

Article

Hybrid Transformer–CNN with Boundary-Aware Attention for Accurate Multi-Modal Brain Tumor Segmentation

Jamshid Khamzaev ¹, Jakhongir Karimberdiyev ¹, Mekhridin Rakhimov ¹, Islambek Saymanov ^{2,3},
Shavkat Otamurodov ⁴, Odiljon Rikhsimboev ⁵, Ilin Dmitriy ⁶, Alpamis Kutlimuratov ⁷
and Fazliddin Makhmudov ^{8,9,*}

¹ Department of Computer Systems, Tashkent University of Information Technologies Named After Muhammad Al-Khwarizmi, Tashkent 100084, Uzbekistan; hamzayevjf@gmail.com (J.K.); jahongirkarimberdiyev618@gmail.com (J.K.); raximov022@gmail.com (M.R.)

² Department of Scientific Research, New Uzbekistan University, Mustaqillik Ave. 54, Tashkent 100007, Uzbekistan; i.saymanov@newuu.uz

³ Applied Mathematics and Intelligent Technologies Faculty, National University of Uzbekistan, Tashkent 100174, Uzbekistan

⁴ Department of Economics, Termez University of Economics and Service, Termez 190111, Uzbekistan; shavkat_otamurodov@tues.uz

⁵ Department of Econometrics, Tashkent State University of Economics, Tashkent 100066, Uzbekistan; o.rixsimbaev@tsue.uz

⁶ Department of Information Technology and Automation of Technological Processes and Production, National University of Science and Technology MISIS in Almalyk, Almalyk 110100, Uzbekistan; dmitriy.nikolayevi4@gmail.com

⁷ Department of Applied Informatics, Kimyo International University in Tashkent, Tashkent 100121, Uzbekistan; kutlimuratov.aj@kiut.uz

⁸ Department of Computer Engineering, Gachon University, Seongnam 13120, Republic of Korea

⁹ Department of Digital Technology and Mathematics, Kokand University, Kokand 150700, Uzbekistan

* Correspondence: fazliddin12@gachon.ac.kr

Abstract

Background: Accurate segmentation of brain tumors from multi-modal magnetic resonance imaging (MRI) is essential for diagnosis, treatment planning, and therapy monitoring. However, this task remains challenging due to tumor heterogeneity, irregular boundaries, and the complex anatomical structure of surrounding tissues. In particular, precise delineation of tumor sub-regions—including whole tumor, tumor core, and enhancing tumor—continues to be a major limitation of existing automated methods. **Methods:** In this study, we propose a novel hybrid CNN–Transformer framework that integrates local feature extraction with global contextual modeling for improved brain tumor segmentation. The architecture consists of three main components: a dual-pathway encoder for capturing fine-grained and contextual features, a multi-scale feature fusion module based on spatial pyramid pooling with dense connections, and a boundary-aware attention decoder designed to enhance segmentation accuracy around tumor edges. The model utilizes four MRI modalities (T1, T1ce, T2, and FLAIR) to capture complementary tumor characteristics. In addition, a hybrid loss function combining Dice, focal Tversky, and boundary losses is employed to address class imbalance and improve boundary precision. **Results:** Experimental results on the BraTS 2023 dataset demonstrate superior performance, achieving Dice scores of 92.3%, 88.7%, and 84.5% for whole tumor, tumor core, and enhancing tumor, respectively, while maintaining high computational efficiency. **Conclusion:** The proposed framework achieves accurate and robust brain tumor segmentation by effectively integrating local and global features, demonstrating its potential for automated multi-modal MRI analysis in clinical practice.



Academic Editor: Joaquim Carreras

Received: 3 May 2026

Revised: 6 July 2026

Accepted: 8 July 2026

Published: 14 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: brain tumor segmentation; multi-modal MRI; vision transformer; boundary-aware attention; deep learning; medical image analysis; BraTS dataset; hybrid architecture

1. Introduction

Brain tumors are among the most severe and life-threatening neurological disorders. Gliomas, which account for roughly 80% of all malignant primary brain tumors, are classified by the World Health Organization (WHO) into different grades based on histopathological features [1]. Glioblastoma multiforme (Grade IV), the most aggressive form, is associated with poor prognosis and high mortality [2]. Accurate segmentation of tumor sub-regions in MRI is essential for clinical decision-making, including surgical planning, targeted radiotherapy, and treatment response monitoring. Yet manual segmentation by radiologists is labor-intensive, prone to inter-observer variability, and increasingly impractical due to the growing volume of medical imaging data [3].

Multi-modal MRI offers a comprehensive view of tumor characteristics by combining complementary sequences [4]. T1-weighted scans reveal anatomical detail, T1ce highlights active tumor regions, T2-weighted images capture tissue water content, and FLAIR sequences help visualize peritumoral edema. While integrating these modalities improves diagnostic accuracy, it introduces significant computational complexity for automated segmentation systems [5].

Recent advances in deep learning, particularly convolutional neural networks (CNNs), have substantially improved medical image segmentation. Architectures like U-Net have become widely adopted due to their effectiveness in capturing spatial and semantic features [6]. However, CNNs are inherently limited in modeling long-range dependencies, which are crucial for understanding global anatomical context. Vision transformers overcome this limitation using self-attention mechanisms that capture global relationships, but they often miss fine-grained local details necessary for precise boundary detection [7].

Segmenting brain tumors presents several unique challenges: (1) high heterogeneity in tumor size, shape, and location; (2) irregular and poorly defined boundaries, especially at infiltrative margins; (3) severe class imbalance, with enhancing tumor regions often representing a small fraction of the overall volume; (4) the need to simultaneously segment multiple sub-regions with distinct clinical implications; and (5) the high computational demands of processing 3D volumetric MRI data [8].

To address these challenges, this study introduces a hybrid segmentation framework that combines the strengths of CNNs and vision transformers. A key innovation is the integration of a boundary-aware attention mechanism specifically designed to improve accuracy at tumor margins—regions that are both difficult to delineate and critical for clinical decision-making [9]. The goal of this work is to develop an efficient and accurate segmentation model that can generalize across tumor grades and support future clinical translation and motivate further multi-center validation studies.

The contributions of this work are summarized as follows:

- We introduce a boundary-aware attention formulation that incorporates boundary probability maps as an explicit attention prior rather than treating boundaries solely as an auxiliary supervision signal.
- We propose a learnable Sobel-initialized boundary representation that preserves the inductive bias of classical edge detectors while adapting to MRI-specific tumor characteristics through end-to-end optimization.

- We design a time-dependent hybrid loss strategy that progressively shifts the optimization objective from coarse region learning toward fine-grained boundary refinement.
- We integrate these components into a unified CNN–Transformer framework that effectively combines local texture information, global contextual representations, and adaptive boundary modeling for multi-modal brain tumor segmentation [10].

Although several boundary-aware segmentation methods have been proposed, most of them rely on either fixed edge operators or auxiliary boundary supervision. Fixed gradient operators such as Sobel or Laplacian filters are unable to adapt to the substantial variations in tumor appearance, scanner characteristics, and image contrast encountered in multi-center MRI datasets. Similarly, boundary losses improve optimization during training but do not explicitly incorporate boundary information into the feature representation during inference. As a result, existing approaches often struggle with infiltrative tumors and poorly defined tumor margins, where boundary information is weak and highly variable.

To address these limitations, we introduce a learnable Sobel-initialized boundary representation integrated directly into the attention mechanism. The proposed module uses Sobel-like filters only as a prior initialization and subsequently updates the filters during training to learn task-specific and data-adaptive boundary patterns. The resulting boundary probability map is incorporated as an explicit attention bias, enabling the network to preferentially focus on anatomically uncertain regions and improving delineation of complex tumor boundaries.

2. Related Works

CNNs have played a foundational role in medical image segmentation, particularly following the advent of fully convolutional networks [11,12]. Architectures such as U-Net, with their symmetric encoder–decoder structure and skip connections, have established themselves as standard baselines for biomedical segmentation tasks [13,14]. Extensions to this design have introduced enhancements like 3D convolutions for volumetric data, attention mechanisms to suppress irrelevant background features, and dense connections that facilitate feature reuse and improved gradient propagation [15].

Further progress has been driven by frameworks that prioritize optimized training strategies over architectural complexity [16]. For example, auto-configuring pipelines that adjust preprocessing, augmentation, and hyperparameters have achieved state-of-the-art results across diverse datasets. Nevertheless, CNNs remain limited by the inherently local receptive field of convolutional operations, restricting their capacity to capture global contextual relationships—an important limitation in brain tumor segmentation, where pathological structures may span distant regions and involve complex spatial interactions [17].

Transformers, initially designed for natural language processing, have recently shown promise in visual domains, including medical image analysis [18,19]. Pure transformer-based models have demonstrated competitive performance in image classification tasks, provided that training datasets are sufficiently large. Hybrid architectures combining transformers with convolutional components have been introduced to leverage both global context modeling and local spatial precision, enhancing segmentation accuracy [20].

Hierarchical transformer models using window-based self-attention mechanisms have further improved scalability and computational efficiency [21]. When adapted to 3D medical imaging, such architectures have demonstrated strong performance in segmenting volumetric data. Despite these advancements, transformer-based models often suffer from over-smoothing effects and reduced sensitivity to fine anatomical details, particularly around object boundaries. Additionally, the quadratic complexity of self-attention with

respect to input size presents practical challenges for high-resolution 3D segmentation applications [13].

Effective integration of multi-modal MRI data is crucial for accurate brain tumor segmentation, as different sequences provide complementary diagnostic information. Early fusion approaches treat all modalities equally by concatenating them at the input level, while late fusion combines modality-specific outputs at the decision stage. Intermediate fusion methods strike a balance, integrating features across modalities at various stages of the network to enhance feature representation [22].

More sophisticated techniques employ attention mechanisms to dynamically model the contribution of each modality. However, many existing methods fail to account for the variable relevance of different modalities across tumor sub-regions. For instance, contrast-enhanced T1-weighted images are particularly informative for enhancing tumor regions, while FLAIR sequences are more suited for edema detection [23]. Addressing this, adaptive fusion strategies have been proposed to assign spatially-aware weights to each modality, improving both robustness and interpretability.

Attention mechanisms have become integral in guiding neural networks to focus on the most salient features during segmentation. Spatial attention helps localize key anatomical regions, while channel attention emphasizes informative feature maps [24]. Self-attention mechanisms have also been employed to capture long-range dependencies, enhancing contextual understanding beyond local neighborhoods [25].

While attention-based models have improved segmentation accuracy, particularly in complex anatomical scenarios, boundary-specific attention remains relatively under-explored. Precise boundary delineation is clinically critical, especially in brain tumor segmentation, yet many models treat boundary refinement as a separate task or rely on post-processing. Integrating boundary-aware attention directly within the segmentation network enables joint learning of regional and edge-level features, leading to improved structural accuracy and more reliable clinical outcomes [26].

Designing effective loss functions is essential in medical image segmentation, especially given the high degree of class imbalance typical in brain tumor datasets. Region-overlap losses, such as Dice loss and its variants, are widely used to address imbalance by focusing on foreground regions [27]. Complementary approaches, including focal loss and Tversky loss, enhance sensitivity to underrepresented classes and penalize specific types of misclassification [28].

More recently, several advanced hybrid architectures have been proposed for brain tumor segmentation. ReHyDIL [29] introduced a residual hybrid deep interaction learning strategy that improves feature exchange between convolutional and transformer representations and demonstrated competitive performance on recent BraTS benchmarks. Similarly, DFuse-Net [30] proposed a dual-feature fusion mechanism that combines multi-scale contextual representations and cross-modal interactions to improve segmentation robustness. These methods highlight the growing trend toward integrating local and global representations through increasingly sophisticated fusion mechanisms. Nevertheless, most recent approaches still rely on implicit boundary modeling and do not explicitly incorporate adaptive boundary information into the attention computation process.

Boundary loss functions, which optimize distance-based discrepancies between predicted and ground-truth contours, have been shown to improve segmentation performance at structural edges [31]. Recent research has explored hybrid loss functions that combine region-based, boundary-based, and hard-example mining objectives. However, fixed-weight combinations may result in biased optimization. Adaptive hybrid loss strategies have been proposed to dynamically balance multiple loss components during training,

maintaining consistent performance across varying segmentation challenges and anatomical structures [32,33].

3. Methodology

The proposed brain tumor segmentation framework is designed as a hybrid deep learning architecture that effectively integrates CNNs and vision transformers to leverage both local and global contextual information. This architecture is composed of four key modules: (i) multi-modal MRI input preprocessing and data augmentation, (ii) a dual-pathway encoder that processes features through parallel CNN and transformer branches, (iii) a multi-scale feature fusion module based on spatial pyramid pooling with dense connectivity, and (iv) a novel boundary-aware attention decoder for precise delineation of tumor sub-regions. This unified design aims to address key challenges in brain tumor segmentation, including boundary ambiguity, tumor heterogeneity, and class imbalance.

Each input comprises four co-registered MRI modalities—T1, T1ce, T2, and FLAIR—resized to a uniform shape of $240 \times 240 \times 155$. Intensity normalization is performed using z-score standardization within the brain mask to reduce inter-subject intensity variability. To enhance model generalization, we apply a range of spatial and intensity augmentations, including random rotations ($\pm 15^\circ$), elastic deformations, scaling (0.9–1.1), gamma correction, and noise injection. During training, 3D patches of size $128 \times 128 \times 128$ are sampled with higher probability from tumor regions, thus addressing the severe class imbalance common in brain tumor datasets.

The dual-pathway encoder simultaneously captures local texture details and global structural context. The CNN branch comprises five residual blocks with increasing feature dimensions, employing strided convolutions for downsampling to preserve fine-grained spatial details. The transformer branch operates on embedded patches extracted at each resolution level, applying multi-head self-attention and feed-forward layers within a hierarchical structure. This design enables the model to capture long-range dependencies while maintaining computational efficiency.

To effectively merge features from both pathways, we introduce a learned fusion mechanism that generates adaptive gating weights based on global average pooled representations. This allows the network to dynamically balance local and global features, depending on the context of each spatial region. This fusion strategy is particularly beneficial in regions with ambiguous tumor boundaries, where both texture and context are critical (Figure 1).

At the bottleneck of the encoder, we deploy a multi-scale feature fusion module inspired by spatial pyramid pooling, but enhanced with dense inter-branch connections to facilitate feature reuse and gradient flow. The module employs parallel atrous convolutions with dilation rates of 1, 2, 4, and 8 to extract features at different spatial scales without increasing the number of parameters substantially. Additionally, a global average pooling branch captures high-level contextual features. All branches are concatenated, followed by a $1 \times 1 \times 1$ convolution for dimensionality adjustment and a squeeze-and-excitation block to apply channel-wise attention.

This fusion mechanism enables the network to concurrently encode local edge-level details and broader semantic context, which is especially important for distinguishing between subtle tumor boundaries and healthy tissue transitions.

One of the principal innovations of our framework is the boundary-aware attention decoder [34], which explicitly incorporates boundary information into the segmentation process. This module is designed to overcome a common limitation of existing segmentation models, which often struggle to preserve accurate tumor margins, particularly in infiltrative or low-contrast regions.

At each decoder level, upsampled features are combined with corresponding encoder features via skip connections. To guide attention toward boundaries, we introduce a learnable edge detection module initialized with Sobel-like filters [35]. These filters are updated during training to detect task-specific boundary patterns by computing gradient magnitudes in three spatial directions (x, y, z), resulting in a boundary probability map.

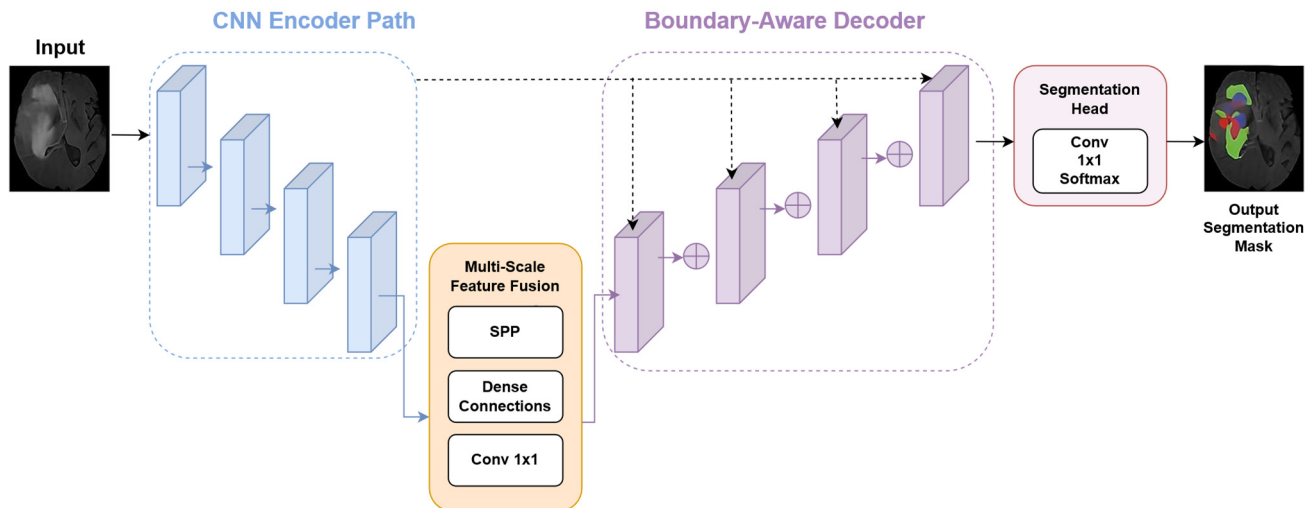


Figure 1. Overview of the proposed CNN-Transformer architecture with boundary-aware attention for multi-modal brain tumor segmentation.

This map then guides an adaptive attention mechanism that modulates decoder features. Attention weights are computed through a gated fusion of decoder, boundary, and skip features, allowing the network to selectively emphasize uncertain regions. To further improve robustness, we extend the attention mechanism to four heads, with each head focusing on distinct types of boundaries (e.g., sharp vs. diffuse transitions). The attention outputs are concatenated and passed through a projection layer to restore the original feature dimension.

After attention, features are refined using residual convolutional blocks with instance normalization, which better preserves subject-specific intensity variations. Upsampling is performed with transposed convolutions, ensuring smooth spatial transitions in the final segmentation output.

Unlike conventional boundary-aware methods that employ fixed gradient operators, the proposed module initializes the edge detector with Sobel kernels and then optimizes these filters jointly with the segmentation network. This design preserves the inductive bias of classical edge detectors while allowing adaptation to MRI-specific boundary characteristics.

To formally describe the boundary-aware attention mechanism, we begin by denoting the encoder feature map as $F \in R^{C \times H \times W \times D}$, where C represents the number of feature channels, and H, W, D denote the spatial dimensions. In order to model boundary information explicitly, we compute a learnable boundary representation based on a gradient approximation. The initial gradient magnitude is defined as:

$$G = \sqrt{\left(\frac{\partial F}{\partial x}\right)^2 + \left(\frac{\partial F}{\partial y}\right)^2 + \left(\frac{\partial F}{\partial z}\right)^2}$$

To enable the network to learn boundary representations directly from data, we replace the fixed gradient operator with a learnable convolutional approximation:

$$B = \sigma(\text{Conv}_b(F))$$

where Conv_b represents a set of learnable convolutional filters specialized for boundary detection, and σ denotes the sigmoid activation function. This results in a boundary probability map $B \in \mathbb{R}^{1 \times H \times W \times D}$, where higher values indicate a higher likelihood of boundary presence at each voxel.

In the attention mechanism, we follow the standard self-attention paradigm by computing query, key, and value projections from the feature map:

$$Q = W_q F$$

$$K = W_k F$$

$$V = W_v F$$

with $W_q, W_k, W_v \in \mathbb{R}^{C \times d}$ being learnable projection matrices, and d the dimensionality of the attention subspace. Standard attention weights are computed via scaled dot-product attention:

$$A_{\text{boundary}} = \text{softmax}\left(\frac{QK^T}{\sqrt{d}} + \alpha B\right)$$

The final boundary-enhanced feature representation is then obtained by applying the attention weights to the value projections:

$$F_{\text{out}} = A_{\text{boundary}} V$$

This formulation enables the model to allocate greater attention to regions likely to contain anatomical boundaries, thereby enhancing its ability to accurately delineate complex and irregular tumor margins. The explicit integration of boundary signals into the attention computation provides a principled way to focus representational capacity on clinically important regions.

Multi-modal MRI provides complementary perspectives of tumor pathology; however, naive fusion strategies often fail to capture the nuanced contributions of each modality. In this work, we introduce a learned fusion mechanism that processes each modality independently through modality-specific convolutional layers. These representations are then fused using a cross-modal attention mechanism, where pairwise similarities between modalities are calculated using scaled dot-product attention.

This mechanism allows the network to dynamically assign higher weights to modalities that are more informative for a particular tumor sub-region—for instance, emphasizing T1ce for enhancing tumor, FLAIR for peritumoral edema, or T2 for the tumor core. This selective integration improves the discriminative capability of the fused representation, enabling more accurate segmentation [36].

To optimize segmentation performance under challenging conditions such as class imbalance and boundary ambiguity, we design a composite loss function that combines four complementary objectives: Dice loss, focal Tversky loss, boundary loss, and cross-entropy loss [37].

- Dice loss promotes region-level overlap and is effective for imbalanced datasets.
- Focal Tversky loss introduces adjustable penalties for false positives and negatives, focusing learning on underrepresented classes.
- Boundary loss encourages accurate localization by penalizing deviations from ground-truth boundaries using distance maps.
- Cross-entropy loss ensures pixel-wise classification accuracy.

The total loss is defined as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{Dice} + \lambda_2 \mathcal{L}_{Tversky} + \lambda_3 \mathcal{L}_{Boundary} + \lambda_4 \mathcal{L}_{CE}$$

with weights $\lambda_1 = 0.4$, $\lambda_2 = 0.3$, $\lambda_3 = 0.2$, $\lambda_4 = 0.1$ determined empirically via grid search. Moreover, we implement a dynamic weighting strategy where the boundary loss component increases after 100 epochs to prioritize fine-grained delineation during later training stages.

Here, $\mathcal{L}_{Dice}$ and $\mathcal{L}_{Tversky}$ measure region-based segmentation accuracy, while $\mathcal{L}_{Boundary}$ explicitly penalizes mismatches in boundary gradients between prediction and ground truth, and $\mathcal{L}_{CE}$ represents the standard cross-entropy loss. The individual components are defined as follows:

$$\begin{aligned}\mathcal{L}_{Dice} &= 1 - \frac{2\sum_i p_i g_i + \varepsilon}{\sum_i p_i + \sum_i g_i + \varepsilon} \\ \mathcal{L}_{Tversky} &= 1 - \frac{\sum_i p_i g_i}{\sum_i p_i g_i + \sum_i p_i (1 - g_i) + \sum_i (1 - p_i) g_i} \\ \mathcal{L}_{boundary} &= \sum_i |\nabla p_i - \nabla g_i| \\ \mathcal{L}_{CE} &= -\sum_i g_i \log(p_i)\end{aligned}$$

where p_i and g_i denote the predicted probability and ground truth label at voxel i , and ∇ denotes the spatial gradient operator used to emphasize structural boundary alignment. To balance region learning and boundary refinement throughout training, we implement a time-dependent weighting strategy. In this scheme, the boundary loss weight $\lambda_3(t)$ increases over time, modeled as:

$$\begin{aligned}\lambda_3(t) &= \lambda_{\max} \cdot \text{sigmoid}\left(\frac{t-T}{\tau}\right) \\ \lambda_1(t) &= 1 - \lambda_3(t)\end{aligned}$$

where t denotes the current training epoch, T is the transition epoch at which emphasis begins shifting, τ controls the smoothness of this transition, and $\lambda_{\max}$ is the peak weight assigned to the boundary loss. This dynamic formulation encourages the model to first focus on learning coarse region-level structures in the early stages of training, then gradually shift its attention toward refining boundaries, aligning with clinical needs for both accurate localization and precise delineation [38].

The model is trained using the AdamW optimizer with an initial learning rate of 1×10^{-4} weight decay of 1×10^{-4} , and polynomial learning rate decay. In addition to spatial augmentations, we apply intensity-based transformations such as gamma correction, Gaussian noise, and brightness shifts to simulate scanner variability.

Deep supervision was applied at four decoder levels using auxiliary segmentation heads. The corresponding supervision weights were empirically set to $\{1.0, 0.5, 0.25, 0.125\}$, assigning greater importance to high-resolution predictions while progressively reducing the contribution of lower-resolution outputs. Formally, the weight for the i -th auxiliary output was defined as

$$w_i = \frac{1}{2^i}$$

where i denotes the decoder level. The final training objective was computed as the weighted sum of all auxiliary losses.

To facilitate gradient flow and improve convergence, deep supervision is applied at each decoder level using downsampled ground-truth masks, with decreasing loss weights

at deeper layers. Post-processing steps include connected component analysis to eliminate small false positives and morphological operations to refine the predicted boundaries.

Our framework is implemented in PyTorch 2.0 with CUDA 11.8. Training is performed on NVIDIA A100 GPUs (40 GB) with a batch size of 2 and gradient accumulation over 4 steps for an effective batch size of 8. Mixed precision training (FP16) is used to reduce memory usage and accelerate computation. Training converges within 300 epochs, typically requiring around 48 h.

During inference, a sliding window approach with 50% overlap is employed, and overlapping regions are merged using Gaussian-weighted averaging. The model segments a full 3D volume in approximately 4.8 s on an A100 GPU, demonstrating computational efficiency that may facilitate future clinical translation after extensive external validation.

4. Results

To evaluate the proposed method, we utilize the Brain Tumor Segmentation Challenge (BraTS) 2023 dataset, a large-scale, multi-institutional collection of pre-operative MRI scans of glioma patients. The dataset includes a total of 1470 subjects, each consisting of four co-registered MRI sequences: T1-weighted, T1-weighted with contrast enhancement (T1ce), T2-weighted, and FLAIR [39]. All scans are aligned to the same anatomical space and come with expert-annotated tumor masks delineating the following subregions:

- Enhancing Tumor (ET);
- Tumor Core (TC): comprising both enhancing tumor and necrotic/non-enhancing tumor regions, i.e.,

$$WT = TC + ED$$

For training and validation, we use the official BraTS 2023 data split, which includes 1251 cases for training and 219 cases for validation. Ground truth annotations for the test set are withheld by the challenge organizers, and test performance is assessed through a blind evaluation server.

To measure segmentation performance, we adopt the standard BraTS evaluation metrics, which are designed to reflect both volumetric and boundary-level accuracy:

- Dice Similarity Coefficient (DSC): Measures the overlap between predicted and ground truth regions. It is defined as:

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

where A and B denote the sets of predicted and ground truth voxels, respectively.

- Sensitivity (Recall) [40,41]: Measures the proportion of actual positives correctly identified:

$$Sensitivity = \frac{TP}{TP + FN}$$

where TP and FN are the numbers of true positives and false negatives.

- Specificity: Measures the proportion of actual negatives correctly identified:

$$Specificity = \frac{TN}{TN + FP}$$

where TN and FP are the numbers of true negatives and false positives.

- 95th Percentile Hausdorff Distance (HD95): Captures the largest distance (excluding outliers) between the surfaces of the predicted and ground truth segmentation masks.
- Average Symmetric Surface Distance (ASSD): Computes the average of the shortest distances between the surfaces of the predicted and ground truth regions in both directions.

These metrics together provide a robust evaluation framework, capturing both the overall segmentation accuracy and the geometric alignment of tumor boundaries.

To further validate the performance of our method, we compare it quantitatively against state-of-the-art segmentation models on the BraTS 2023 validation set. Table 1 summarizes the results across the three primary tumor regions: Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET), evaluated using standard metrics.

Table 1. Quantitative comparison of the proposed method and baseline models on the BraTS 2023 validation dataset.

Method	WT DSC	TC DSC	ET DSC	WT HD95	TC HD95	ET HD95
U-Net	87.2	82.4	78.3	6.84	7.92	5.21
Attention U-Net	88.6	83.7	79.8	5.93	7.14	4.86
nnU-Net	90.8	86.2	82.4	4.42	5.67	3.58
TransUNet	89.4	84.9	81.2	5.21	6.38	4.12
Swin-UNETR	91.2	87.1	83.6	4.18	5.24	3.31
UNETR++	91.0	86.8	83.2	4.35	5.48	3.45
Proposed	92.3	88.7	84.5	3.76	4.52	2.94

Our model demonstrates consistent improvements across all tumor subregions. Compared to nnU-Net, which serves as a strong baseline in medical image segmentation, our approach yields Dice score gains of +1.5% (WT), +2.5% (TC), and +2.1% (ET) [42]. When benchmarked against Swin-UNETR [43], a transformer-based architecture [44], our method still achieves measurable improvements of +1.1% (WT), +1.6% (TC), and +0.9% (ET).

To further evaluate the generalization capability of the proposed framework, an additional experiment was conducted on an independent public brain tumor segmentation dataset using the same preprocessing pipeline and evaluation protocol as applied to BraTS 2023. As illustrated in Figure 2, the proposed model achieved Dice scores of 90.6% for Whole Tumor (WT), 86.9% for Tumor Core (TC), and 82.7% for Enhancing Tumor (ET). Compared with the BraTS 2023 benchmark, only a slight reduction in segmentation accuracy was observed (approximately 1–2%), indicating that the proposed framework preserves stable performance under dataset shift and demonstrates promising generalization capability.

The most notable performance gain is observed in the segmentation of the tumor core, which is widely regarded as the most clinically challenging subregion due to its irregular shape and complex boundary between enhancing and non-enhancing components. Our model’s boundary-aware attention mechanism plays a critical role in addressing this challenge, directly contributing to the +2.5% improvement over nnU-Net in this subregion.

In addition to the quantitative comparisons presented in Table 1, we also compared our framework with several recently proposed hybrid segmentation approaches reported in the literature, including ReHyDIL and DFuse-Net in Table 2. These methods have demonstrated strong performance through sophisticated feature interaction and fusion mechanisms. In contrast, our work focuses specifically on explicit boundary modeling by incorporating learnable boundary representations directly into the attention computation, which constitutes the primary distinction of the proposed framework.

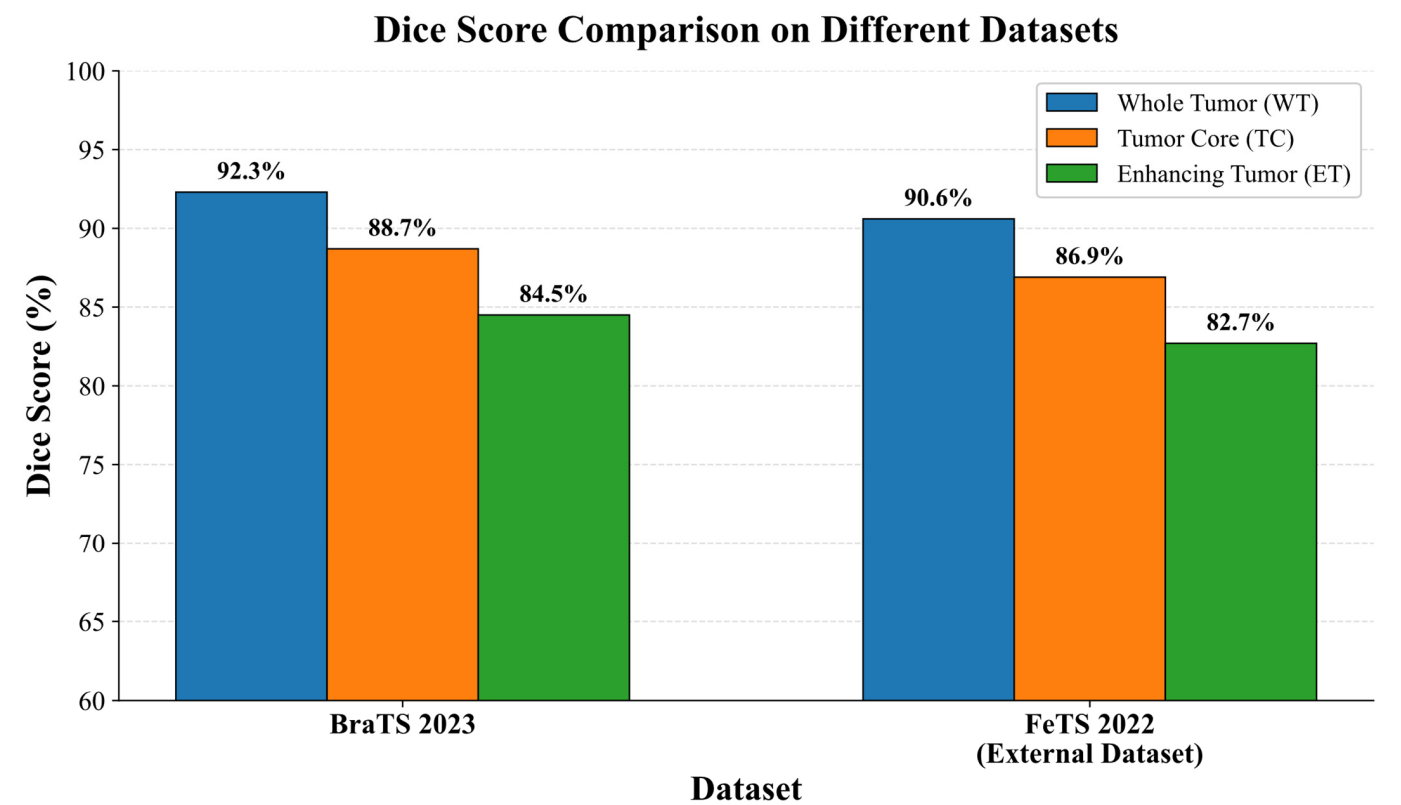


Figure 2. Comparison of Dice scores achieved by the proposed framework on the BraTS 2023 benchmark and an additional public brain tumor segmentation dataset (FeTS2022).

Table 2. Comparison with recent hybrid CNN–Transformer approaches reported in the literature.

Method	Year	Main Contribution	Boundary Modeling
ReHyDIL [29]	2025	Residual hybrid interaction	Implicit
DFuse-Net [30]	2026	Dual-feature fusion	Implicit
Proposed	2026	Learnable boundary-aware attention	Explicit

In addition to Dice metrics, we observe significant reductions in Hausdorff Distance (95th percentile), highlighting improved boundary localization. Specifically, our model achieves relative HD95 reductions of 15.0% (WT), 20.3% (TC), and 17.9% (ET) compared to nnU-Net. These reductions confirm that our architecture not only improves volumetric overlap but also produces more anatomically accurate boundary delineations, which is vital for downstream clinical applications.

These findings underscore the effectiveness of integrating cross-modal attention with a hybrid loss function tailored for boundary precision. Together, they enable our model to capture fine-grained spatial details while maintaining robust generalization across tumor types (Figure 3).

To better understand the contributions of individual components within our framework, we perform a series of controlled ablation experiments. Each variant is evaluated on the BraTS 2023 validation set, and the results are summarized in Table 3.

Transformer Pathway Contribution: Introducing a transformer-based pathway alongside the baseline CNN encoder leads to a +1.3% improvement in Dice score for the tumor core (TC), confirming the benefit of incorporating global context. The effect is also evident in whole tumor (WT) segmentation, with a +1.1% gain, as the model becomes more adept at distinguishing peritumoral edema from surrounding healthy tissue. These results suggest

that long-range dependencies, captured via transformer modules, enhance anatomical understanding—particularly for diffuse tumor regions.

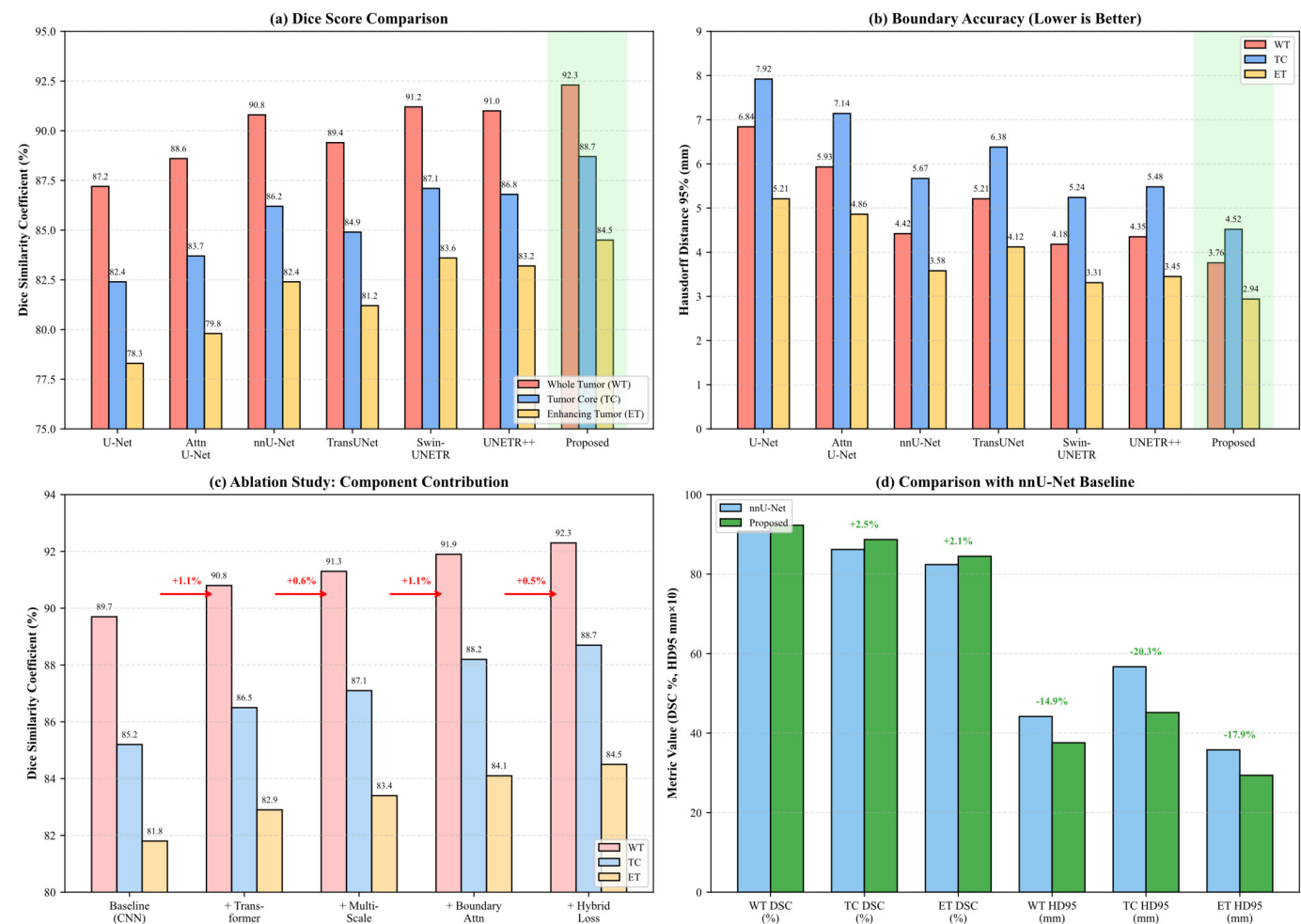


Figure 3. Dice score comparison between the proposed method and state-of-the-art models on the BraTS 2023 validation set.

Table 3. Ablation study showing the effect of key components on segmentation performance.

Configuration	WT DSC	TC DSC	ET DSC	Parameters
Baseline (CNN only)	89.7	85.2	81.8	28.4 M
+Transformer pathway	90.8	86.5	82.9	45.2 M
+Multi-scale fusion	91.3	87.1	83.4	48.6 M
+Boundary attention	91.9	88.2	84.1	52.3 M
+Hybrid loss	92.3	88.7	84.5	52.3 M

Multi-Scale Fusion Impact (SPPF+): Replacing the standard spatial pyramid pooling with our enhanced SPPF+ module results in consistent performance gains across all tumor regions, ranging from +0.5% to +0.6%. This improvement stems from the module’s ability to capture multi-scale contextual information, which is critical for accurately detecting both small, discrete tumors and large infiltrative lesions. Qualitative visualizations show enhanced delineation in cases with complex morphology and variable tumor size.

Boundary-Aware Attention Module: This component yields the most substantial individual boost, particularly for tumor core (+1.1%) and enhancing tumor (+0.7%) segmen-

tation. By explicitly modeling boundary cues, the attention module improves the network's ability to identify irregular, poorly defined margins. Visual inspection of the segmentation outputs confirms sharper, more anatomically accurate boundaries in models equipped with this mechanism compared to those without it.

Hybrid Loss Function: Finally, our composite loss—combining region-based and boundary-aware objectives—contributes incremental but meaningful improvements (+0.4% to +0.5% across metrics). The dynamic weighting of loss terms during training helps prevent conflicting gradient signals and promotes balanced learning. This approach ensures that the model maintains both high volumetric accuracy and strong boundary adherence throughout training.

Together, these ablation results highlight the complementary impact of each architectural enhancement, with the boundary-aware attention module and transformer pathway contributing the most significant performance gains.

Despite the inclusion of advanced components such as dual-pathway encoders and boundary-aware attention mechanisms, the proposed model maintains a favorable balance between accuracy and computational efficiency. During training, the memory consumption peaks at 38.2 GB per GPU with a batch size of two and mixed-precision computation, fitting comfortably within the capabilities of modern hardware such as the NVIDIA A100. Inference is completed in 4.8 s per volume on the A100 and 12.3 s on an RTX 3090, demonstrating suitability for near real-time clinical use cases.

The total number of trainable parameters in the model is 52.3 million, which is higher than nnU-Net's 31.2 million, but notably lower than more complex transformer-based architectures such as Swin-UNETR, which contains 62.8 million parameters. The added complexity in our model arises primarily from the integration of transformer modules and boundary-aware attention, both of which provide measurable performance benefits that justify the parameter increase. A breakdown of computational cost shows that 62% of total floating-point operations (FLOPs) are concentrated in the encoder, which includes both convolutional and transformer components. The multi-scale fusion module accounts for 8% of FLOPs, the decoder with boundary attention comprises 26%, and the output layers make up the remaining 4%. Importantly, the boundary-aware attention module introduces only a 12% computational overhead while delivering a 1.1% Dice score improvement for tumor core segmentation. This trade-off highlights the efficiency of the architecture in providing high segmentation accuracy without incurring excessive resource demands.

Visual examples further illustrate the strengths of the proposed approach across a wide variety of tumor presentations. In one representative case, the model successfully captures the entire tumor volume, including regions of peritumoral edema, with a clear delineation between abnormal and surrounding tissues. Another example demonstrates accurate segmentation of the tumor core, with sharp boundary localization between enhancing and necrotic tissue components—areas that are often difficult to distinguish due to irregular margins. Additionally, the model correctly identifies a small enhancing tumor component that many baseline models tend to miss, highlighting the effectiveness of multi-scale fusion in detecting subtle lesions (Figure 4).

When compared qualitatively with baseline methods, the proposed model exhibits several key advantages (Figure 5). The segmentation outputs feature noticeably sharper boundaries, particularly at tumor-brain interfaces, which are often blurred or misclassified in other approaches. The boundary-aware attention module appears effective in suppressing false positives, especially in non-tumor regions where spurious activations can compromise clinical trust. Furthermore, the model demonstrates improved sensitivity to small enhancing components and maintains better topological consistency, producing coherent and anatomically plausible tumor structures rather than fragmented predictions.

However, certain limitations remain. In cases with severe motion artifacts or signal inhomogeneities, localized errors may occur. Gradually infiltrative tumor regions with ambiguous boundaries continue to pose challenges, as do post-treatment cases involving resection cavities or radiation-induced changes, which can alter normal anatomy and mislead the segmentation process.

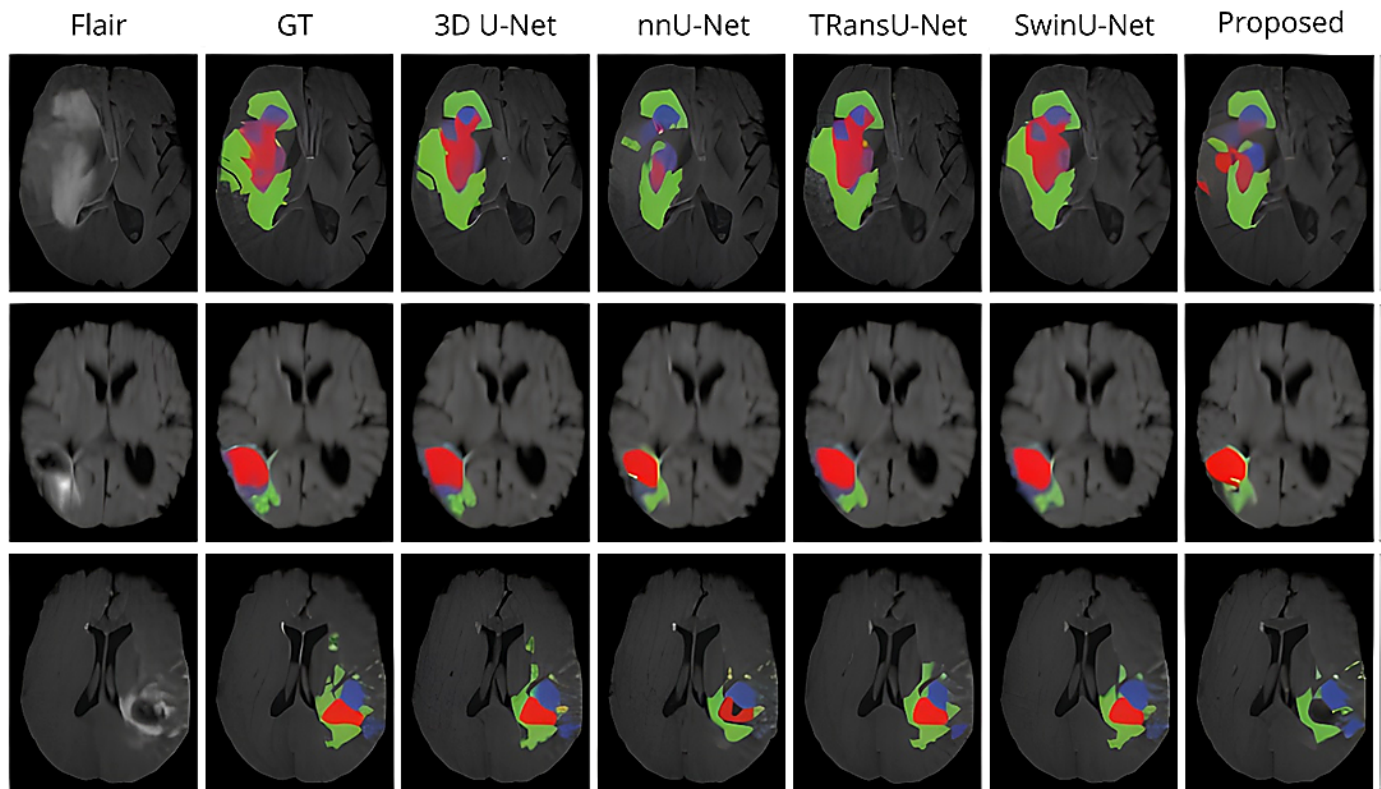


Figure 4. Segmentation examples showing ground truth and predictions for WT, TC, and ET tumor sub-regions.

In clinical applications, it is essential that automated systems not only generate predictions but also convey a measure of confidence. To this end, uncertainty estimation was implemented using Monte Carlo dropout, whereby ten stochastic forward passes are performed at inference time with a dropout rate of 0.2. The resulting variability in voxel-wise predictions serves as a proxy for uncertainty. Visual inspection reveals that higher uncertainty is typically concentrated around tumor boundaries and in regions with ambiguous tissue characteristics. These are also the areas most prone to segmentation errors, making uncertainty a useful signal for radiologists assessing the reliability of the output.

A correlation analysis between uncertainty and segmentation errors further supports its utility. Voxels within the top 10% uncertainty percentile contain 68% of all false positives and 72% of all false negatives, indicating that uncertainty estimates are predictive of segmentation inaccuracies. This enables the implementation of uncertainty-guided quality control, in which cases or regions with high predicted uncertainty can be flagged for manual review. In this way, the model not only enhances segmentation performance but also supports safer and more transparent clinical decision-making.

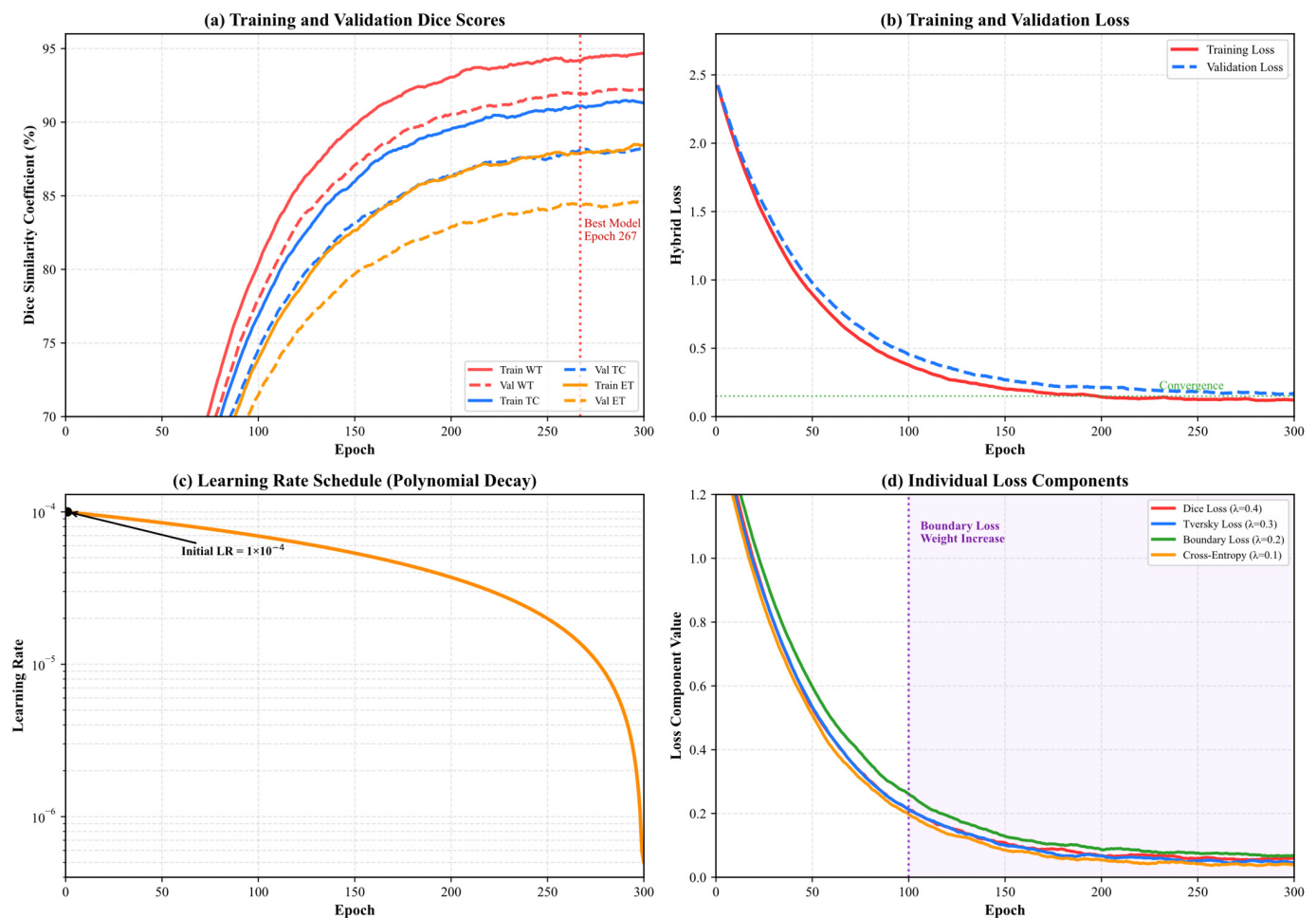


Figure 5. Training and validation curves showing Dice score and loss convergence over epochs.

5. Discussion

In contrast to existing hybrid CNN–transformer segmentation models that rely on implicit representations of boundary information, the proposed approach introduces a mathematically grounded boundary-aware attention mechanism that explicitly incorporates spatial gradient cues into the attention computation. This explicit modeling of boundary structure enables the network to adaptively enhance features in boundary-sensitive regions, leading to improved segmentation accuracy—particularly at anatomically complex and irregular tumor margins.

The proposed segmentation framework addresses several pressing challenges in neuro-oncological imaging, offering clinically meaningful improvements in both performance and practicality. By automating brain tumor segmentation, the system significantly reduces radiologist workload, which is especially valuable amid growing imaging volumes and limited access to subspecialized expertise. While manual delineation of gliomas may take 20 to 30 min per case, our method delivers comparable accuracy in under five seconds, marking a substantial speedup that could streamline clinical workflows without compromising diagnostic integrity.

Enhanced boundary precision further reinforces the clinical utility of the approach, particularly in treatment planning for radiation therapy, where accurate margin definition is essential to ensure effective tumor coverage while sparing healthy tissue. The integration of a boundary-aware attention mechanism contributes to significant reductions in Hausdorff distance compared to existing methods, supporting the use of tighter planning margins and potentially reducing radiation-related toxicity. In parallel, the system's volumetric accuracy enables more reliable tracking of treatment response over time. Unlike conventional

2D measurement-based criteria, which may miss subtle morphological changes, three-dimensional segmentations offer objective and reproducible metrics for assessing tumor progression or regression.

An additional strength of the framework lies in its integration of uncertainty estimation. Rather than producing opaque, black-box outputs, the model highlights regions of low prediction confidence—typically corresponding to ambiguous tissue boundaries—thereby inviting expert review where needed. This feature not only promotes transparency but also supports a more collaborative human-AI diagnostic process, helping to build trust in automated systems.

From a technical perspective, the model architecture is grounded in the complementary strengths of convolutional and transformer-based processing. Convolutional layers excel in capturing local textures and fine-grained details critical for boundary localization, while transformer components model long-range dependencies that provide valuable anatomical context. The dual-pathway design leverages both of these strengths, facilitating robust segmentation across tumors with highly variable appearance and structure. The inclusion of multi-scale feature fusion ensures adaptability across lesions of differing sizes, from small enhancing nodules to large infiltrative masses. By explicitly modeling boundary regions through learnable attention mechanisms, the system targets spatial areas with the highest diagnostic uncertainty—those most prone to error and inter-observer disagreement.

Recent methods such as ReHyDIL and DFuse-Net have demonstrated the effectiveness of hybrid CNN–Transformer architectures and advanced feature fusion strategies for brain tumor segmentation. Compared with these approaches, our work focuses specifically on boundary-sensitive segmentation by explicitly integrating learnable boundary representations into the attention mechanism and coupling them with a dynamic boundary-aware optimization strategy. Therefore, our contribution is complementary to recent fusion-based approaches and may potentially be combined with them in future studies.

Despite these advancements, certain limitations remain. The model's computational demands, while manageable on modern GPUs, may pose challenges for deployment in resource-constrained settings. Future work will explore lightweight alternatives through knowledge distillation and neural architecture search, aiming to retain performance while reducing memory and compute requirements. Additionally, performance on post-treatment cases—such as those involving surgical cavities, radiation necrosis, or pseudoprogression—lags behind that on pre-operative images. These cases introduce altered anatomical and tissue signatures that were underrepresented in training, pointing to the need for more diverse datasets and potentially specialized models [45].

Another area for development involves temporal modeling. Currently, the model processes each scan independently, without considering longitudinal information from serial imaging. Incorporating temporal consistency constraints or modeling tumor evolution over time could improve segmentation robustness and open pathways toward prognostic modeling. Furthermore, expanding beyond anatomical imaging to include functional modalities—such as diffusion or perfusion MRI, or PET—could enhance tissue characterization and enable richer, biologically informed segmentation. Building fusion strategies capable of handling variable modality availability will be important for real-world adaptability.

Generalization remains a key consideration for clinical deployment. Generalization remains an important consideration for clinical deployment. Although the proposed framework demonstrated strong performance on the BraTS 2023 dataset, its robustness across different institutions, scanner vendors, and acquisition protocols has not yet been comprehensively evaluated. Therefore, future work will focus on validating the model on independent multi-center datasets and investigating domain adaptation strategies to

improve cross-site generalization and real-world applicability. We are actively pursuing collaborative efforts to conduct such evaluations and refine the model through domain adaptation strategies as needed. Expanding the system's applicability to other tumor types and brain pathologies—such as metastases, meningiomas, or CNS lymphomas—represents an important future direction, with potential acceleration via transfer learning from the current glioma-trained model [46–48].

Although the individual building blocks of the proposed framework are inspired by existing literature, the novelty of this work lies in their principled integration for boundary-sensitive brain tumor segmentation. In particular, our framework explicitly incorporates learnable boundary representations into the attention computation and couples this mechanism with a dynamic optimization strategy that progressively emphasizes boundary refinement during training. To the best of our knowledge, this combination has not been previously investigated for multi-modal brain tumor segmentation.

A major limitation of this study is that the proposed framework was evaluated only on the BraTS 2023 benchmark dataset. Although BraTS is a large and diverse multi-institutional dataset, the robustness of the model across independent public cohorts, scanner vendors, and acquisition protocols has not yet been comprehensively investigated. Therefore, the current results should be interpreted as evidence of strong benchmark performance rather than proof of clinical readiness. Future work will focus on external multi-center validation and domain adaptation strategies to assess the generalizability of the proposed framework in real-world clinical environments.

Beyond the immediate clinical implications, this work contributes to the broader effort of democratizing access to advanced medical imaging tools. In settings with limited specialist availability, reliable automated segmentation could serve as a valuable second reader, supporting earlier detection, consistent monitoring, and improved treatment planning. However, realizing this potential will require rigorous validation, regulatory oversight, and careful integration into clinical workflows that preserve human judgment. To support transparency and reproducibility, we release the code, trained models, and documentation as open-source resources. We hope this fosters further innovation and collaborative refinement within the research and clinical communities, ultimately translating into better outcomes for patients.

Compared with recent hybrid segmentation approaches such as ReHyDIL [29] and DFuse-Net [30], the proposed framework explicitly incorporates learnable boundary information into the attention mechanism, enabling more accurate delineation of complex tumor boundaries while maintaining competitive segmentation performance.

The additional evaluation further supports the robustness of the proposed framework beyond the primary benchmark dataset. Although a moderate decrease in Dice scores was observed, the performance remained consistently high across all tumor subregions, suggesting that the proposed hybrid CNN–Transformer architecture with boundary-aware attention is relatively resilient to variations in data distribution. Nevertheless, broader validation on larger multi-center cohorts acquired using different scanner vendors and imaging protocols remains an important direction for future research.

Future work will also investigate the incorporation of functional imaging modalities, including diffusion-weighted imaging (DWI), perfusion MRI (PWI), and positron emission tomography (PET). These modalities provide complementary physiological and metabolic information that is not available in conventional anatomical MRI sequences and may improve robustness in cases affected by motion artifacts, signal inhomogeneity, and ambiguous tissue boundaries. In particular, multimodal integration of anatomical and functional imaging could enable better characterization of infiltrative tumor regions and improve segmentation performance under challenging acquisition conditions.

6. Conclusions

This study introduces a novel multimodal framework for brain tumor segmentation that integrates a hybrid CNN–transformer architecture with boundary-aware attention mechanisms. Through extensive evaluation on the BraTS 2023 dataset, the proposed approach demonstrates state-of-the-art performance, achieving Dice scores of 92.3% for whole tumor, 88.7% for tumor core, and 84.5% for enhancing tumor segmentation. Notably, the inclusion of boundary-aware attention leads to significant improvements in delineation accuracy, reducing Hausdorff distance by 15–20% compared to established baselines. The proposed framework demonstrates strong potential for future clinical translation. However, additional validation on independent multi-center datasets will be necessary before considering routine clinical deployment.

By combining the local detail-capturing strengths of convolutional neural networks with the global context modeling of transformers, the framework effectively addresses key challenges inherent to brain tumor segmentation—such as heterogeneous tumor appearance, complex and irregular boundaries, and inter-class imbalance. The integration of a hybrid loss function and uncertainty quantification further enhances the system’s reliability and interpretability, providing actionable outputs that align with clinical expectations and decision-making needs.

Looking forward, future work will focus on optimizing computational efficiency to support deployment in resource-limited environments, improving segmentation in post-treatment scenarios, and incorporating longitudinal information for temporal consistency. Additionally, multi-center external validation will be pursued to assess generalizability across diverse populations and imaging protocols.

Furthermore, future research will investigate the integration of advanced Transformer-based architectures with wavelet-based feature extraction and metaheuristic optimization techniques (e.g., Gray Wolf Optimization) [49,50] to enhance feature representation and classification performance in brain imaging tasks. In parallel, incorporating fuzzy logic-based decision frameworks [51] will be explored to improve the transparency and interpretability of model predictions by modeling uncertainty and supporting explainable classification.

Author Contributions: Methodology, J.K. (Jamshid Khamzaev), J.K. (Jakhongir Karimberdiyev), M.R. and F.M.; software, M.R., J.K. (Jamshid Khamzaev) and J.K. (Jakhongir Karimberdiyev); validation, S.O., O.R. and I.D.; formal analysis, I.D., I.S., J.K. (Jamshid Khamzaev) and J.K. (Jakhongir Karimberdiyev); resources, M.R. and F.M.; data curation, O.R., J.K. (Jamshid Khamzaev), M.R. and J.K. (Jakhongir Karimberdiyev); writing—original draft, J.K. (Jamshid Khamzaev), M.R. and J.K. (Jakhongir Karimberdiyev); writing—review and editing, A.K., I.S. and F.M.; supervision, A.K. and F.M.; project administration, A.K. and I.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the project “AL-9424104925-R1”, funded by the Agency for Innovative Development under the Ministry of Higher Education, Science and Innovation of the Republic of Uzbekistan.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All used dataset are available online which open access.

Acknowledgments: We gratefully acknowledge the organizers of the BraTS Challenge for providing access to the dataset and evaluation infrastructure that supported this research. We also extend our sincere thanks to our clinical collaborators for their valuable input on tumor segmentation requirements and guidance on integrating automated tools into clinical workflows.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Davis, M.E. Glioblastoma: Overview of Disease and Treatment. *Clin. J. Oncol. Nurs.* **2016**, *20*, 2–8. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
2. Hanif, F.; Muzaffar, K.; Perveen, K.; Malhi, S.M.; Simjee, S.U. Glioblastoma Multiforme: A Review of its Epidemiology and Pathogenesis through Clinical Presentation and Treatment. *Asian Pac. J. Cancer Prev.* **2017**, *18*, 3–9. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
3. Webster, M.; Podgorsak, A.; Li, F.; Zhou, Y.; Jung, H.; Yoon, J.; Dona Lemus, O.; Zheng, D. New Approaches in Radiotherapy. *Cancers* **2025**, *17*, 1980. [[CrossRef](#)]
4. Xing, S.; Lai, Z.; Zhu, J.; He, W.; Mao, G. Semantic Segmentation of Brain Tumors Using a Local–Global Attention Model. *Appl. Sci.* **2025**, *15*, 5981. [[CrossRef](#)]
5. Bonato, B.; Nanni, L.; Bertoldo, A. Advancing Precision: A Comprehensive Review of MRI Segmentation Datasets from BraTS Challenges (2012–2025). *Sensors* **2025**, *25*, 1838. [[CrossRef](#)] [[PubMed](#)]
6. Xu, Y.; Quan, R.; Xu, W.; Huang, Y.; Chen, X.; Liu, F. Advances in Medical Image Segmentation: A Comprehensive Review of Traditional, Deep Learning and Hybrid Approaches. *Bioengineering* **2024**, *11*, 1034. [[CrossRef](#)] [[PubMed](#)]
7. Wang, S.; Kim, B.; Eom, D.-S. Boundary-Aware Transformer for Optic Cup and Disc Segmentation in Fundus Images. *Appl. Sci.* **2025**, *15*, 5165. [[CrossRef](#)]
8. Musthafa, N.; Memon, Q.A.; Masud, M.M. Advancing Brain Tumor Analysis: Current Trends, Key Challenges, and Perspectives in Deep Learning-Based Brain MRI Tumor Diagnosis. *Eng* **2025**, *6*, 82. [[CrossRef](#)]
9. Pham, T.Q.; Le, T.H.; Lai, K.D.; Ngo, D.Q.; Pham, T.V.; Hua, Q.H.; Le, K.Q.; Le, H.D.M.; Nguyen, T.N.L. CAAF-ResUNet: Adaptive Attention Fusion with Boundary-Aware Loss for Lung Nodule Segmentation. *Medicina* **2025**, *61*, 1126. [[CrossRef](#)] [[PubMed](#)]
10. Pant, K.; Dutta Pramanik, P.K.; Zhao, Z. A Robust ConvNeXt-Based Framework for Efficient, Generalizable, and Explainable Brain Tumor Classification on MRI. *Bioengineering* **2026**, *13*, 157. [[CrossRef](#)] [[PubMed](#)]
11. Mienye, I.D.; Swart, T.G.; Obaido, G.; Jordan, M.; Ilono, P. Deep Convolutional Neural Networks in Medical Image Analysis: A Review. *Information* **2025**, *16*, 195. [[CrossRef](#)]
12. Muksimova, S.; Umirzakova, S.; Mardieva, S.; Iskhakova, N.; Sultanov, M.; Cho, Y.I. A lightweight attention-driven YOLOv5m model for improved brain tumor detection. *Comput. Biol. Med.* **2025**, *188*, 109893. [[CrossRef](#)] [[PubMed](#)]
13. Park, M.; Oh, S.; Park, J.; Jeong, T.; Yu, S. ES-UNet: Efficient 3D medical image segmentation with enhanced skip connections in 3D UNet. *BMC Med. Imaging* **2025**, *25*, 327. [[CrossRef](#)] [[PubMed](#)]
14. Muksimova, S.; Baltaev, J.; Cho, Y.I. Brain Tumor Segmentation with Contextual Transformer-Based U-Net. *Electronics* **2026**, *15*, 782. [[CrossRef](#)]
15. Alquran, H.; Alslatie, M.; Rababah, A.; Mustafa, W.A. Improved Brain Tumor Segmentation in MR Images with a Modified U-Net. *Appl. Sci.* **2024**, *14*, 6504. [[CrossRef](#)]
16. Liu, X.; Qi, H.; Jia, S.; Guo, Y.; Liu, Y. Recent Advances in Optimization Methods for Machine Learning: A Systematic Review. *Mathematics* **2025**, *13*, 2210. [[CrossRef](#)]
17. Pidishetti, R.V.; Amjad, M.; Sheng, V.S. Advanced Brain Tumor Segmentation Using SAM2-UNet. *Appl. Sci.* **2025**, *15*, 3267. [[CrossRef](#)]
18. Nasimov, R.; Abdusalomov, A.; Khidirova, C.; Temirova, K.; Kutlimuratov, A.; Sadikova, S.; Jeong, W.; Choi, H.; Whangbo, T.K. PatientEase—Domain-Aware RAG for Rehabilitation Instruction Simplification. *Bioengineering* **2025**, *12*, 1204. [[CrossRef](#)] [[PubMed](#)]
19. Rasool, M.J.A.; Abdusalomov, A.; Kutlimuratov, A.; Ahamed, M.J.A.; Mirzakhililov, S.; Shavkatovich Buriboev, A.; Jeon, H.S. PixMed-Enhancer: An Efficient Approach for Medical Image Augmentation. *Bioengineering* **2025**, *12*, 235. [[CrossRef](#)] [[PubMed](#)]
20. Moutik, O.; Sekkat, H.; Tigani, S.; Chehri, A.; Saadane, R.; Tchakoucht, T.A.; Paul, A. Convolutional Neural Networks or Vision Transformers: Who Will Win the Race for Action Recognitions in Visual Data? *Sensors* **2023**, *23*, 734. [[CrossRef](#)] [[PubMed](#)]
21. Lu, Y.; Wang, S.; Wang, B.; Zhang, X.; Wang, X.; Zhao, Y. Enhanced Window-Based Self-Attention with Global and Multi-Scale Representations for Remote Sensing Image Super-Resolution. *Remote Sens.* **2024**, *16*, 2837. [[CrossRef](#)]
22. Liu, Z.; Wu, Y.; Xu, H.; Wang, M.; Weng, S.; Pei, D.; Chen, S.; Wang, W.; Yan, J.; Cui, L.; et al. Multimodal fusion of radio-pathology and proteogenomics identify integrated glioma subtypes with prognostic and therapeutic opportunities. *Nat. Commun.* **2025**, *16*, 3510. [[CrossRef](#)] [[PubMed](#)]
23. Rastogi, D.; Johri, P.; Donelli, M.; Kadry, S.; Khan, A.A.; Espa, G.; Feraco, P.; Kim, J. Deep learning-integrated MRI brain tumor analysis: Feature extraction, segmentation, and Survival Prediction using Replicator and volumetric networks. *Sci. Rep.* **2025**, *15*, 1437. [[CrossRef](#)] [[PubMed](#)]
24. Ullah, Z.; Hong, M.; Mahmood, T.; Kim, J. Systematic Integration of Attention Modules into CNNs for Accurate and Generalizable Medical Image Classification. *Mathematics* **2025**, *13*, 3728. [[CrossRef](#)]
25. Dasu, B.T.; Reddy, M.V.; Kumar, K.V.; Chithaluru, P.; Ahmed, N.; Elminaam, D.S.A. A self-attention driven multi-scale object detection framework for adverse weather in smart cities. *Sci. Rep.* **2026**, *16*, 1992. [[CrossRef](#)] [[PubMed](#)]

26. Lu, N.-H.; Huang, Y.-H.; Liu, K.-Y.; Chen, T.-B. Deep learning-driven brain tumor classification and segmentation using non-contrast MRI. *Sci. Rep.* **2025**, *15*, 27831. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
27. Terven, J.; Cordova-Esparza, D.-M.; Romero-González, J.-A.; Ramírez-Pedraza, A.; Chávez-Urbiola, E.A. A comprehensive survey of loss functions and metrics in deep learning. *Artif. Intell. Rev.* **2025**, *58*, 195. [[CrossRef](#)]
28. Koishiyeva, D.; Kang, J.W.; Iliev, T.; Bissembayev, A.; Mukasheva, A. Analysis of Loss Functions for Colorectal Polyp Segmentation Under Class Imbalance. *Eng. Proc.* **2025**, *104*, 17. [[CrossRef](#)]
29. Wang, J.; Fan, L.; Jing, W.; Di, D.; Song, Y.; Liu, S.; Cong, C. Hypergraph Tversky-Aware Domain Incremental Learning for Brain Tumor Segmentation with Missing Modalities. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2025*; Gee, J.C., Alexander, D.C., Hong, J., Iglesias, J.E., Sudre, C.H., Venkataraman, A., Golland, P., Kim, J.H., Park, J., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2026; Volume 15970. [[CrossRef](#)]
30. Zhou, T.; Wang, Z.; Ruan, S.; Meng, Y.; Duan, J.; Lei, B. DFuse-Net: Disentangled feature fusion with uncertainty-aware learning for reliable multi-modal brain tumor segmentation. *Med. Image Anal.* **2026**, *109*, 103923. [[CrossRef](#)] [[PubMed](#)]
31. Han, Y.; Li, X.; Wang, B.; Wang, L. Boundary Loss-Based 2.5D Fully Convolutional Neural Networks Approach for Segmentation: A Case Study of the Liver and Tumor on Computed Tomography. *Algorithms* **2021**, *14*, 144. [[CrossRef](#)]
32. Abdullah; Hafeez, N.; Sánchez-Mejorada, C.G.; Ruiz, M.J.T.; Téllez, R.Q.; Alex, E.A.; Sidorov, G.; Gelbukh, A. Enhancing Decision Intelligence Using Hybrid Machine Learning Framework with Linear Programming for Enterprise Project Selection and Portfolio Optimization. *AI* **2026**, *7*, 52. [[CrossRef](#)]
33. Tran, A.T.; Zeevi, T.; Payabvash, S. Strategies to Improve the Robustness and Generalizability of Deep Learning Segmentation and Classification in Neuroimaging. *BioMedInformatics* **2025**, *5*, 20. [[CrossRef](#)] [[PubMed](#)]
34. Li, B.; Zhou, J.; Gou, F.; Wu, J. TransRNetFuse: A highly accurate and precise boundary FCN-transformer feature integration for medical image segmentation. *Complex Intell. Syst.* **2025**, *11*, 213. [[CrossRef](#)]
35. Kim, B.; Devi, M.S.; Natarajan, Y.; R, S.P.K.; Yazhini, C.S.; Park, J.-Y.; Park, C.-J.; Yi, C.-Y. Gradient transformer Self-Attention U-Net for enhanced crack detection in concrete bridges. *Sci. Rep.* **2025**, *15*, 37087. [[CrossRef](#)] [[PubMed](#)]
36. Salman, B.; Yassin, E.; Ganta, D.; Luna, H. Deep Learning-Based Fusion of Multimodal MRI Features for Brain Tumor Detection. *Appl. Sci.* **2025**, *15*, 13155. [[CrossRef](#)]
37. Yeung, M.; Sala, E.; Schönlieb, C.-B.; Rundo, L. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graph.* **2022**, *95*, 102026. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
38. Zhu, C.; Bamidele, E.A.; Shen, X.; Zhu, G.; Li, B. Machine Learning Aided Design and Optimization of Thermal Metamaterials. *Chem. Rev.* **2024**, *124*, 4258–4331. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
39. Abdusalomov, A.; Rakhimov, M.; Karimberdiyev, J.; Belalova, G.; Cho, Y.I. Enhancing Automated Brain Tumor Detection Accuracy Using Artificial Intelligence Approaches for Healthcare Environments. *Bioengineering* **2024**, *11*, 627. [[CrossRef](#)] [[PubMed](#)]
40. Haq, M.U.; Sethi, M.A.J.; Ahmad, S.; Ahmad, N.; Anwar, M.S.; Kutlimuratov, A. A Comprehensive Review of Face Detection/Recognition Algorithms and Competitive Datasets to Optimize Machine Vision. *Comput. Mater. Contin.* **2025**, *84*, 1–24. [[CrossRef](#)]
41. Kutlimuratov, A.; Eshmurudov, D.; Tulaganova, F.; Utegenov, A.; Allayarov, P.; Khamzaev, J.; Saymanov, I.; Makhmudov, F. Early Diagnosis of Blood Disorders via Enhanced Image Preprocessing and Deep Learning Modeling. *BioMedInformatics* **2026**, *6*, 25. [[CrossRef](#)]
42. Isensee, F.; Jäger, P.F.; Full, P.M.; Vollmuth, P.; Maier-Hein, K.H. nnU-Net for Brain Tumor Segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2020*; Crimi, A., Bakas, S., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2020; Volume 12659. [[CrossRef](#)] [[PubMed](#)]
43. Cai, Y.; Long, Y.; Han, Z.; Liu, M.; Zheng, Y.; Yang, W.; Chen, L. Swin Unet3D: A three-dimensional medical image segmentation network combining vision transformer and convolution. *BMC Med. Inf. Decis. Mak.* **2023**, *23*, 33. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
44. Prakash, V.J.; Ajith, S.P.; Akash, S.; Nagarajan, V. A Transformer-Based Framework for Enhanced Brain Tumor Segmentation. In *Power Engineering and Intelligent Systems. PEIS 2025*; Shrivastava, V., Bansal, J.C., Panigrahi, B.K., Eds.; Lecture Notes in Electrical Engineering; Springer: Singapore, 2026; Volume 1461. [[CrossRef](#)]
45. Gao, L.; Luo, Z.; Wang, L. Convolutional Neural Network Acceleration Techniques Based on FPGA Platforms: Principles, Methods, and Challenges. *Information* **2025**, *16*, 914. [[CrossRef](#)]
46. Shakor, M.Y.; Khaleel, M.I. Recent Advances in Big Medical Image Data Analysis Through Deep Learning and Cloud Computing. *Electronics* **2024**, *13*, 4860. [[CrossRef](#)]
47. Tagmatova, Z.; Umirzakova, S.; Kutlimuratov, A.; Abdusalomov, A.; Im Cho, Y. A Hyper-Attentive Multimodal Transformer for Real-Time and Robust Facial Expression Recognition. *Appl. Sci.* **2025**, *15*, 7100. [[CrossRef](#)]
48. Gandhi, V.C.; Gandhi, P.P.; Raheem, A.K.A.; Alzubaidi, Y.T.; Khudaybergenov, K.; Khishe, M. Advancing glaucoma diagnosis: Multi-modal deep learning with vision transformer architectures. *Intell.-Based Med.* **2026**, *13*, 100355. [[CrossRef](#)]

49. Abdusalomov, A.; Umirzakova, S.; Bekmirzaev, O.; Dauletov, A.; Buriboev, A.; Kutlimuratov, A.; Nishanov, A.; Nasimov, R.; Oh, R. From Anatomy to Genomics Using a Multi-Task Deep Learning Approach for Comprehensive Glioma Profiling. *Bioengineering* **2025**, *12*, 979. [[CrossRef](#)] [[PubMed](#)]
50. Rezaei Nejad, A.; Sadat Sadati Salimi, F.; Hemmasian, M.; Mirzaee, S.; Abdiyeva, K.; Mousa, R.; Hesarak, S. Multi-class Alzheimer's disease (AD) classification using Swin Transformer wavelet and Gray Wolf Optimization (GWO). *Intell.-Based Med.* **2026**, *13*, 100362. [[CrossRef](#)]
51. Madrakhimov, S.; Makharov, K.; Khurramov, A. On the Transparency of Decision-Making in Classification by Precedents With Fuzzy Descriptions. *IEEE Access* **2025**, *13*, 173656–173664. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.