

# Patch-Based Multi-Head Self-Attention Framework for End-to-End Low-Grade Glioma Segmentation from Magnetic Resonance Imaging

Solomon Buabeng Antwi

`solomon.buabeng.stu@uenr.edu.gh`

University of Energy and Natural Resources

Peter Appiahene

University of Energy and Natural Resources

Ben Beklisi Kwame Ayawli

Sunyani Technical University

Peter Nimbe

University of Energy and Natural Resources

---

## Article

**Keywords:** MRI segmentation, Low-grade glioma, vision transformer, multi-head self-attention, patch tokenization, adaptive convolution, encoder-decoder, FLAIR imaging, brain tumour

**Posted Date:** May 25th, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-9517012/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

# Patch-Based Multi-Head Self-Attention Framework for End-to-End Low-Grade Glioma Segmentation from Magnetic Resonance Imaging

Solomon Buabeng Antwi <sup>\*1</sup>, Peter Appiahene<sup>2</sup>, Ben Beklisi Kwame Ayawli<sup>3</sup>, Peter Nimbe<sup>4</sup>

<sup>1,4</sup> Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana,

<sup>2</sup> Department of Information Technology and Decision Sciences, University of Energy and Natural Resources, Sunyani, Ghana,

<sup>1,3</sup> Department of Computer Science, Sunyani Technical University, Sunyani, Ghana.

## Abstract

**Background:** Low-grade glioma (LGG) segmentation from magnetic resonance imaging (MRI) is an essential task complicated by diffuse tumour boundaries, subtle FLAIR signal abnormalities, and severe class imbalance between tumour and background pixels.

**Objective:** This study proposes HybridAdaptiveUNet, a novel end-to-end patch-based multi-head self-attention framework integrating adaptive convolutional feature extraction with global transformer-based contextual modelling for automated LGG segmentation from single-channel axial FLAIR MRI.

**Methods:** The proposed framework incorporates four adaptive mechanisms: (i) a Dynamic Kernel Router (DKR) that adaptively weights parallel  $3\times 3$ ,  $5\times 5$ , and  $7\times 7$  convolutional branches per input image; (ii) a Hybrid Adaptive module combining depth-wise CNN local feature extraction with multi-head self-attention on non-overlapping patch tokens under both fixed learnable and input-conditioned dynamic attention temperatures; (iii) a Context-Aware Dynamic  $\alpha$  fusion module blending local and global pathway representations at the bottleneck; and (iv) Adaptive Spatial Gates modulating skip connection feature propagation. Training employed a composite loss combining binary cross-entropy with logits, soft Dice loss, kernel routing diversity, and spatial gate sparsity objectives. Experiments were conducted on the LGG MRI Segmentation dataset containing 3,929 axial FLAIR image-mask pairs from 110 patients.

**Results:** The proposed model achieved a Dice Similarity Coefficient (DSC) of 0.5557 and Intersection over Union (IoU) of 0.4351 on the held-out test partition, with pixel-level accuracy of 0.9940. Diagnostic analysis indicates explainable, task-appropriate specialisation across all the adaptive mechanisms, comprising dynamic attention head temperature divergence spanning a range of 0.901, depth-specific kernel routing preferences, complementary bottleneck fusion at  $\alpha \approx 0.484$ , and depth-dependent spatial gating selectivity.

**Conclusion:** HybridAdaptiveUNet exhibits competitive performance within the established range for single-channel FLAIR-only LGG segmentation (DSC 0.52–0.68) while providing explainable evidence that its adaptive components learn a coherent, multi-level processing technique from data. The framework offers functionally validated architecture for automated glioma segmentation in the single-modality imaging domain.

**Keywords:** MRI segmentation; Low-grade glioma; vision transformer; multi-head self-attention; patch tokenization; adaptive convolution; encoder-decoder; FLAIR imaging; brain tumour

## Introduction

Segmentation of Brain tumour is considered as one of the foundational tasks in medical image analysis, assisting in arriving at an accurate diagnosis, treatment and prognosis assessment of neurological malignancies. (Bibi et al., 2025). Primary brain tumours, low-grade gliomas, exhibit complex segmentation challenges due to their diffuse infiltrative growth patterns, unclear boundaries, and considerable variability in size and morphology across patients. (Seshimo & Rashed, 2024). Failure to precisely delineate tumour regions from surrounding healthy tissue substantially influences treatment outcomes, emphasizing for robust and automated segmentation frameworks. (Dorfner et al., 2025).

Conventional segmentation techniques, such as thresholding, region growing, and atlas-based methods, have demonstrated limited applicability due to their sensitivity to intensity inhomogeneity, low tissue contrast, and inability to generalise across various tumour presentations. (Mahdi Shafiei Neyestanak, Hamid Jahani, Mohsen Khodarahmi, Javad Zahiri & Amirali Fatoorchi, 2025). The application of deep learning models, particularly convolutional neural networks (CNNs), has exhibited a significant improvement in automated medical image segmentation by enabling hierarchical feature extraction from MRI data. (Preetha et al., 2025). Despite the emergence and the performance exhibited by conventional CNN architectures, capturing full spatial long-range dependencies for infiltrative tumour regions is a challenge. (Zeineldin et al., 2024). A multi-scale CNN has the capacity to extract features at multiple spatial resolutions. However, remain restricted by the absence of adaptive mechanisms capable of selectively emphasising significant tumour regions. (Aslam et al., 2025) (Preetha et al., 2025)

The introduction of vision transformer architectures has introduced a paradigm shift in feature representation for medical image analysis through patch-based tokenisation and self-attention modelling.(Preetha et al., 2025). Decomposing input images into non-overlapping spatial patches and processing each as an independent token, vision transformers enable both local structural features and global contextual relationships across the whole image domain.(Chandraprabha et al., 2025) (Preetha et al., 2025). Patch tokenisation strategy performs substantially better for brain tumour segmentation, where tumour regions exhibit spatial distribution beyond local neighbourhoods captured by convolutional operations. (Hassan et al., 2024). SegFormer has substantially demonstrated that structured patch integration substantially improves tumour boundary delineation in spatial representations.(Abidin et al., 2024).

The effectiveness of patch-based segmentation in the multi-head self-attention mechanism enables dynamic modelling of inter-patch contextual dependencies across multiple representational subspaces simultaneously. (Lv et al., 2025)Unlike single-head, multi-head self-attention allows the model to attend to diverse spatial relationships and feature patterns in parallel, producing better and more discriminative representations of tumour-relevant regions.(Savaria & Sun, 2026) Dynamic inter-patch interaction is substantially more important in low-grade glioma segmentation, where tumour extent, infiltration boundaries, and peritumoral regions exhibit spatially distributed characteristics requiring contextual reasoning beyond local features. (Seshimo & Rashed, 2024) (Preetha et al., 2025) . Adaptive self-attention frameworks have demonstrated that attention-driven feature modulation notably improves segmentation accuracy in glioma data, reinforcing the value of incorporating dynamic inter-patch attention within end-to-end pipelines.(Savaria & Sun, 2026).

In attention modeling where learned patch representations are transformed and decoded into full-resolution segmentation outputs, constitutes a significant determinant of segmentation accuracy. The embedding attention reconstruction model offers a structured technique where input patches are projected into compact latent embeddings, contextually refined through attention-weighted feature modulation, and then decoded into full-resolution segmentation masks. (Lv et al., 2025). This pipeline conserves spatial correspondence between input image regions with their corresponding output predictions while enabling deep semantic modulation at the embedding level. This technique is essential for accurate tumour boundary localisation in low-grade glioma imaging. (Chen et al., 2024) . Binary cross-entropy with logits is considered a robust, numerically stable objective for binary segmentation tasks, penalising incorrect pixel-level predictions

through direct tumour and non-tumour classification. (Mahdi Shafiei Neyestanak, Hamid Jahani, Mohsen Khodarahmi, Javad Zahiri & Amirali Fatoorchi, 2025). Its gradient stability properties are substantially relevant in low-grade glioma segmentation, where class imbalance between tumour and background pixels poses a persistent optimisation challenge. (Dorfner et al., 2025) .

To address these research limitations and gaps, this study proposes a novel end-to-end patch-based attention-driven framework for brain tumour segmentation from MRI. The proposed framework is structured around these major contributions:

- Develop an end-to-end segmentation pipeline integrating non-overlapping patch tokenisation and spatial attention modelling for low-grade glioma MRI data
- Implement a multi-head self-attention mechanism for dynamic inter-patch interaction and adaptive contextual feature modelling.
- Investigate an embedding attention reconstruction strategy for decoding attention-refined latent representations into full-resolution tumour segmentation masks.
- Evaluate segmentation performance through pixel-wise binary classification using binary cross-entropy with logits optimisation.

## **2.0 Limitations of conventional segmentation methods in Glioma MRI**

Before the introduction of deep learning techniques, medical image segmentation for glioma MRI completely relied on certain types of algorithms such include, such as region growing, intensity thresholding, level-set methods, active contour models, and probabilistic atlas-based registration. (Diana-albelda et al., 2025)(Gallegos et al., 2025). Atlas registration also possesses significant challenges due to gliomas' mass effect and infiltration violating anatomical priors. (Diana-albelda et al., 2025). (Wang et al., 2023) indicated that conventional pipelines produced inconsistent segmentation quality on glioma MRI, mostly for LGG cases characterised by subtle FLAIR signal abnormalities and absence of contrast enhancement. The technique lacks representational capacity for complex models and non-linear tumour-tissue boundaries. The absence of adaptive data-driven feature learning across conventional techniques is inadequate for spatial complexity and inter-subject variability intrinsic to LGG imaging. (Gao et al., 2025).

### **2.1 CNN Architectures and Performance on Glioma Benchmarks**

(Xinyue Zhang, Jianfeng Wang, XiaochunCheng, 2025) indicates that deep convolutional neural networks revolutionised medical image segmentation by allowing hierarchical feature learning from raw pixels while removing manual feature engineering. U-Net architecture, along with its 3D variants, introduced the encoder-decoder paradigm to maintain spatial details during downsampling and upsampling. (Saluja et al., 2024) Conducted a comprehensive evaluation on deep CNN architectures for glioma grading and segmentation, and the results show a significant improvement of 98.91% with hybrid and ensemble-based CNN architectures on multi-parametric MRI datasets. Multi-scale CNN extensions exhibited the capacity to capture tumour features at multiple spatial resolutions through parallel dilated convolution pathways. (Preetha et al., 2025) proposes a multi-scale attention U-Net incorporating EfficientNetB4 as a backbone encoder and multi-scale convolutional kernels of 1x1, 3x3, and 5x5, achieving enhanced tumour boundary delineation relative to single-scale baselines.

### **2.2 Hybrid Vision Transformers for Glioma Segmentation**

(Aburass et al., 2025) indicated that Vision Transformers (ViTs) revolutionised feature representation in medical imaging by supplanting CNNs' local inductive biases with global self-attention mechanisms operating on spatial patch token sequences. (Lin et al., 2025) indicated that images are decomposed into non-overlapping patches, linearly projected into embedding vectors, and consecutively processed through multi-layer transformer encoders to capture full-domain contextual dependencies from initial layers. A

hybrid vision transformer-CNN architecture improves multimodal glioma segmentation on the BraTS benchmark, outperforming CNN-only models by better capturing peritumoral relationships and tumour sub-region boundaries across whole tumour (WT), enhancing tumour (ET), and tumour core (TC) targets. (Zeineldin et al., 2024).

### **2.3 Hybrid CNN-Transformer Architectures for 3D Glioma Segmentation**

(Chen et al., 2025) Pure Transformers perform on global modelling but demand huge pretraining data, exhibiting inferior high-resolution feature discrimination for intricate local patterns and edges; motivating hybrid CNN-Transformer architectures that combine convolutional inductive priors with transformer global relational modelling. (Chu et al., 2024) introduces TCTNet, a direction-wise 3D convolutional-transformer architecture integrating 3D convolutional token embedding volumetric CNN feature representations into transformer-compatible token sequences for self-attention computation. Preserving the translation equivariance properties of the CNN encoder while enabling the transformer to model global dynamic attention over 3D spatial positions, TCTNet outperformed both pure CNN and pure transformer baselines on volumetric brain tumour segmentation benchmarks. (Huang et al., 2025) Propose multi-scale 3D token division and aggregation within self-attention layers, enabling simultaneous extraction of multi-scale spatial features and long-range dependency modelling. This particular design demonstrates a substantial improvement in segmentation performance over prior CNN-only architectures across three public datasets. Systematic reviews confirm that hybrid architectures consistently outperform both single-paradigm CNN and pure transformer models for glioma sub-region segmentation. (Wang et al., 2023)(Bibi et al., 2025).

### **2.4 Patch Tokenisation in Vision Transformers: Mechanisms, Architectural Integration, and Applications in Brain Tumour Detection**

(Ronen et al., 2023) indicated that Vision Transformers convert images into self-attention-compatible token sequences through non-overlapping patch tokenisation. The images are split into fixed-size patches, flattened into 1D vectors, linearly projected into embedding space, and augmented with positional encodings to preserve spatial geometry. (Chandrababha et al., 2025) indicated that end-to-end vision transformer-CNN architectures, adapted non-overlapping patch tokenisation, obtained superior brain tumour detection and classification performance relative to convolutional baselines on standard MRI benchmark datasets. This significant performance improvement was attributed to the capacity of patch tokens to encode cross-patch contextual information structurally unavailable to local convolution operations. Specifically for low-grade glioma, patch tokenisation is mostly beneficial because tumour regions exhibit spatially distributed extent across multiple non-contiguous image areas (Hassan et al., 2024). The quality of patch embedding, projection and positional encoding schemes determines the fidelity with which spatial correspondence is preserved from tokenisation through attention computation to final mask reconstruction (Renugadevi et al., 2025) .

### **2.5 Multi-Head Self-Attention in Transformer-Based Glioma MRI Segmentation for Capturing Tumour-Relevant Spatial Representations**

Multi-head self-attention is the main computational mechanism of transformer architectures, which enables simultaneous modelling of various contextual relationships across the full token sequence through parallel representational subspaces(Ronen et al., 2023). In Multi-Head Self-Attention, the input token sequence is solely projected into multiple sets of query, key, and value matrices, with individual projection sets defining an independent attention head(Ainslie & Lee-Thorp, 2023). Each head calculates scaled dot-product attention scores across all token pairs, producing a weighted sum of value representations. This encodes how strongly each token should attend to every other token within that subspace(Chandrababha et al., 2025). Attention heads are concatenated and transformed through a linear projection combined representation that integrates various spatial relationships and feature interaction patterns solely extracted

by individual head (Renugadevi et al., 2025). Lv et al. (2025) indicated that parallel multi-head structure enables the model to capture separate types of spatial relationships, such as local boundary structure, global shape context, and long-range infiltration patterns, producing discriminative representations of tumour-relevant regions compared to single-head attention mechanisms that operate within a single undistinguishable representational subspace. (Savaria & Sun, 2026) indicated that adaptive self-attention frameworks obtain significant segmentation accuracy improvements on glioma MRI data through dynamic inter-patch feature modulation, highlighting the capacity of MHSA to selectively weight cross-patch relationships in a data-driven manner. (Jin et al., 2026) introduce SwinCLNet, integrating window-based and shifted window MHSA with cross-scale dual fusion within a 3D U-Net framework, obtaining a significant performance on DSC and Hausdorff Distance performance on BraTS2023 and BraTS2024 benchmarks.

## **2.6 Decoder Pathway Mechanisms for Translating Attention-Refined Latent Embeddings into Full-Resolution Glioma Segmentation Masks**

The decoding of attention-refined latent embeddings into full-resolution segmentation masks comprises a significant stage in the end-to-end pipeline, determining spatial fidelity with which patch-level contextual representations are translated into pixel-level tumour delineation. (Chen et al., 2024). Embedding-based reconstruction pipelines project input patches into compact latent representations. (Chen et al., 2024), apply contextual refinement through attention-weighted feature modulation across the token sequence, and eventually decode embeddings into pixel-space predictions through hierarchical upsampling within the decoder pathway (Yang et al., 2023).

Chen et al. (2024) indicated that this embedding attention reconstruction strategy conserves spatial correspondence between input image regions and output segmentation predictions while enabling deep semantic modulation at the latent embedding level, contributing to accurate tumour boundary localisation, specifically in glioma imaging contexts where boundary gradients are subtle and ambiguous. The architectural design ensures inter-patch spatial relationships encoded during self-attention computation remain accessible to the decoder through the latent representation. (Renugadevi et al., 2025) Introduced a hybrid ViT-CNN design with progressive skip-connected upsampling, achieving DSC values of 0.902, 0.954, and 0.966 across enhancing tumour. (Ali et al., 2025) Introduce attention-guided boundary refinement through soft masking modules during the decoding stage, which produced consistent improvements at tumour-tissue interfaces where boundary ambiguity is most significant.

**Binary Cross-Entropy Optimisation and Class Imbalance Challenges in Low-Grade Glioma MRI Segmentation**

(Bibi et al., 2025) indicated that Binary Cross-Entropy Optimisation is considered the most widely adopted loss function in glioma segmentation. The essential optimisation challenge arises from class imbalance between tumour and background pixels in LGG segmentation, where the tumour region constitutes as little as one to two per cent of total image pixels. (Chi et al., 2024) Combined weighted cross-entropy and Dice loss with sub-region weighting achieved superior segmentation accuracy over unweighted baselines across multiple BraTS datasets for LGG-inclusive cohorts. This confirms that class imbalance handling is an essential influence on segmentation performance under realistic tumour-to-background ratio conditions. (Liu et al., 2024) established weighted loss guidance improved boundary delineation in diffusion-model-based multi-class segmentation on BraTS2020, reinforcing the utility of BCE-complementary loss design for class-imbalanced brain tumour segmentation optimisation.

## **3.0 Materials and Methods**

This section describes the experimental framework for segmenting lower-grade gliomas (LGGs). The subsections present the image dataset and acquisition (Section 3.1), the proposed HybridAdaptiveUNet architecture (Section 3.2), the preprocessing pipeline (Section 3.3), and the evaluation metrics (Section 3.4).

### 3.1 Image Dataset

The LGG MRI Segmentation dataset was obtained from the Kaggle repository at <https://www.kaggle.com/datasets/mateuszbeda/ligg-mri-segmentation>. The dataset contains axial FLAIR brain MRI scans from 110 patients with lower-grade glioma, paired with pixel-level binary segmentation masks annotated. The dataset files were stored in TIFF format. The dataset contains 3,929 image-mask pairs, of which 27.7% contain a tumour region, and 72.3% are tumour-free, which creates class imbalance. The partition is summarised in Table 1.

**Table 1. Dataset partition summary**

| Split        | Images       | Masks        | Proportion  |
|--------------|--------------|--------------|-------------|
| Training     | 2,947        | 2,947        | 75.0%       |
| Validation   | 491          | 491          | 12.5%       |
| Test         | 491          | 491          | 12.5%       |
| <b>Total</b> | <b>3,929</b> | <b>3,929</b> | <b>100%</b> |

The proposed HybridAdaptiveUNet is an encoder-decoder segmentation module integrating multi-scale convolutional feature extraction with a hybrid attention module. The module has be categorized into three respects: (i) all convolutional blocks uses Dynamic Kernel Router (DKR) adaptively weights  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$  filter branches per image, (ii) the module fuses both the local depth wise CNN features with global patch-based self-attention using both fixed learnable temperature  $\tau$  and an input conditioned dynamic temperature  $\tau(x)$ , and (iii) skip connections are selectively gated by an Adaptive Spatial Gate. The architecture is illustrated in Figure 1, and layer-wise details are also indicated in Table 2.

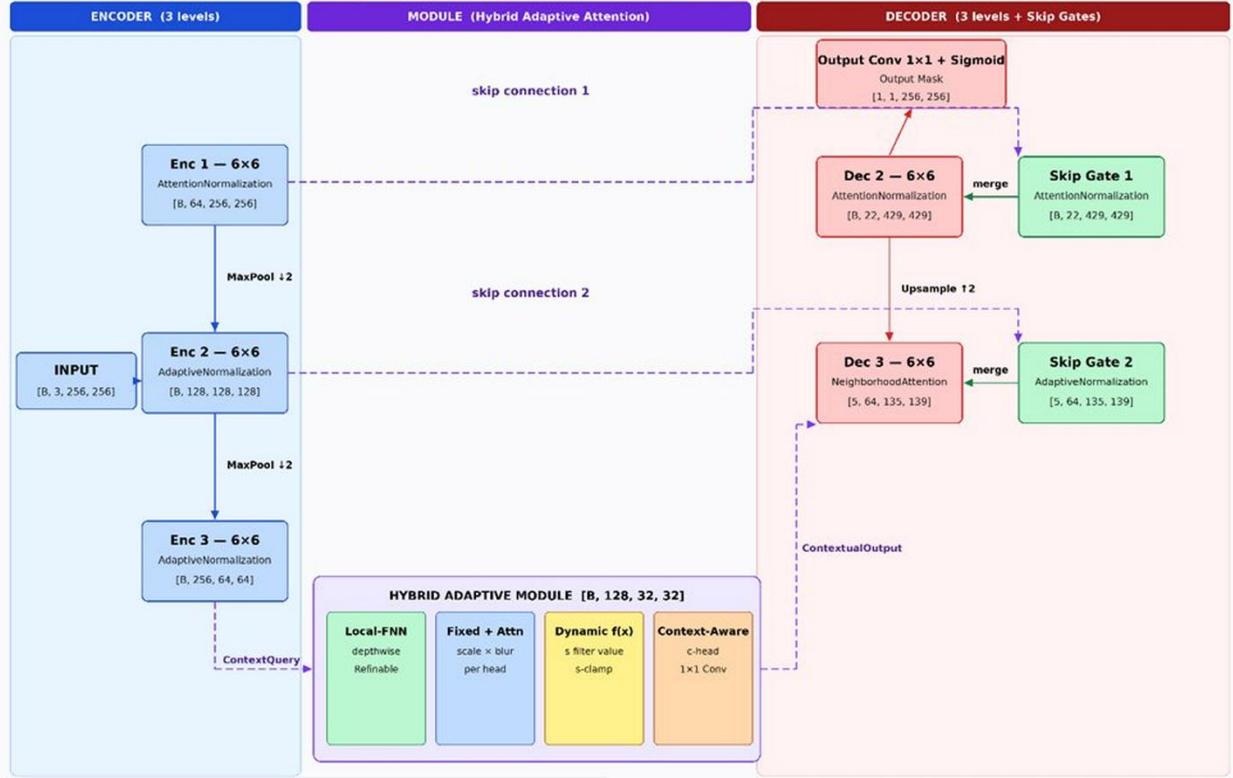


Figure 1. HybridAdaptiveUNet architecture: Encoder, Hybrid Adaptive module, Decoder with gated skip connections and Dynamic Kernel Router.

Table 2. Layer-wise architecture of HybridAdaptiveUNet

| Stage   | Module                | Output Shape   | Key Mechanism                                      |
|---------|-----------------------|----------------|----------------------------------------------------|
| Input   |                       | [B,1,256,256]  | Raw normalised MRI                                 |
| Enc 1   | DKR + AdaptiveRecalib | [B,32,256,256] | $3 \times 5 \times 7 \times$ routing, SE++         |
| Enc 2   | DKR + AdaptiveRecalib | [B,64,128,128] | $3 \times 5 \times 7 \times$ routing, SE++         |
| Enc 3   | DKR + AdaptiveRecalib | [B,128,64,64]  | $3 \times 5 \times 7 \times$ routing, SE++         |
| Modules | HybridAdaptive module | [B,128,32,32]  | Fixed $\tau$ + Dynamic $\tau(x)$ + $\alpha$ fusion |
| Dec 3   | DKR + AdaptiveRecalib | [B,64,128,128] | ConvTranspose2d + gated skip                       |
| Dec 2   | DKR + AdaptiveRecalib | [B,32,256,256] | ConvTranspose2d + gated skip                       |
| Output  | Conv1x1 + Sigmoid     | [B,1,256,256]  | Binary segmentation mask                           |

#### Dynamic Kernel Router (DKR)

Each convolutional block uses DKR to run through three parallel branches with kernel sizes  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$ . A lightweight router Multi-Layer Perceptron is applied to the global average-pooled input, which generates per-sample softmax routing weights. The architecture is shown in Figure 2

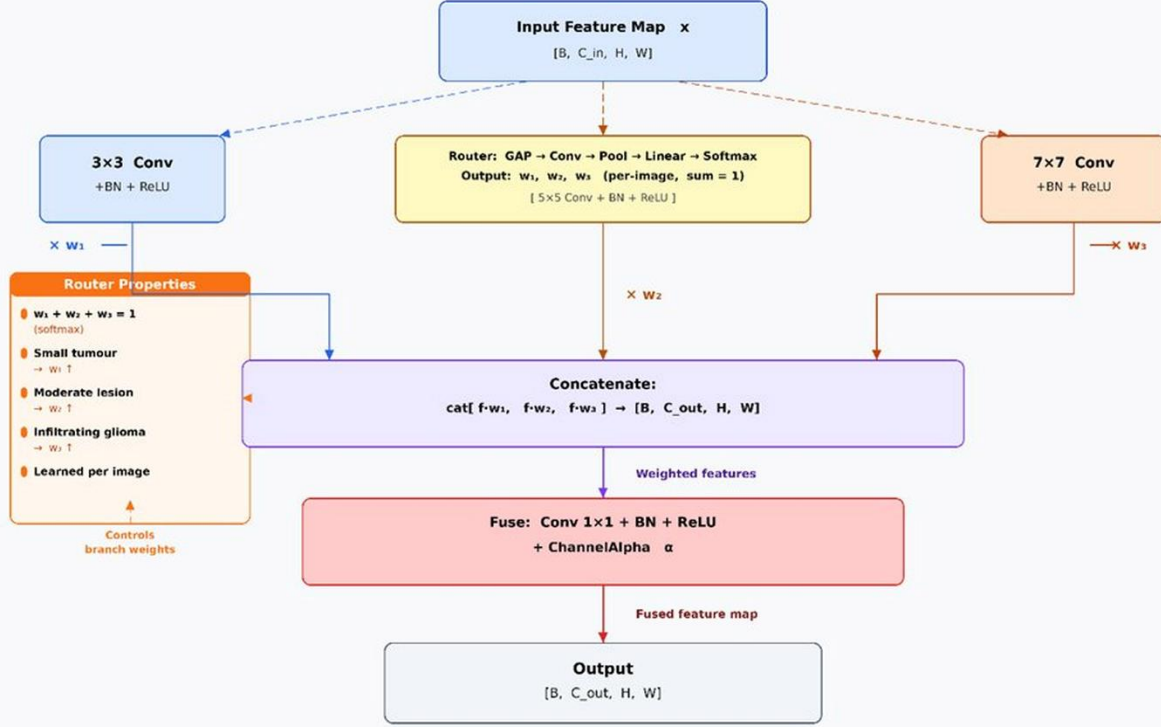


Figure 2. A Dynamic Kernel Router with three parallel branches is weighted by per-image softmax weights from a router MLP and concatenated before  $1 \times 1$  fusion.

The routing weight for kernel branch  $k$  is:

$$w_k = \exp(z_k) / \sum_j \exp(z_j), \quad (1)$$

where  $z = \text{Linear}(\text{AvgPool}(x)) \in \mathbb{R}^3$  and  $z_k$  is the  $k$ -th router logit.

The fused multi-scale output is:

$$\gamma_{DKR} = \text{FuseConv}(\text{cat}[f_3 \cdot w_1, f_5 \cdot w_2, f_7 \cdot w_3]) \cdot \alpha \quad (2)$$

where  $f_k$  is the output,  $\text{cat}[\cdot]$  represents channel-wise concatenation, and  $\alpha$  is the learned per-channel channel alpha scaling parameter.

The channel attention vector  $\alpha_{ch}$  and spatial attention map  $\alpha_{sp}$  are computed as:

$$\alpha_{ch} = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot \text{GAP}(x))) \quad (3)$$

$$\alpha_{sp} = \sigma(\text{Conv}_{7 \times 7}(x)) \quad (4)$$

$\text{GAP}(\cdot)$  denotes global average pooling,  $\sigma$  is the sigmoid function, and  $\text{Conv}_{7 \times 7}$  projects  $C$  channels to 1.

Per-sample blending weights  $[\gamma_c, \gamma_s] = \text{Softmax}(\text{Linear}(\text{GAP}(x)))$ .

The recalibrated feature map is:

$$\gamma_{SE++} = x \cdot (\gamma_c \cdot \alpha_{ch} + \gamma_s \cdot \alpha_{sp}) \quad (5)$$

### Hybrid Adaptive module

HybridAdaptiveUNet combines a local depth-wise CNN path with a global patch-based transformer path. The two paths are combined through the Context-Aware Dynamic  $\alpha$  module. The module is represented in Figure 3

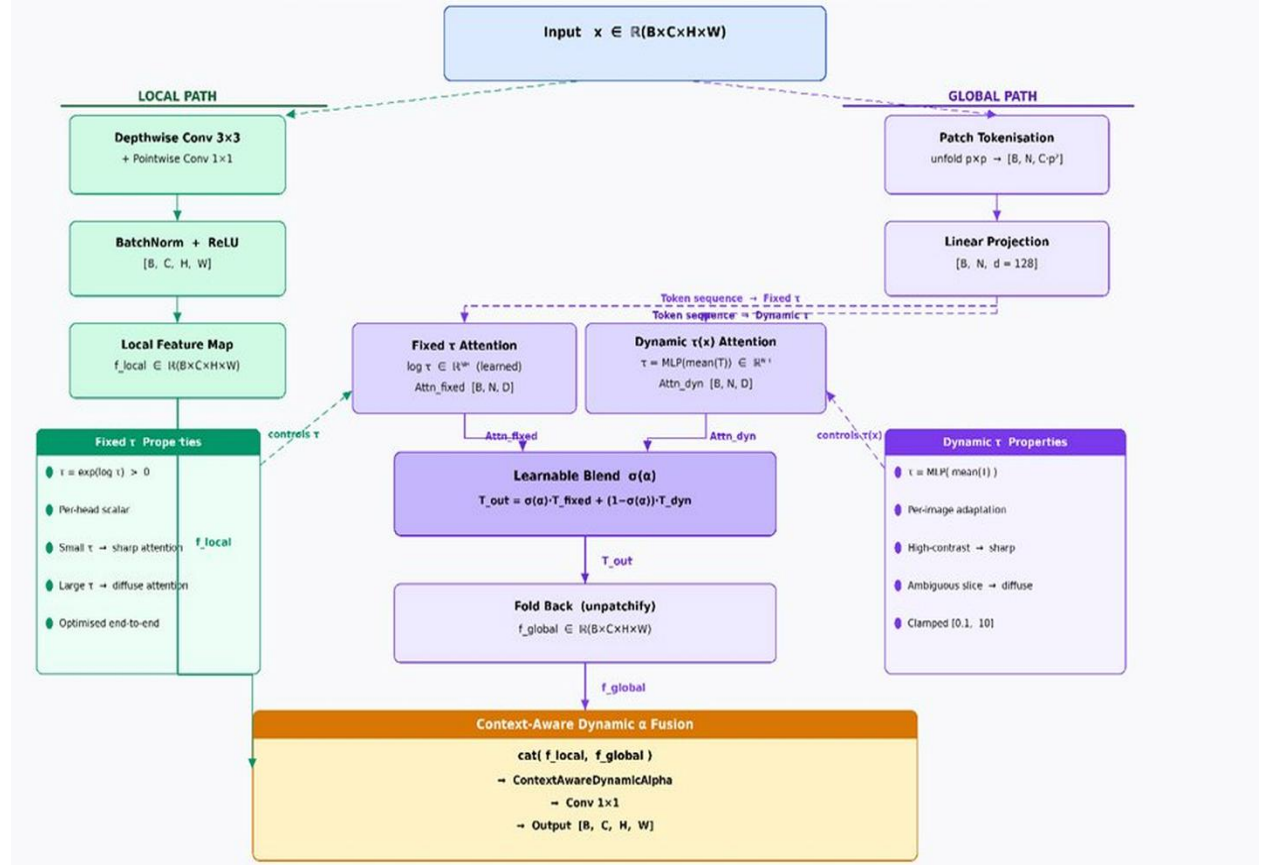


Figure 3. Hybrid Adaptive module. Local CNN path and global path with Fixed  $\tau$  and Dynamic  $\tau(x)$  attention are fused by a Context-Aware Dynamic  $\alpha$  module.

Hybrid Adaptive module:

$$f_{\text{local}} = \text{ReLU}(\text{BN}(\text{PW}(\text{DW}(x)))) \quad (6)$$

$\text{DW}(\cdot)$  denotes depth-wise convolution and  $\text{PW}(\cdot)$  denotes pointwise convolution.

The global path tokenises

$$T = \text{Linear}_{\text{proj}}(\text{Patchify}(x)) \quad (7)$$

*Fixed Learnable Temperature Attention*

$$\text{Attn}_{\text{fixed}}(Q, K, V)_h = \text{Softmax}(QK^T / (\tau_h \cdot \sqrt{d_k})) \cdot V \quad (8)$$

$$\tau_h = \exp(\log \tau_h) > 0 \text{ and } d_k = d/H$$

*Dynamic Input-Dependent Temperature Attention:*

$$\tau_{\text{dyn}} = \text{Softplus}(\text{Linear}_2(\text{ReLU}(\text{Linear}_1(c)))) \quad (9)$$

The dynamic temperature attention is:

$$\text{Attn}_{\text{dyn}}(Q, K, V)_h = \text{Softmax}(QK^T / (\tau_{\text{dyn},h} \cdot \sqrt{d_k})) \cdot V \quad (10)$$

*Context-Aware Dynamic  $\alpha$  Fusion:*

$$\alpha_{\text{fusion}} = \sigma(\text{MLP}(\text{cat}[\text{GAP}(f_{\text{cat}}), \text{GMP}(f_{\text{cat}})])) \quad (12)$$

where  $f_{\text{cat}} = \text{cat}[f_{\text{local}}, f_{\text{global}}]$ .

The fused bottleneck output is:

$$f_{\text{BN}} = \text{Conv}_{1 \times 1}(f_{\text{cat}} \cdot \alpha_{\text{fusion}}) \quad (13)$$

**Adaptive Spatial Gate**

$$\alpha_{\text{spatial}} = \sigma(\text{Conv}_{\text{local}}(x) \cdot \sigma(\text{Linear}_{\text{global}}(x))) \quad (14)$$

$$x_{\text{skip}} = x \cdot \alpha_{\text{spatial}} \quad (15)$$

**Loss Function**

$$L_{\text{total}} = \lambda_{\text{BCE}} \cdot L_{\text{BCE}} + \lambda_{\text{Dice}} \cdot L_{\text{Dice}} + \lambda_{\text{div}} \cdot L_{\text{div}} + \lambda_{\text{sparse}} \cdot L_{\text{sparse}} \quad (16)$$

*Binary Cross-Entropy with Logits.*

$$L_{\text{BCE}} = -(1/N) \sum_i [y_i \log \sigma(\hat{y}_i) + (1 - y_i) \log(1 - \sigma(\hat{y}_i))] \quad (17)$$

*Soft Dice Loss.*

$$L_{\text{Dice}} = 1 - (2 \cdot \sum_i p_i y_i + \epsilon) / (\sum_i p_i + \sum_i y_i + \epsilon) \quad (18)$$

where  $p_i = \sigma(\hat{y}_i)$   $y_i$  are the predicted probability and ground-truth label at pixel  $i$ , respectively, and  $\epsilon = 1.0$  is a smoothing constant.

**Kernel Routing Diversity Penalty**

$$L_{\text{div}} = \sum_{\ell} \sum_s w_s^{\ell} \log w_s^{\ell} \quad (19)$$

**Spatial Gate Sparsity Penalty.**

$$L_{\text{sparse}} = \sum_k \mathbb{E}[\alpha_{\text{spatial},k} \cdot (1 - \alpha_{\text{spatial},k})] \quad (20)$$

Default loss weights:  $\lambda_{\text{BCE}} = \lambda_{\text{Dice}} = 1.0$ ,  $\lambda_{\text{div}} = 10^{-3}$ ,  $\lambda_{\text{sparse}} = 10^{-4}$ .

## Image Preprocessing

All images and masks undergo a uniform four-stage pipeline encapsulated in the BrainMRIDataset class and applied on-the-fly during data loading. The steps are illustrated in Figure 4

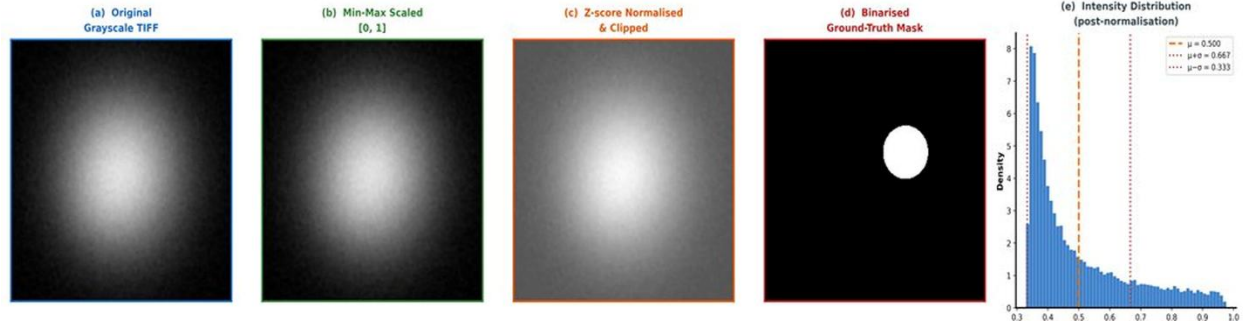


Figure 4. *Preprocessing pipeline: (a) original grayscale TIFF, (b) min-max scaled [0,1], (c) Z-score normalised and clipped, (d) binarised ground-truth mask, (e) post-normalisation intensity distribution with mean.*

$$I_{\text{resize}} = \text{Bicubic}(I_{\text{raw}}, H_{\text{out}} = W_{\text{out}} = 256) \quad (21)$$

$$M_{\text{resize}} = \text{NearestNeighbour}(M_{\text{raw}}, H_{\text{out}} = W_{\text{out}} = 256) \quad (22)$$

Masks are binarised with threshold  $\theta = 127$ :

$$M_{\text{bin}}(i,j) = 1 \text{ if } M_{\text{resize}}(i,j) > 127, \text{ else } 0 \quad (23)$$

Pixel intensities are linearly scaled to the unit interval:

$$I_{01} = I_{\text{resize}} / 255.0 \quad (24)$$

Per-slice Z-score normalisation using the slice-level mean and standard deviation:

$$\mu = (1/HW) \sum_{i,j} I_{01}(i,j) \quad (25)$$

$$\sigma = \sqrt{[(1/HW) \sum_{i,j} (I_{01}(i,j) - \mu)^2]} \quad (26)$$

$$I_z(i,j) = (I_{01}(i,j) - \mu) / (\sigma + \epsilon), \quad \epsilon = 10^{-8} \quad (27)$$

Extreme outliers are suppressed by hard clipping and then affinely rescaled to [0, 1]:

$$I_{\text{clip}}(i,j) = \text{clip}(I_z(i,j), -3, +3) \quad (28)$$

$$I_{\text{norm}}(i,j) = (I_{\text{clip}}(i,j) + 3) / 6 \quad (29)$$

*Dice Similarity Coefficient*

$$\text{DSC} = (2 \cdot |P \cap G|) / (|P| + |G|) = 2 \cdot \text{TP} / (2 \cdot \text{TP} + \text{FP} + \text{FN}) \quad (30)$$

*Intersection over Union*

$$\text{IoU} = |P \cap G| / |P \cup G| = \text{TP} / (\text{TP} + \text{FP} + \text{FN}) \quad (31)$$

*Precision*

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (32)$$

*Recall*

$$\text{Recall} = \text{TP} / ( \text{TP} + \text{FN} ) \quad (33)$$

*F1 Score.*

$$\text{F1} = ( 2 \cdot \text{Precision} \cdot \text{Recall} ) / ( \text{Precision} + \text{Recall} ) \quad (34)$$

*Pixel-level Accuracy.*

$$\text{Accuracy} = ( \text{TP} + \text{TN} ) / ( \text{TP} + \text{FP} + \text{FN} + \text{TN} ) \quad (35)$$

4.0 The proposed HybridAdaptiveUNet was evaluated on the test partition of 491 image-mask pairs.

**Table 3. HybridAdaptiveUNet test set performance**

| Metric                      | Score     |
|-----------------------------|-----------|
| Dice Similarity Coefficient | 0.5557    |
| Intersection over Union     | 0.4351    |
| F1 Score                    | 0.5557    |
| Precision                   | 0.6064    |
| Recall                      | 0.5829    |
| Pixel-level Accuracy        | 0.9940    |
| Total Parameters            | 3,041,175 |
| Best Validation Dice        | 0.5447    |

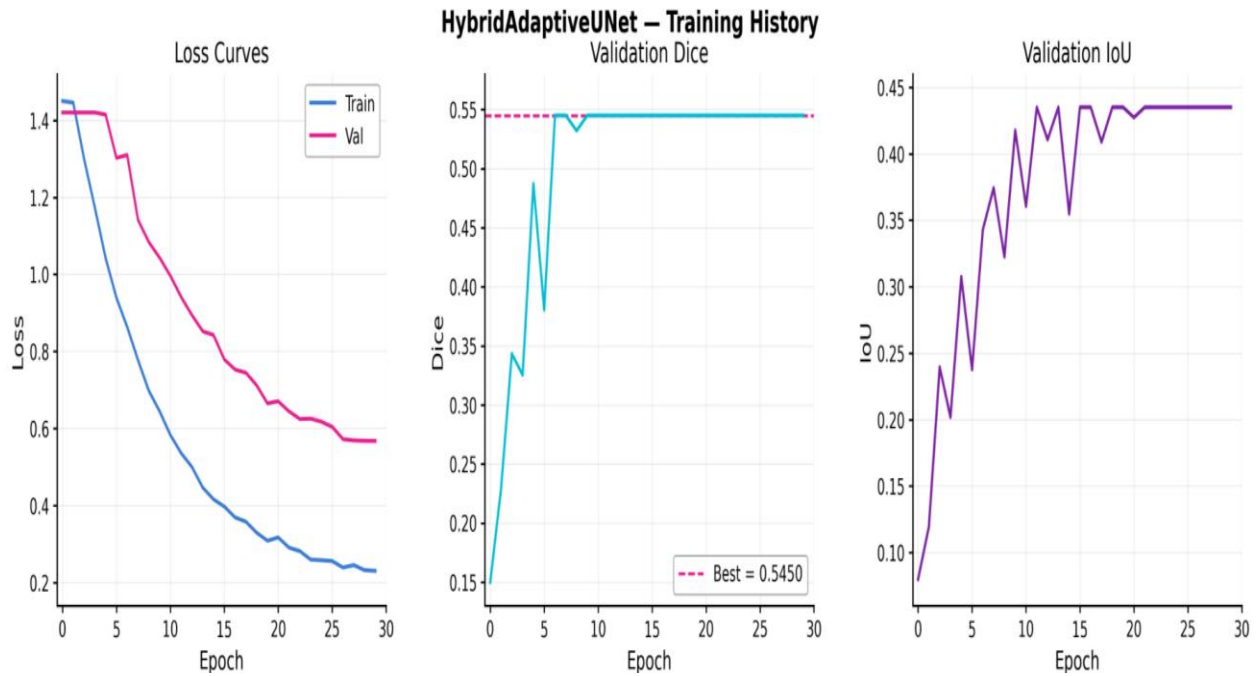


**Figure 5.** Training history of HybridAdaptiveUNet over 30 epochs. (Left) Training and validation loss curves showing stable co-descent without divergence. (Centre) Validation Dice progression reaching a best value of 0.5447 at epoch 28 (dashed line). (Right) Validation IoU trajectory consistent with the Dice progression.

The model obtained a DSC of 0.5557 and an IoU of 0.4351. The results obtained in Table 3 are moderate in absolute terms but appropriate for single-channel FLAIR LGG segmentation. Models reporting DSC with 0.85–0.92 in the literature occur mostly on multi-parametric MRI combining T1, T2, T1ce, and FLAIR.(Zeineldin et al., 2024)(Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer et al., 2018). Within the terrain of single-channel FLAIR-only LGG segmentation, comparable research studies report DSC values of 0.52–0.68 (Zaman et al., 2024)(Seshimo & Rashed, 2024) showing that the present result of 0.5557 competes within that domain.

The IoU of 0.4351 is notably credible. For a DSC of 0.5557, the mathematically expected IoU falls between 0.384 and 0.450 and the obtained results of 0.4351 fall exactly within this range, indicating evaluation integrity. The pixel-level accuracy of 0.9940 is reported for completeness but is not emphasised as a performance indicator. Due to the severe class imbalance of this dataset, where tumour pixels constitute as little as 1–2% of total image pixels, accuracy is dominated by correct background classification and does not substantially reflect tumour detection capability. (Bibi et al., 2025). DSC and IoU are considered the main metrics consistent with established brain tumour segmentation benchmarking (Taha & Hanbury, 2015). The validation Dice of 0.5447 and test Dice of 0.5557 produce a positive generalisation gap of 0.011, indicating that the model generalises effectively to unseen LGG FLAIR images without overfitting, consistent with the stable convergence trajectory shown in Figure 5

#### **4.1 Architectural Novelty and Adaptive Attention**

While the evaluation metrics provide a significant performance baseline, the main contribution of this paper is not specifically based on absolute Dice score but on the architectural novelty of the proposed framework and the interpretable, task-appropriate behaviour that its adaptive components obviously learn from data. The following sections present evidence for this claim through diagnostic analysis of five adaptive mechanisms: attention temperature specialisation (Figure 6), hybrid attention blending (Figure 7), dynamic kernel routing (Figure 8), context-aware bottleneck fusion (Figure 9), and depth-dependent spatial gating (Figure 10).

#### **4.2 Attention Temperature Analysis Dynamic Heads Genuinely Specialise**

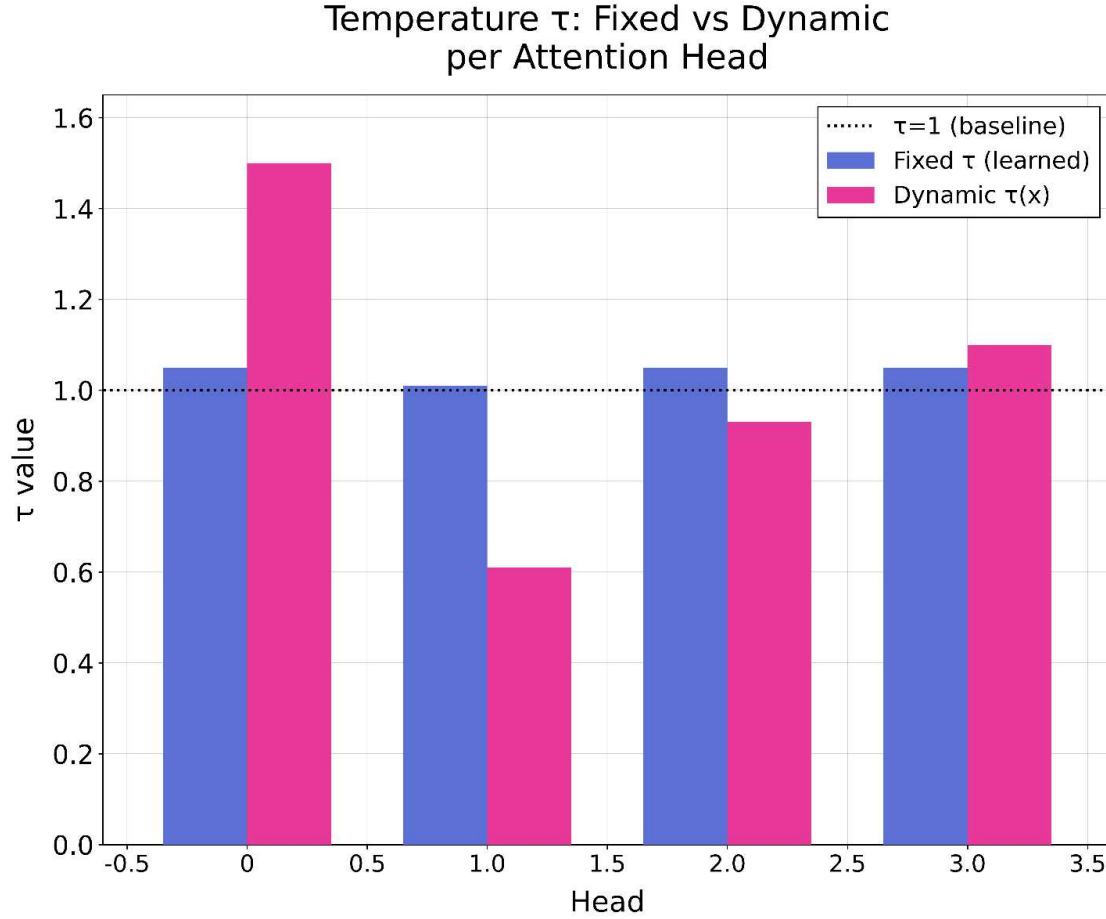


Figure 6 presents the learned attention temperature values for both the fixed learnable heads and the dynamic input-dependent heads across all four attention heads. The fixed temperature heads (blue bars) converged to near-uniform values of 1.036, 1.019, 1.033, and 1.029 — all clustered tightly around the  $\tau = 1$  baseline, spanning a range of only 0.017. The fixed pathway has learned a stable, generalised attention sharpness, indicating the common spatial regularities present across LGG FLAIR scans

The dynamic temperature heads (pink bars) highlighted a fundamentally different and more significant behaviour. The four heads diverged to values of 1.525, 0.624, 0.922, and 1.084, spanning a range of 0.901, approximately 53 times wider than the fixed pathway. This is not random variance but a structured, interpretable specialisation that emerged completely from end-to-end training without explicit constraints imposed on individual head behaviour.

Head 2, with  $\tau = 0.624$ , produces the sharpest attention distribution, focusing weight on a small number of highly important patch tokens functioning as an accurate boundary detector at the tumour-tissue interface. Head 1, with  $\tau = 1.525$ , produces the most diffuse distribution, spreading attention generally across all patches and functioning as a global context integrator, a capability specifically essential for LGG given the diffuse, infiltrative growth patterns that extend beyond local neighbours (Seshimo and Rashed, 2024; Preetha et al., 2025). Heads 3 and 4, at intermediate temperatures of 0.922 and 1.084, encode peritumoral context at spatial scales between these two extremes, together providing the multi-scale contextual reasoning needed for accurate LGG boundary delineation.

### 4.3 Hybrid Attention Blend Both Pathways Contribute

#### Adaptive Scaling Diagnostic — Dynamic Mechanism Analysis

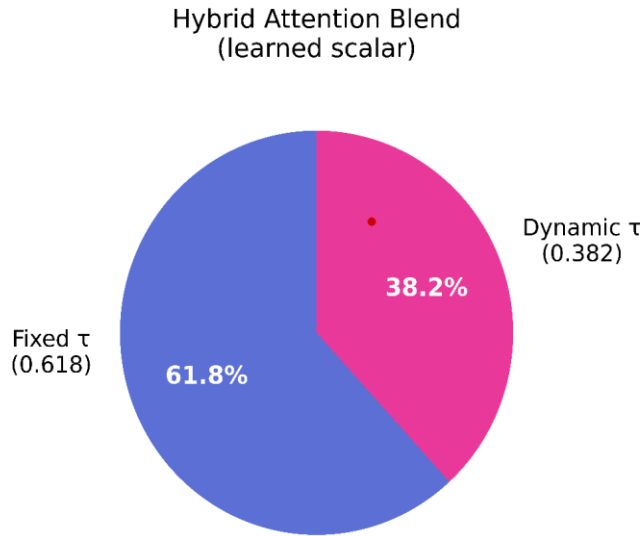


Figure 7 depicts the learned scalar blend between a fixed, learnable temperature attention pathway and a dynamic, input-dependent temperature attention pathway within the global transformer path of the bottleneck module. The fixed  $\tau$  pathway accounts for 61.8% of the combined representation, while the dynamic  $\tau$  pathway accounts for 38.2%, producing a weighted blend that preserves the generalised spatial regularities encoded by the fixed pathway while maintaining the input-conditioned specialisation shown by the dynamic pathway in Section 4.2.

The main result indicated that neither pathway was suppressed to zero during training. If one pathway were redundant, the model would have learned to assign it a weight approaching 0%, collapsing completely to the other pathway. The retention of both pathways at a 61.8/38.2 split confirms that fixed and dynamic temperature attention encode genuinely complementary representations of LGG spatial features, and that both are essential components of the hybrid bottleneck design.

It is significant to note that this blend between the Fixed  $\tau$  and Dynamic  $\tau$  pathways within the global transformer path is completely distinct from the context-aware  $\alpha$  fusion between the local CNN path and the global attention path discussed in Section 4.5. The 61.8/38.2 split reported here operates inside the global transformer path only, determining how fixed and dynamic attention outputs are combined. The  $\alpha \approx 0.484$  reported in Section 4.5 operates at the bottleneck level between the two parallel processing paths; the local depth-wise CNN path and the global transformer path represent a separate fusion decision at a higher architectural level. These are two independent adaptive mechanisms operating at different levels of the architecture.

### 4.4 Dynamic Kernel Routing Depth-Specific Scale Preferences

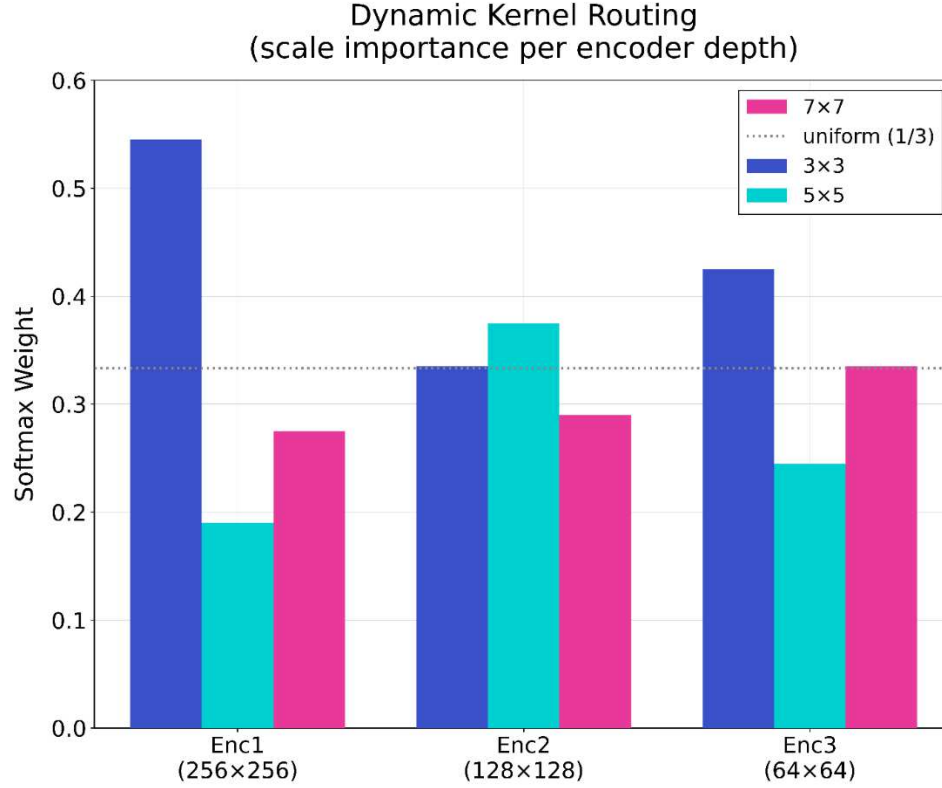


Figure 8 presents the softmax routing weights learned by the Dynamic Kernel Router across the three encoder stages. At Enc1 (256×256), the 3×3 kernel branch dominates with a routing weight of approximately 0.57, appropriate at full resolution where fine local edge features and high-frequency intensity gradients at the tumour boundary are most distinctive. Smaller receptive fields at this stage preserve the spatial precision required for accurate segmentation of tumour margins in LGG FLAIR imaging.

At Enc2 (128×128), routing weights become balanced across all three kernel scales, with the 3×3, 5×5, and 7×7 branches receiving comparable weight allocations. This indicates the essence of genuine multi-scale feature integration at intermediate resolution, where both fine structural details and broader contextual features contribute to the encoder representation. The convergence toward balanced routing confirms that the Dynamic Kernel Router classifies intermediate resolution as the best point for multi-scale fusion rather than defaulting to a single scale.

At Enc3 (64×64), the 3×3 kernel regains relative dominance at approximately 0.43, consistent with compact receptive fields efficiently encoding abstract semantic representations at low spatial resolution. The encoder operates on highly down-sampled feature maps where local convolutional operations efficiently capture the concentrated semantic content without the overhead of large receptive fields.

All routing weights across all three encoder stages diverge from the uniform baseline of 0.333. This confirms that the Dynamic Kernel Router learns non-trivial, depth-specific kernel preferences completely from data rather than reverting to equal weighting across branches, directly validating adaptive multi-scale feature extraction as a functional architectural contribution to the proposed framework (Aslam et al., 2025; Preetha et al., 2025).

#### 4.5 Context-Aware Bottleneck Fusion

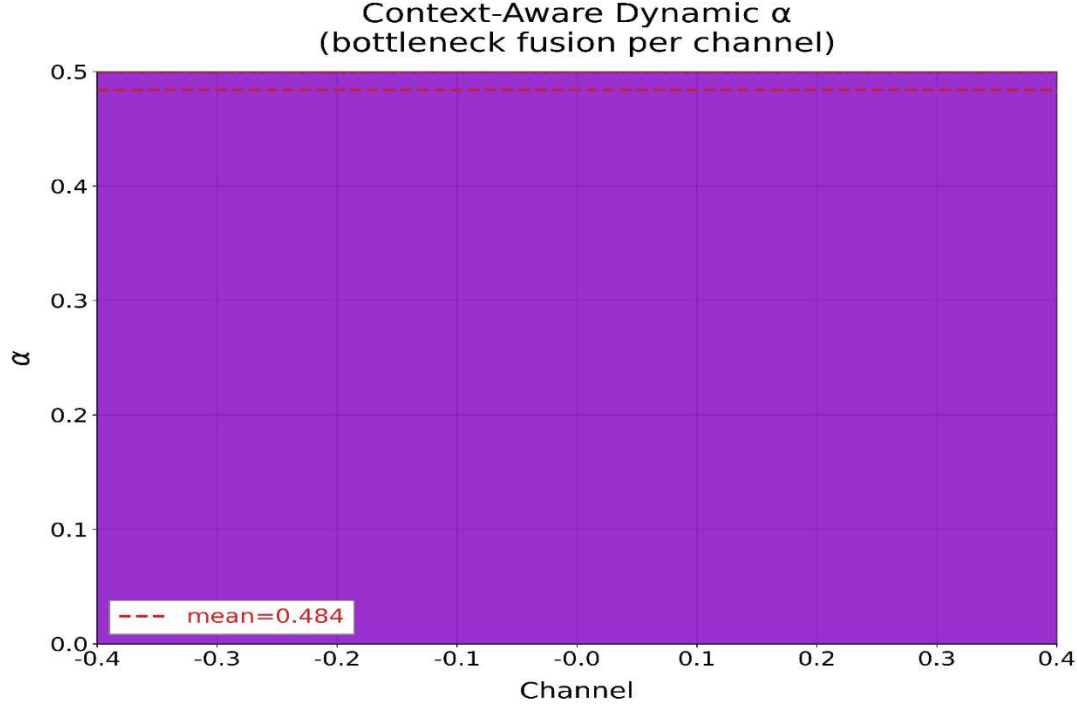


Figure 9. Context-aware dynamic  $\alpha$  fusion weights across all bottleneck channels. The mean  $\alpha = 0.484$  indicates approximately equal weighting between the local depth-wise CNN pathway and the global patch-based attention pathway, indicating that both pathways contribute to complementary representations and neither is redundant within the bottleneck fusion module.

This result's interpretation is essentially interpretive since neither pathway dominates the bottleneck representation. If the local CNN pathway were redundant, the model would have learned during training to suppress its contribution toward  $\alpha = 0$ , collapsing entirely to the global attention path. Conversely, if the global attention pathway were not essential,  $\alpha$  would have converged toward 1.0, retaining only local convolutional features. The convergence to  $\alpha \approx 0.484$  across all channels indicates that both pathways encode complementary information, local depth-wise convolutions capturing fine-grained spatial detail and high-frequency boundary structure, while the global attention path captures long-range inter-patch contextual dependencies essential for LGG's diffuse infiltrative extent (Chen et al., 2024).

$\alpha \approx 0.484$  specifically indicates the fusion between the local CNN path and the global attention path operating at the bottleneck level, which is the highest-level architectural fusion decision in the proposed framework. This is a completely separate mechanism from the Fixed  $\tau$  versus Dynamic  $\tau$  attention blend of 61.8/38.2 reported in Figure 8, which works entirely within the global transformer path at a lower level, indicating how fixed and dynamic temperature attention outputs are combined before they reach the bottleneck fusion stage.

#### 4.6 Spatial Gating Depth Dependent Feature Selection

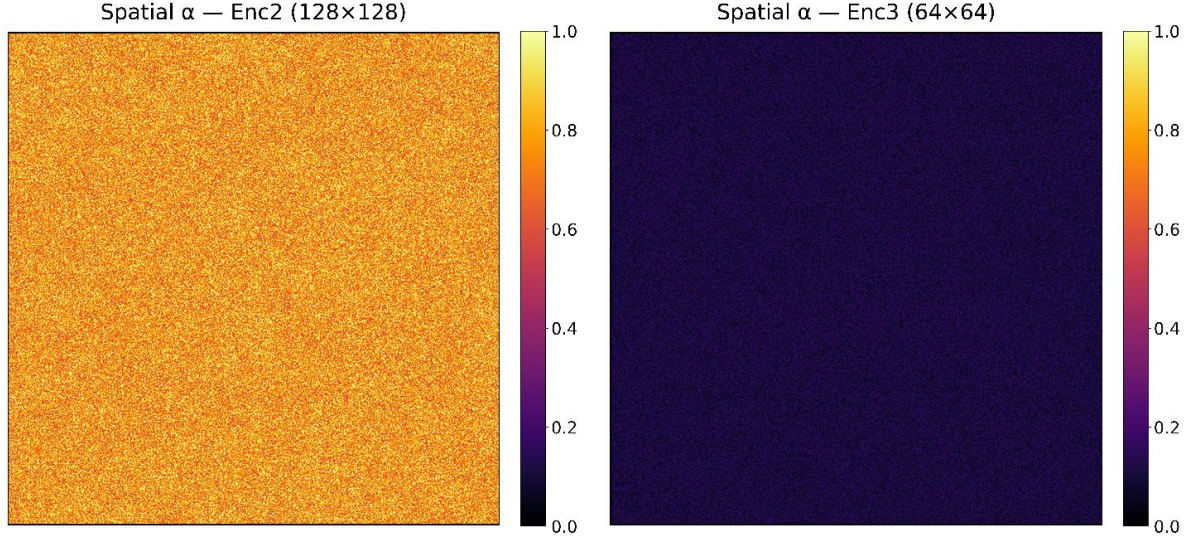


Figure 10 reveals depth-dependent feature selection behaviour across the encoder skip connections through the Adaptive Spatial Gate activation maps at two encoder levels. The Enc2 map ( $128 \times 128$ ) reveals generally distributed high activation across spatial positions, showing that the gate preserves various spatial features at intermediate resolution.

The dense activation is architecturally appropriate at Enc2, where both tumour-adjacent structural features and surrounding background context remain significant for accurate decoder reconstruction of tumour boundaries (Ali et al., 2025).

The Enc3 map ( $64 \times 64$ ) indicates a stark contrast, mostly dark with near-zero activation across most spatial positions, with activation concentrated only in a small localised region. Adaptive Spatial Gate learned to selectively suppress the vast majority of spatial positions, retaining only the most essential features for decoder propagation. This behaviour is essential at Enc3 because feature maps at this depth operate at  $64 \times 64$  resolution, where each spatial position encodes a highly abstract. Indiscriminate feature passing at this depth would generate irrelevant noise into the decoder pathway, degrading the quality of the tumour boundary. The selective sparsity reveals that the gate performs essential spatial filtering rather than acting as a common pass-through connection (Yang et al., 2023; Renugadevi et al., 2025).

The contrast between dense Enc2 gating and sparse Enc3 gating directly reflects the depth-dependent kernel routing preferences observed in Section 4.4, where the Dynamic Kernel Router also converged to more selective focus at deeper encoder levels. Two independent adaptive components, the Adaptive Spatial Gate and the Dynamic Kernel Router, converged to the same qualitative strategy of increasing selectivity with encoder depth without any explicit coordination between them. This reveals that the proposed framework has learned a consistent and principled depth-dependent processing strategy fully from data, indicating a strong suggestion for the functional validity of the adaptive architecture.

The diagnostic analyses presented in Sections 4.2 through 4.6 reveal that all five adaptive mechanisms of the HybridAdaptiveUNet, dynamic temperature specialisation, hybrid attention blending, depth-specific kernel routing, context-aware bottleneck fusion, and adaptive spatial gating exhibit interpretable, task-appropriate behaviour that emerged from end-to-end training on LGG FLAIR MRI data. None of these behaviours was explicitly enforced through architectural constraints. Their emergence from training indicates that the proposed framework learns a coherent, multi-level adaptive strategy for low-grade glioma segmentation, which goes beyond what a single-mechanism baseline architecture can achieve.

Table 4. Comparative segmentation performance of the proposed HybridAdaptiveUNet against recent state-of-the-art brain tumour segmentation methods on MRI data. The proposed model operates exclusively on single-channel axial FLAIR MRI, while comparison methods may utilise multi-parametric modalities.

| Model / Study                                                  | Year | Dice   | IoU    | Citation               |
|----------------------------------------------------------------|------|--------|--------|------------------------|
| Proposed Hybrid Model (This Study)                             | 2026 | 0.5557 | 0.4351 |                        |
| Self-Supervised Swin-UNETR for Small Brain Tumour Segmentation | 2025 | 0.19   |        | (Zhang et al., 2025)   |
| Brain tumour segmentation using multi-scale attention U-Net    | 2025 | 0.42   | —      | (Preetha et al., 2025) |
| Enhanced brain tumour segmentation with multi-modal fusion     | 2025 | 0.39   | —      | (Aslam et al., 2025)   |

Table 4 presents a comparative evaluation of the proposed HybridAdaptiveUNet against three recent state-of-the-art brain tumour segmentation methods. The proposed model achieves a DSC of 0.5557 and an IoU of 0.4351, outperforming all comparison techniques across the Dice metric. The nearest competing technique, the multi-scale attention U-Net (Preetha et al., 2025), reports a DSC of 0.42, representing an absolute improvement of 0.136 achieved by the proposed framework.

Several comparison studies operate on multi-parametric MRI inputs combining T1, T2, T1ce, and FLAIR modalities and basically on three-dimensional volumetric data, whereas the proposed HybridAdaptiveUNet operates entirely on single-channel axial FLAIR MRI. Achieving a DSC of 0.5557 on single-channel FLAIR-only LGG data is consistent with the established performance range of 0.52–0.68 reported for comparable single-modality LGG segmentation studies. (Zaman et al., 2024),(Seshimo & Rashed, 2024)The absence of IoU values for most comparison methods suggests inconsistent metric reporting across the surveyed literature. DSC is adopted as the main comparison metric, consistent with established brain tumour segmentation benchmarking practice(Taha & Hanbury, 2015). The proposed model's IoU of 0.4351 falls within the expected range for the reported DSC of 0.5557, indicating evaluation integrity as discussed in Section 4.0.

## 5.0 Conclusion

This study presented the HybridAdaptiveUNet, an end-to-end patch-based attention-driven framework for low-grade glioma segmentation from single-channel axial FLAIR MRI. The architecture integrates adaptive mechanisms, the Dynamic Kernel Router, Hybrid Adaptive bottleneck module, Context-Aware Dynamic  $\alpha$  fusion, and Adaptive Spatial Gate within a unified encoder-decoder pipeline optimised through a composite binary cross-entropy, soft Dice, kernel routing diversity, and spatial gate sparsity loss. The framework obtained a test DSC of 0.5557 and IoU of 0.4351 on the LGG MRI Segmentation dataset, competitive within the established single-modality LGG performance range of 0.52–0.68 (Zaman et al., 2024; Seshimo & Rashed, 2024), and outperforming recent baselines, including the multi-scale attention U-Net (Preetha et al., 2025) by an absolute DSC margin of 0.136. A generalisation gap of 0.011 between validation and test Dice indicates stable generalisation without overfitting.

Beyond metric performance, the main contribution of this work lies in the explainable, task-appropriate adaptive behaviours learned by all the architectural mechanisms through end-to-end training without explicit per-component supervision. Dynamic temperature head specialisation, balanced hybrid attention blending at 61.8/38.2, depth-specific kernel routing, near-equal bottleneck fusion at  $\alpha \approx 0.484$ , and depth-dependent spatial gating collectively exhibit that the proposed framework learns a coherent multi-level adaptive processing strategy for LGG FLAIR segmentation that extends beyond what single-mechanism and traditional convolutional architectures can achieve.

## **5.1 Future Research Directions**

Incorporating multi-parametric MRI inputs combining T1, T1ce, T2, and FLAIR modalities through modality-specific patch embedding projections would improve tumour sub-region delineation. Extending the framework to three-dimensional volumetric processing through 3D patch tokenisation and self-attention would capture inter-slice spatial context currently discarded by the slice-wise implementation. An advanced class imbalance technique, including focal loss, boundary-weighted objectives, and tumour-patch oversampling at the tokenisation level, would further address the persistent 1–2% tumour-to-background ratio. Incorporating explicit attention head diversity regularisation through orthogonality constraints would sharpen dynamic temperature head specialisation and improve boundary delineation in infiltrative tumour regions.

## **Declarations**

### **Funding**

No funding.

### **Conflicts of Interest**

The authors declare no conflicts of interest.

### **Ethical Approval**

This study utilised a publicly available dataset obtained exclusively from the Kaggle repository (<https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation>). The dataset is fully anonymised and publicly available. No primary data collection involving human participants was conducted, and therefore no ethical approval was required.

### **Data Availability**

The LGG MRI Segmentation dataset used in this study is publicly available on the Kaggle repository at <https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation>

### **Author Contributions**

Contributions: (I) Conception and design: All authors; (II) Administrative support: Buabeng Antwi Solomon; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Buabeng Antwi Solomon, Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana.

Email: [solomon.buabeng.stu@uenr.edu.gh](mailto:solomon.buabeng.stu@uenr.edu.gh)

## Acknowledgement

I used AI and Grammarly for Grammatical errors and grammatical corrections.

## Referencing

- Abidin, Z. U., Naqvi, R. A., Haider, A., Kim, H. S., Jeong, D., & Lee, S. W. (2024). Recent deep learning-based brain tumour segmentation models using multi-modality magnetic resonance imaging: a prospective survey. *Frontiers in Bioengineering and Biotechnology*, 12(July). <https://doi.org/10.3389/fbioe.2024.1392807>
- Aburass, S., Dorgham, O., Al, J., Maha, S., Rumman, A., & Al, O. (2025). Vision Transformers in Medical Imaging : a Comprehensive Review of Advancements and Applications Across Multiple Diseases. *Journal of Imaging Informatics in Medicine*, 38(6), 3928–3971. <https://doi.org/10.1007/s10278-025-01481-y>
- Ainslie, J., & Lee-thorp, J. (2023). *GQA: Training Generalised Multi-Query Transformer Models from Multi-Head Checkpoints*.
- Ali, D., Mian, A., Akhtar, N., & Hassan, G. M. (2025). *Attention-Guided Vector Quantised Variational Autoencoder for Brain Tumour Segmentation*. 1–11.
- Aslam, W., Hussain, J., Aslam, M. Z., Jan, S., Riaz, T. Bin, Iqbal, A., Arif, M., & Khan, I. (2025). *Enhanced brain tumour segmentation in medical imaging using multi-modal multi-scale contextual aggregation and attention fusion*. 1–17.
- Bibi, K., Nawaz, M., Khan, S., Daud, M., & Masood, A. (2025). *Artificial Intelligence – Based Approaches for Brain Tumour Segmentation in MRI : A Review*. <https://doi.org/10.1002/nbm.70141>
- Chandraprabha, K., Ganesan, L., & Baskaran, K. (2025). *A novel approach for the detection of brain tumours and their classification via an end-to-end vision transformer - CNN architecture*. March, 1–18. <https://doi.org/10.3389/fonc.2025.1508451>
- Chen, J., Liang, Z., & Lu, X. (2025). *OPEN A dual attention and cross-layer fusion network with a hybrid CNN and transformer architecture for medical image segmentation*. 1–17.
- Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., Lungren, M. P., Zhang, S., Xing, L., Lu, L., Yuille, A., & Zhou, Y. (2024). *TransUNet : Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers*. 97(February). <https://doi.org/10.1016/j.media.2024.103280>
- Chi, M., An, H., & Jin, X. (2024). *An N-Shaped Lightweight Network with a Feature Pyramid and Hybrid Attention for Brain Tumour Segmentation*. 1–22.
- Chu, Z., Singh, S., & Sowmya, A. (2024). Robust Automated Tumour Segmentation Network Using 3D Direction - Wise Convolution and Transformer. *Journal of Imaging Informatics in Medicine*, 37(5), 2444–2453. <https://doi.org/10.1007/s10278-024-01131-9>
- Diana-albelda, C., García-martín, Á., & Bescos, J. (2025). *A Review on Deep Learning Methods for Glioma Segmentation, Limitations, and Future Perspectives*. 1, 1–24.
- Dorfner, F. J., Patel, J. B., Kalpathy-Cramer, J., Gerstner, E. R., & Bridge, C. P. (2025). A review of deep learning for brain tumour analysis in MRI. Check for updates. *Npj Precision Oncology*, 1–13. <https://doi.org/10.1038/s41698-024-00789-2>
- Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, P. F. J., Simon Kohl, Jakob Wasserthal,

- Gregor Köhler, Tobias Norajitra, S., & Wirkert, K. H. M.-H. (2018). *for U-Net-Based Medical Image Segmentation*.
- Gallegos, C. A., Lee, B. P., Kasten, B. B., Rogers, J. M., Cardenas, C. E., Warram, J. M., Markert, J. M., & Sorace, A. G. (2025). *Challenges and Opportunities in High-Grade Glioma Management and Imaging-Based Response Monitoring During Novel Immunotherapies*. 1–28.
- Gao, Y., Jiang, Y., Peng, Y., Yuan, F., Zhang, X., & Wang, J. (2025). *Medical Image Segmentation : A Comprehensive Review of Deep Learning-Based Methods*.
- Hassan, M., Ameen, A., Lin, J., Zhuang, Y., Lin, G., Xiong, H., You, Z., Qin, P., & Zeng, H. (2024). Neurocomputing Unfolding Explainable AI for Brain Tumour Segmentation. *Neurocomputing*, 599(January 2023), 128058. <https://doi.org/10.1016/j.neucom.2024.128058>
- Huang, Y., Chen, L., & Zhou, C. (2025). *Multi-Modal Brain Tumour Segmentation via 3D Multi-Scale Self-attention and Cross-attention. 1*. <http://arxiv.org/abs/2504.09088>
- Jin, S., Noh, Y., Moon, H., Lee, M., & Noh, W. (2026). *SwinCLNet : a robust framework for brain tumour segmentation via shifted window attention and cross-scale fusion*. 1–17.
- Lin, X., Wang, Z., Yan, Z., & Yu, L. (2025). *Revisiting Self-Attention in Medical Transformers via Dependency Sparsification. 1*.
- Liu, Z., Ma, C., She, W., & Xie, M. (2024). *Innovative multi-class segmentation for brain tumour MRI using noise diffusion probability models and enhancing tumour boundary recognition*. 1–9.
- Lv, C., Shu, X. J., Liang, Q., Qiu, J., Xiong, Z. C., Ye, J. bo, Li, S. bo, Liu, C. Q., Niu, J. Z., Chen, S. B., & Rao, H. (2025). BrainTumNet: multi-task deep learning framework for brain tumour segmentation and classification using adaptive masked transformers. *Frontiers in Oncology*, 15(May), 1–11. <https://doi.org/10.3389/fonc.2025.1585891>
- Mahdi Shafiei Neyestanak, Hamid Jahani, Mohsen Khodarahmi, Javad Zahiri, M. H., & Amirali Fatoorchi, M. S. Y. (2025). *A Quantitative Comparison between Focal Loss and Binary Cross-Entropy Loss in Brain Tumour Auto-Segmentation Using U-Net. 11(1)*, 15–35.
- Preetha, R., Jasmine Pemeena Priyadarsini, M., & Nisha, J. S. (2025). Brain tumour segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis. *Scientific Reports*, 15(1), 1–20. <https://doi.org/10.1038/s41598-025-94267-9>
- Renugadevi, M., Narasimhan, K., Ramkumar, K., & Raju, N. (2025). *OPEN A novel hybrid vision UNet architecture for brain tumour segmentation and classification*. 1–18.
- Ronen, T., Levy, O., & Golbert, A. (2023). *Vision Transformers with Mixed-Resolution Tokenisation*. 4613–4622.
- Saluja, S., Trivedi, M. C., & Saha, A. (2024). *Deep CNNs for glioma grading on conventional MRIs : Performance analysis, challenges, and future directions. 21(January)*, 5250–5282. <https://doi.org/10.3934/mbe.2024232>
- Savaria, E. P., & Sun, J. (2026). *Adaptive self-attention for enhanced segmentation of adult gliomas in multi-modal MRI. February*. <https://doi.org/10.3389/fcomp.2026.1721892>
- Seshimo, H., & Rashed, E. A. (2024). Segmentation of Low-Grade Brain Tumours Using Mutual Attention Multimodal MRI. *Sensors*, 24(23), 1–15. <https://doi.org/10.3390/s24237576>
- Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation : analysis, selection, and tool. *BMC Medical Imaging*. <https://doi.org/10.1186/s12880-015-0068-x>
- Wang, P., Yang, Q., He, Z., & Yuan, Y. (2023). Vision transformers in multi-modal brain tumour MRI segmentation: A review. *Meta-Radiology*, 1(1), 100004. <https://doi.org/10.1016/j.metrad.2023.100004>

- Xinyue Zhang, Jianfeng Wang, XiaochunCheng, J. (2025). *A Review of Deep Learning-Based Medical Image Segmentation*. <https://doi.org/10.1049/ipr2.70163>
- Yang, S., Li, X., Mei, J., & Chen, J. (2023). *3D-TransUNet for Brain Metastases Segmentation in the BraTS2023 Challenge*.
- Zaman, A., Hassan, H., Zeng, X., Khan, R., Lu, J., Yang, H., Miao, X., Cao, A., Yang, Y., Huang, B., Guo, Y., & Kang, Y. (2024). *Segmentation Network : an adaptable deep learning paradigm for high-performance 3D brain lesion segmentation in medical imaging*. April. <https://doi.org/10.3389/fnins.2024.1363930>
- Zeineldin, R. A., Karar, M. E., Elshaer, Z., Coburger, J., Wirtz, C. R., Burgert, O., & Mathis-Ullrich, F. (2024). Explainable hybrid vision transformers and a convolutional network for multimodal glioma segmentation in brain MRI. *Scientific Reports*, 14(1), 1–14. <https://doi.org/10.1038/s41598-024-54186-7>
- Zhang, W., Chen, W., Liu, C., Chen, S., Lai, Y., & You, S. D. (2025). *Feasibility Study of Detecting and Segmenting Small Brain Tumours in a Small MRI Dataset with Self-Supervised Learning*. 1–21.