ISSN 2616-5775 Vol. 9, Issue 2: 23-29, DOI: 10.25236/AJCIS.2026.090204

Research on Transformer-Based Brain Tumor Image Segmentation

Dai Lina[1,a,*], Zheng Donghua[1], Xiu Weirong[1], Ye Lizhu[1]

[1]*School of Information Technology and Engineering, Guangzhou College of Commerce, Guangzhou, China*
[a] *dailina27@163.com*
[*]*Corresponding author*

Abstract: *Accurate segmentation of brain tumors is a critical component of image-guided diagnosis and treatment. However, this task faces substantial challenges due to highly irregular morphology, ill-defined boundaries, strong tissue heterogeneity of tumor subregions, as well as distribution discrepancies and missing modalities in multi-modal MRI data. In recent years, Transformer-based methods, driven by global self-attention and flexible cross-modal interaction mechanisms, have achieved remarkable progress in brain tumor segmentation. This paper systematically reviews representative Transformer-based approaches for multi-modal MRI brain tumor segmentation over the past five years, focusing on architectural design, problem–solution correspondence, applicable scenarios, and quantitative metrics. The analysis reveals two primary research trajectories: one centered on improving segmentation accuracy and boundary quality via multi-scale attention, cross-modal fusion, and explicit boundary modeling; the other oriented toward clinical needs, emphasizing interpretable outputs, robustness under missing modalities, and integrated segmentation–diagnosis pipelines. Evaluations on public datasets such as BraTS demonstrate that Transformers excel in global consistency modeling and small-lesion perception, while boundary errors (HD95) and stability under worst-case modality combinations remain major bottlenecks. Future research should prioritize multi-center domain generalization, lightweight deployment, and trustworthy segmentation mechanisms to drive the transition from benchmark leadership to meaningful clinical deployment.*

Keywords: *Medical image segmentation; Transformer; Deep learning; U-Net; Hybrid model*

1. Introduction

Accurate segmentation of brain tumors such as gliomas is a core component of image-guided diagnosis and treatment, directly supporting key clinical workflows including tumor burden assessment, delineation of surgical and radiotherapy target volumes, longitudinal treatment response monitoring, and prognosis analysis. Compared with general organ segmentation, brain tumor segmentation poses more pronounced challenges: tumor subregions (e.g., enhancing tumor, tumor core, and whole tumor) typically exhibit highly irregular morphology, ambiguous boundaries, and strong tissue heterogeneity. Meanwhile, multimodal MRI sequences (T1, T1ce/T1Gd, T2, and FLAIR) provide complementary contrast mechanisms but differ substantially in intensity distributions, requiring models to jointly achieve effective cross-modality fusion and globally consistent structural modeling. In recent years, the BraTS series challenges have continuously driven this research direction toward more clinically realistic scenarios. In particular, BraTS 2024 extends its focus to post-treatment MRI, introducing more complex anatomical structures and variation patterns such as resection cavities, thereby imposing higher requirements on algorithmic robustness, generalization capability, and clinical applicability [1].

Within the deep learning paradigm, convolutional neural networks (CNNs), especially U-shaped encoder–decoder architectures, have long dominated this field. However, due to the inherently local receptive fields of convolution and the stacked propagation mechanism, CNN-based models often struggle to reliably capture long-range dependencies across spatial regions. This limitation is typically manifested as insufficient joint modeling of global tumor shape consistency, long-distance contextual constraints, and fine-grained boundary details. Transformers, driven by self-attention mechanisms, enable more direct global interactions in the feature space and thus provide a new modeling pathway for multimodal brain tumor segmentation. Along this trend, prior studies have systematically reviewed the application landscape and development directions of vision Transformers in multimodal brain tumor MRI segmentation, highlighting their considerable potential in global context modeling and cross-modality information fusion [2].

Since 2023, Transformer-based approaches for brain tumor segmentation have demonstrated a dual trajectory of "task-specific architectural specialization" and "clinical problem-oriented design." On the one hand, researchers have begun to develop attention mechanisms tailored to key challenges in tumor segmentation (e.g., boundary delineation and small-volume subregions). For example, EoFormer, proposed at MICCAI 2023, explicitly emphasizes edge information modeling to enhance the discrimination of irregular tumor boundaries[3]. On the other hand, model development has gradually shifted from merely "improving benchmark scores" toward "enhancing trustworthiness and usability." For instance, TransXAI (2024) integrates hybrid Transformer architectures with explainability mechanisms, aiming to improve the interpretability of segmentation decisions via visual heatmaps and thereby alleviate clinical concerns regarding "black-box" models during deployment[4].

In parallel, to address the complexity of real-world clinical data, Transformer-based methods are increasingly strengthening their capability to handle issues such as missing modalities, degraded image quality, and domain shifts. AMGFormer (2026) explicitly targets prediction stability under missing-modality settings as a central objective, reducing performance fluctuations induced by different modality combinations through adaptive fusion and quality-aware enhancement modules. This reflects a broader transition of the field from idealized experimental settings toward clinically deployable configurations[5]. From a more macro-level perspective, a systematic review in 2025 further categorizes Transformer-based brain tumor analysis tasks across detection, segmentation, classification, and prognosis, suggesting that future research should, beyond architectural innovation, place greater emphasis on data-related bias control, generalization evaluation, and clinically critical factors such as interpretability and uncertainty estimation[6].

Based on the above background, this paper reviews Transformer-related studies in brain tumor image segmentation over the past five years, structured around three main themes: (1) the architectural characteristics and design motivations of different Transformer-based segmentation algorithms; (2) the strengths and limitations of representative methods; and (3) performance comparisons and trend analysis under unified or comparable experimental settings. In addition, key evaluation metrics are summarized to facilitate rapid understanding of research progress and reproducible implementation details.

2. Datasets and Evaluation Metrics

This section focuses on the most commonly used data sources, data organization strategies, and evaluation protocols for brain tumor segmentation—primarily in BraTS-related tasks—in recent studies. This provides a unified reference framework for the subsequent comparison of model architectures and segmentation performance.

2.1. Overview of Public Datasets

Brain tumor image segmentation tasks place high demands on dataset quality and annotation accuracy. To ensure the reliability and comparability of experimental results, many researchers choose the widely influential and authoritative BraTS (Brain Tumor Segmentation) public dataset series, which is commonly used in the medical image segmentation field, for experimental validation. The BraTS datasets are jointly released by multiple international research institutions and include multimodal MRI scan data along with finely detailed expert annotations, making them one of the most widely used standard benchmark datasets for brain tumor segmentation research. Among them, the BraTS 2019 and BraTS 2020 versions have been extensively adopted for evaluating the effectiveness and generalization ability of brain tumor image segmentation methods.

2.1.1. BraTS 2019: An Extended Classic Benchmark for Preoperative Multimodal Glioma Segmentation

BraTS 2019 continued the benchmark setting centered on multi-institutional, preoperative multimodal MRI glioma data and segmentation tasks, providing standardized evaluation for both glioblastoma/high-grade glioma (GBM/HGG) and low-grade glioma (LGG). At the challenge level, BraTS 2019 further highlighted the clinical relevance of segmentation: in addition to segmentation, overall survival (OS) prediction and segmentation uncertainty estimation were introduced as important extensions. This promoted a transition from evaluation based solely on "pure segmentation metrics" toward a more clinically oriented comprehensive assessment framework that better supports decision-

ISSN 2616-5775 Vol. 9, Issue 2: 23-29, DOI: 10.25236/AJCIS.2026.090204

making[7].

2.1.2. BraTS 2020: Strengthening Application-Oriented Evaluation (e.g., Survival and Pseudoprogression) While Maintaining Segmentation as the Core Task

BraTS 2020 maintained the main track of multi-institutional, preoperative multimodal MRI glioma segmentation (with segmentation remaining the core task). Meanwhile, the challenge design more explicitly incorporated extended evaluations related to clinical management (e.g., survival prediction and pseudoprogression-related task settings). As a result, competing and research methods were not only expected to achieve strong segmentation performance measured by Dice and HD95, but were also increasingly required to provide generalizable modeling and validation strategies for more complex clinical outcomes or longitudinal follow-up scenarios[8].

2.2. Evaluation Metrics and Comparison Principles

To ensure a comprehensive and fair evaluation of segmentation performance, multiple evaluation metrics are employed. These metrics assess the results from different perspectives, including volumetric overlap and boundary accuracy, thereby enabling a more objective and systematic comparison of various methods. The combination of overlap-based and distance-based measurements helps to better reflect the overall segmentation quality as well as the geometric precision of tumor delineation.

2.2.1. Dice Similarity Coefficient (DSC)

Dice is the primary metric for measuring segmentation overlap and is suitable for evaluating overall volumetric similarity. BraTS 2024 explicitly designates lesion-wise Dice as one of the core standardized evaluation metrics, enabling a more sensitive assessment of algorithm performance in scenarios involving multiple lesions or small lesions [1].

2.2.2. Hausdorff Distance (HD95)

HD95 emphasizes boundary error and geometric consistency, and is particularly important when tumor boundaries are ambiguous or shapes are irregular. BraTS 2024 also adopts lesion-wise Hausdorff Distance (commonly HD95) as one of the core evaluation metrics. It complements Dice: HD95 focuses more on robust statistics of the worst boundary errors, while Dice emphasizes voxel-level overlap.Use 18-point font for the title of article, aligned to the left and font bold, with single linespace and all the initial letters capitalized. No formulas or special characters of any form or language are allowed in the title.

3. Transformer-based Brain Tumor Segmentation Methods

Transformer-based approaches for brain tumor segmentation (primarily on BraTS multimodal MRI) have formed a relatively clear technical evolution. Early studies mainly adopted hybrid architectures in which CNNs capture local features while Transformers complement global context modeling. This paradigm has gradually developed into hierarchical U-shaped networks with Transformers as the main backbone, as well as convolution–attention interleaved designs tailored for 3D volumetric data. More recently, Transformer-based research has further expanded toward clinically critical directions, including robustness to missing modalities, boundary and small-target enhancement, and interpretability.

3.1. Hybrid CNN–Transformer Methods (CNN Encoder + Transformer Bottleneck/Enhancement)

Hybrid CNN–Transformer frameworks typically employ 3D CNNs to stably extract local textures and boundary details. High-level features are then tokenized and fed into a Transformer module to model long-range dependencies, after which a decoder restores the spatial resolution. The main advantage of this paradigm lies in preserving the strong inductive bias of CNNs for small medical datasets and local morphological patterns, while leveraging attention mechanisms to improve global consistency.

TransBTS is a representative framework following the pipeline of “3D CNN encoder to tokenization to Transformer-based global modeling to decoder”. It was proposed for multimodal MRI brain tumor segmentation, emphasizing the complementarity between local 3D contextual information and global dependencies, and its effectiveness was validated on BraTS 2019 and BraTS 2020[9]. Hybrid CNN–Transformer methods generally exhibit the following characteristics: (1) local features are robustly extracted by convolutional operations; (2) the Transformer is placed at the bottleneck stage, resulting in relatively controllable computational cost; and (3) multimodal inputs are usually handled via channel

ISSN 2616-5775 Vol. 9, Issue 2: 23-29, DOI: 10.25236/AJCIS.2026.090204

concatenation with shared encoding or feature-level fusion. These methods tend to be stable during training, less sensitive to dataset scale, and easy to integrate with nnU-Net-style engineering details. However, in most hybrid designs, the Transformer is still treated as an auxiliary module, and its global modeling capacity is constrained by the number and resolution of bottleneck tokens. Moreover, improvements in fine-grained boundary delineation may be less significant compared with architectures specifically designed for boundary modeling.

3.2. U-shaped Architectures with Transformer Encoders (ViT/Hierarchical Transformer Encoder + CNN Decoder)

The core idea of U-shaped architectures with Transformer encoders is to treat 3D volumetric data as sequences or hierarchical tokens, where the Transformer is primarily responsible for learning multi-scale global representations. On the decoding side, U-Net-style progressive upsampling is maintained, and multi-scale skip connections are used to recover fine-grained details.

Swin UNETR (2022) employs window-based attention and hierarchical features from Swin Transformer as the encoder, combined with a CNN-based decoder for multi-scale reconstruction. It demonstrated strong competitiveness on brain tumor segmentation tasks such as BraTS 2021[10]. The main advantages of Transformer-encoder U-shaped structures include more systematic global-context modeling and multi-scale representation learning. They also show greater potential for capturing tumor-level structural consistency and aligning cross-modality semantics. However, for 3D inputs, these models typically require substantial GPU memory and incur high training costs, and their performance is more sensitive to pretraining, data augmentation, and inference strategies (e.g., sliding-window inference and test-time augmentation).

3.3. Convolution–Attention Interleaved Architectures for Volumetric Data

In 3D brain tumor segmentation, pure attention mechanisms incur enormous computational costs at the voxel level and lack the local inductive bias inherent to convolutional operations. Conversely, purely convolutional models struggle to stably capture long-range inter-regional dependencies and overall structural consistency. Therefore, recent mainstream approaches no longer treat the Transformer as an auxiliary add-on module, but instead tightly interleave or integrate convolution and attention mechanisms within the backbone network. Convolutions are employed to stabilize local texture and boundary representations, while attention mechanisms establish long-range voxel interactions and facilitate multi-scale information flow.

HCA-former proposes a Hybrid Convolution–Attention Transformer tailored for 3D medical segmentation. Within a U-shaped encoder–decoder framework, it emphasizes the effective fusion of convolutional and Transformer features, and employs attention designs targeting multi-channel and multi-scale interactions to reduce the semantic gap between encoder and decoder. Its validation experiments span multiple 3D segmentation tasks, including MSD brain tumor, demonstrating that this convolution–attention fused backbone exhibits strong generalization ability for volumetric data[11]. In 3D scenarios, HCA-former more systematically balances local inductive bias and global relational modeling, improving both Dice scores and boundary-related metrics such as HD95. However, due to its relatively complex architectural components, it introduces higher engineering complexity during training and inference. Moreover, its performance is more sensitive to hyperparameters (e.g., patch size, sliding-window stride, and regularization) across different datasets.

VSmTrans introduces the concept of a Variable-Shape Mixed Transformer. Its key innovation lies in embedding self-attention and convolution tightly within a unified hybrid paradigm rather than simply arranging them in parallel branches. By adopting variable-shape or variable-window attention mechanisms, it enlarges the receptive field without significantly increasing computational cost, thereby jointly capturing local details and global context. This method has been validated on multiple 3D segmentation datasets, including BraTS 2021 MRI, demonstrating the adaptability of such 3D hybrid backbones to brain tumor segmentation tasks[12]. Compared with fixed-window attention, variable-shape attention is better suited for handling tumors with irregular morphology and large scale variations, while convolutional enhancement benefits boundary delineation and small-volume region representation. Nevertheless, the implementation and reproducibility of this approach rely on more refined module design, and additional robustness strategies are still required under cross-center domain shifts or missing-modality scenarios.

4. Comparative Performance Analysis

This section focuses on the lineage of Transformer-based brain tumor segmentation methods mentioned earlier, selecting representative Transformer-based approaches from recent years for comparative performance analysis. The analysis revolves around four key questions: "What critical challenge does the method address, why can it solve it, in which scenarios is it more applicable, and what is the final outcome?" Considering potential differences in dataset versions, training strategies (e.g., sliding window inference, TTA), and evaluation scripts across studies, this section prioritizes horizontal comparisons under the same dataset or similar settings, and explicitly notes cases where direct comparisons are not feasible.

Table 1 summarizes, from a methodological perspective, the task-specific challenges each model targets, key mechanisms, and applicable scenarios, highlighting the logic of "structure design–problem correspondence." Table 2 provides a quantitative evaluation, compiling the Dice coefficient and Hausdorff distance of each model on public datasets to demonstrate the practical benefits of different structures and mechanisms in terms of segmentation accuracy and boundary errors. Through the dual-table presentation of "mechanism summarization + metric comparison," this section aims to offer a more intuitive and reproducible reference for subsequent model selection and improvement directions.

Table 1: Comparison of Model Mechanisms.

Model	Challenges Addressed	Key Mechanisms	Applicable Scenarios
EA-DFFTU-Net[13]	Unstable segmentation under missing modalities and loss of boundary information	EFM (Edge Feature Module) + DFFM (Dynamic Feature Fusion Module)	Multi-modal MRI brain tumor segmentation, especially with missing modality inputs
TransXAI[4]	etability, low clinical trust	CNN+ViT hybrid segmentation; post-hoc XAI (Grad-CAM) generates interpretable heatmaps without altering network structure; analyzes contribution of each MRI modality	Multi-modal MRI glioma segmentation, emphasizing interpretability and clinical usability
multiPI-TransBTS[14]	Fusion interference caused by multi-modal MRI heterogeneity; insufficient boundary voxel context	Multi-branch modality-specific extraction + AFF attention fusion + TSFI/PCI task-specific introduction	BraTS multi-modal MRI tumor segmentation, suitable for scenarios requiring modality discrepancy modeling and task-specific decoding
BrainTumNet[15]	Integrated segmentation + classification to avoid single-task fragmentation	CNN encoder-decoder + Adaptive Masked Transformer (dynamic mask sparse attention) + Pyramid multi-scale fusion + Skip connection CBAM	T1-enhanced (CE-T1) brain tumors: joint diagnosis of segmentation + pathological type classification (glioma/metastasis/meningioma)
TMA-TransBTS[16]	Difficulty in modeling 3D long-range dependencies; multi-scale lesions prone to missed detection	CNN-Transformer hybrid U-shape; TMSM multi-scale self-attention (multi-scale token partitioning/aggregation) + TMCM multi-scale cross-attention (replacing skip connections) + Deep supervision	3D medical image segmentation, suitable for scenarios requiring global context modeling and multi-scale feature learning

Table 2: Performance metrics comparison.

Model	Dataset	Dice (WT / TC / ET)	Hausdorff Distance (WT / TC / ET)
EA-DFFTU-Net[13]	BraTS 2019	95.6 / 96.5 / 92.6	3.7 / 3.2 / 2.1
EA-DFFTU-Net[13]	BraTS 2020	93.7 / 85.5 / 80.6	3.7 / 3.2 / 2.5
TransXAI[4]	BraTS 2019	88.2 / 78.2 / 74.5	6.4 / 7.9 / 4.3
multiPI-TransBTS[14]	BraTS 2019	95.3 / 94.4 / 90.4	2.2 / 1.5 / 1.3
multiPI-TransBTS[14]	BraTS2020	95.4 / 94.7 / 90.9	2.2 / 1.5 / 1.3
TMA-TransBTS[16]	BraTS 2019	91.4 / 78.7 / 72.5	5.7 / 9.4 / 9.1

From the above analysis, it can be seen that recent research on Transformer-based brain tumor segmentation has followed two relatively distinct trajectories: one focused on improving segmentation accuracy and boundary quality, emphasizing multi-scale attention, cross-modal interaction, and selective feature fusion; the other targeting clinical usability, highlighting interpretable outputs and integrated “segmentation + diagnosis” pipelines.

ISSN 2616-5775 Vol. 9, Issue 2: 23-29, DOI: 10.25236/AJCIS.2026.090204

To address the challenges of missing modalities and loss of boundary information, EA-DFFTU enhances stability through edge feature extraction and selective fusion[13]; for multi-modal heterogeneity and fusion interference, multiPI-TransBTS adopts modality-specific extraction and adaptive fusion to achieve more robust overall performance[14]; TMA-TransBTS leverages 3D multi-scale self-attention and cross-attention to strengthen long-range dependency modeling and small lesion perception, systematically reporting Dice and HD95 on multi-year BraTS datasets, facilitating correlation analysis between architecture and metrics[16]. Overall, multi-scale attention and cross-modal interaction have become the dominant mechanisms for improving segmentation accuracy and boundary quality (HD/HD95) in 3D tumor segmentation.

Some studies have shifted focus from optimizing metrics toward clinical applicability. TransXAI emphasizes post-hoc interpretability (heatmaps) to enhance model auditability[4]; BrainTumNet couples segmentation and classification to better align with clinical workflows[15]. In summary, comparative performance analysis suggests that future research should, while maintaining standardized evaluation protocols, also address robustness under missing modalities, boundary reliability, and interpretable/trustworthy outputs, thereby closing the loop from laboratory metrics to clinical deployment.

5. Challenges and Future Trends

(1) From Full-Modality SOTA to Robustness Under Missing Modalities and Real-World Clinical Conditions. Clinical data often suffer from missing modalities, quality degradation, and variations in scanning protocols. For models, performance and stability under the “worst-case modality combination” are more critical than average Dice scores. Trends indicate that Transformer-based methods are shifting from simple feature concatenation toward cross-attention fusion, quality-aware fusion, and robustness evaluation (e.g., cross-modality combination variance, worst-case metrics) to enhance practical usability.

(2) Boundaries and Small Targets (e.g., ET) Remain Performance Bottlenecks. The blurred boundaries of brain tumor subregions, small volume of enhancing tumors, and their susceptibility to confusion with blood vessels or necrosis make global context alone insufficient. Future designs are more likely to incorporate explicit boundary evidence modeling (e.g., edge attention, boundary branches) combined with surface distance-driven optimization (e.g., HD95/NSD-guided losses) to simultaneously improve overlap accuracy and geometric error.

(3) 3D Computational Cost and Deployment Constraints Drive Efficient Attention Mechanisms. Under 3D volumetric data, attention computation and memory consumption remain limiting factors. This has spurred research into window-based attention, sparse attention, deformable attention, and CNN-attention hybrid architectures, alongside distillation, pruning, or lightweight Transformers to meet clinical inference latency and hardware conditions.

(4) Trustworthy Segmentation: Interpretability and Uncertainty Will Become “Mandatory”. For clinical deployment, models must provide evidence for “why this segmentation” and highlight low-confidence regions. Uncertainty estimation, failure case analysis, and interpretable heatmaps will evolve alongside segmentation backbones, gradually shifting from “paper demonstrations” to standardized evaluation (e.g., calibration, risk-benefit trade-offs).

6. Conclusion

This paper systematically reviews recent advances in Transformer-based multi-modal MRI brain tumor segmentation, organizing the datasets and evaluation metrics, and categorizing representative methods according to architectural paradigms and task-specific challenges. Overall, Transformers—through global dependency modeling and more flexible cross-modal interaction mechanisms—offer an effective pathway for improving tumor subregion consistency and segmentation performance under complex scenarios. Meanwhile, multi-scale attention, selective fusion, and hybrid backbone designs constitute the primary sources of current performance gains.

Future efforts should place greater emphasis on real-world clinical challenges such as missing modalities and multi-center domain generalization, while incorporating boundary quality, stability, and trustworthy outputs (interpretability/uncertainty) as core evaluation dimensions. With continued advances in efficient 3D attention, robust fusion, and trustworthy segmentation mechanisms, Transformers are poised to drive brain tumor segmentation from "benchmark leadership" toward meaningful clinical deployment.

ISSN 2616-5775 Vol. 9, Issue 2: 23-29, DOI: 10.25236/AJCIS.2026.090204

Acknowledgements

This work is funded by the following projects: the 2024 scientific research project of Guangzhou College of Commerce (2024XJYB024); the Education Science Planning Project of the Guangdong Provincial Department of Education (2024GXJK452); the Higher Education Research Project of the Guangdong Higher Education Association (24GYB74); the Guangdong Higher Vocational Education Teaching Quality and Teaching Reform Project (2023JG452); the Education Science Planning Project (Higher Education Special Program) of Guangdong Province (2025GXJK0643); the "14th Five-Year Plan" 2024 Higher Education Research Project of the Guangdong Higher Education Association (24GYB75); the 2024 Quality Engineering Project of Guangzhou College of Commerce (2024ZLGC11); and the First-Class Course Project of Guangzhou College of Commerce (2020XJYLKC06).

References

[1] de Verdier, M.C., et al., The 2024 brain tumor segmentation (brats) challenge: Glioma segmentation on post-treatment mri. arXiv preprint arXiv:2405.18368, 2024.
[2] Wang, P., et al., Vision transformers in multi-modal brain tumor MRI segmentation: A review. Meta-Radiology, 2023. 1(1): p. 100004.
[3] She, D., et al. Eoformer: Edge-oriented transformer for brain tumor segmentation. in International Conference on Medical Image Computing and Computer-Assisted Intervention. 2023. Springer.
[4] Zeineldin, R.A., et al., Explainable hybrid vision transformers and convolutional network for multimodal glioma segmentation in brain MRI. Scientific reports, 2024. 14(1): p. 3713.
[5] Guo, C., et al., AMGFormer: Adaptive Multi-Granular Transformer for Brain Tumor Segmentation with Missing Modalities. arXiv preprint arXiv:2601.19349, 2026.
[6] Kumar, A., A comprehensive review of transformer models in brain tumor analysis. Expert Systems with Applications, 2025: p. 130509.
[7] Bonato, B., L. Nanni, and A. Bertoldo, Advancing precision: A comprehensive review of MRI segmentation datasets from brats challenges (2012–2025). Sensors (Basel, Switzerland), 2025. 25(6): p. 1838.
[8] Mehta, R., et al., QU-BraTS: MICCAI BraTS 2020 challenge on quantifying uncertainty in brain tumor segmentation-analysis of ranking scores and benchmarking results. The journal of machine learning for biomedical imaging, 2022. 2022: p. https://www. melba-journal. org/papers/2022: 026. html.
[9] Wang, W., et al. Transbts: Multimodal brain tumor segmentation using transformer. in International conference on medical image computing and computer-assisted intervention. 2021. Springer.
[10] Hatamizadeh, A., et al. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. in International MICCAI brainlesion workshop. 2021. Springer.
[11] Yang, F., et al., HCA-former: Hybrid convolution attention transformer for 3D medical image segmentation. Biomedical Signal Processing and Control, 2024. 90: p. 105834.
[12] Liu, T., et al., VSmTrans: A hybrid paradigm integrating self-attention and convolution for 3D medical image segmentation. Medical image analysis, 2024. 98: p. 103295.
[13] Jagadeesh, B. and G.A. Kumar, Brain tumor segmentation with missing MRI modalities using edge aware discriminative feature fusion based transformer U-net. Applied Soft Computing, 2024. 161: p. 111709.
[14] Zhu, H., et al., multiPI-TransBTS: A multi-path learning framework for brain tumor image segmentation based on multi-physical information. Computers in Biology and Medicine, 2025. 191: p. 110148.
[15] Lv, C., et al., BrainTumNet: multi-task deep learning framework for brain tumor segmentation and classification using adaptive masked transformers. Frontiers in Oncology, 2025. 15: p. 1585891.
[16] Huang, Y., L. Chen, and C. Zhou, Multi-Modal Brain Tumor Segmentation via 3D Multi-Scale Self-attention and Cross-attention. arXiv preprint arXiv:2504.09088, 2025.