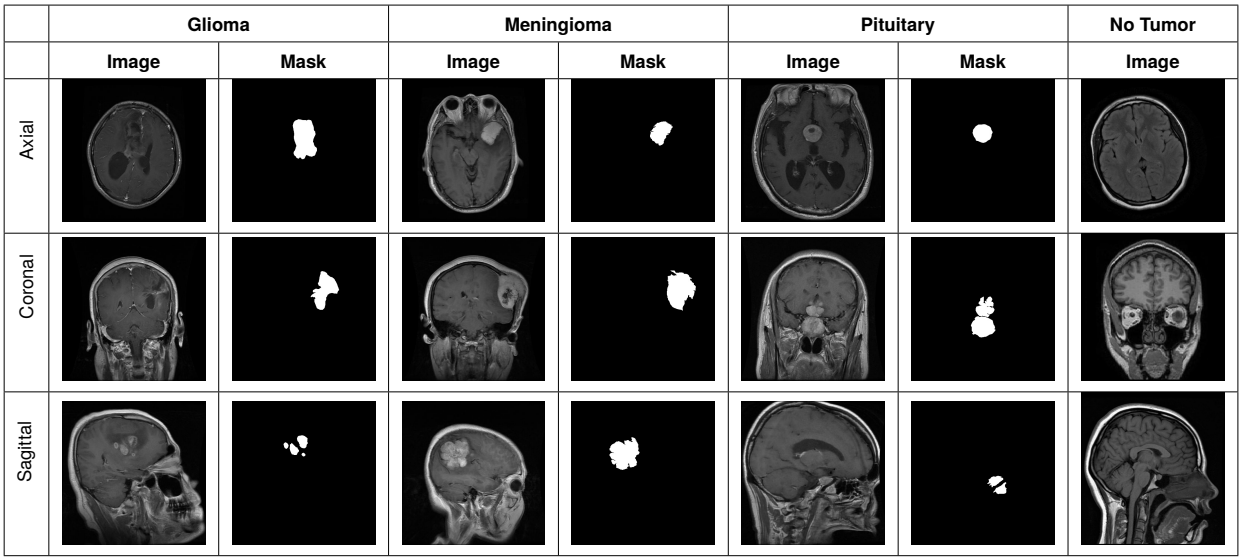

Figure 5: Representative T1-weighted MRI scans and corresponding expert-annotated segmentation masks from the BRISC dataset. The samples illustrate the three tumor types (Glioma, Meningioma, and Pituitary) alongside healthy control (No Tumor) cases across axial, coronal, and sagittal planes.

5.1. Dataset

We conducted experiments using the BRISC dataset, a large-scale, expert-annotated brain tumor MRI collection for segmentation and classification [60]. BRISC contains approximately 6,000 contrast-enhanced T1-weighted scans across four categories: glioma, meningioma, pituitary tumor, and no tumor, with pixel-level segmentation masks verified by radiologists. As illustrated in Fig. 5, the dataset includes images from axial, coronal, and sagittal planes, supporting robust model evaluation across different orientations. Its balanced class distribution and high-quality annotations make it a reliable benchmark for developing and testing brain tumor analysis models.

To assess the generalizability of our proposed method, we additionally utilized the Brain Tumors 256×256 dataset [61]. This enhanced dataset builds upon the "Uncovering Knowledge: A Clean Brain Tumor Dataset for Advanced Medical Research." Also, to further evaluate the segmentation robustness of Swin-HAFNet, we also conducted experiments on the Kaggle Brain Tumor Segmentation dataset [62]. This dataset consists of 3,064 T1-weighted contrast-enhanced MRI images with corresponding binary masks for three types of brain tumors: glioma, meningioma, and pituitary tumor. The inclusion of this additional benchmark allows for a more comprehensive validation of the model's pixel-level delineation capabilities across different imaging sources.

5.2. Evaluation Metrics

5.2.1. Segmentation Metric

In this part, we detail the evaluation metric employed to assess the performance of segmentation models. These metrics provide comprehensive insights into the efficacy of the models.

Intersection over Union (IoU) Intersection over Union (IoU), also known as the Jaccard Index, is a fundamental metric for evaluating binary segmentation tasks. It quantifies the overlap between the predicted tumor regions and the ground truth, normalized by their union [63]. For binary segmentation, IoU is computed as shown in Equation 20.

$$\text{IoU} = \frac{\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i - \sum_{i=1}^{N} y_i \hat{y}_i + \epsilon}, \tag{20}$$

where $y_i \in \{0, 1\}$ denotes the ground truth label, $\hat{y}_i \in \{0, 1\}$ represents the predicted label (after thresholding), and ϵ is a small constant added for numerical stability.

As shown in Equation 20, this formulation captures the pixel-wise overlap between the predicted and actual tumor regions and is particularly effective for evaluating segmentation quality, especially along object boundaries.

5.2.2. Classification Metrics

As commonly employed in the evaluation of multi-class classification models, metrics such as Accuracy, Precision, Recall, and F1-Score are widely utilized due to their effectiveness in assessing performance across diverse tasks [64, 65, 66].

Accuracy Accuracy measures the overall correctness of predictions across all four classes. It is defined as:

$$\text{Accuracy} = \frac{\sum_{i=1}^{C} \text{Correct Predictions for Class } i}{\text{Total Samples}} \tag{21}$$

where C denotes the total number of classes, and "Correct Predictions for Class i" represents the samples correctly classified as class i.

Precision Precision quantifies the proportion of correctly predicted positive instances for each class. For class i, Precision is defined as:

$$\text{Precision}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \tag{22}$$

In multi-class classification, Precision is averaged using either macro-averaging or weighted-averaging:

$$\text{Macro Precision} = \frac{1}{C} \sum_{i=1}^{C} \text{Precision}_i \tag{23}$$

$$\text{Weighted Precision} = \frac{\sum_{i=1}^{C} w_i \cdot \text{Precision}_i}{\sum_{i=1}^{C} w_i} \tag{24}$$

where w_i represents the proportion of samples in class i.

Recall Recall, or Sensitivity, measures the proportion of actual positive instances correctly identified by the model. For class i, Recall is defined as:

$$\text{Recall}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \tag{25}$$

For multi-class classification, Recall is averaged similarly to Precision:

$$\text{Macro Recall} = \frac{1}{C} \sum_{i=1}^{C} \text{Recall}_i \tag{26}$$

$$\text{Weighted Recall} = \frac{\sum_{i=1}^{C} w_i \cdot \text{Recall}_i}{\sum_{i=1}^{C} w_i} \tag{27}$$

F1-Score The F1-Score is the harmonic mean of Precision and Recall. For class i, it is defined as:

$$\text{F1-Score}_i = 2 \cdot \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{28}$$

For multi-class classification, F1-Score is averaged as follows:

$$\text{Macro F1-Score} = \frac{1}{C} \sum_{i=1}^{C} \text{F1-Score}_i \tag{29}$$

$$\text{Weighted F1-Score} = \frac{\sum_{i=1}^{C} w_i \cdot \text{F1-Score}_i}{\sum_{i=1}^{C} w_i} \tag{30}$$

By calculating these metrics per class and aggregating them through macro- or weighted-averaging, we ensure a detailed evaluation of model performance, particularly in datasets with imbalanced class distributions.

5.2.3. Segmentation results

To evaluate the effectiveness of our proposed **Swin-HAFNet**, we conducted a comparative study against a diverse set of brain tumor segmentation models, including traditional convolutional architectures, attention-enhanced methods, and transformer-based approaches. The baselines include UNet [67], UNet++ [68], LinkNet [69], MANet [70], DeepLabV3+ [71], PAN [72], EINet [73], EU-Net [74], DAD [75], and BASNet [76], as well as two recent transformer-enhanced models, SaberNet [77] and ABANet [78].

Each model was evaluated using the mean Intersection over Union (mIoU) metric for three tumor types: *Glioma*, *Meningioma*, and *Pituitary*. Furthermore, we report a *weighted mIoU*, which is calculated based on the proportion of samples belonging to each tumor type, providing a more representative performance indicator across the dataset.

As summarized in Table 1, our proposed Swin-HAFNet achieves the highest mIoU scores across all tumor types, particularly excelling in segmenting *Meningioma* and *Pituitary* tumors. It also outperforms all competing methods in terms of weighted mIoU, demonstrating the strength of our architectural design in capturing multi-scale contextual features and handling inter-class variability in medical image segmentation.

In particular, our proposed method achieves the highest weighted mIoU of 82.4%, surpassing the next best-performing model (Saber et al. [77]) by a margin of 2.6%. This improvement underscores the robustness of our approach in handling heterogeneous tumor types. Importantly, the reported weighted mIoU is calculated as a weighted average based on the number of samples in each tumor class to provide a more realistic assessment under dataset imbalance. Unlike simple arithmetic means, the weighted mean better reflects the overall segmentation performance in real-world clinical distributions.

Moreover, our model consistently outperforms existing baselines across all tumor categories, with notable improvements observed in the segmentation of glioma and meningioma tumors. These gains can be attributed to the architectural choices that enhance multi-scale feature extraction and contextual representation, which are particularly beneficial for capturing diverse morphological structures in brain tumors.

To further validate the robustness and generalization of Swin-HAFNet, we conducted additional benchmarking on the Kaggle Brain Tumor Segmentation dataset, as detailed in Table 2. Our proposed method achieved a state-of-the-art IoU of 0.7142 and a Dice score of 0.8294, consistently outperforming a wide array of recent custom implementations. Notably, Swin-HAFNet demonstrated a significant performance gain over standard transformer-based architectures like UNETR and even more recent hybrid models such as MHA-UNet SegGAN. These results confirm that the integration of the HAF module and the CBE allows the model to effectively adapt to different data distributions while maintaining high-fidelity boundary delineation, even when compared to models utilizing adversarial learning or complex attention-enhanced residual structures.

5.2.4. Classification Results

To evaluate the classification performance on BRISC dataset, we conducted a comprehensive analysis of several baseline models alongside our Proposed Method for classifying brain tumor types: *Glioma*, *Meningioma*, *Pituitary*,

Table 1
IoU (%) for Brain Tumor Segmentation Models on Different Tumor Types. Weighted mIoU is calculated as a weighted average based on the number of samples per tumor type: Glioma, Meningioma, Pituitary.

Model	mIoU Glioma	mIoU Meningioma	mIoU Pituitary	Weighted mIoU
UNet [67]	69.7	77.1	79.3	75.7
UNet++ [68]	71.7	74.2	79.7	75.3
MANet [70]	72.4	77.5	78.0	76.2
LinkNet [69]	71.7	74.8	79.0	75.3
DeepLabV3+ [71]	72.0	77.5	78.7	76.3
PAN [72]	72.0	74.5	80.7	75.9
EINet [73]	73.6	78.4	80.3	77.7
EU-Net [74]	71.7	76.1	78.3	75.6
DAD [75]	75.2	80.4	82.3	79.5
BASNet [76]	74.0	77.5	81.7	77.9
ABANet [78]	72.4	80.4	84.7	79.5
SaberNet [77]	74.0	82.4	84.3	80.6
Swin-HAFNet (our)	**76.0**	**85.0**	**85.3**	**82.4**

Table 2
Comparison of segmentation performance on the Kaggle Brain Tumor Segmentation dataset [62]. The proposed Swin-HAFNet is compared against various custom implementations.

Model Name	Year	IoU (Jaccard)	Dice (F1)
Custom U-Net (ResNet-34) [79]	2023	0.6700	-
Custom U-Net [80]	2024	0.6622	0.7461
Custom UNETR [81]	2024	0.5091	0.6042
Custom UNETR [82]	2024	0.5323	0.6852
Attention ResUNet [83]	2025	0.6637	0.7443
Custom U-Net [84]	2025	0.6249	0.7682
MHA-UNet SegGAN [85]	2025	0.6961	0.8208
Swin-HAFNet (our)	**2026**	**0.7142**	**0.8294**

and *non-tumorous*. The evaluated models include ResNet50, ResNet101, DenseNet121, DenseNet169, MobileNetV2, MobileNetV3, EfficientNetB0, EfficientNetB1, EfficientNetB2, Xception, VGG16, VGG19, InceptionV3, and our Proposed Method. Each model was trained and tested three times to ensure robust and reliable results, with performance reported as the mean and standard deviation of key metrics: Precision, Recall, F1-Score, and Accuracy.

The evaluation metrics were computed per class, alongside macro and weighted averages, to provide a comprehensive view of model performance across diverse tumor types. The macro average treats all classes equally, while the weighted average accounts for class imbalance by weighting each class's contribution based on the number of samples, offering a realistic assessment of performance in clinical scenarios where tumor type distributions may vary.

As presented in Table 3, our proposed method achieves the highest overall performance, with a weighted average F1-Score of 0.9963 ± 0.0015 and an accuracy of 0.9963 ± 0.0015, demonstrating exceptional consistency and robustness. Notably, it achieves near-perfect performance across all classes, with an F1-Score of 0.9988 ± 0.0021 for non-tumorous and 0.9961 ± 0.0020 for Glioma, surpassing all baseline models. These gains can be attributed to architectural

Table 3
Per-Class and Average Classification Performance (%) for Brain Tumor Classification Models. Metrics are reported as mean ± standard deviation over three runs.

Model	Class	Precision	Recall	F1-Score	Accuracy
ResNet50	glioma	0.9868 ± 0.0098	0.9751 ± 0.0164	0.9808 ± 0.0103	-
	meningioma	0.9815 ± 0.0150	0.9673 ± 0.0226	0.9741 ± 0.0087	-
	no_tumor	0.9906 ± 0.0107	0.9952 ± 0.0083	0.9929 ± 0.0035	-
	pituitary	0.9756 ± 0.0285	0.9967 ± 0.0000	0.9859 ± 0.0146	-
	Macro Avg	0.9836 ± 0.0064	0.9836 ± 0.0077	0.9834 ± 0.0072	-
	Weighted Avg	0.9823 ± 0.0076	0.9820 ± 0.0080	0.9820 ± 0.0080	0.9820 ± 0.0080
ResNet101	glioma	0.9726 ± 0.0347	0.9869 ± 0.0082	0.9794 ± 0.0138	-
	meningioma	0.9879 ± 0.0081	0.9575 ± 0.0279	0.9722 ± 0.0106	-
	no_tumor	0.9883 ± 0.0107	0.9905 ± 0.0165	0.9893 ± 0.0037	-
	pituitary	0.9793 ± 0.0113	0.9956 ± 0.0020	0.9874 ± 0.0066	-
	Macro Avg	0.9820 ± 0.0074	0.9826 ± 0.0093	0.9821 ± 0.0086	-
	Weighted Avg	0.9815 ± 0.0085	0.9810 ± 0.0092	0.9810 ± 0.0092	0.9809 ± 0.0092
DenseNet121	glioma	0.4838 ± 0.4753	0.5879 ± 0.5095	0.5197 ± 0.4731	-
	meningioma	0.6640 ± 0.5751	0.2800 ± 0.4569	0.3178 ± 0.4966	-
	no_tumor	0.5324 ± 0.4095	0.6976 ± 0.4871	0.4721 ± 0.4204	-
	pituitary	0.4502 ± 0.4128	0.6422 ± 0.5574	0.5259 ± 0.4678	-
	Macro Avg	0.5326 ± 0.4550	0.5519 ± 0.3376	0.4589 ± 0.4310	-
	Weighted Avg	0.5356 ± 0.4650	0.5253 ± 0.3850	0.5253 ± 0.3850	0.4531 ± 0.4390
DenseNet169	glioma	0.9543 ± 0.0690	0.3543 ± 0.5356	0.3840 ± 0.5173	-
	meningioma	0.7522 ± 0.2090	0.8007 ± 0.2607	0.7560 ± 0.2021	-
	no_tumor	0.4841 ± 0.4507	0.9738 ± 0.0393	0.5754 ± 0.3763	-
	pituitary	0.3333 ± 0.5774	0.3322 ± 0.5754	0.3328 ± 0.5764	-
	Macro Avg	0.6310 ± 0.3116	0.6152 ± 0.3305	0.5121 ± 0.4160	-
	Weighted Avg	0.6404 ± 0.3020	0.5710 ± 0.3689	0.5710 ± 0.3689	0.5093 ± 0.4169
MobileNetV2	glioma	0.3026 ± 0.0558	0.8517 ± 0.1229	0.4418 ± 0.0494	-
	meningioma	0.6667 ± 0.5774	0.0120 ± 0.0180	0.0233 ± 0.0348	-
	no_tumor	0.5343 ± 0.3066	0.6548 ± 0.1750	0.5249 ± 0.0899	-
	pituitary	0.1412 ± 0.2445	0.0400 ± 0.0693	0.0623 ± 0.1080	-
	Macro Avg	0.4112 ± 0.2078	0.3896 ± 0.0345	0.2631 ± 0.0307	-
	Weighted Avg	0.3980 ± 0.2379	0.3237 ± 0.0264	0.3237 ± 0.0264	0.2115 ± 0.0366
MobileNetV3	glioma	0.8912 ± 0.0353	0.9777 ± 0.0082	0.9321 ± 0.0154	-
	meningioma	0.9755 ± 0.0050	0.8639 ± 0.0334	0.9160 ± 0.0175	-
	no_tumor	0.9445 ± 0.0308	1.0000 ± 0.0000	0.9713 ± 0.0165	-
	pituitary	0.9679 ± 0.0073	0.9733 ± 0.0208	0.9706 ± 0.0138	-
	Macro Avg	0.9448 ± 0.0148	0.9537 ± 0.0113	0.9475 ± 0.0140	-
	Weighted Avg	0.9475 ± 0.0123	0.9447 ± 0.0140	0.9447 ± 0.0140	0.9442 ± 0.0142
EfficientNetB0	glioma	0.9960 ± 0.0000	0.9882 ± 0.0000	0.9921 ± 0.0000	-
	meningioma	0.9934 ± 0.0000	0.9869 ± 0.0000	0.9902 ± 0.0000	-
	no_tumor	0.9929 ± 0.0000	1.0000 ± 0.0000	0.9964 ± 0.0000	-
	pituitary	0.9868 ± 0.0000	0.9967 ± 0.0000	0.9917 ± 0.0000	-
	Macro Avg	0.9923 ± 0.0000	0.9929 ± 0.0000	0.9926 ± 0.0000	-
	Weighted Avg	0.9920 ± 0.0000	0.9920 ± 0.0000	0.9920 ± 0.0000	0.9920 ± 0.0000
EfficientNetB1	glioma	0.9987 ± 0.0023	0.9921 ± 0.0000	0.9954 ± 0.0011	-
	meningioma	0.9933 ± 0.0001	0.9750 ± 0.0136	0.9840 ± 0.0070	-
	no_tumor	0.9976 ± 0.0041	1.0000 ± 0.0000	0.9988 ± 0.0021	-
	pituitary	0.9773 ± 0.0116	1.0000 ± 0.0000	0.9885 ± 0.0059	-
	Macro Avg	0.9918 ± 0.0036	0.9918 ± 0.0034	0.9917 ± 0.0036	-
	Weighted Avg	0.9905 ± 0.0040	0.9903 ± 0.0042	0.9903 ± 0.0042	0.9903 ± 0.0042
EfficientNetB2	glioma	0.9919 ± 0.0040	0.9712 ± 0.0164	0.9814 ± 0.0091	-
	meningioma	0.9699 ± 0.0128	0.9782 ± 0.0105	0.9740 ± 0.0084	-
	no_tumor	0.9906 ± 0.0107	1.0000 ± 0.0000	0.9953 ± 0.0054	-
	pituitary	0.9879 ± 0.0082	0.9922 ± 0.0077	0.9900 ± 0.0017	-
	Macro Avg	0.9851 ± 0.0054	0.9854 ± 0.0047	0.9852 ± 0.0051	-
	Weighted Avg	0.9838 ± 0.0049	0.9837 ± 0.0049	0.9837 ± 0.0049	0.9837 ± 0.0049
Xception	glioma	0.0847 ± 0.1466	0.3333 ± 0.5774	0.1350 ± 0.2339	-
	meningioma	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	-
	no_tumor	0.0933 ± 0.0808	0.6667 ± 0.5774	0.1637 ± 0.1418	-
	pituitary	0.0000 ± 0.0000	0.0000 ± 0.0000	0.0000 ± 0.0000	-
	Macro Avg	0.0445 ± 0.0165	0.2500 ± 0.0000	0.0747 ± 0.0230	-
	Weighted Avg	0.0346 ± 0.0259	0.1780 ± 0.0658	0.1780 ± 0.0658	0.0572 ± 0.0395
VGG16	glioma	0.9803 ± 0.0150	0.9396 ± 0.0741	0.9582 ± 0.0335	-
	meningioma	0.9427 ± 0.0509	0.9684 ± 0.0019	0.9549 ± 0.0257	-
	no_tumor	0.9790 ± 0.0068	0.9976 ± 0.0041	0.9882 ± 0.0020	-
	pituitary	0.9867 ± 0.0099	0.9822 ± 0.0117	0.9844 ± 0.0051	-
	Macro Avg	0.9722 ± 0.0132	0.9720 ± 0.0172	0.9714 ± 0.0162	-
	Weighted Avg	0.9706 ± 0.0157	0.9693 ± 0.0177	0.9693 ± 0.0177	0.9692 ± 0.0178
VGG19	glioma	0.9484 ± 0.0351	0.9541 ± 0.0421	0.9502 ± 0.0081	-
	meningioma	0.9624 ± 0.0130	0.9434 ± 0.0068	0.9527 ± 0.0030	-
	no_tumor	0.9725 ± 0.0231	0.9976 ± 0.0041	0.9848 ± 0.0111	-
	pituitary	0.9744 ± 0.0287	0.9745 ± 0.0267	0.9739 ± 0.0039	-
	Macro Avg	0.9645 ± 0.0061	0.9674 ± 0.0041	0.9654 ± 0.0047	-
	Weighted Avg	0.9639 ± 0.0039	0.9630 ± 0.0044	0.9630 ± 0.0044	0.9629 ± 0.0044
InceptionV3	glioma	0.6564 ± 0.5686	0.5315 ± 0.4983	0.5778 ± 0.5128	-
	meningioma	0.8972 ± 0.1694	0.6645 ± 0.5132	0.6401 ± 0.4458	-
	no_tumor	0.5887 ± 0.4180	0.9905 ± 0.0165	0.6719 ± 0.3794	-
	pituitary	0.6629 ± 0.5741	0.5722 ± 0.5136	0.6104 ± 0.5343	-
	Macro Avg	0.7013 ± 0.3670	0.6897 ± 0.3750	0.6250 ± 0.4676	-
	Weighted Avg	0.7225 ± 0.3490	0.6487 ± 0.4312	0.6487 ± 0.4312	0.6198 ± 0.4797
Swin-HAFNet (our)	glioma	0.9974 ± 0.0023	0.9948 ± 0.0023	0.9961 ± 0.0020	-
	meningioma	0.9946 ± 0.0018	0.9945 ± 0.0068	0.9945 ± 0.0025	-
	no_tumor	0.9976 ± 0.0041	1.0000 ± 0.0000	0.9988 ± 0.0021	-
	pituitary	0.9967 ± 0.0033	0.9978 ± 0.0019	0.9972 ± 0.0009	-
	Macro Avg	0.9966 ± 0.0018	0.9968 ± 0.0013	0.9967 ± 0.0016	-
	Weighted Avg	0.9963 ± 0.0015	0.9963 ± 0.0015	0.9963 ± 0.0015	**0.9963 ± 0.0015**

innovations that enhance multi-scale feature extraction and contextual representation, which are particularly effective for capturing the diverse morphological structures of brain tumors.

Among the baseline models, EfficientNetB0 performs strongly, with a weighted average F1-score of 0.9920 ± 0.0000 and an accuracy of 0.9920 ± 0.0000, achieving perfect recall (1.0000 ± 0.0000) for the *non-tumorous* class.

Table 4
Comparison of classification performance on the Brain Tumors 256×256 Kaggle dataset. The proposed method is compared against top-performing public kernels and baseline models.

Model Name	Accuracy (%)	F1-Score (%)
DenseNet121	93.00	93.00
MobileNetV2	92.00	92.00
VGG19	91.29	-
InceptionV3	89.00	90.00
ResNet50	65.32	-
DeiT-Tiny + Fuzzy Attention [86]	95.81	96.18
Custom CNN [87]	89.07	-
Custom CNN [88]	89.00	90.00
Custom CNN [89]	88.89	-
Swin Transformer [90]	96.00	96.00
ViT [91]	93.00	94.00
Ensemble Learning [92]	77.00	77.00
Swin-HAFNet (our)	**97.60**	**97.64**

EfficientNetB1 follows closely with a weighted F1-score of 0.9903 ± 0.0042, while ResNet50 and MobileNetV3 deliver competitive results (weighted F1-scores of 0.9820 ± 0.0080 and 0.9447 ± 0.0140, respectively).

In contrast, Xception exhibits the lowest performance, with a weighted F1-score of 0.1780 ± 0.0658, failing entirely on meningioma and pituitary (F1-score: 0.0000 ± 0.0000). Similarly, DenseNet121 and DenseNet169 show unstable performance, with high standard deviations, indicating limited generalizability. MobileNetV2 also struggles, particularly with meningioma (recall: 0.0120 ± 0.0180), likely due to insufficient model capacity.

The VGG variants (VGG16 and VGG19) achieve moderate performance, with weighted F1-scores of 0.9693 ± 0.0177 and 0.9630 ± 0.0044, respectively, while InceptionV3 shows inconsistent results (weighted F1-score: 0.6487 ± 0.4312), reflecting challenges in handling complex tumor morphology or class imbalances.

This evaluation underscores the strong performance of EfficientNet models, particularly EfficientNetB0, which combines high accuracy with remarkable stability across all tumor types. The results validate the utility of our dataset for developing reliable diagnostic tools, while the stark performance differences across architectures emphasize the importance of model selection in medical imaging tasks, where precision and consistency are critical. This work establishes a robust benchmark for brain tumor classification and provides a foundation for future research to explore diverse models and training protocols using this dataset.

To further validate the robustness of Swin-HAFNet, we compared its performance against state-of-the-art models on the Brain Tumors dataset [61]. As detailed in Table 4, our method was evaluated against several architectures including DeiT-Tiny with Fuzzy Attention, Swin Transformer, ViT, and various custom CNN implementations.

Our proposed method achieved the highest performance with an accuracy of 97.60% and an F1-score of 97.64%, surpassing the top-performing baseline, DeiT-Tiny + Fuzzy Attention [86], which achieved an F1-score of 96.18%. Notably, our model outperformed the standard Swin Transformer implementation by Dubail [90] (96.0% F1-score) and the Vision Transformer (ViT) [91] (94.0% F1-score). Conventional CNN-based approaches, such as VGG19 and custom CNN architectures [87, 88, 89], yielded lower accuracies ranging from 88.0% to 91.0%, while ensemble learning methods [92] struggled significantly with an F1-score of 77.0%. These results confirm that Swin-HAFNet maintains superior generalization capabilities across different datasets, effectively handling the specific preprocessing and characteristics of the Brain Tumors 256×256 benchmark.

5.3. Ablation study

5.3.1. Classification Ablation Study

To validate the architectural design of our classification branch, we conducted a step-by-step ablation study. This analysis investigates the impact of the proposed classification head, the dimension and spatial reduction mechanisms, and the multi-scale feature fusion strategy. The results are summarized in Table 5.

We established a baseline using a Simple classifier, consisting of a global average pooling layer followed by a single linear layer, applied directly to the deepest feature map of the backbone. This baseline achieved an accuracy of

Table 5
Ablation study of the proposed Classification Branch. The impact of the Classifier Head architecture, Dimension Reduction, Spatial Reduction, and multi-scale feature fusion is evaluated incrementally.

Classifier Head	Dim. Red.	Spat. Red.	Multi-Scale	Precision	Recall	F1-Score	Accuracy
Simple	–	–	–	96.5	96.4	96.4	96.4
Head Block	–	–	–	97.9	97.8	97.8	97.8
Head Block	✓	–	–	98.1	98.1	98.1	98.1
Head Block	–	✓	–	98.7	98.7	98.7	98.7
Head Block	✓	✓	–	99.0	99.0	99.0	99.0
Head Block	✓	✓	✓	**99.6**	**99.6**	**99.6**	**99.6**

96.4%. Replacing this simple linear layer with our proposed Head Block resulted in a significant performance boost, increasing the F1-score to 97.8%. This highlights the importance of proper feature normalization and non-linearity in the final classification stage.

We then evaluated the feature reduction modules independently. Introducing Dimension Reduction alone improved the F1-score to 98.1%, while applying Spatial Reduction alone proved even more effective, reaching 98.7%. When both reduction mechanisms were combined, the model achieved a robust 99.0% accuracy, demonstrating that compact and spatially aligned feature representations are crucial for efficient learning.

Finally, the integration of multi-scale features(aggregating information from all four stages of the Swin Transformer backbone) yielded the highest performance. By fusing low-level texture details with high-level semantic context, the complete proposed method achieved a peak F1-score and accuracy of 99.6%. These results confirm that a holistic view of the feature hierarchy, enabled by our unified reduction and fusion strategy, is essential for accurate brain tumor grading. These results also shown in Figure 6.

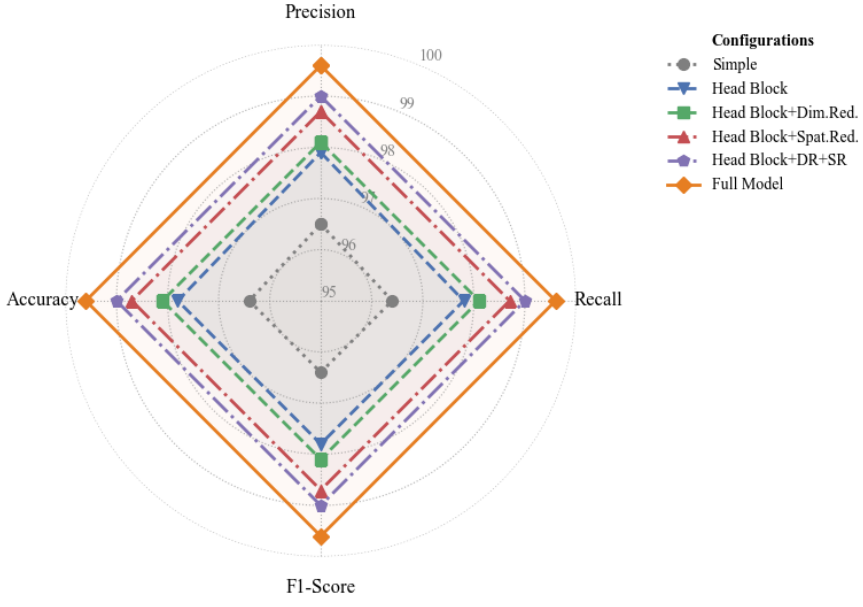

Figure 6: The performance plot of different configuration of Swin-HAFNet in classification task

Table 6
The performance of different configuration of Swin-HAFNet in segmentation task.

baseline	HAF	CBE	mIoU Glioma	mIoU Meningioma	mIoU Pituitary	Weighted mIoU
✓			73.6	82.4	82.3	79.8
✓	✓		75.2	83.5	83.6	81.1
✓		✓	74.4	83.8	84.7	81.3
✓	✓	✓	76.0	85.0	85.3	82.4

5.3.2. Segmentation Ablation Study

To evaluate the contribution of each component in the proposed Swin-HAFNet architecture for segmentation task, we conducted an ablation study. Four configurations were tested by progressively integrating the HAF module and the CBE into the baseline. The results are reported in Table 6.

As shown in Figure 7 the baseline model includes a simple Swin-UNet structure without the HAF or CBE modules. Adding the HAF module alone leads to a notable improvement across all tumor types, increasing the weighted mIoU from 79.8% to 81.1%. This indicates that the hierarchical attention fusion effectively enhances the representation of skip connections.

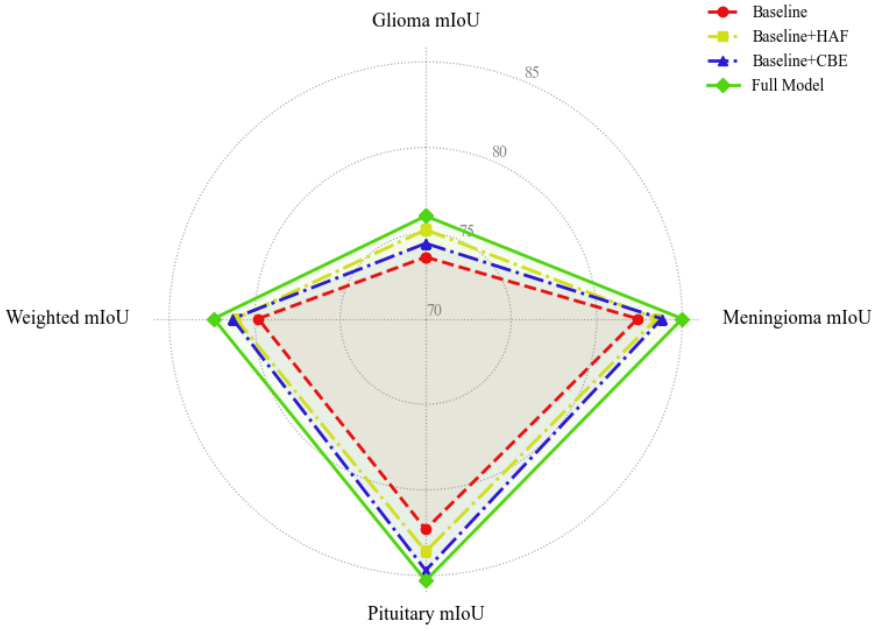

Figure 7: The performance plot of different configuration of Swin-HAFNet in segmentation task

Incorporating the CBE module without HAF further improves performance, yielding a weighted mIoU of 81.3%. This demonstrates the benefit of contextual feature extraction and enrichment in the encoder path.

Finally, the complete model that includes both HAF and CBE achieves the best performance with a weighted mIoU of 82.4%. The consistent gains across all tumor categories (particularly glioma, which is typically more challenging) highlight the complementary strengths of the proposed components and their combined effectiveness in accurate tumor segmentation.

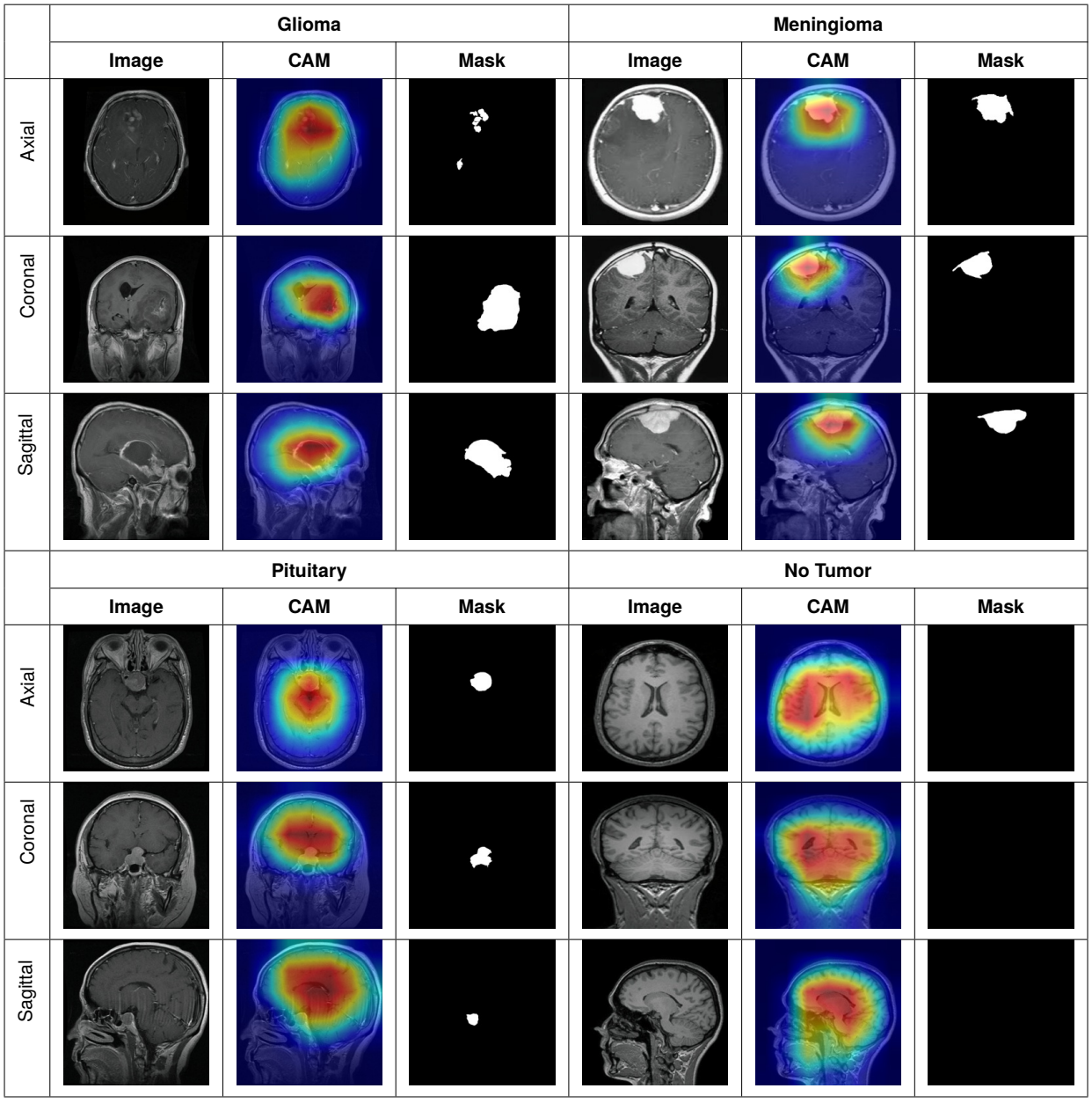

Figure 8: Grad-CAM visualization of the classification attention maps across the four diagnostic categories. The columns represent the input T1-weighted MRI scan, the generated Class Activation Map (CAM) overlaid on the image, and the corresponding ground truth segmentation mask. The visualization demonstrates that for Glioma, Meningioma, and Pituitary cases, the model's attention is precisely focused on the tumor region, aligning with the ground truth masks. Conversely, for the No Tumor case, the attention is diffusely distributed across the brain tissue, indicating that the model relies on global context to confirm the absence of pathologies rather than focusing on irrelevant artifacts.

5.4. Interpretability and Visual Analysis

To validate that the high classification accuracy achieved by Swin-HAFNet is driven by clinically relevant features rather than background noise, we employed Grad-CAM to visualize the model's decision-making process. Figure 8 illustrates the attention heatmaps generated by the classification branch across the four diagnostic categories: Glioma, Meningioma, Pituitary, and No Tumor.

As observed in the visual results, for the pathological classes (Glioma, Meningioma, and Pituitary), the model's attention is highly concentrated on the specific tumor regions. By comparing the "CAM" column with the ground truth "Mask" column, it is evident that the high-activation regions (indicated by red and yellow heatmaps) align almost perfectly with the annotated tumor boundaries. This alignment suggests that the shared backbone successfully extracts spatial features that are mutually beneficial: the precise localization required for the segmentation task effectively guides the classification head to focus on the lesion itself.

Furthermore, the visualization of the "No Tumor" class reveals a distinct and critical behavior. In the absence of pathological lesions, the model does not fixate on a specific focal point. Instead, the attention map is diffusely distributed across the entire brain tissue, appearing as a broad, balanced activation over the anatomical structure. This behavior indicates that the model is actively scanning the global context to verify the absence of anomalies, rather than hallucinating features or overfitting to irrelevant background artifacts. This "equal attention" strategy for healthy scans confirms the robustness of the global representation learned by the Swin-HAFNet, ensuring that negative classifications are based on a holistic assessment of the brain volume.

6. Conclusion

In this paper, we proposed a transformer-based segmentation model for brain tumor MRI analysis, designed to effectively capture multi-scale and contextual features in complex medical images. Experimental results demonstrate that the method achieves competitive segmentation accuracy, particularly in challenging tumor categories, and confirms the importance of architectural designs that leverage both local and global information. Despite these promising results, several limitations remain. The model's performance can be affected by small or irregularly shaped tumors, variations in imaging protocols, and the inherent class imbalance present in clinical data. Additionally, the computational cost of transformer-based architectures may limit their applicability in real-time or resource-constrained settings. Future work will focus on addressing these challenges by exploring more efficient and adaptive network designs, incorporating multi-modal imaging information, and integrating clinically informed evaluation protocols. We also aim to extend the model to joint segmentation and classification tasks, enabling more comprehensive analysis and improved clinical decision support. Overall, the proposed approach provides a solid foundation for advancing automated brain tumor analysis, and the insights gained from our experiments offer valuable guidance for the development of more robust and generalizable medical imaging models.

References

[1] Ruiquan Ge, Qingsong Wang, Xin Lin, Xinyang Li, Changmiao Wang, Zhipeng Wang, Yuqing Peng, Xiang Wan, and Ahmed Elazab. Segmentation-guided multi-modal brain tumors survival prediction model using pseudo-labelling approach. *Computerized Medical Imaging and Graphics*, page 102724, 2026.

[2] Ri Jin, Hu-Ying Tang, Qian Yang, and Wei Chen. La-resunet: Attention-based network for longitudinal liver tumor segmentation from ct images. *Computerized Medical Imaging and Graphics*, 123:102536, 2025.

[3] Zhuonneng Zhang, Luyi Han, Dengqiang Jia, Tianyu Zhang, Zehui Lin, Kahou Chen, Jiaju Huang, Shaobin Chen, Xiangyu Xiong, Sio-Kei Im, et al. Sgafnet: Robust brain tumor segmentation via learnable sequence-guided adaptive fusion in available mri acquisitions. *Computerized Medical Imaging and Graphics*, page 102703, 2026.

[4] Chengcheng Jin, Nor Safira Elaina Mohd Noor, Theam Foo Ng, Mohd Shahrimie Mohd Asaari, and Haidi Ibrahim. Transformer-based architectures in mri brain tumor segmentation: a review. *Computerized Medical Imaging and Graphics*, page 102729, 2026.

[5] Shadi Dorosti, Thomas Landry, Kimberly Brewer, Alyssa Forbes, Christa Davis, and Jeremy Brown. High-resolution ultrasound data for ai-based segmentation in mouse brain tumor. *Scientific Data*, 12(1):1322, 2025.

[6] Ayse Bastug Koc and Devrim Akgun. Lcbts-net: A lightweight cascaded 3d brain tumor segmentation network in magnetic resonance imaging. *Computerized Medical Imaging and Graphics*, page 102727, 2026.

[7] Chenjun Li, Dian Yang, Shun Yao, Shuyue Wang, Ye Wu, Le Zhang, Qiannuo Li, Kang Ik Kevin Cho, Johanna Seitz-Holland, Lipeng Ning, et al. Ddevenet: Evidence-based ensemble learning for uncertainty-aware brain parcellation using diffusion mri. *Computerized Medical Imaging and Graphics*, 120:102489, 2025.

[8] Sadjad Rezvani, Mansoor Fateh, Yeganeh Jalali, and Amirreza Fateh. Fusionlungnet: Multi-scale fusion convolution with refinement network for lung ct image segmentation. *Biomedical Signal Processing and Control*, 107:107858, 2025.

[9] Fatemeh Askari, Amirreza Fateh, and Mohammad Reza Mohammadi. Enhancing few-shot image classification through learnable multi-scale embedding and attention mechanisms. *Neural Networks*, 187:107339, 2025.

[10] Qiong Zhang, Yiliu Hang, Jianlin Qiu, and Hao Chen. Application of u-net network utilizing multiattention gate for mri segmentation of brain tumors. *Journal of Computer Assisted Tomography*, 48(6):991–997, 2024.

[11] Amirreza Fateh, Mohammad Reza Mohammadi, and Mohammad Reza Jahed Motlagh. Msdnet: Multi-scale decoder for few-shot semantic segmentation via transformer-guided prototyping. *Image and Vision Computing*, page 105672, 2025.

[12] Abdullah Almuhaimeed, Anas Bilal, Abdulkareem Alzahrani, Malek Alrashidi, Mansoor Al-Ghamdi, and Raheem Sarwar. Brain tumor classification using gan-augmented data with autoencoders and swin transformers. *Frontiers in Medicine*, 12:1635796, 2025.

[13] Ameer Hamza and Robertas Damaševičius. Deep learning for brain tumor segmentation and classification: a systematic review of methods and trends. *Computers, materials and continua.*, 86(1):1–41, 2026.

[14] Serena Grazia De Benedictis, Grazia Gargano, and Gaetano Settembre. Enhanced mri brain tumor detection and classification via topological data analysis and low-rank tensor decomposition. *Journal of Computational Mathematics and Data Science*, 13:100103, 2024.

[15] Maria Correia de Verdier, Rachit Saluja, Louis Gagnon, Dominic LaBella, Ujjwall Baid, Nourel Hoda Tahon, Martha Foltyn-Dumitru, Jikai Zhang, Maram Alafif, Saif Baig, et al. The 2024 brain tumor segmentation (brats) challenge: Glioma segmentation on post-treatment mri. *arXiv preprint arXiv:2405.18368*, 2024.

[16] Antonio MP Omuro, Claudia C Leite, Karima Mokhtari, and Jean-Yves Delattre. Pitfalls in the diagnosis of brain tumours. *The Lancet Neurology*, 5(11):937–948, 2006.

[17] Baiju Karun, Arunprasath Thiyagarajan, Pallikonda Rajasekaran Murugan, Natarajan Jeyaprakash, Kottaimalai Ramaraj, and Rakhee Makreri. Advanced hybrid brain tumor segmentation in mri: elephant herding optimization combined with entropy-guided fuzzy clustering. *Mathematical and Computational Applications*, 30(1):1, 2024.

[18] Mushtaq Mahyoob Saleh, Musab Elkheir Salih, Mohamed AA Ahmed, and Altahir Mohamed Hussein. From traditional methods to 3d u-net: A comprehensive review of brain tumour segmentation techniques. *Journal of Biomedical Science and Engineering*, 18(1):1–32, 2025.

[19] Beatrice Bonato, Loris Nanni, and Alessandra Bertoldo. Advancing precision: A comprehensive review of mri segmentation datasets from brats challenges (2012–2025). *Sensors (Basel, Switzerland)*, 25(6):1838, 2025.

[20] Phuoc-Nguyen Bui, Duc-Tai Le, Junghyun Bum, and Hyunseung Choo. Multi-scale feature enhancement in multi-task learning for medical image analysis. *arXiv preprint arXiv:2412.00351*, 2024.

[21] Rabeea Fatma Khan, Mu Sook Lee, and Byoung-Dai Lee. Harnessing transformer-based attention mechanisms for multi-scale feature fusion in medical image segmentation. *Applied Intelligence*, 55(17):1120, 2025.

[22] Kenneth Aldape, Kevin M Brindle, Louis Chesler, Rajesh Chopra, Amar Gajjar, Mark R Gilbert, Nicholas Gottardo, David H Gutmann, Darren Hargrave, Eric C Holland, et al. Challenges to curing primary brain tumours. *Nature reviews Clinical oncology*, 16(8):509–520, 2019.

[23] Ebrahim Mohammed Senan, Mukti E Jadhav, Taha H Rassem, Abdulaziz Salamah Aljaloud, Badiea Abdulkarem Mohammed, and Zeyad Ghaleb Al-Mekhlafi. Early diagnosis of brain tumour mri images using hybrid techniques between deep and machine learning. *Computational and Mathematical Methods in Medicine*, 2022(1):8330833, 2022.

[24] Sarah Lapointe, Arie Perry, and Nicholas A Butowski. Primary brain tumours in adults. *The Lancet*, 392(10145):432–446, 2018.

[25] Stefan Sunaert. Presurgical planning for tumor resectioning. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 23(6):887–905, 2006.

[26] Amirreza Fateh, Mohsen Rezvani, Alireza Tajary, and Mansoor Fateh. Providing a voting-based method for combining deep neural network outputs to layout analysis of printed documents. *Journal of Machine Vision and Image Processing*, 9(1):47–64, 2022.

[27] Amit Mehndiratta and Frederik L Giesel. *Brain tumour imaging*. INTECH Open Access Publisher, 2011.

[28] Arie Perry and Pieter Wesseling. Histologic classification of gliomas. *Handbook of clinical neurology*, 134:71–95, 2016.

[29] Kristjan R Jessen. Glial cells. *The international journal of biochemistry & cell biology*, 36(10):1861–1867, 2004.

[30] Ali-Reza Fathi and Ulrich Roelcke. Meningioma. *Current neurology and neuroscience reports*, 13:1–8, 2013.

[31] Krishnakali Dasgupta and Juhee Jeong. Developmental biology of the meninges. *genesis*, 57(5):e23288, 2019.

[32] Christine Marosi, Marco Hassler, Karl Roessler, Michele Reni, Milena Sant, Elena Mazza, and Charles Vecht. Meningioma. *Critical reviews in oncology/hematology*, 67(2):153–171, 2008.

[33] Peng-Fei Yan, Ling Yan, Ting-Ting Hu, Dong-Dong Xiao, Zhen Zhang, Hong-Yang Zhao, and Jun Feng. The potential value of preoperative mri texture and shape analysis in grading meningiomas: a preliminary investigation. *Translational oncology*, 10(4):570–577, 2017.

[34] Rodrigo E Bancalari, Louise C Gregory, Mark J McCabe, and Mehul T Dattani. Pituitary gland development: an update. *Endocr Dev*, 23(1), 2012.

[35] Sylvia L Asa and Shereen Ezzat. The pathogenesis of pituitary tumours. *Nature Reviews Cancer*, 2(11):836–849, 2002.

[36] Gerald Raverot, Mirela Diana Ilie, Helene Lasolle, Vincent Amodru, Jacqueline Trouillas, Frederic Castinetti, and Thierry Brue. Aggressive pituitary tumours and pituitary carcinomas. *Nature Reviews Endocrinology*, 17(11):671–684, 2021.

[37] Adam Kelly. Neurological therapeutics: Principles andpractice, secondedition, 2008.

[38] Arun Reang. Clinico-radiologic profile of intracranial space occupying lesion imaged with mri and spectroscopy in a tertiary care centre. *International Journal of Life Sciences, Biotechnology and Pharma Research*, 14, 2025.

[39] Kenhub. Anatomical terminology: Planes, directions & regions, 2025. URL https://www.kenhub.com/en/library/anatomy/anatomical-terminology. Accessed: 2025-05-31.

[40] Catherine Westbrook and John Talbot. *MRI in Practice*. John Wiley & Sons, 2018.

[41] Peng-Fei Yan, Ling Yan, Zhen Zhang, Adnan Salim, Lei Wang, Ting-Ting Hu, and Hong-Yang Zhao. Accuracy of conventional mri for preoperative diagnosis of intracranial tumors: A retrospective cohort study of 762 cases. *International Journal of Surgery*, 36:109–117, 2016.

[42] Shoffan Saifullah and Rafał Drężewski. Ga-unet: Genetic algorithm-optimized lightweight u-net architecture for multi-sequence brain tumor mri segmentation. *IEEE Access*, 2025.

[43] Agnesh Chandra Yadav, Maheshkumar H Kolekar, and Mukesh Kumar Zope. Modified recurrent residual attention u-net model for mri-based brain tumor segmentation. *Biomedical Signal Processing and Control*, 102:107220, 2025.

[44] Usman Amjad, Asif Raza, Muhammad Fahad, Doaa Farid, Adnan Akhunzada, Muhammad Abubakar, and Hira Beenish. Context aware machine learning techniques for brain tumor classification and detection–a review. *Heliyon*, 11(2), 2025.

[45] Takowa Rahman, Md Saiful Islam, and Jia Uddin. Mri-based brain tumor classification using a dilated parallel deep convolutional neural network. *Digital*, 4(3):529–554, 2024.

[46] Kondra Pranitha and Vuda Sreenivasa Rao. Enhanced brain tumor classification using optimized u-net segmentation by epo and hybrid feature extraction using fo. 2025.

[47] Namya Musthafa, Qurban A Memon, and Mohammad M Masud. Advancing brain tumor analysis: Current trends, key challenges, and perspectives in deep learning-based brain mri tumor diagnosis. *Eng*, 6(5):82, 2025.

[48] Delaram J Ghadimi, Amir M Vahdani, Hanie Karimi, Pouya Ebrahimi, Mobina Fathi, Farzan Moodi, Adrina Habibzadeh, Fereshteh Khodadadi Shoushtari, Gelareh Valizadeh, Hanieh Mobarak Salari, et al. Deep learning-based techniques in glioma brain tumor segmentation using multi-parametric mri: A review on clinical applications and future outlooks. *Journal of Magnetic Resonance Imaging*, 61(3):1094–1109, 2025.

[49] Chengcheng Jin, Theam Foo Ng, and Haidi Ibrahim. Advancements in semi-supervised deep learning for brain tumor segmentation in mri: A literature review. *AI*, 6(7):153, 2025.

[50] Muhammad Adeel Abid and Kashif Munir. A systematic review on deep learning implementation in brain tumor segmentation, classification and prediction. *Multimedia Tools and Applications*, pages 1–40, 2025.

[51] Dichao Pan, Jianguo Shen, Zaid Al-Huda, and Mohammed AA Al-Qaness. Vcanet: Vision transformer with fusion channel and spatial attention module for 3d brain tumor segmentation. *Computers in Biology and Medicine*, 186:109662, 2025.

[52] Minye Shao, Zeyu Wang, Haoran Duan, Yawen Huang, Bing Zhai, Shizheng Wang, Yang Long, and Yefeng Zheng. Rethinking brain tumor segmentation from the frequency domain perspective. *IEEE Transactions on Medical Imaging*, 2025.

[53] M Renugadevi, Venkateswarlu Gonuguntla, Ihssan S Masad, G Venkat Babu, and K Narasimhan. Gliosurvqnet: A duelcontextattn dqn framework for brain tumor prognosis with metaheuristic optimization. *Diagnostics*, 15(18):2304, 2025.

[54] Pardhu Thottempudi, Biswaranjan Acharya, Srilakshmi Aouthu, Narra Dhanalakshmi, B Mathura Bai, K Reddy Madhavi, K Swaraja, and Saurav Malik. An explainable hybrid cnn–transformer framework with aquila optimization for mri-based brain tumor classification. *medRxiv*, pages 2025–10, 2025.

[55] Sofia El Amoury, Youssef Smili, and Youssef Fakhri. Simulated annealing-based hyperparameter optimization of a convolutional neural network for mri brain tumor classification. *Machine Learning and Knowledge Extraction*, 7(2):50, 2025.

[56] Saif Ur Rehman Khan, Sohaib Asif, Ming Zhao, Wei Zou, Yangfan Li, and Chenggen Xiao. Shallowmri: A novel lightweight cnn with novel attention mechanism for multi brain tumor classification in mri images. *Biomedical Signal Processing and Control*, 111:108425, 2026.

[57] Tathagat Banerjee, Prachi Chhabra, Manoj Kumar, Abhay Kumar, Kumar Abhishek, and Mohd Asif Shah. Pyramidal attention-based t network for brain tumor classification: a comprehensive analysis of transfer learning approaches for clinically reliable and reliable ai hybrid approaches. *Scientific Reports*, 15(1):28669, 2025.

[58] Omran Azeez and Adnan Abdulazeez. Classification of brain tumor based on machine learning algorithms: A review. *Journal of Applied Science and Technology Trends*, 6(1):01–15, 2025.

[59] Nan-Han Lu, Yung-Hui Huang, Kuo-Ying Liu, and Tai-Been Chen. Deep learning-driven brain tumor classification and segmentation using non-contrast mri. *Scientific reports*, 15(1):27831, 2025.

[60] Amirreza Fateh, Yasin Rezvani, Sara Moayedi, Sadjad Rezvani, Fatemeh Fateh, and Mansoor Fateh. Brisc: Annotated dataset for brain tumor segmentation and classification with swin-hafnet. *arXiv preprint arXiv:2506.14318*, 2025.

[61] Thomas Dubail. Brain tumors 256x256. https://www.kaggle.com/datasets/thomasdubail/brain-tumors-256x256/data, 2024. Accessed: 2024.

[62] Nikhil Tomar. Brain tumor segmentation, 2023. URL https://www.kaggle.com/datasets/nikhilroxtomar/brain-tumor-segmentation.

[63] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 658–666, 2019.

[64] Ian Goodfellow. Deep learning, 2016.

[65] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[66] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[67] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[68] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support: 4th international workshop, DLMIA 2018, and 8th international workshop, ML-CDS 2018, held in conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, proceedings 4*, pages 3–11. Springer, 2018.

[69] Abhishek Chaurasia and Eugenio Culurciello. Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE visual communications and image processing (VCIP)*, pages 1–4. IEEE, 2017.

[70] Tongle Fan, Guanglei Wang, Yan Li, and Hongrui Wang. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, 8:179656–179665, 2020.

[71] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[72] Hanchao Li, Pengfei Xiong, Jie An, and Lingxue Wang. Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180*, 2018.

[73] Chen Li and Ge Jiao. Einet: camouflaged object detection with pyramid vision transformer. *Journal of Electronic Imaging*, 31(5):053002–053002, 2022.

[74] Krushi Patel, Andrés M Bur, and Guanghui Wang. Enhanced u-net: A feature enhancement network for polyp segmentation. In *2021 18th conference on robots and vision (CRV)*, pages 181–188. IEEE, 2021.
[75] Jiepan Li, Wei He, and Hongyan Zhang. Towards complex backgrounds: A unified difference-aware decoder for binary segmentation. *arXiv preprint arXiv:2210.15156*, 2022.
[76] Xuebin Qin, Deng-Ping Fan, Chenyang Huang, Cyril Diagne, Zichen Zhang, Adrià Cabeza Sant'Anna, Albert Suarez, Martin Jagersand, and Ling Shao. Boundary-aware segmentation network for mobile and web applications. *arXiv preprint arXiv:2101.04704*, 2021.
[77] Alireza Saber, Amirreza Fateh, Pouria Parhami, Alimohammad Siahkarzadeh, Mansoor Fateh, and Saideh Ferdowsi. Efficient and accurate pneumonia detection using a novel multi-scale transformer approach. *Sensors*, 25(23):7233, 2025.
[78] Sadjad Rezvani, Mansoor Fateh, and Hossein Khosravi. Abanet: Attention boundary-aware network for image segmentation. *Expert Systems*, 41(9):e13625, 2024.
[79] Md Younus Ahamed. Brain tumor segmentation using u-net, 2023. URL https://www.kaggle.com/code/shuvostp/brain-tumor-segmentation-using-u-net.
[80] Oussama Slimani. Mri brain tumor segmentation u-net, 2024. URL https://www.kaggle.com/code/oussamaslmani/mri-brain-tumor-segmentation-unet.
[81] Divya Nayan. U-net tensorflow implementation, 2024. URL https://www.kaggle.com/code/dnayan/u-net-tensorflow.
[82] Dhruvil Joshi. Unetr implementation on figshare dataset, 2024. URL https://www.kaggle.com/code/dhruviljoshi1910/unetr-implementation-on-figshare-dataset.
[83] Ayisha Zakia. Attention resunet for brain tumor segmentation, 2025. URL https://www.kaggle.com/code/ayishazakia/attention-resunet-dcs-80.
[84] Sara Zahran. Brain tumor segmentation kernel, 2025. URL https://www.kaggle.com/code/sarazahran1/brain-tumor-segmentation.
[85] Fariska Ratna. Mha-unet seggan evaluation, 2025. URL https://www.kaggle.com/code/fariskar/mha-unet-seggan.
[86] Dheeresh Chandra. Gdoel's deit fuzzy. https://www.kaggle.com/code/dheereshc/gdoel-s-deit-fuzzy, 2024. Kaggle Notebook.
[87] Rahaf Fayez. Brain tumor detection. https://www.kaggle.com/code/rahaffayez/brain-tumor-detection, 2024. Kaggle Notebook.
[88] Mohamed Haitham. Brain tumor classification using cnn. https://www.kaggle.com/code/mohamedhaithamyamani/brain-tumor-classification-using-cnn, 2024. Kaggle Notebook.
[89] Abdelrahman Said. Brain tumor detector. https://www.kaggle.com/code/abdosaid123/brain-tumor-detector, 2024. Kaggle Notebook.
[90] Thomas Dubail. Brain tumor swinv2 (96 https://www.google.com/search?q=https://www.kaggle.com/code/thomasdubail/brain-tumor-swinv2-96-acc, 2024. Kaggle Notebook.
[91] Thomas Dubail. Brain tumor vit (93 https://www.kaggle.com/code/thomasdubail/brain-tumor-vit-93-acc, 2024. Kaggle Notebook.
[92] Farhaan Ali. Ml project image classification. https://www.kaggle.com/code/farhaanalisiddiqui/ml-project-image-classification, 2024. Kaggle Notebook.