

# From U-Net to Swin-Unet Transformers: The Next-Generation Advances in Brain Tumor Segmentation with Deep Learning

Mushtaq Mahyoob Saleh, Bharat B. Biswal

Biomedical Engineering Department, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China

**Correspondence to:** Mushtaq Mahyoob Saleh, [mushtaqbme2022@gmail.com](mailto:mushtaqbme2022@gmail.com), [mmsbme2@gmail.com](mailto:mmsbme2@gmail.com)

**Keywords:** Brain Tumor Segmentation, Deep Learning, 3D U-Net, Vision Transformers, Federated Learning, Generative Models

**Received:** July 6, 2025

**Accepted:** August 15, 2025

**Published:** August 18, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## ABSTRACT

Brain tumor segmentation is a vital step in diagnosis, treatment planning, and prognosis in neuro-oncology. In recent years, deep learning approaches have revolutionized this field, evolving from the foundational 3D U-Net architecture widely regarded as a benchmark for volumetric medical image segmentation to more advanced transformer-based models such as Swin UNET Transformers. Despite the success of 3D U-Net, challenges remain due to its reliance on local dependencies, which limit its ability to capture global context. Additional difficulties include accurately delineating complex tumor boundaries, addressing class imbalance across tumor subregions, and ensuring stable training of deep networks. This review comprehensively surveys the evolution of brain tumor segmentation techniques, emphasizing the transition from conventional U-Net models to cutting-edge Swin UNET transformer architectures. We discuss the impact of novel activation functions on improving gradient stability and segmentation accuracy. Furthermore, we explore complementary advancements, including weakly supervised learning, transformer-based frameworks for enhanced global context modeling, generative models for data augmentation, explainable AI for increased interpretability, and federated learning approaches for privacy-preserving collaboration. Key public datasets, such as BraTS, and standard evaluation metrics are reviewed to contextualize model performance. Lastly, we address ongoing challenges such as data heterogeneity, real-time clinical applicability, and integration barriers, and propose future directions for developing robust, interpretable, and scalable brain tumor segmentation systems. This review aims to provide researchers with a clear perspective on the next-generation evolution of deep learning methods that are shaping the future of brain tumor segmentation.

## 1. INTRODUCTION

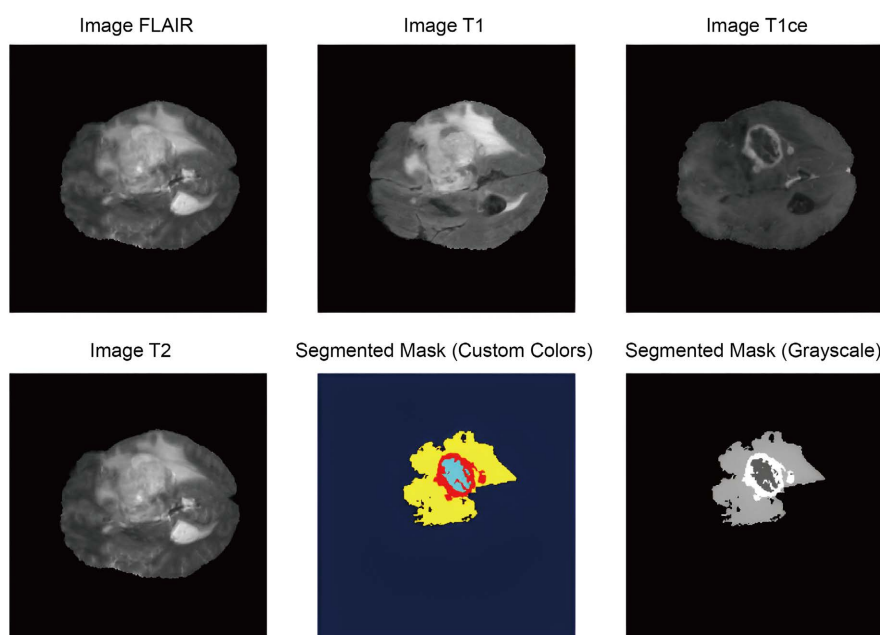
Brain tumors are among the most life-threatening neurological disorders, significantly impacting morbidity and mortality worldwide. According to the World Health Organization (WHO), brain tumors are classified into various grades (I - IV) based on their malignancy, with glioblastoma multiform (GBM) being the most aggressive [1, 2]. Early and accurate diagnosis is crucial for treatment planning, surgical intervention, and patient prognosis.

Magnetic Resonance Imaging (MRI) is the primary diagnostic tool for brain tumor assessment due to its superior soft-tissue contrast and non-invasive nature [3]. However, manual segmentation of tumor regions by radiologists is time-consuming, subjective, and prone to inter-observer variability. This has led to the development of automated and semi-automated brain tumor segmentation techniques to improve efficiency and reproducibility.

Brain tumor segmentation involves delineating different tumor subregions, such as the enhancing tumor (ET), peritumoral edema (ED), and necrotic core (NCR), from multimodal MRI scans (T1, T1c, T2, FLAIR).

Figure 1 showcases the integration of multiple MRI modalities FLAIR, T1, T1ce, and T2 to visualize different tissue characteristics and tumor components in the brain. Each modality highlights specific pathological features: FLAIR emphasizes edema, T1 provides anatomical details, T1ce highlights actively enhancing tumor regions, and T2 assists in differentiating tumor from healthy tissue. The segmented masks, displayed in both custom colors and grayscale, represent the output of an automated segmentation model, classifying tumor subregions such as enhancing tumor, tumor core, and edema with distinct labels.

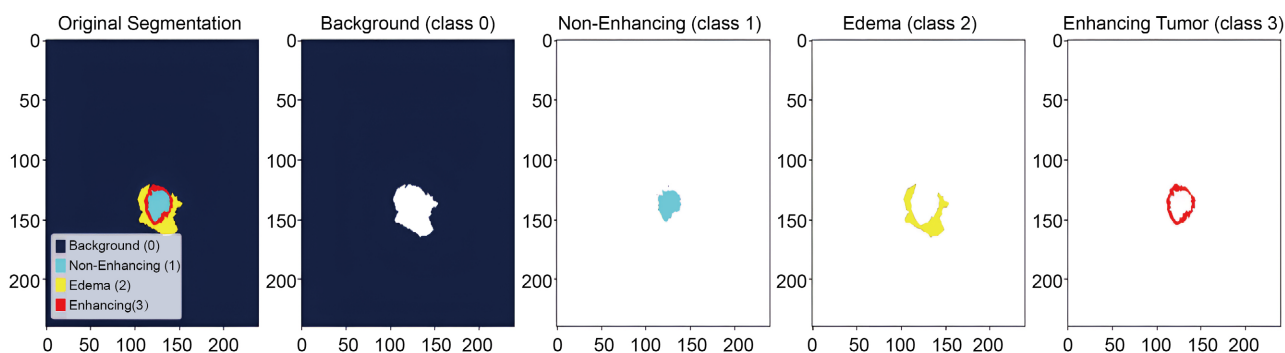
The 3D U-Net model plays a critical role in this process by processing the entire volumetric MRI data to produce accurate voxel-wise segmentation. Its encoder-decoder architecture captures spatial context in three dimensions, preserving detailed anatomical structures through skip connections. By effectively combining the complementary information from multiple MRI sequences, 3D U-Net generates detailed, multi-class segmentation maps that are essential for precise tumor localization, treatment planning, and prognosis evaluation in clinical practice.



**Figure 1.** MRI modalities (Flair, T1, T1ce, T2), segmented color and grayscale mask. Color-coded: (background) in dark blue, (non-enhancing tumor) in cyan, (edema) in yellow, and (enhancing tumor) in red.

Over the past decade, advancements in machine learning (ML) and deep learning (DL) have revolutionized segmentation accuracy. Traditional methods, such as thresholding, region-growing, and clustering (K-means, Fuzzy C-means), have been increasingly replaced by convolutional neural networks (CNNs) and transformer-based architectures.

**Figure 2** visualizes the segmentation of a brain tumor into its constituent subregions using a multi-class labeling scheme. The first panel (“Original Segmentation”) shows the complete labeled mask, where different tumor components are color-coded: class 0 (background) in dark blue, class 1 (non-enhancing tumor) in cyan, class 2 (edema) in yellow, and class 3 (enhancing tumor) in red. The subsequent panels isolate each class for clearer interpretation. Panel 2 displays the non-tumor region (class 0), while panels 3 to 5 separately highlight the non-enhancing core (class 1), the surrounding edema (class 2), and the enhancing tumor core (class 3), respectively. This breakdown aids in analyzing tumor heterogeneity and is crucial for diagnosis, treatment planning, and model evaluation.



**Figure 2.** Illustrates the classification of brain tumor segmentation into four categories: class 0—Non-tumor region, class 1—Non-enhancing tumor, class 2—Edema, and class 3—Enhancing tumor.

Despite significant progress, challenges remain, including heterogeneous tumor appearance, class imbalance, and limited annotated datasets. Publicly available datasets like the BraTS (Brain Tumor Segmentation Challenge) have played a pivotal role in benchmarking algorithms [4, 5]. Recent trends include the integration of attention mechanisms, 3D CNNs, and hybrid models to enhance segmentation performance [6].

In recent years, deep learning (DL)-based segmentation has dominated the field, surpassing traditional machine learning (ML) techniques such as random forests, support vector machines (SVMs), and atlas-based methods. The introduction of U-Net [7] revolutionized medical image segmentation, and its 3D variants (e.g., 3D U-Net, V-Net) further improved volumetric tumor analysis.

This review paper provides a comprehensive analysis of state-of-the-art brain tumor segmentation techniques, discussing their strengths, limitations, and future directions. We cover traditional ML approaches, deep learning models, evaluation metrics, and emerging trends in the field.

## 2. MATERIALS AND METHODS

### 2.1. Review Methodology

This review aims to comprehensively analyse the current state of brain tumour segmentation using deep learning techniques, with a particular focus on advanced techniques such as U-Net architectures and transformers based models. To ensure a systematic and thorough examination of the literature, a structured search was conducted using electronic databases, including PubMed, IEEE Xplore, Science Direct, and Google Scholar.

**Search Strategy:** The search terms used included “brain tumour segmentation,” “deep learning,” “U-Net,” “MRI,” “BraTS challenge,” “glioma segmentation,” “convolutional neural networks,” “activation func-

tions,” “transformers”, and “medical image analysis.” The review studies were considered to capture the most recent advancements, especially those related to the BraTS challenges. Only articles published in English were included.

**Inclusion Criteria:** The review included peer-reviewed journal articles and conference papers that focused on brain tumour segmentation using deep learning techniques. Research involving U-Net architectures or their variants was prioritised. Papers discussing next generation advances techniques in brain tumor segmentations.

**Exclusion Criteria:** Studies not related to brain tumour segmentation or those not utilising deep learning methods were excluded. Non-English publications were also excluded.

**Data Extraction and Synthesis:** Relevant information from the selected studies was extracted, including the proposed methods, datasets used, CNN, activation functions, performance metrics, and key findings. Emphasis was placed on studies that provided critical insights into the advancements, challenges, and future directions of U-Net architectures, and transformers in brain tumour segmentation.

The BraTS (Brain Tumor Segmentation) challenges have served as a pivotal benchmark for evaluating AI-driven segmentation methods, catalyzing remarkable progress in the field [3, 5, 8]. Traditional machine learning techniques, while initially useful, struggled with the heterogeneous appearance of tumors in BraTS multi-modal MRI datasets, achieving limited Dice scores (typically 60% - 75%). The introduction of deep learning, particularly U-Net variants [9], dramatically improved performance (Dice ~85% - 90%) by automatically learning discriminative features across T1, T2, FLAIR, and T1ce sequences. Subsequent BraTS editions witnessed transformer-based models like Swin UNETR [10] pushing boundaries further (Dice > 90%) through global context modeling, while diffusion models enhanced edge detection in tumor sub-regions. The challenges also spurred innovations in federated learning [11] to address data privacy concerns and weakly supervised techniques [12] to mitigate annotation bottlenecks. Notably, BraTS-2023 highlighted how ensemble methods combining CNNs and transformers achieved state-of-the-art results (Dice ~92%), demonstrating the synergistic potential of hybrid architectures [6]. These methodological advances, rigorously tested through BraTS, have not only improved algorithmic performance but also translated to more reliable clinical decision-support systems, reducing inter-rater variability from 15% - 20% to under 5% in tumor volume estimation.

## 2.2. Traditional Machine Learning (ML) Approaches

Before the deep learning era, classical ML techniques formed the foundation of brain tumor analysis by leveraging statistical models and manually engineered features.

### 2.2.1. Feature-Based Methods

Texture and intensity-based approaches were pivotal in early tumor characterization. Haralick features and Gabor filters extracted textural patterns from MRIs, while Local Binary Patterns (LBP) captured local contrast variations [13]. Histogram-based methods analyzed intensity distributions across T1, T2, and FLAIR sequences to identify abnormal tissue. For morphological analysis, Active Contours (Snakes) and Level Sets evolved initial contours to match tumor boundaries [14], with Region Growing techniques propagating seeds based on intensity similarity [15].

### 2.2.2. Supervised Learning Classifiers

Supervised methods enabled automated tumor classification using labeled data. Support Vector Machines (SVMs) with RBF kernels effectively separated tumor and healthy tissue by maximizing margin hyperplanes in high-dimensional feature spaces [16]. Random Forests improved robustness through ensemble decision trees that handled multi-class segmentation tasks [17]. While simpler, k-Nearest Neighbors (k-NN) provided baseline performance by classifying voxels based on neighboring annotated samples [18].

### 2.2.3. Unsupervised Learning Methods

These techniques identified tumor regions without prior labels. K-means and Fuzzy C-Means (FCM)

clustered voxels based on intensity similarity, with FCM allowing partial membership to account for tissue heterogeneity [19]. Gaussian Mixture Models (GMMs) offered probabilistic segmentation by fitting MRI intensities to multiple Gaussian distributions, particularly effective for differentiating tumor sub-regions.

### 2.3. Deep Learning (DL) Approaches

Deep Learning has revolutionized brain tumor segmentation through its ability to automatically learn hierarchical features from medical images without manual feature engineering.

#### 2.3.1. Convolutional Neural Networks (CNNs)

2D CNNs established foundational architectures for medical image analysis. The U-Net architecture [7] became the gold standard by combining an encoder-decoder structure with skip connections to preserve spatial details. SegNet [20] improved computational efficiency by using pooling indices for precise upsampling. For volumetric analysis, 3D U-Net [21] extended this approach to process whole MRI volumes, while V-Net [22] incorporated residual connections to enhance gradient flow in deep networks. DeepMedic [23] introduced parallel processing pathways to capture multi-scale tumor features simultaneously.

#### 2.3.2. Advanced CNN Architectures

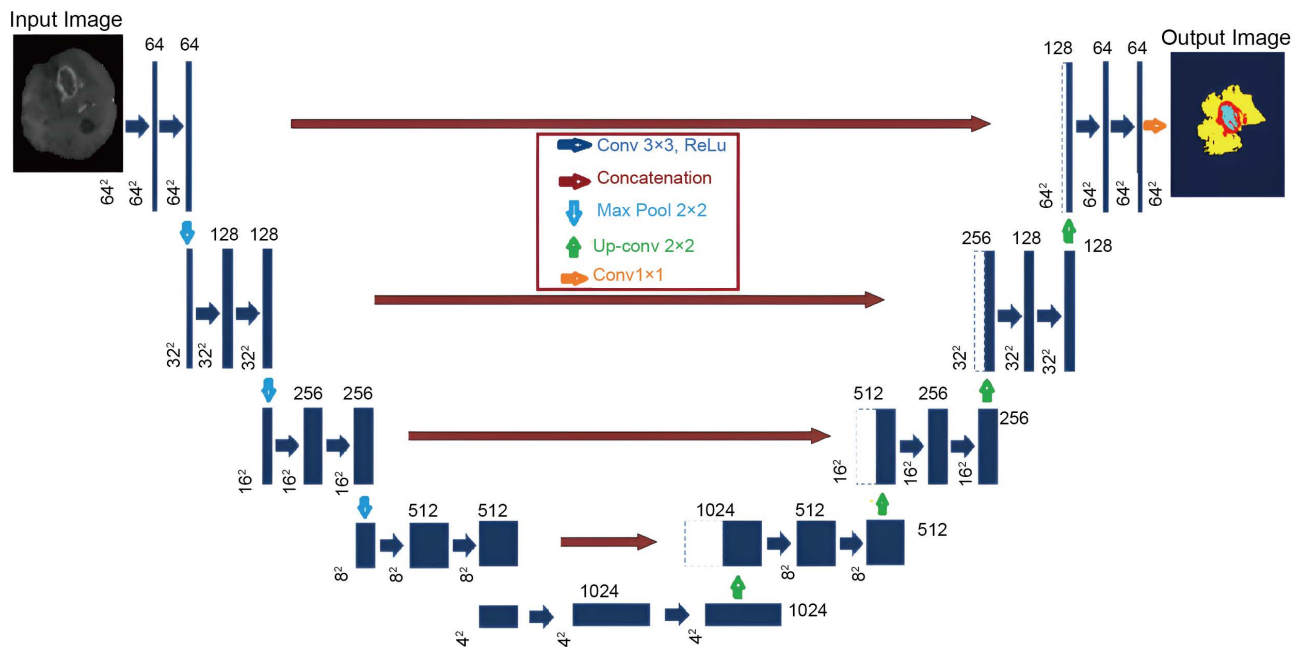
Attention mechanisms significantly improved segmentation precision. Attention U-Net [24] learned to focus computational resources on tumor regions while suppressing irrelevant areas, enhancing the localization of complex tumor boundaries. Squeeze-and-Excitation blocks [25] dynamically recalibrate channel-wise feature responses, allowing the network to emphasize informative features while diminishing noise. Hybrid architectures like DenseUNet [26, 27] employed pyramid pooling to capture contextual information at various receptive fields, which is essential in recognizing tumors of varying sizes and shapes. These designs offer richer semantic understanding and more robust feature representation. Further advancements incorporated residual connections, dilated convolutions, and deep supervision to enhance learning stability and accuracy.

Networks like DeepLabV3+, and High Resolution Network (HRNet) pushed the boundaries by integrating high-resolution representations and spatial hierarchies, improving both edge delineation and intratumoral heterogeneity recognition. Collectively, these innovations in CNN architecture have led to notable improvements in segmentation accuracy, sensitivity, and clinical applicability, marking a significant leap from early deep learning models.

#### 2.3.3. U-Net Architecture

U-Net is a convolutional neural network (CNN) architecture designed for biomedical image segmentation [7], featuring a symmetric encoder-decoder structure with skip connections to preserve spatial details.

**Figure 3** illustrates a 3D U-Net architecture for image segmentation, detailing the input and output dimensions along with the constraints and costs associated with the network's operations. The input image undergoes a series of transformations through the network, with dimensions progressively changing from  $64 \times 64$  to  $1024 \times 1024$  and then downsampling back to  $64 \times 64$ . The network employs upsampling (denoted as "Up") with a cost of  $2 \times 2$  and standard convolutions with a cost of  $1 \times 1$ . A key constraint is that the architecture avoids  $1 \times 2$  operations, ensuring specific dimensional transformations. The output image mirrors certain dimensions of the input (e.g.,  $256 \times 256$ ,  $32 \times 32$ , etc.), highlighting the U-Net's symmetric structure where high-resolution features are combined with upsampled outputs to achieve precise segmentation. This design is critical for tasks requiring detailed spatial accuracy, such as medical image segmentation, where the U-Net's ability to capture both local and global features is essential. The inclusion of cost metrics suggests an emphasis on computational efficiency, which is vital for practical implementations. Overall, the image encapsulates the U-Net's hierarchical approach, balancing depth, resolution, and computational constraints to optimize performance.



**Figure 3.** The U-Net architecture for brain tumor segmentation in MRI, illustrating the input MRI scan and corresponding segmented output.

### 2.3.4. Activation Functions

Activation functions are crucial components of deep learning models, introducing non-linearity that allows networks to approximate complex, non-linear functions and learn intricate patterns from input data [28, 29]. In the context of medical image segmentation, the choice of activation function can significantly influence model performance, convergence speed, and the ability to mitigate issues like the vanishing gradient problem [30, 31].

**Table 1** shows the comparative analysis for various activation functions used in deep learning such as ReLU, Leaky ReLU, Swish, Mish, ELiSH, HardELiSH, Softsign, and Tanh, and these are among the most widely used in neural network models. The table highlights their individual strengths, such as computational efficiency (ReLU), smooth gradient flow (Swish and Mish), and ability to avoid vanishing gradients (HardELiSH). Limitations like dead neuron problems (ReLU) or computational cost (Mish) are also noted. The table further outlines their contributions to model performance in terms of convergence speed, stability, and accuracy.

These functions are critical for shaping the learning dynamics and overall performance of neural networks, influencing the convergence speed, gradient flow, and ultimately the model's accuracy.

Mushtaq Salih *et al.* (2019) conducted a comparative study on activation functions and concluded that the HardELiSH activation function outperformed ReLU, particularly in addressing the vanishing gradient problem. Their findings demonstrated that HardELiSH not only mitigated this issue more effectively than ReLU but also led to an overall improvement in detection accuracy [37].

These advanced activation functions enable deep learning models to overcome the limitations of traditional functions, such as Sigmoid or Tanh, which suffer from the vanishing gradient problem in deeper networks. In the context of brain tumour segmentation, they provide essential benefits:

- **Improved Gradient Flow:** By mitigating the vanishing gradient problem, advanced functions ensure that deeper layers in models like 3D U-Net continue learning effectively, resulting in better segmentation performance.
- **Enhanced Feature Extraction:** These functions allow for more nuanced feature mapping, critical for detecting tumour boundaries and distinguishing different tissue types in MRI data.

**Table 1.** Comparative analysis of activation functions: strengths and limitations in deep learning for brain tumor segmentation.

Activation Function	Strengths	Limitations	Contribution to Brain Tumor Segmentation
ReLU [28]	Simple and fast; effective in shallow networks	Dying ReLU problem (zero gradient); ignores negative input	Used as a baseline in early U-Net models; limited capacity in handling complex tumor boundaries
Leaky ReLU [32]	Allows gradient for negative inputs; fixes dying ReLU	Still linear in nature; slight performance gain only; costlier to compute	Improves segmentation robustness over ReLU; better handling of low-intensity tumor regions
Swish [30]	Smooth and non-monotonic; promotes better generalization	May slow training in resource-limited settings	Enhances model expressiveness; improves Dice score by capturing complex tumor patterns
ELiSH [33]	Strong non-linearity; effective in both positive and negative ranges	Computationally complex; slower convergence	Provides better boundary delineation; improves gradient flow in deep U-Nets
HardELiSH [33]	Combines Swish & ELU benefits; fast; excellent gradient flow	Relatively new; needs fine-tuning	Shows superior segmentation accuracy, mitigates vanishing gradient in deep models
Mish [34]	Smooth & non-monotonic; good generalization; outperforms Swish in some tasks	High computational cost; may be unstable in very deep networks	Demonstrates high segmentation accuracy, effective in capturing fine tumor details
Softsign [35]	Smooth & continuous; bounded output $[-1, 1]$ ; simple math	Vanishing gradient; limited dynamic range	Rarely used in modern models; limited impact on tumor segmentation performance
Tanh [36]	Bounded output $[-1, 1]$ ; smooth & differentiable	Severe vanishing gradient; saturates quickly; slow convergence	Historically used in early models; largely replaced by advanced activations

- **Computational Efficiency:** Functions like ReLU and HardELiSH offer efficiency, enabling models to process high-dimensional MRI data faster, which is crucial for clinical applications.

Incorporating these advanced activation functions in brain tumour segmentation models can lead to more accurate, robust, and clinically viable results, making them a vital component in modern medical imaging techniques. Continued research into novel activation functions remains crucial for further improving the performance and efficiency of deep learning models in medical image analysis, Mushtaq *et al.* (2025) [38] provided a detailed review of various activation functions, including their mathematical formulations and graphical representations, offering clear insights into their operational behaviors.

## 2.4. Transformer-Based Models

The advent of Vision Transformers (ViTs) has revolutionized medical imaging by overcoming the limitations of traditional CNNs, particularly in capturing long-range dependencies and global context. Swin UNETR [10] pioneered the use of hierarchical Swin Transformers for 3D medical segmentation, enabling more precise delineation of complex anatomical structures through shifted window-based self-attention. TransBTS [39] further bridged the gap between CNNs and Transformers, synergizing the local feature extraction strength of convolutional networks with the global contextual reasoning of Transformers, leading to robust performance in tumor and lesion segmentation. Meanwhile, the Medical Transformer (MedT) [40] introduced a breakthrough with gated axial attention, significantly improving computational efficiency while processing high-resolution MRI scans, making it feasible to handle large volumetric data without compromising accuracy. These innovations underscore the transformative potential of Transformer-based models in medical imaging, paving the way for more interpretable, scalable, and high-performance AI-driven diagnostic systems.

### Swin 3D U-Net

The Swin 3D U-Net is an advanced deep learning architecture that integrates the hierarchical Swin Transformer with the 3D U-Net framework to enhance volumetric medical image segmentation [41]. Unlike traditional U-Nets that rely solely on convolutional operations, this hybrid model leverages the self-attention mechanism of Swin Transformers to capture long-range dependencies in 3D medical scans (e.g., MRI, CT) while maintaining the U-Net's ability to preserve spatial hierarchies through its encoder-decoder structure. The shifted window (Swin) mechanism improves computational efficiency by processing non-overlapping local windows in 3D space, reducing memory overhead compared to standard Vision Transformers (ViTs). By combining shifted window-based self-attention with 3D convolutions, the Swin 3D U-Net achieves superior performance in tasks like tumor segmentation, organ delineation, and multimodal image analysis, offering better scalability and accuracy than purely convolutional or transformer-based approaches. This architecture is particularly effective for high-resolution 3D datasets where global context and fine-grained localization are critical.

The Swin 3D U-Net architecture consists of an encoder, bottleneck, and decoder with specialized layers for hierarchical feature learning in volumetric medical images.

**Encoder:** Processes input patches ( $224 \times 224$ , patch size 4) using Swin Transformer blocks, maintaining feature resolution while progressively downsampling via patch merging layers ( $2 \times$  reduction in resolution,  $2 \times$  increase in dimension). This repeats three times.

**Bottleneck:** Two Swin Transformer blocks capture deep features without altering resolution or dimension.

**Decoder:** Uses patch expanding layers to upsample features ( $2 \times$  resolution increase, halving dimension) and skip connections to fuse multi-scale encoder features, preserving spatial details [42].

## 2.5. Generative Models

Generative approaches have significantly expanded the possibilities of tumor analysis by leveraging advanced deep learning techniques to improve accuracy and realism in medical imaging. SegAN [43] introduced an adversarial training framework to generate more realistic and precise segmentation masks, enhancing the delineation of tumor boundaries in MRI and CT scans. Meanwhile, CycleGAN [44] addressed the challenge of unpaired data by enabling image-to-segmentation translation without requiring exact correspondences between input and output images, thus facilitating domain adaptation in heterogeneous datasets. More recently, diffusion models like DDPM [45] have achieved state-of-the-art results in tumor segmentation and synthesis by employing iterative denoising processes, which enhance the model's ability to capture fine-grained details and reduce artifacts. These generative approaches not only improve diagnostic accuracy but also enable synthetic data generation for training robust models in scenarios where annotated medical data is scarce. Furthermore, their application extends to treatment planning, where realistic synthetic

images can aid in simulating tumor progression and therapeutic responses. As these methods continue to evolve, they hold great promise for advancing personalized medicine and improving clinical decision-making in oncology.

## 2.6. Weakly/Self-Supervised Learning

These methods addressed data scarcity challenges. Contrastive learning frameworks [46] are effective pre-training on unlabeled MRI datasets. Pseudo-labeling and scribble learning approaches [12] significantly reduced annotation requirements while maintaining competitive performance.

## 2.7. Federated Learning (FL)

Federated learning has emerged as a groundbreaking paradigm for collaborative model development while addressing data privacy concerns in healthcare. FedAvg (Federated Averaging) and FedBN (Federated Batch Normalization) [11] enable multiple medical institutions to jointly train segmentation models without sharing raw patient data. These approaches work by distributing model training across institutions and aggregating only the learned parameters, not the sensitive imaging data. This is particularly valuable for brain tumor segmentation, where datasets are often small and fragmented across hospitals. Recent implementations have demonstrated that FL can achieve comparable performance to centralized training while complying with strict medical data regulations like Health Insurance Portability and Accountability Act (HIPAA), and General Data Protection Regulation (GDPR). Advanced variants now incorporate differential privacy and secure multi-party computation to further enhance data protection during the federated training process.

## 2.8. Explainable AI (XAI) for Clinical Trust

The increasing complexity of deep learning models has necessitated the development of explainability techniques to facilitate clinical adoption. Grad-CAM (Gradient-weighted Class Activation Mapping) and SHAP (SHapley Additive exPlanations) [47, 48] provide intuitive visualizations that highlight which image regions most influenced the model's segmentation decisions. These methods help clinicians understand why a model classified certain areas as tumorous, enabling them to verify the algorithm's reasoning against their medical expertise. Recent advances in XAI for medical imaging now combine attention maps with uncertainty quantification, providing not only localization of important features but also confidence estimates in the predictions. This dual approach has been shown to improve radiologists' trust and diagnostic efficiency when working with AI-assisted segmentation systems.

## 2.9. Real-Time Models for Clinical Deployment

The translation of segmentation algorithms into clinical practice requires models that can operate efficiently on medical hardware. Lightweight architectures like MobileNet [49] and EfficientNet [50] have been specifically adapted for this purpose through techniques such as depthwise separable convolutions and neural architecture search. These optimizations enable near real-time tumor segmentation on standard hospital workstations and even mobile devices, with inference times often under one second per MRI slice. Recent work has focused on developing hybrid models that maintain this efficiency while incorporating 3D contextual information crucial for accurate tumor volume estimation. Such models are now being integrated into surgical navigation systems and intraoperative MRI suites, providing surgeons with continuously updated tumor delineations during procedures.

## 2.10. Multimodal Fusion for Comprehensive Analysis

Modern brain tumor characterization benefits immensely from combining information across multiple imaging modalities. Early fusion networks process concatenated MRI, PET, and DTI inputs through shared feature extractors, while late fusion approaches combine predictions from modality-specific networks [51,

52]. The latest architectures employ attention-based fusion mechanisms that dynamically weight the contribution of each modality based on contextual relevance. Advanced implementations now incorporate cross-modal contrastive learning during pretraining to better align feature spaces across modalities. This multi-modal approach has proven particularly valuable for distinguishing tumor recurrence from radiation necrosis and for precisely delineating infiltrative tumor margins that appear ambiguous in single-modality scans.

**Table 2** presents a comprehensive comparison of techniques used in brain tumor segmentation across various categories. Traditional machine learning methods like SVM and Random Forests are interpretable but limited by manual feature extraction. CNN architectures such as U-Net and 3D U-Net improve spatial understanding, though they require high computational resources. Advanced CNNs (e.g., Attention U-Net) introduce focus mechanisms but add complexity. Transformer-based models like Swin UNETR capture global context yet demand large datasets. Generative models enhance realism but suffer from training instability. Weak supervision reduces labeling costs at the expense of accuracy. Federated learning ensures privacy in multi-center settings but involves communication challenges. Explainable AI tools foster clinical trust, though their outputs are sometimes unreliable. Real-time models and multimodal fusion enable mobile deployment and cross-modality synergy, respectively, but face trade-offs in accuracy and data alignment.

**Table 2. Review summary of related works: techniques, strengths and limitations.**

Category	Key Paper	Technique	Strengths	Limitations
Traditional ML	Caulier <i>et al.</i> (2011) [53]	Haralick features, Gabor filters	Interpretable, works on small datasets	Manual feature engineering, poor generalization
	Zahra <i>et al.</i> (2021) [54]	Active Contours, Level Sets	Precise boundary detection	Sensitive to initialization
	Zhang <i>et al.</i> (2015) [55]	SVM with RBF kernel	Effective for binary classification	Struggles with multi-class tasks
	Geremia <i>et al.</i> (2013) [56]	Random Forests	Handles multi-class segmentation	Limited to extracted features
CNN Architectures	Ronneberger <i>et al.</i> (2015) [7]	U-Net	Skip connections preserve spatial details	2D version loses volumetric context
	Çiçek <i>et al.</i> (2016) [21]	3D U-Net	Volumetric processing	High memory requirements
	Milletari <i>et al.</i> (2016) [22]	V-Net	Residual connections improve gradient flow	Computationally intensive
Advanced CNNs	Oktay <i>et al.</i> (2018) [24]	Attention U-Net	Focuses on relevant regions	Additional parameters to train
	Hu <i>et al.</i> (2018) [25]	Squeeze and Excitation Blocks	Channel-wise feature recalibration	Minor computational overhead
Transformers	Hatamizadeh <i>et al.</i> (2021) [10]	Swin UNETR	Captures long-range dependencies	Requires large datasets
	Wang <i>et al.</i> (2021) [39]	TransBTS	Combines CNN + Transformer strengths	Complex architecture

Continued

Generative Models	Xue <i>et al.</i> (2018) [43]	SegAN	Produces realistic segmentations	Training instability
	Pinaya <i>et al.</i> (2022) [45]	DDPM	High-precision iterative refinement	Slow inference time
Weak Supervision	Zhou <i>et al.</i> (2018) [12]	Scribble learning	Reduces annotation burden	Lower accuracy than full supervision
Federated Learning	Li <i>et al.</i> (2021) [11]	FedAvg, FedBN	Privacy-preserving multi-center collaboration	Communication overhead
Explainable AI	Santos <i>et al.</i> (2024) [48]	Grad-CAM, SHAP	Increases clinical trust	Explanations sometimes unreliable
Real-Time Models	Howard <i>et al.</i> (2017) [49]	MobileNet	Mobile/edge device deployment	Reduced accuracy
Multimodal Fusion	Zhou <i>et al.</i> (2023) [52]	Attention fusion	Leverages complementary modality information	Requires co-registered data

### 3. EVALUATION METRICS IN BRAIN TUMOR SEGMENTATION

Tumor segmentation in medical imaging relies on robust evaluation metrics to assess model performance accurately. The most common metrics include the Dice Similarity Coefficient (DSC) [57], which measures overlap between predicted and ground truth segmentations, and the Hausdorff Distance (HD) [58], which evaluates boundary agreement. Sensitivity (Recall) and Specificity assess detection accuracy for tumor vs. non-tumor regions, while Precision minimizes false positives. The Jaccard Index (IoU) complements DSC by measuring intersection-over-union [57].

For clinical relevance, Volume Difference (VD) quantifies size discrepancies, and Accuracy provides overall pixel-wise correctness. Advanced metrics like Normalized Mutual Information (NMI) and Receiver Operating Characteristic (ROC) curves help evaluate probabilistic predictions. Recent challenges (e.g., BraTS), which are summarized from 2014 to 2023 [59], also emphasize Uncertainty Quantification to gauge model confidence. These metrics collectively ensure segmentation models meet diagnostic precision and reliability standards in oncology.

**Table 3** summarizes both standard and emerging evaluation metrics used in tumor segmentation. The Dice Similarity Coefficient (DSC) is the most widely adopted, capturing overall segmentation accuracy. Hausdorff Distance (HD) complements DSC by highlighting boundary mismatches, crucial in clinical edge detection. Sensitivity and specificity evaluate a model's ability to detect tumor tissue accurately and avoid false alarms, respectively. Precision is especially critical in high-risk diagnoses. Volumetric similarity assesses how well the predicted tumor size matches the actual, which is vital for planning treatments. The Jaccard Index (IoU) offers a stricter alternative to DSC, while Jaccard Distance provides a complementary view for identifying segmentation errors. These metrics collectively ensure robust and clinically meaningful model evaluation.

**Table 4** compares several deep learning architectures on the BraTS2021 validation dataset using key metrics such as parameter count, model size, and Dice scores across tumor subregions: Enhancing Tumor (ET), Tumor Core (TC), and Whole Tumor (WT).

**Table 3. Standard and emerging metrics for evaluating tumor segmentation.**

Metric	Description	Clinical Relevance
Dice Similarity Coefficient (DSC) [57]	Measures overlap between predicted and ground truth regions, $DSC = \frac{2\ A \cap B\ }{\ A\  + \ B\ }$	Most commonly used; reflects segmentation accuracy
Hausdorff Distance (HD) [58]	Measures the maximum distance between the boundary points of two sets	Sensitive to boundary errors
Sensitivity (Recall) [60]	True Positives/(True Positives + False Negatives)	Reflects ability to detect tumor pixels
Specificity [61]	True Negatives/(True Negatives + False Positives)	Reflects ability to avoid false tumor detection
Precision [62]	True Positives/(True Positives + False Positives)	Important in high-risk clinical decisions
Volumetric Similarity	Compares total segmented volume vs. actual tumor volume	Key in treatment planning and prognosis estimations
Jaccard Index (IoU) [57]	Ratio of intersection to union: $IoU = \frac{\ A \cap B\ }{\ A \cup B\ }$	Similar to DSC but stricter (always $\leq$ DSC); widely used in detection tasks.
Jaccard Distance (Jarrad)	Complement of IoU: $Jarrad = 1 - IoU$	Quantifies dissimilarity; useful for error analysis.

**Table 4. Presents the parameter counts, memory footprint, and segmentation performance (Mean Dice, ET, TC, WT) of several state-of-the-art models. Statistical significance is assessed for each model in comparison to Swin-Unet3D using the Wilcoxon signed-rank test [44].**

Model Name	Params (M)	Param Size (MB)	Mean Dice	ET Dice	TC Dice	WT Dice
3D U-Net	7.9	15.834	0.825	0.825	0.844	0.900
V-Net	45.6	182.432	0.815	0.815	0.840	0.751
UnetR	102	204.899	0.842	0.842	0.853	0.905
TransBTS	33.0	65.975	0.824	0.824	0.843	0.889
SwinBTS	35.7	71.394	0.828	0.828	0.843	0.896
Attention U-Net	23.6	47.257	0.841	0.841	0.851	0.870
Swin Pure Unet3D	33.6	67.163	0.817	0.817	0.822	0.885
Swin Unet3D	33.7	67.403	0.834	0.834	0.866	0.905

Swin-Unet3D achieved the highest Dice scores for TC (0.866) and WT (0.905), indicating superior segmentation accuracy, particularly for complete tumor regions. UnetR also showed strong performance across all metrics, especially with the highest mean Dice (0.842). Lightweight models like 3D U-Net had fewer parameters and smaller memory usage but performed moderately in Dice scores. V-Net had a relatively large parameter size but underperformed in WT segmentation (Dice: 0.751). Transformer-based models like TransBTS and SwinBTS demonstrated a good trade-off between performance and complexity. Statistical significance (Wilcoxon test) showed that Swin-Unet3D outperformed most models significantly in at least one of the tumor subregions.

Hatamizadeh *et al.* (2022) have introduced Swin UNETR, a novel transformer-based model for 3D brain tumor segmentation in multi-modal MRI, addressing the limitations of traditional FCNNs (e.g., U-Net) in capturing long-range dependencies due to their restricted kernel sizes. By reformulating segmentation as a sequence-to-sequence task, Swin UNETR leverages a hierarchical Swin transformer encoder to process input data as 1D embeddings, extracting multi-scale features through shifted-window self-attention, while a CNN-based decoder connected via skip connections refines the output. This hybrid architecture effectively combines the strengths of transformers (long-range modeling) and CNNs (local feature extraction), achieving state-of-the-art performance in the BraTS 2021 challenge and demonstrating the potential of transformers in medical image analysis [63].

Overall, Swin-Unet3D and UNETR appear to be leading models in terms of both accuracy and generalization across tumor structures, while maintaining reasonable parameter sizes.

**Table 5** demonstrates that Swin UNETR achieves the highest overall performance among the evaluated models, with an average Dice score of 0.913, outperforming nnU-Net, SegResNet, and TransBTS across all tumor regions (Enhancing Tumor, Whole Tumor, and Tumor Core). Both nnU-Net and SegResNet exhibit nearly identical performance, each with an average Dice score above 0.907, indicating their consistent and robust segmentation capabilities. TransBTS, while still competitive, trails behind the other models with a lower average Dice score of 0.891, suggesting reduced effectiveness in capturing fine-grained tumor structures. Overall, transformer-based Swin UNETR shows the most promising segmentation performance in this five-fold cross-validation setting.

**Table 5.** 5-Fold cross-validation (mean dice scores) [63].

Model	Dice Score			Avg. Dice
	ET	WT	TC	
Swin UNETR	0.891	0.933	0.917	0.913
nnU-Net	0.883	0.927	0.913	0.908
SegResNet	0.883	0.927	0.913	0.907
TransBTS	0.868	0.911	0.898	0.891

Wang *et al.* (2021) [39] have proposed TransBTS, which is a hybrid encoder-decoder architecture that integrates convolutional neural networks (CNNs) with transformer blocks. It leverages CNNs for local feature extraction and transformers for modeling long-range dependencies, enabling accurate brain tumor segmentation with improved contextual understanding.

**Table 6** shows that among the compared methods, TransBTS, particularly with test-time augmentation (TTA), achieved the highest Dice scores and the lowest Hausdorff distances across all tumor subregions on the BraTS 2020 validation set. This demonstrates its superior ability to segment tumors accurately while preserving boundary precision, outperforming traditional 3D U-Net and V-Net variants.

**Table 6.** Performance comparison on BraTS 2020 validation set [39].

Method	Dice Score (%)			Hausdorff Distance (mm)		
	ET	WT	TC	ET	WT	TC
3D U-Net [6]	68.76	84.11	79.06	50.98	13.37	13.61
Basic V-Net [11]	61.79	84.63	75.26	47.70	20.41	12.18
Deeper V-Net [11]	68.97	86.11	77.90	43.52	14.50	16.15
Residual 3D U-Net	71.63	82.46	76.47	37.42	12.34	13.11
TransBTS (w/o TTA)	78.50	89.00	81.36	16.72	6.47	10.47
TransBTS (w/ TTA)	78.73	90.09	81.73	17.95	4.96	9.77

## 4. DISCUSSION

Brain tumor segmentation has undergone remarkable advancements, transitioning from traditional machine learning approaches to sophisticated deep learning architectures and hybrid techniques. This discussion synthesizes the key developments, highlights their clinical and technical impacts, and identifies critical challenges that must be addressed to facilitate broader adoption in clinical practice.

### 4.1. The Transition from Traditional ML to Deep Learning: A Paradigm Shift

Traditional machine learning methods, such as SVM [64], Random Forests [65], and texture-based feature extraction [66], laid the foundation for automated brain tumor segmentation. These approaches were interpretable and computationally efficient, but their reliance on handcrafted features limited their ability to generalize across diverse datasets [5]. The introduction of deep learning, particularly U-Net [7] and its 3D variants [21] marked a turning point by enabling end-to-end learning of hierarchical features directly from imaging data. This shift significantly improved segmentation accuracy, with Dice scores rising from ~70% (traditional ML) to over 85% (DL) on BraTS benchmarks [9]. However, early CNNs faced challenges in handling multi-modal MRI inconsistencies and required large annotated datasets for training, a limitation partially addressed by data augmentation and transfer learning.

### 4.2. The Rise of Transformers and Hybrid Architectures

The introduction of Vision Transformers (ViTs) and their medical adaptations (e.g., Swin UNETR [63], TransBTS [39]) addressed CNNs' inability to model long-range spatial dependencies [10]. These models excel in capturing global context, making them particularly effective for segmenting diffuse tumor margins in glioblastoma [67]. Hybrid architectures, such as CNN-Transformer ensembles, further bridged the gap between local feature extraction and global reasoning, achieving Dice scores > 90% on BraTS 2021 [8]. Despite these advances, Transformers demand substantial computational resources and training data, limiting their accessibility for smaller institutions.

### 4.3. Generative Models and Weak Supervision: Mitigating Data Scarcity

Generative approaches, including GANs [43] and diffusion models [45], have enhanced segmentation by generating synthetic training data or refining predictions through iterative denoising. Weakly supervised methods [12] reduced reliance on pixel-level annotations by leveraging scribbles or bounding boxes, making AI more feasible for rare tumor subtypes. However, GANs suffer from training instability, while diffusion models are computationally expensive for real-time applications.

## Effectiveness of Generative Models and Weakly Supervised Learning

Manual annotation of 3D brain tumor volumes is time-consuming and requires domain expertise, limiting the scalability of supervised learning. Generative models such as GANs and VAEs have been successfully employed to synthesize realistic tumor images and augment training data, improving model generalization and robustness. For instance, Mok & Chung (2018) [68] introduced a coarse-to-fine GAN-based augmentation strategy (CB-GAN) that improved segmentation performance by 3.5% Dice score over traditional augmentation on the BraTS15 dataset. Their two-stage GAN was effective in generating anatomically plausible tumors with refined boundaries, reducing dependence on manual labeling.

In parallel, weakly supervised learning offers a practical alternative by utilizing partial annotations such as image-level labels or scribbles. Mlynarski *et al.* (2019) [69] demonstrated that combining a small number of fully annotated MRI scans (e.g., 5 or 15) with a larger set of weakly labeled data could achieve up to 78.3% Dice score, closely matching models trained on extensive full supervision. This underscores the potential of mixed supervision in minimizing annotation costs while maintaining segmentation accuracy.

Together, these strategies offer scalable and efficient training solutions, particularly valuable for clinical environments with limited annotation resources.

## 4.4. Federated Learning and Explainability: Toward Clinical Translation

Federated learning [11] has emerged as a privacy-preserving solution for multi-institutional collaborations, crucial for rare tumor types. This innovation is especially vital for brain tumor segmentation, where access to large, diverse, and annotated datasets is limited, particularly for rare subtypes such as pediatric gliomas or atypical meningiomas. By harnessing the statistical power of geographically distributed data, FL mitigates bias associated with single-institution models and fosters the development of robust, generalizable algorithms applicable across a wide range of patient populations and imaging protocols [70].

## Performance Variation Across Brain Tumor Subtypes

Deep learning models, particularly those trained on standard datasets like BraTS, have demonstrated strong performance in segmenting high-grade gliomas (HGGs). However, their effectiveness varies significantly when applied to different brain tumor subtypes, such as pediatric gliomas and atypical meningiomas, which often exhibit distinct morphological and radiographic characteristics.

Pediatric gliomas often present with distinct imaging characteristics, such as more diffuse growth patterns and lower contrast on MRI compared to adult high-grade gliomas (HGGs). These differences contribute to challenges in segmentation, including decreased Dice scores and higher boundary uncertainty. Liu *et al.* (2023) [71] investigated this issue and reported that models trained on adult glioma datasets exhibit a Dice score reduction of over 10% when applied to pediatric brain tumor cases, underscoring a significant domain shift. Similarly, atypical meningiomas, which are less frequent and show considerable heterogeneity, present additional challenges for model generalization due to limited annotated data and intra-class variability. These findings emphasize the need for subtype-specific training approaches, domain adaptation methods, and more diverse datasets to enhance the accuracy and generalizability of brain tumor segmentation models across different tumor types.

In parallel, the push for explainable AI (XAI) has become indispensable for clinical adoption. Tools such as Gradient-weighted Class Activation Mapping (Grad-CAM), SHAP (SHapley Additive exPlanations), and Integrated Gradients allow clinicians to peer inside the “black box” of deep neural networks, offering intuitive visualizations of which regions in the MRI contribute most to the model’s prediction [47, 48]. These methods have increased clinician trust and enabled model auditing, especially in high-stakes applications like preoperative planning and radiotherapy targeting. However, FL introduces communication overhead, and XAI methods sometimes produce unreliable explanations for complex models.

## 4.5. Real-Time and Multimodal Systems: The Next Frontier

Lightweight deep learning models such as MobileNet [49], EfficientNet [50], and their derivatives are

playing a pivotal role in enabling real-time intraoperative segmentation of brain tumors. These architectures, optimized for computational efficiency and low-latency inference, allow deployment on edge devices and integration into surgical navigation systems, thereby assisting neurosurgeons in achieving maximal tumor resection while preserving healthy tissue [49, 50]. For example, MobileNetV3 and EfficientNet-Lite have demonstrated promising accuracy-speed trade-offs when applied to real-time segmentation tasks on resource-constrained hardware like portable workstations or intraoperative imaging consoles [72].

Concurrently, the integration of multimodal imaging, such as combining structural MRI with functional modalities like Positron Emission Tomography (PET), Diffusion Tensor Imaging (DTI), and MR spectroscopy, has emerged as a powerful strategy for comprehensive tumor characterization [73]. These multimodal approaches leverage complementary features: while MRI offers high-resolution anatomical details, PET reveals metabolic activity, and DTI provides insight into white matter tract integrity.

Deep learning models equipped with multimodal fusion layers, such as attention-guided or cross-modal transformers, are showing promise in exploiting these heterogeneous data streams to improve segmentation precision and tumor sub-region delineation [74]. However, a major challenge lies in the harmonization of imaging protocols across different scanners, vendors, and institutions. Variability in acquisition parameters, field strengths, and contrast agent usage can significantly degrade model generalizability. To address this, techniques such as domain adaptation, intensity normalization, and synthetic modality generation using GANs are actively being explored [75]. Moreover, federated learning paradigms are gaining traction as a means to train models collaboratively across institutions without sharing raw patient data, thereby enhancing data diversity while preserving privacy [70].

#### 4.6. Key Challenges and Future Directions

Despite significant advancements in brain tumor segmentation, several critical challenges persist that hinder widespread clinical adoption. A primary obstacle is the inherent data heterogeneity stemming from variations in MRI scanner protocols and acquisition parameters across institutions, which compromises model generalizability [3]. Additionally, class imbalance remains a persistent issue, as certain tumor subregions like necrotic cores are often underrepresented in training datasets, leading to biased predictions. Perhaps most crucially, the majority of segmentation models have not undergone rigorous prospective clinical validation, creating a translational gap between algorithmic performance and real-world clinical utility. These limitations underscore the need for more robust and standardized approaches to ensure reliable deployment in healthcare settings.

Looking ahead, the field must prioritize several key directions to address these challenges:

- First, expanding standardized benchmarks like BraTS to include underrepresented tumor types, such as pediatric and diffuse midline gliomas, would enhance model versatility.
- Second, developing efficient models through techniques like neural architecture search and quantization will be essential for real-time deployment on edge devices in surgical settings.
- Third, advancing robust explainable AI (XAI) methods will be critical for regulatory approval and clinician trust.
- Finally, multimodal federated learning approaches that can harmonize disparate data sources while preserving patient privacy represent a promising avenue for creating more comprehensive and generalizable models.

These strategic focuses will be instrumental in bridging the current gaps between technical innovation and clinical application in neuro-oncology.

The evolution of brain tumor segmentation techniques has brought forth significant strengths that have transformed neuro-oncology research and clinical practice. Deep learning approaches, particularly U-Net and its variants, have demonstrated remarkable capabilities in automatically extracting hierarchical features from multi-modal MRI data, achieving unprecedented segmentation accuracy with Dice scores exceeding 90% in some cases [9]. The advent of transformer-based architectures has further enhanced performance by effectively capturing long-range spatial dependencies in volumetric scans, while generative models have

shown promise in addressing data scarcity through synthetic data generation [10, 45]. These technical advancements have been complemented by the development of privacy-preserving federated learning frameworks and explainable AI techniques, which facilitate multi-institutional collaboration and improve clinical interpretability [20, 48]. However, these approaches are not without limitations that require careful consideration.

- The computational complexity of advanced architectures, particularly 3D CNNs and transformers, poses significant challenges for clinical deployment due to their high memory requirements and inference times.
- Data heterogeneity across imaging protocols and scanners remains a persistent obstacle to model generalizability, while class imbalance in tumor subregions continues to affect segmentation accuracy [3].
- Furthermore, the black-box nature of many deep learning models and the lack of standardized clinical validation protocols hinder their widespread adoption in healthcare settings.
- Addressing these limitations requires a multi-faceted approach:
- Computational challenges can be mitigated through model compression techniques such as quantization and knowledge distillation, which can reduce model size without significant performance degradation [49].
- Data heterogeneity could be overcome by developing advanced normalization techniques and domain adaptation methods that account for scanner-specific variations.
- To tackle class imbalance, innovative loss functions like focal loss and tailored data augmentation strategies could be employed to better represent rare tumor subregions.
- The interpretability gap might be bridged by integrating attention mechanisms with clinically meaningful feature visualizations, while prospective multicenter trials could establish standardized validation protocols [48]. Federated learning frameworks, combined with synthetic data generation, offer a promising solution to data scarcity while maintaining patient privacy.

The path forward for brain tumor segmentation lies in developing more efficient, interpretable, and clinically validated models that can seamlessly integrate into existing healthcare workflows. Future research should focus on creating adaptive systems that can learn from limited annotations, generalize across diverse imaging protocols, and provide clinically actionable insights with measurable confidence intervals. By addressing these challenges through collaborative efforts between computer scientists and clinicians, the next generation of segmentation tools could significantly improve diagnostic accuracy, treatment planning, and patient outcomes in neuro-oncology.

## 5. CONCLUSIONS

Brain tumor segmentation has undergone a profound evolution from classical machine learning with handcrafted features to modern deep learning frameworks that incorporate transformers, generative modeling, and federated learning. These advancements have significantly improved segmentation accuracy, interpretability, and real-time clinical applicability. U-Net and its variants enabled volumetric analysis, while transformers like Swin UNETR brought long-range context modeling. Federated learning has enhanced data privacy and cross-institutional training, while explainable AI techniques have improved clinician trust. However, challenges such as data heterogeneity, class imbalance, and computational demands still hinder real-world deployment. Future research should prioritize lightweight models, robust validation protocols, and clinically meaningful, multimodal systems that integrate seamlessly into neuro-oncology workflows. Bridging the gap between algorithmic performance and clinical reliability remains essential for the successful translation of AI into routine medical practice.

The journey toward fully automated, clinically validated brain tumor segmentation systems is ongoing, but the remarkable progress to date offers compelling evidence of AI's potential to transform neuro-oncology. By continuing to bridge the gap between technical innovation and clinical needs, researchers can deliver tools that not only achieve high performance on benchmarks but also provide tangible benefits to patients and healthcare providers worldwide. The future of brain tumor segmentation lies in creating adaptable,

transparent, and clinically relevant solutions that can keep pace with the evolving landscape of precision medicine.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this paper.

## REFERENCES

1. Hanif, F., Muzaffar, K., Perveen, K., Malhi, S.M. and Simjee, S.U. (2017) Glioblastoma Multiforme: A Review of its Epidemiology and Pathogenesis through Clinical Presentation and Treatment. *Asian Pacific Journal of Cancer Prevention*, **18**, 3-9.
2. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., *et al.* (2021) The 2021 WHO Classification of Tumors of the Central Nervous System: A Summary. *Neuro-Oncology*, **23**, 1231-1251. <https://doi.org/10.1093/neuonc/noab106>
3. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., *et al.* (2015) The Multimodal Brain Tumor Image Segmentation Benchmark (Brats). *IEEE Transactions on Medical Imaging*, **34**, 1993-2024. <https://doi.org/10.1109/tmi.2014.2377694>
4. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M. and Maier-Hein, K.H. (2018) Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge. In: Crimi, A., Bakas, S., Kuijff, H., Menze, B. and Reyes, M., Eds., *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2017*, Springer, 287-297. [https://doi.org/10.1007/978-3-319-75238-9\\_25](https://doi.org/10.1007/978-3-319-75238-9_25)
5. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., *et al.* (2017) Advancing the Cancer Genome Atlas Glioma MRI Collections with Expert Segmentation Labels and Radiomic Features. *Scientific Data*, **4**, Article No. 170117. <https://doi.org/10.1038/sdata.2017.117>
6. Myronenko, A. (2019) 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. In: Crimi, A., Bakas, S., Kuijff, H., Keyvan, F., Reyes, M. and van Walsum, T., Eds., *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer, 311-320. [https://doi.org/10.1007/978-3-030-11726-9\\_28](https://doi.org/10.1007/978-3-030-11726-9_28)
7. Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Springer, 234-241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
8. Baid, U., *et al.* (2021) The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification. arXiv: 2107.02314v2. <https://arxiv.org/pdf/2107.02314>
9. Isensee, F., Jäger, P.F., Full, P.M., Vollmuth, P. and Maier-Hein, K.H. (2021) Nnu-Net for Brain Tumor Segmentation. In: Crimi, A. and Bakas, S., Eds., *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer, 118-132. [https://doi.org/10.1007/978-3-030-72087-2\\_11](https://doi.org/10.1007/978-3-030-72087-2_11)
10. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., *et al.* (2022) UNETR: Transformers for 3D Medical Image Segmentation. 2022 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 3-8 January 2022, 1748-1758. <https://doi.org/10.1109/wacv51458.2022.00181>
11. Li, X., Jiang, M., Zhang, X., Kamp, M. and Dou, Q. (2021) FedBN: Federated Learning on Non-IID Features via Local Batch Normalization. arXiv: 2102.07623v2. <https://arxiv.org/pdf/2102.07623>
12. Zhou, Z. (2017) A Brief Introduction to Weakly Supervised Learning. *National Science Review*, **5**, 44-53. <https://doi.org/10.1093/nsr/nwx106>
13. Liu, L., Fieguth, P., Pietikainen, M. and Lao, S. (2015) Median Robust Extended Local Binary Pattern for Texture Classification. 2015 *IEEE International Conference on Image Processing (ICIP)*, Quebec City, 27-30 September

2015, 2319-2323. <https://doi.org/10.1109/icip.2015.7351216>

14. Almijalli, M., Almusayib, F.A., Albugami, G.F., Aloqalaa, Z., Altwijri, O. and Saad, A.S. (2025) Automatic Active Contour Algorithm for Detecting Early Brain Tumors in Comparison with AI Detection. *Processes*, **13**, Article 867. <https://doi.org/10.3390/pr13030867>
15. Verma, O.P., Hanmandlu, M., Susan, S., Kulkarni, M. and Jain, P.K. (2011) A Simple Single Seeded Region Growing Algorithm for Color Image Segmentation Using Adaptive Thresholding. 2011 *International Conference on Communication Systems and Network Technologies*, Katra, 3-5 June 2011, 500-503. <https://doi.org/10.1109/csnt.2011.107>
16. Zhang, X., Zhao, J. and Lecun, Y. (2015) Character-Level Convolutional Networks for Text Classification. arXiv: 1509.01626v3. <https://arxiv.org/pdf/1509.01626>
17. Katuwal, R. and Suganthan, P.N. (2018) Enhancing Multi-Class Classification of Random Forest using Random Vector Functional Neural Network and Oblique Decision Surfaces. 2018 *International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, 8-13 July 2018, 1-8
18. Raja, S.G. and Nirmala, K. (2020) Detection of Brain Tumor Using K-Nearest Neighbor (KNN) Based Classification Model and Self Organizing Map (SOM) Algorithm. *International Journal of Neural Networks and Advanced Applications*, **7**, 42-48. <https://doi.org/10.46300/91016.2020.7.6>
19. Bhardwaj, M., Khan, N.U., Baghel, V., Vishwakarma, S.K. and Bashar, A. (2023) Brain Tumor Image Segmentation Using K-Means and Fuzzy C-Means Clustering. In: Rajput, S.S., *et al.*, Eds., *Digital Image Enhancement and Reconstruction*, Elsevier, 293-316. <https://doi.org/10.1016/b978-0-32-398370-9.00020-2>
20. Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. <https://doi.org/10.1109/tpami.2016.2644615>
21. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T. and Ronneberger, O. (2016) 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G. and Wells, W., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016*, Springer, 424-432. [https://doi.org/10.1007/978-3-319-46723-8\\_49](https://doi.org/10.1007/978-3-319-46723-8_49)
22. Milletari, F., Navab, N. and Ahmadi, S. (2016) V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 *Fourth International Conference on 3D Vision (3DV)*, Stanford, 25-28 October 2016, 565-571. <https://doi.org/10.1109/3dv.2016.79>
23. Kamnitsas, K., Ferrante, E., Parisot, S., Ledig, C., Nori, A.V., Criminisi, A., *et al.* (2016) Deepmedic for Brain Tumor Segmentation. In: Crimi, A., Menze, B., Maier, O., Reyes, M., Winzeck, S. and Handels, H., Eds., *BrainLesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer, 138-149. [https://doi.org/10.1007/978-3-319-55524-9\\_14](https://doi.org/10.1007/978-3-319-55524-9_14)
24. Oktay, O., *et al.* (2018) Attention U-Net: Learning Where to Look for the Pancreas. arXiv: 1804.03999v3. <https://arxiv.org/pdf/1804.03999>
25. Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. <https://doi.org/10.1109/cvpr.2018.00745>
26. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C. and Heng, P. (2018) H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes. *IEEE Transactions on Medical Imaging*, **37**, 2663-2674. <https://doi.org/10.1109/tmi.2018.2845918>
27. Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J. (2017) Pyramid Scene Parsing Network. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 6230-6239.

<https://doi.org/10.1109/cvpr.2017.660>

28. Fred Agarap, A.M. (2018) Deep Learning Using Rectified Linear Units (ReLU). arXiv: 1803.08375v2. <https://arxiv.org/abs/1803.08375v2>
29. Lederer, J. (2021) Activation Functions in Artificial Neural Networks: A Systematic Overview. arXiv: 2101.09957v1. <https://arxiv.org/abs/2101.09957v1>
30. Ramachandran, P., Zoph, B. and Le Google Brain, Q.V. (2017) Searching for Activation Functions. arXiv: 1710.05941v2. <https://arxiv.org/abs/1710.05941v2>
31. Pratiwi, H., Windarto, A.P., Susliansyah, S., Aria, R.R., Susilowati, S., Rahayu, L.K., *et al.* (2020) Sigmoid Activation Function in Selecting the Best Model of Artificial Neural Networks. *Journal of Physics: Conference Series*, **1471**, Article ID: 012010. <https://doi.org/10.1088/1742-6596/1471/1/012010>
32. Xu, J., Li, Z., Du, B., Zhang, M. and Liu, J. (2020) Reluplex Made More Practical: Leaky ReLU. 2020 *IEEE Symposium on Computers and Communications (ISCC)*, Rennes, 7-10 July 2020, 1-7. <https://doi.org/10.1109/iscc50000.2020.9219587>
33. Basirat, M. and Roth, P.M. (2018) The Quest for the Golden Activation Function. arXiv: 1808.00783v1. <http://arxiv.org/abs/1808.00783v1>
34. Misra, D. (2020) Mish: A Self Regularized Non-Monotonic Activation Function. <https://github.com/digantamisra98/Mish>
35. Szandała, T. (2020) Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks. arXiv: 2010.09458.
36. Shen, S., Zhang, N., Zhou, A. and Yin, Z. (2022) Enhancement of Neural Networks with an Alternative Activation Function Tanhlu. *Expert Systems with Applications*, **199**, Article ID: 117181. <https://doi.org/10.1016/j.eswa.2022.117181>
37. Salih, M.M., Salih, M.E. and Ahmed, M.A.A. (2019) Enhancement of U-Net Performance in MRI Brain Tumour Segmentation Using HardELISH Activation Function. 2019 *International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, Khartoum, 21-23 September 2019, 1-5. <https://doi.org/10.1109/icceee46830.2019.9071235>
38. Saleh, M.M., Salih, M.E., Ahmed, M.A.A. and Hussein, A.M. (2025) From Traditional Methods to 3D U-Net: A Comprehensive Review of Brain Tumour Segmentation Techniques. *Journal of Biomedical Science and Engineering*, **18**, 1-32. <https://doi.org/10.4236/jbise.2025.181001>
39. Wang, W., Chen, C., Ding, M., Yu, H., Zha, S. and Li, J. (2021) TransBTS: Multimodal Brain Tumor Segmentation Using Transformer. In: de Bruijne, M., *et al.*, Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021*, Springer, 109-119. [https://doi.org/10.1007/978-3-030-87193-2\\_11](https://doi.org/10.1007/978-3-030-87193-2_11)
40. Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I. and Patel, V.M. (2021) Medical Transformer: Gated Axial-Attention for Medical Image Segmentation. In: de Bruijne, M., *et al.*, Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021*, Springer, 36-46. [https://doi.org/10.1007/978-3-030-87193-2\\_4](https://doi.org/10.1007/978-3-030-87193-2_4)
41. Cai, Y., Long, Y., Han, Z., Liu, M., Zheng, Y., Yang, W., *et al.* (2023) Swin Unet3D: A Three-Dimensional Medical Image Segmentation Network Combining Vision Transformer and Convolution. *BMC Medical Informatics and Decision Making*, **23**, Article No. 33. <https://doi.org/10.1186/s12911-023-02129-z>
42. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., *et al.* (2023) Swin-UNet: UNet-Like Pure Transformer for Medical Image Segmentation. In: Karlinsky, L., Michaeli, T. and Nishino, K., Eds., *Computer Vision—ECCV 2022 Workshops*, Springer, 205-218. [https://doi.org/10.1007/978-3-031-25066-8\\_9](https://doi.org/10.1007/978-3-031-25066-8_9)
43. Xue, Y., Xu, T., Zhang, H., Long, L.R. and Huang, X. (2018) Segan: Adversarial Network with Multi-Scale L1 Loss for Medical Image Segmentation. *Neuroinformatics*, **16**, 383-392. <https://doi.org/10.1007/s12021-018-9377-x>

44. Zhu, J., Park, T., Isola, P. and Efros, A.A. (2017) Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2242-2251. <https://doi.org/10.1109/iccv.2017.244>
45. Pinaya, W.H.L., Tudosiu, P., Dafflon, J., Da Costa, P.F., Fernandez, V., Nachev, P., *et al.* (2022) Brain Imaging Generation with Latent Diffusion Models. In: Mukhopadhyay, A., Oksuz, I., Engelhardt, S., Zhu, D. and Yuan, Y., Eds., *Deep Generative Models*, Springer, 117-126. [https://doi.org/10.1007/978-3-031-18576-2\\_12](https://doi.org/10.1007/978-3-031-18576-2_12)
46. Chen, T., Kornblith, S., Norouzi, M. and Hinton, G. (2020) A Simple Framework for Contrastive Learning of Visual Representations. <https://proceedings.mlr.press/v119/chen20j.html>
47. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017) Grad-Cam: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 618-626. <https://doi.org/10.1109/iccv.2017.74>
48. Santos, M.R., Guedes, A. and Sanchez-Gendriz, I. (2024) Shapley Additive Explanations (SHAP) for Efficient Feature Selection in Rolling Bearing Fault Diagnosis. *Machine Learning and Knowledge Extraction*, **6**, 316-341. <https://doi.org/10.3390/make6010016>
49. Howard, A.G., *et al.* (2017) MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv: 1704.04861v1. <https://arxiv.org/pdf/1704.04861>
50. Tan, M. and Le, Q.V. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv: 1905.11946. <https://arxiv.org/pdf/1905.11946>
51. Calhoun, V.D. and Sui, J. (2016) Multimodal Fusion of Brain Imaging Data: A Key to Finding the Missing Link(s) in Complex Mental Illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, **1**, 230-244. <https://doi.org/10.1016/j.bpsc.2015.12.005>
52. Zhou, Y., Guo, J., Sun, H., Song, B. and Yu, F.R. (2023) Attention-Guided Multi-Step Fusion: A Hierarchical Fusion Network for Multimodal Recommendation. *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Chinese Taipei, 23-27 July 2023, 1816-1820. <https://doi.org/10.1145/3539618.3591950>
53. Caulier, Y. and Stolz, C. (2011) Optimal Gabor Filters and Haralick Features for the Industrial Polarization Imaging. In: Gagalowicz, A. and Philips, W., Eds., *Computer Vision/Computer Graphics Collaboration Techniques. MIRAGE 2011*, Springer, 122-132. [https://doi.org/10.1007/978-3-642-24136-9\\_11](https://doi.org/10.1007/978-3-642-24136-9_11)
54. Shahvaran, Z., Kazemi, K., Fouladivanda, M., Helfroush, M.S., Godefroy, O. and Aarabi, A. (2021) Morphological Active Contour Model for Automatic Brain Tumor Extraction from Multimodal Magnetic Resonance Images. *Journal of Neuroscience Methods*, **362**, Article ID: 109296. <https://doi.org/10.1016/j.jneumeth.2021.109296>
55. Zhang, X. and Song, Q. (2015) A Multi-Label Learning Based Kernel Automatic Recommendation Method for Support Vector Machine. *PLOS ONE*, **10**, e0120455. <https://doi.org/10.1371/journal.pone.0120455>
56. Geremia, E., Menze, B.H. and Ayache, N. (2013) Spatially Adaptive Random Forests. 2013 *IEEE 10th International Symposium on Biomedical Imaging*, San Francisco, 7-11 April 2013, 1344-1347. <https://doi.org/10.1109/isbi.2013.6556781>
57. Thada, V. and Jaglan, V. (2013) Comparison of Jaccard, Dice, Cosine Similarity Coefficient to Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm. *International Journal of Engineering and Innovative Technology*, **2**, 202-205.
58. Aydin, O.U., Taha, A.A., Hilbert, A., Khalil, A.A., Galinovic, I., Fiebach, J.B., *et al.* (2021) On the Usage of Average Hausdorff Distance for Segmentation Performance Assessment: Hidden Error When Used for Ranking. *European Radiology Experimental*, **5**, Article No. 4. <https://doi.org/10.1186/s41747-020-00200-2>
59. Saleh, M., Salih, M., Ahmed, M. and Hussein, A. (2025) From Traditional Methods to 3D U-Net: A Comprehen-

sive Review of Brain Tumour Segmentation Techniques. *Journal of Biomedical Science and Engineering*, **18**, 1-32. <https://www.scirp.org/journal/papercitationdetails?PaperID=140886&JournalID=30>

60. (2024) Sensitivity and Specificity. Wikipedia. [https://en.wikipedia.org/w/index.php?title=Sensitivity\\_and\\_specificity&oldid=1245547015](https://en.wikipedia.org/w/index.php?title=Sensitivity_and_specificity&oldid=1245547015)
61. 10.1 Sensitivity and Specificity—Foundations of Biomedical Science: Quantitative Literacy: Theory and Problems. <https://oercollective.caul.edu.au/foundations-of-biomedical-science/chapter/10-1-sensitivity-and-specificity/>
62. Precision and Recall. Wikipedia. [https://en.wikipedia.org/wiki/Precision\\_and\\_recall](https://en.wikipedia.org/wiki/Precision_and_recall)
63. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R. and Xu, D. (2022) Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. arXiv: 2201.01266. <https://monai.io/research/swin-unetr>
64. Ayachi, R. and Ben Amor, N. (2009) Brain Tumor Segmentation Using Support Vector Machines. In: Sossai, C. and Chemello, G., Eds., *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Springer, 736-747. [https://doi.org/10.1007/978-3-642-02906-6\\_63](https://doi.org/10.1007/978-3-642-02906-6_63)
65. Soltaninejad, M., Zhang, L., Lambrou, T., Yang, G., Allinson, N. and Ye, X. (2019) MRI Brain Tumor Segmentation using Random Forests and Fully Convolutional Networks. arXiv: 1909.06337.
66. McIntyre, L. and Tuba, E. (2023) Brain Tumor Segmentation and Classification Using Texture Features and Support Vector Machine. 2023 11th International Symposium on Digital Forensics and Security (ISDFS), Chattanooga, 11-12 May 2023, 1-5. <https://doi.org/10.1109/isdfs58141.2023.10131719>
67. Khodadadi Shoushtari, F., Sina, S. and Dehkordi, A.N.V. (2022) Automatic Segmentation of Glioblastoma Multiform Brain Tumor in MRI Images: Using Deeplabv3+ with Pre-Trained Resnet18 Weights. *Physica Medica*, **100**, 51-63. <https://doi.org/10.1016/j.ejmp.2022.06.007>
68. Mok, T.C.W. and Chung, A.C.S. (2019) Learning Data Augmentation for Brain Tumor Segmentation with Coarse-To-Fine Generative Adversarial Networks. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M. and van Walsum, T., Eds., *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer, 70-80. [https://doi.org/10.1007/978-3-030-11723-8\\_7](https://doi.org/10.1007/978-3-030-11723-8_7)
69. Mlynarski, P., Delingette, H., Criminisi, A. and Ayache, N. (2019) Deep Learning with Mixed Supervision for Brain Tumor Segmentation. *Journal of Medical Imaging*, **6**, Article ID: 034002. <https://doi.org/10.1117/1.jmi.6.3.034002>
70. Sheller, M.J., Edwards, B., Reina, G.A., Martin, J., Pati, S., Kotrotsou, A., *et al.* (2020) Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations without Sharing Patient Data. *Scientific Reports*, **10**, Article No. 12598. <https://doi.org/10.1038/s41598-020-69250-1>
71. Liu, X., Bonner, E.R., Jiang, Z., Roth, H., Packer, R., Bornhorst, M., *et al.* (2023) From Adult to Pediatric: Deep Learning-Based Automatic Segmentation of Rare Pediatric Brain Tumors. *Medical Imaging 2023: Computer-Aided Diagnosis*, **12465**, Article ID: 1246505. <https://doi.org/10.1117/12.2654245>
72. Pacal, I., Celik, O., Bayram, B. and Cunha, A. (2024) Enhancing Efficientnetv2 with Global and Efficient Channel Attention Mechanisms for Accurate MRI-Based Brain Tumor Classification. *Cluster Computing*, **27**, 11187-11212. <https://doi.org/10.1007/s10586-024-04532-1>
73. Zhu, Z., He, X., Qi, G., Li, Y., Cong, B. and Liu, Y. (2023) Brain Tumor Segmentation Based on the Fusion of Deep Semantics and Edge Information in Multimodal MRI. *Information Fusion*, **91**, 376-387. <https://doi.org/10.1016/j.inffus.2022.10.022>
74. Shi, J., *et al.* (2023) H-DenseFormer: An Efficient Hybrid Densely Connected Transformer for Multimodal Tumor Segmentation. arXiv: 2307.01486. <https://github.com/shijun18/H-DenseFormer>
75. Karani, N., Chaitanya, K., Baumgartner, C. and Konukoglu, E. (2018) A Lifelong Learning Approach to Brain MR

Segmentation across Scanners and Protocols. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C. and Fichtinger, G., Eds., *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018*, Springer, 476-484. [https://doi.org/10.1007/978-3-030-00928-1\\_54](https://doi.org/10.1007/978-3-030-00928-1_54)