The Most Important Models, Articles, Datasets and Challenges In Brain Tumor Segmentation

Halil İbrahim AKÇA[1]

[1*]Software Engineering, Mugla Sıtkı Kocman University, Turkey.

Contributing authors: halilibrahimakca@posta.mu.edu.tr;

Abstract

Brain tumor segmentation is a critical process that enables accurate identification of brain tumors in MRI scans and plays an important role in treatment planning and surgical interventions. Instead of traditional manual methods, deep learning (DL)-based automatic segmentation models, including U-Net and its variants, offer the potential to provide consistent and efficient results. U-Net-based models have been particularly successful in medical image segmentation due to their encoder-decoder architecture and ability to preserve spatial features through skip connections. Deep learning models such as Convolutional Neural Networks (CNNs) and Image Transformers (ViTs) have also been shown to be successful in accurately segmenting complex tumor structures. While the integration of multimodal MRI data increases segmentation accuracy, challenges such as missing modalities, limited labeled data, tumor heterogeneity, and class imbalance remain important problems to be solved. In addition, federated learning offers a more robust and privacy-preserving approach to brain tumor segmentation by enabling collaborative model training across institutions without data sharing.

Keywords: Brain Tumor Segmentation, MRI Image Analysis, Federated Learning for Medical Imaging

1 - Introduction

Brain tumor segmentation is a critical process that involves accurately determining the location, size, and shape of brain tumors in magnetic resonance imaging (MRI) scans.[17][19] This process is vital for treatment planning, disease monitoring, and surgical interventions. Traditionally, brain tumor segmentation was performed manually by radiologists; however, this method is time-consuming and subjective.[27]

In recent years, advancements in deep learning (DL) algorithms have shown that automated methods for brain tumor segmentation hold great potential. Deep learning-based segmentation models have the potential to reduce human intervention and provide consistent, efficient results.[23] Convolutional Neural Networks (CNNs) have achieved significant success in this field. Architectures such as UNet and Mask R-CNN have proven effective in accurately segmenting complex tumor structures. Additionally, Vision Transformer (ViT) models have successfully overcome the local feature-focused limitations of CNNs by using self-attention mechanisms and capturing global contextual information.[21][22] Hybrid models aim to combine the strengths of both CNNs and transformers, balancing local and global information to enhance segmentation accuracy. The use of multimodal MRI can significantly improve brain tumor segmentation accuracy. Different MRI modalities, such as T1-weighted, T2-weighted, contrast-enhanced T1-weighted, and FLAIR, provide complementary information about the tumor's location, shape, size, intensity, and contrast patterns.[20] The integration of these modalities enhances the comprehensive and accurate representation of the tumor, improving the performance of the segmentation model. However, there are still several key research challenges to address in deep learning-based brain tumor segmentation. Issues such as missing modalities,[24] limited labeled data,[25] tumor heterogeneity, class imbalance, and model efficiency are important areas of focus in this field. Class weighting strategies and specialized loss functions can be used to address these challenges. Additionally, federated learning has the potential to enhance performance in brain tumor segmentation by enabling collaborative training of deep learning models across multiple institutions without sharing data.[26] In conclusion, while deep learning-based brain tumor segmentation has made significant progress, research should continue to overcome existing challenges and develop more robust, accurate, and interpretable models for clinical applications. The integration of multimodal data, advancements in hybrid models, and the adoption of federated learning are important factors that will shape the future of brain tumor segmentation.

2 - Models And Methods

2.1 The Role of CNN-Based Models in Brain Tumor Segmentation

In brain tumor segmentation, deep learning-based models, particularly CNN-based methods, play a crucial role. Syazwany et al. [1] introduced the MM-BiFPN model, which uses the MICCAI 2018–2020 dataset.[29] This model processes MRI data from different modalities through separate encoders and then merges the information using late fusion with a Bidirectional Feature Pyramid Network (Bi-FPN).[28] This method significantly enhances segmentation accuracy by modeling cross-scale relationships, which is particularly beneficial in complex tasks like brain tumor segmentation. By integrating multimodal data, MM-BiFPN improves segmentation mask accuracy and highlights the effectiveness of late fusion and Bi-FPN in medical imaging tasks.[30] Similarly, Dingwen Zhang et al. [2] proposed a method to handle missing modalities in brain tumor segmentation through multimodal feature fusion. Their approach combines data from various MRI modalities, including T1, T1c, T2, and FLAIR, to improve segmentation accuracy. The method utilizes two fusion strategies: early fusion, which integrates data at the initial layers, and late fusion, which processes each modality separately before merging. An innovative cross-modal framework, using Cross-Modal Feature Transition (CMFT) and Cross-Modal Feature Fusion (CMFF), leverages GANs (Generative Adversarial Networks) to capture inter-modality patterns and improve segmentation performance. This method highlights the importance of multimodal fusion in achieving reliable tumor characterization. Furthermore, Wang et al. [3] developed an end-to-end modality matching learning approach to address missing modality challenges, validated on the BraTS 2020 dataset.[31] Built on a 3D U-Net framework, this method processes modalities in two groups: Group 1 (FLAIR, T2) emphasizes tumor structure and edema, while Group 2 (T1, T1ce) focuses on contrast-enhanced and necrotic regions. The Modality Matching Loss reduces discrepancies between the branches, enhancing consistency. Additional techniques such as model ensembling and post-processing further refine segmentation accuracy. The method achieved high Dice scores, ranking 2nd out of 78 teams, demonstrating the effectiveness of modality matching in improving MRI-based segmentation.

2.2 The Use of Image Transformer-Based Models in Brain Tumor Segmentation

Image Transformer (ViT) [32] models are leading a significant shift in image analysis architecture, particularly due to their self-attention mechanisms and ability to capture global contextual information.[33] These features enable ViT models to enhance segmentation accuracy and robustness, especially when combined with multimodal MRI data for brain tumor segmentation. Unlike traditional CNNs that focus on local features, ViT models utilize self-attention mechanisms to capture the relationships between different regions of an image, providing a more comprehensive understanding of the global context. This global perspective allows for better interpretation of complex tumor shapes and their interactions with surrounding tissues, which leads to more accurate segmentations.

One example of the application of ViT in medical imaging is the ViTBIS model, proposed by Sagar et al. [4]. This Vision Transformer (ViT) is specifically designed for biomedical image segmentation, where it divides input feature maps into three parts using 1x1, 3x3, and 5x5 convolutions in both the encoder and decoder. Trained on datasets such as the Synapse Multi-Organ Segmentation and Brain Tumor MRI Segmentation Datasets,[34] ViTBIS employs a multiscale architecture that captures features at multiple scales, combining fine details like tumor boundaries with broader contextual information such as organ positions. This approach significantly improves segmentation accuracy and robustness, especially for structures of varying sizes and shapes. By integrating multiscale convolution with attention mechanisms,[33] the ViTBIS model achieves more precise and reliable segmentations.

In addition to ViTBIS, Xing et al. (2022) [5] introduced the Nested Modality Awareness Transformer (NestedFormer) for brain tumor segmentation using multimodal MRIs. The NestedFormer model is designed to capture both intra- and inter-modality relationships through the Nested Modality-Aware Feature Aggregation

(NMaFA) module, which facilitates precise segmentation. Key components of NestedFormer include the Three-Way Spatial-Attentive Transformer (Ttsa) for capturing intra-modality correlations,[35] the Cross-Modality Attentive Transformer (Tcma) for modeling inter-modality relationships, and the Global Poolformer Encoder for efficient volumetric feature extraction. Additionally, Modality-sensitive gating (MSG) enhances feature reuse during decoding. The NestedFormer has shown superior performance over state-of-the-art methods, offering flexibility in fusing multiple modalities and enhanced computational efficiency. However, its fusion process could still be optimized for broader MRI applications.

Lastly, Liu et al. (2022c) [6] proposed SFusion, a self-attention-based fusion block designed to address the challenge of missing MRI modalities in brain tumor segmentation. Although not specifically developed for this task, SFusion serves as a versatile fusion strategy that can be integrated into various deep learning models. It includes a Correlation Extraction (CE) module,In CE Each layer includes a multi-head attention (MHA) block and a fully connected feed-forward network (FFN), [37] which uses self-attention layers to capture multi-modal correlations from the available modalities, and a Modal Attention (MA) module, which determines the contribution of each modality at each voxel for weighted fusion. SFusion is particularly effective in handling missing modalities, without the need for simulation or synthesis of the missing data. When evaluated on the BraTS2020 dataset, SFusion outperformed the Convolutional Gated Fusion (GFF) method, [36] especially excelling in tumor core segmentation. Its flexibility makes it suitable for various real-world applications, demonstrating accurate and reliable segmentation even in the presence of missing modalities.

2.3 Hybrid Models in Brain Tumor Segmentation

Hybrid models combine the advantages of CNNs and transformers, resulting in more comprehensive and effective segmentation techniques. These models balance the ability of CNNs to extract local features with the capacity of transformers to capture global contextual information, [38] allowing for a better understanding of complex tumor shapes and their interactions with surrounding tissues. This synergy between local and global processing leads to more accurate and reliable segmentation results.

One such hybrid model, TransBTS, was introduced by Wang et al. (2021a) [7] and combines 3D CNNs and transformers for multimodal MRI brain tumor segmentation. The framework uses 3D CNNs to capture local contextual information and a transformer to model global semantic correlations. Initially, the 3D CNNs generate compact feature maps that encode spatial and depth information, which are then processed by a transformer encoder with Multi-Head Attention (MHA) and Feed-Forward Network (FFN) layers to learn long-range dependencies.[37] Finally, a 3D CNN decoder upsamples the features to produce pixel-level segmentation outputs. This model has been evaluated on the BraTS 2019 and BraTS 2020 datasets, demonstrating the effectiveness of combining 3D CNNs and transformers for both local and global contextual modeling in brain tumor segmentation.

Another hybrid model, mmFormer, introduced by Zhang et al. (2022a) [8], is designed to handle missing modalities in medical imaging. mmFormer integrates a hybrid modality-specific encoder for capturing local and global contexts,[39] an inter-modality transformer for synchronizing long-range correlations, and a convolutional decoder for generating robust segmentation outputs. This model is particularly suited for brain tumor segmentation, addressing challenges like tumor heterogeneity, overlapping boundaries, and incomplete imaging data. The key components of mmFormer include CNN-transformer integration, hybrid encoders, and a modality-agnostic encoder for robustness. When evaluated on the BraTS 2018 dataset,[40] mmFormer demonstrated significant improvements, with a 19.07 increase in Dice scores in single-modality scenarios, showcasing the model's ability to handle missing modalities and produce accurate segmentations.

Lastly, a study by Jinglin Yuan [9] proposes a model that combines Mask RCNN with attention mechanisms and a ResNet architecture for MRI brain tumor segmentation. The model integrates channel (SENet) and spatial (CANet) [41] attention modules to focus on crucial regions, improving segmentation accuracy. The backbone network, ResNet50, is enhanced with attention modules and Feature Pyramid Networks (FPN) for effective feature extraction.[42] Additionally, the Region of Interest Alignment (RoIAlign) technique is used to refine the alignment of irregular regions identified by the Region Proposal Network (RPN). Evaluated on the BraTS2020 dataset, this model achieved high accuracy, effectively distinguishing tumor features under challenging conditions like low resolution, noise, and low contrast. Ablation experiments confirmed the significant impact of the attention mechanisms on performance. This method demonstrates strong potential for clinical applications, offering robust and precise tumor segmentation in MRI scans.

2.4 Federated Learning Techniques: Brain Tumor Segmentation while Protecting Patient Privacy

Federated Learning (FL) is a promising method for training deep learning models across multiple institutions while preserving patient data privacy. In medical imaging, where data sharing is often restricted by privacy regulations, FL trains local models at each institution and sends their parameters to a central server, which constructs a global model.[43] This process, repeated without centralizing data, ensures privacy. In brain tumor segmentation, FL is particularly valuable, as it enables the use of diverse institutional data, overcoming the challenge of limited labeled data and leading to more robust and generalizable deep learning models for tumor identification in MR images.

HDC-Net (Hierarchically Decoupled Convolutional Network), developed by Luo et al. (2021) [10], offers a lightweight and efficient approach for brain tumor segmentation by reducing computational complexity. This is achieved through the use of Hierarchically Decoupled Convolution (HDC) modules, which decompose convolutions across spatial and channel domains to efficiently capture multi-scale and multi-view spatial contexts. Key features include HDC modules that capture spatial and channel information, [44] enabling efficient segmentation of tumor subregions in a single pass. HDC-Net demonstrated state-of-the-art accuracy on the BraTS 2018 and 2017 datasets, offering superior performance with reduced computational requirements, making it highly practical for clinical use. Future directions for HDC-Net include better fusion of multimodal MRI data, incorporating uncertainty estimation for clinical reliability, and developing lightweight models for resource-constrained environments. Overall, HDC-Net has advanced brain tumor segmentation by improving both accuracy and efficiency, with the potential to enhance diagnosis and patient care.

Ranjbarzadeh et al. (2022) [11] proposed a deep learning-based approach for brain tumor segmentation using MRI multi-modality brain images from the BRATS 2018 dataset. Traditional methods, like manual segmentation, are time-consuming and subjective, while machine learning approaches face limitations in handling ambiguous tumor boundaries. In contrast, deep learning models, particularly fully convolutional networks (FCNs) [47] and U-Net derivatives, have significantly improved segmentation accuracy and efficiency. Ranjbarzadeh et al.'s method utilized four MRI modalities and incorporated preprocessing to remove irrelevant areas, a hierarchical CNN to capture both local and global features, and a distance-wise attention (DWA) mechanism to enhance accuracy based on tumor size and location. The approach demonstrated superior performance in tumor core and enhancing region segmentation, being computationally efficient. Future improvements include better multimodal data integration, uncertainty modeling, and optimization for resource-limited environments, advancing clinically viable solutions for brain tumor segmentation.

Ghaffari et al. (2020) [12] analyzed the development of automated brain tumor segmentation models using multi-modal MRI images, comparing various methods and evaluating models on the BraTS 2012–2018 challenges. Key advancements include the shift from Random Forests to CNN-based models,[45] with U-Net, and DenseNet architectures improving segmentation accuracy. The study highlights methods such as multi-scale analysis, ensemble learning, patch-based processing, and attention mechanisms to enhance model performance. Preprocessing techniques, including noise reduction and intensity normalization, and postprocessing with Conditional Random Fields (CRF),[46] were also discussed. The BraTS competition has driven significant improvements in segmentation accuracy, with top models achieving high Dice scores over the years, such as 0.91 for whole tumor and 0.82 for active tumor in BraTS 2018.

2.5 U-net Models in Brain Tumor Segmentation

TSEUnet is a neural network architecture developed for 3D brain tumor MRI segmentation [13], combining Transformer and SE-Attention mechanisms. Unlike traditional CNN-based models, which struggle to capture long-range dependencies and global contexts, TSEUnet integrates the strengths of Transformer and SE-Attention mechanisms to address these limitations.[47] The Transformer block, which operates in parallel with the encoder layers, utilizes multi-head self-attention and MLP modules to effectively extract both local features and long-range dependencies. Since convolutional blocks already encode local features and spatial information, TSEUnet does not require positional embedding. In the decoder, the SE-Attention mechanism highlights relevant features, enhancing focus on key areas like tumor regions while suppressing irrelevant information. Evaluated on the BraTS 2018 dataset, TSEUnet demonstrated improved segmentation accuracy by effectively capturing long-range dependencies and refining results through the SE-Attention mechanism.

BiTr-Unet, introduced by Jia et al. (2021) [14], combines CNN and Transformer architectures for brain tumor segmentation. Like TSEUnet, BiTr-Unet uses CNNs to extract local features and Transformers to capture long-range dependencies. The model includes a 3D CNN encoder-decoder for hierarchical feature extraction,[48] Transformer layers for self-attention modeling, and skip connections to facilitate information flow between the encoder and decoder, enhancing multi-scale feature integration. Key improvements over other models, such as TransBTS,[7] include two sets of Vision Transformer (ViT) layers, attention modules, and a deeper network structure, which lead to better feature extraction and segmentation performance. By combining CNN and Transformer strengths, BiTr-Unet captures both local and global contexts, achieving high accuracy and efficiency. Its design ensures strong generalization across datasets, positioning it as a powerful tool for brain tumor segmentation.

ST-Unet, introduced by Zhang et al. (2023) [15], addresses the limitations of CNNs in capturing long-range dependencies, which can lead to segmentation errors. By incorporating the Swin Transformer Encoder, ST-Unet extracts hierarchical visual features while modeling global dependencies. [50] The model also improves segmentation accuracy through the Cross-Layer Feature Enhancement (CLFE) Module, which replaces traditional merging methods with a more effective combination of multi-scale feature representations. Evaluated on datasets such as Synapse and ISIC 2018, ST-Unet outperforms state-of-the-art methods, demonstrating the effectiveness of combining CNNs and Transformers for enhanced medical image segmentation.

For volumetric segmentation, Vision Transformer (ViT) models have shown great promise, as demonstrated by Peiris et al. (2022a) [16]. Their VT-UNet architecture is designed for 3D tumor segmentation without relying on convolutional modules. It processes 3D data to capture inter-slice dependencies using Swin Transformer blocks in the encoder to model both local and global contexts. The decoder employs parallel attention mechanisms, including cross-attention and self-attention, to maintain global context and internal dependencies. A fusion module utilizing Fourier Positional Encoding (FPE) enhances feature conditioning.[52] VT-UNet outperforms existing methods in brain tumor segmentation on the BraTS dataset, offering high computational efficiency and robustness to data corruptions like motion and shading, making it ideal for large-scale 3D medical datasets.

DenseUNet+, introduced by Cetiner et al. (2023) [17], is designed for automatic brain tumor segmentation using multimodal MRI images. The model utilizes dense blocks and weighted nodes for robust feature integration and a preprocessing method that identifies regions of interest (ROI) across multiple MRI modalities.[53] Key features include dense blocks for multimodal information integration, weighted skip nodes for enhancing encoder-decoder connections, and add and res blocks for improved edge definition and feature extraction. DenseUNet+ achieves high segmentation accuracy, surpassing other models in performance, and offers gradient stability, reducing issues with vanishing gradients. While the model performs well in multimodal segmentation, it may have longer training times due to its increased complexity. Despite this, DenseUNet+ demonstrates strong potential for clinical decision-making tools and real-world applications, with opportunities for optimization and adaptation to various MRI use cases.

3 - Recent Challenges and Proposed Solutions in Deep Learning Methods for Brain Tumor Segmentation

1. Data Scarcity and Diversity

Challenge: The limited availability of labeled medical imaging data particularly hinders the success of deep learning models. Additionally, the heterogeneity in data collected from different MRI devices, scanning protocols, and patient groups makes it difficult for models to generalize. [12] In some studies, overfitting issues can arise due to the scarcity of training data.[11] Missing modalities, such as the absence of all MRI imaging types (T1, T1c, T2, FLAIR) in some patients, also pose a significant problem.[8]

Proposed Solutions:

- **Data Augmentation:** Data augmentation techniques such as rotation, cropping, and shifting are applied to artificially increase training data.[8] This approach helps models generalize better.[4]
- **Synthetic Data Generation:** Synthetic images similar to real data are created to compensate for data scarcity.[8]
- **Multimodal Fusion:** Methods that combine information from existing modalities are used to address missing modalities. Approaches like self-attention-based fusion blocks (SFusion) can work with a variable number of input data and reduce the impact of missing modalities.[6]

2. Complex Tumor Structures

Challenge: Brain tumors can exhibit inconsistent locations, blurred boundaries, and irregular shapes in MRI images.[1] Segmenting different tissues within the tumor (edema, necrosis, tumor core, enhancing tumor) is challenging due to their similar appearance. Detecting smaller tumors and different tumor parts can also be difficult. Moreover, the ambiguity of tumor boundaries and the presence of diffuse structures complicate segmentation.[11]

Proposed Solutions:

- **Advanced Convolutional Layers:** Techniques like atrous convolutions allow feature extraction over a wider area.[17]
- **Attention Mechanisms:** Attention mechanisms help the model focus on critical regions, enabling more accurate segmentation of challenging tumor parts. For example, distance-based attention (DWA) improves segmentation by considering the tumor's central position.[11]
- **Hybrid Models:** Combining CNNs and Transformer models allows for modeling both local details and the overall tumor structure, effectively addressing the tumor's complex nature.[11]

3. Improving Segmentation Performance

Challenge: It can be difficult to achieve satisfactory performance in segmentation metrics (e.g., Dice score, Hausdorff distance).[17] Imbalanced data distribution (e.g., the ratio of tumor regions to the entire image) can negatively affect segmentation accuracy. In some cases, models may under-segment parts of the tumor or over-segment areas outside the tumor.

Proposed Solutions:

- **Improved Loss Functions:** Using combinations of loss functions such as Dice loss, focal loss, and cross-entropy loss can enhance segmentation performance.[8]
- **Post-Processing Methods:** Post-processing techniques, such as cleaning small erroneous regions and correcting connected components, are applied to improve segmentation results.
- **Multiple Model Integration:** The outputs of different models are combined to achieve more accurate segmentation results.[7]
- **Feature Selection:** Attention mechanisms highlighting important features are used to improve tumor classification.
- **Validation Techniques:** Techniques such as cross-validation are used to assess model performance on different data subsets, leading to more generalized models.[8]

Model Type	Core Approach	Advantages	Disadvantages
CNN-Based	Captures local features.	Fast training, fewer parameters, good at capturing local details.	Lack of global context, limited spatial awareness.
ViT-Based	Uses attention mechanisms to model long-range relationships.	Global context capture, parallel processing, flexible architecture.	High computational cost, more data needed, lack of local detail.
Hybrid	Combines CNNs and Transformers to capture both local and global features.	Both local and global features, high accuracy, more robust.	More complex, higher computational cost.
Federated	Uses data from different institutions while protecting data privacy.	Data privacy, large data usage, better generalization.	Complex training process, communication cost, data heterogeneity.
UNet-Based	Uses encoder-decoder architecture and skip connections for medical image segmentation.	Effective segmentation, simple and widespread, adaptability.	Lack of global context, lack of local detail.

Fig. 1 The advantages and disadvantages of the models..

Model	Dataset	WT Dice Score	TC Dice Score	ET Dice Score	Publication
CNN Based Models					
MM-BiFPN	BraTS 2018 BraTS 2020	81.07% 83.58%	77.73% 81.47%	73.52% 77.95%	IEEE Access, 2021
Cross-Modality Model	BraTS 2017-2018	0.898	0.823	0.762	Pattern Recognition, 2021
Modality Pairing Model	BraTS 2020	0.891	0.842	0.816	BrainLes 2020, MICCAI
ViT Based Models					
ViTBIS	BraTS 2019 Cross-Validation	90.28% 79.52%	82.23% 76.11%	79.24% 60.90%	MICCAI Workshop, 2021
Nested Modality-Aware Transformer	BraTS 2020	0.920	0.864	0.800	MICCAI, 2022
SFusion	BraTS 2020	0.889	0.822	0.738	arXiv, 2022
Hybrid Models					
TransBTS	BraTS 2019	0.900	0.819	0.789	MICCAI, 2021
MMFormer	BraTS 2018	0.896	0.858	0.776	MICCAI, 2022
RCNN	BraTS 2020	Precision: 90.72%	Recall: 91.68%	mIoU: 94.56%	Scientific Reports, 2024
Federated Learning					
HDC-Net	BraTS 2018	89.7%	84.7%	80.9%	IEEE JBHI, 2021
Cascade-CNN (CConvNet)	BraTS 2018	0.9203	0.8726	0.9113	Scientific Reports, 2021
Automated Brain Tumor Segmentation Survey					
CNN-based	BraTS 2014	0.87	0.73	0.77	IEEE RBME, 2019
InputCascadeCNN	BraTS 2015	0.88	0.79	0.73	IEEE RBME, 2019
DeepMedic	BraTS 2016	0.89	0.76	0.72	IEEE RBME, 2019
EMMA	BraTS 2017	0.90	0.82	0.75	IEEE RBME, 2019
Asymmetric Encoder-Decoder	BraTS 2018	0.91	0.86	0.82	IEEE RBME, 2019
U-Net Based Models					
DenseUNet+	BraTS 2021	95%	93%	95%	Journal of King Saud University – Computer and Information Sciences, 2023
BiTr-Unet	BraTS 2021 Validation BraTS 2021 Testing	0.9335 0.9257	0.9304 0.9350	0.8899 0.8874	MICCAI BrainLes 2021
ST-Unet	Synapse, ISIC 2018	78.86%	N/A	N/A	Computers in Biology and Medicine, 2022
VT-UNet	MSD BraTS Dataset	91.9%	82.2%	87.2%	-
TSEUnet	BraTS 2018	91.05%	87.25%	82.35%	IEEE CBMS, 2022

Fig. 2 Brain Tumor Segmentation Models, Performance Metrics, and Publications

Model Performance Evaluation

The performance of various models utilized for brain tumor segmentation reveals distinct trends and critical findings, highlighting the strengths and limitations of each approach and contributing to the literature.

1. CNN-Based Models

CNN-based approaches generally demonstrate high performance in **WT (Whole Tumor)** and **TC (Tumor Core)** segmentation when evaluated across a broad range of datasets. For instance, the **Cross-Modality Model** achieves **89.8% accuracy for WT** and **82.3% for TC**, while its performance for **ET (Enhancing Tumor)** drops to **76.2%**. Similarly, the **Modality Pairing Model** achieves **89.1% for WT**, **84.2% for TC**, and **81.6% for ET**. However, the lower performance of CNN-based models in ET segmentation underscores the need for further optimization in this region.

2. ViT-Based Models
Transformer-based models exhibit notable accuracy, especially in WT and TC segmentation. The **Nested Modality-Aware Transformer model** achieves **92% accuracy for WT** and **86.4% for TC**, while its ET accuracy (**80%**) is comparatively lower. The **SFusion model** demonstrates a reduced ET accuracy of **73.8%**. These findings indicate the necessity for further research to improve the performance of ViT-based models in ET segmentation.

3. Hybrid Models
Hybrid approaches that combine the strengths of CNN and Transformer architectures offer balanced performance. For example, the **TransBTS model** achieves accuracies of **90.0% for WT**, **81.9% for TC**, and **78.9% for ET**. Similarly, the **MMFormer model** achieves **89.6% for WT** and **85.8% for TC**. Although hybrid models generally perform well in WT and TC segmentation, further improvements are needed for ET accuracy.

4. Federated Learning Models
Federated learning-based models are notable for their ET segmentation accuracy. The **Cascade-CNN model** achieves **91.13% accuracy for ET**, one of the highest performances reported. The **HDC-Net model** also demonstrates strong performance with **80.9% accuracy for ET**. These models show the potential to enhance segmentation accuracy through decentralized data learning.

5. U-Net-Based Models
U-Net-based models, particularly in WT and TC segmentation, achieve superior results. The **DenseUNet+ model** achieves **95% accuracy for WT**, **93% for TC**, and **95% for ET**, making it the best-performing model in this analysis. Similarly, the **BiTr-Unet model** achieves **93.35% for WT** and **93.04% for TC**, establishing U-Net-based approaches as reliable and robust methods for brain tumor segmentation.

General Observations: - WT Performance: ViT-based and U-Net models exceed **90% accuracy** in WT segmentation. **- TC Performance:** CNN, ViT, and U-Net models achieve comparable accuracies (**80–93%**). **- ET Performance:** Federated learning and certain U-Net models demonstrate high performance (e.g., **Cascade-CNN: 91.13%**).

Recommendations for Future Work: - Improving ET Segmentation: Development of new algorithms and approaches to enhance ET accuracy. This comprehensive evaluation systematically highlights the strengths and weaknesses of the models employed in brain tumor segmentation, providing valuable insights for further advancements in the field.

4 - Dataset Statistics: Datasets Used in Brain Tumor Segmentation

The effective evaluation of deep learning-based models for brain tumor segmentation depends on the availability of multimodal MRI datasets. Since 2012, the Medical Image Computing and Computer-Assisted Intervention Society (MICCAI) has organized the annual BraTS competition.[19] This long-standing competition has played a significant role in promoting research and establishing a benchmark for evaluating brain tumor segmentation methods. The BraTS competition provides a standardized multimodal MRI dataset that includes four different scans—T1, T1c, T2, and FLAIR. These modalities together offer a comprehensive view of brain anatomy and pathology, allowing researchers to develop and assess deep learning-based brain tumor segmentation methods. The impact of the BraTS competition on research methodologies is profound, and most studies prefer to use the BraTS datasets for training and testing segmentation approaches.

5 - Evaluation Metrics for Brain Tumor Segmentation

Dice Similarity Coefficient (DSC) is a crucial performance evaluation metric widely used in medical imaging and segmentation tasks, particularly in applications like brain tumor segmentation, where it assesses the accuracy of a segmentation model by evaluating the similarity between two sets and specifically measuring their overlap. DSC expresses how consistent the model's segmentation is with the ground truth, typically provided by experts. It offers several advantages, such as objective evaluation, as it provides a numerical assessment of segmentation accuracy for objective comparison; simplicity and interpretability, as it is easy to calculate and the results are straightforward to understand; and versatility, as it is applied not only in brain tumor segmentation but also in various medical imaging tasks involving different organs or tumors. As a key metric in brain tumor segmentation and other medical imaging applications, DSC plays a vital role in evaluating model performance and is fundamental in the development and improvement of segmentation models. However, it should be used alongside other metrics to ensure a more comprehensive evaluation, especially in cases of class imbalance or when boundary precision is critical.

6 - Conclusion

This study highlights the transformative role of deep learning (DL) and federated learning (FL) in advancing brain tumor segmentation using multi-modal MRI. By systematically reviewing recent developments, we emphasize the significant advantages of DL models, such as saving time, eliminating human bias, and minimizing errors. Among DL architectures, U-Net-based models stand out due to their encoder-decoder architecture and skip connections, which enable precise localization and feature extraction. Hybrid models that combine U-Net, CNN, and Vision Transformers also show exceptional performance by leveraging the strengths of each approach. Cascaded networks and ensemble techniques further enhance segmentation accuracy by effectively extracting and combining features. Our findings underscore the dominance of the BraTS dataset in this field, providing comprehensive multimodal information and including HGG and LGG samples. The Dice coefficient remains the primary evaluation metric, ensuring consistency and comparability across studies. Furthermore, fusion and attention mechanisms, particularly those integrated into U-Net and hybrid models, have demonstrated great potential in handling missing modality information and enhancing segmentation accuracy by extracting invariant features from the available data. Federated learning emerges as a promising solution for developing effective, secure, and unbiased models while preserving data privacy. It addresses the challenges posed by limited datasets during training and testing, offering practical applications in the healthcare sector. Future advancements in FL are expected to lead to the development of more sophisticated protocols that ensure enhanced security and privacy, setting new standards in medical imaging and patient care. This study provides a comprehensive evaluation of DL and FL architectures, datasets, and experimental parameters, identifies existing research gaps, and proposes promising directions for future research in brain tumor segmentation

References

[1] Syazwany NS, Nam JH, Lee SC. Mm-bifpn: multi-modality fusion network with bi-fpn for MRI brain tumor segmentation. IEEE Access. 2021;9:160708–160720. DOI: https://doi.org/10.1109/ACCESS.2021.3132050

[2] Zhang D, Huang G, Zhang Q, Han J, Han J, Yu Y. Cross-modality deep feature learning for brain tumor segmentation. Pattern Recognition. 2021;110:107562. DOI: https://doi.org/10.48550/arXiv.2201.02356

[3] Wang Y, Zhang Y, Hou F, Liu Y, Tian J, Zhong C, et al. Modality pairing learning for brain tumor segmentation. BrainLesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in conjunction with MICCAI 2020, Lima, Peru. DOI: https://doi.org/10.48550/arXiv.2010.09277

[4] Sagar A. ViTBIS: Vision Transformer for biomedical image segmentation. MICCAI Workshop on Distributed and Collaborative Learning. Springer; 2021. p. 34–45. DOI: https://doi.org/10.48550/arXiv.2201.05920.

[5] Xing Z., Yu L., Wan L., Han T., Zhu L., "NestedFormer: Nested Modality-Aware Transformer for Brain Tumor Segmentation," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2022, pp. 140–150. DOI: https://doi.org/10.48550/arXiv.2208.14876.

[6] Liu Y., Zhang X., Chen Z., "SFusion: A Self-Attention-Based Fusion Block for Automatic Combination of Missing Modalities Without Zero-Padding," *arXiv preprint*, 2022. DOI: 10.48550/arXiv.2208.12776.

[7] Wenxuan W., Chen C., Meng D., Hong Y., Sen Z., Jiangyun L., "TransBTS: Multimodal Brain Tumor Segmentation Using Transformer," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I*, Springer, 2021, pp. 109–119. DOI: https://doi.org/10.48550/arXiv.2103.0443010.48550/arXiv.2103.04430.

[8] Zhang Y., He N., Yang J., Li Y., Wei D., Huang Y., et al., "MMFormer: Multimodal Medical Transformer for Incomplete Multimodal Learning of Brain Tumor Segmentation," *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2022, pp. 107–117. DOI: 10.48550/arXiv.2206.02425.

[9] Yuan J., "Brain tumor image segmentation method using hybrid attention module and improved mask RCNN," *Scientific Reports*, 2024; 14: Article 20615. DOI: 10.1038/s41598-024-71250-4.

[10] Luo Z., Jia Z., Yuan Z., Peng J., "HDC-Net: Hierarchical Decoupled Convolution Network for Brain Tumor Segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 737–748, 2021. DOI: 10.1109/JBHI.2020.2998146.

[11] Ranjbarzadeh R., Bagherian Kasgari A., Jafarzadeh Ghoushchi S., Anari S., Naseri M., Bendechache M., "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images," *Scientific Reports*, vol. 11, Article 10930, 2021. DOI: 10.1038/s41598-021-90428-8.

[12] Ghaffari M., Sowmya A., Oliver R., "Automated brain tumor segmentation using multimodal brain scans: A survey based on models submitted to the BraTS 2012–2018 challenges," *IEEE Reviews in Biomedical Engineering*, vol. 13, pp. 156–168, 2019. DOI: 10.1109/RBME.2019.2946868.

[13] Chen Y., Wang J., "TSEUnet: A 3D neural network with fused Transformer and SE-Attention for brain tumor segmentation," *2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS)*, Shenzhen, China, 2022. DOI: 10.1109/CBMS55023.2022.00030.

[14] Jia Q., Shu H., "BiTr-Unet: a CNN-Transformer Combined Network for MRI Brain Tumor Segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, LNCS 12963, pp. 3-14, 2022. DOI: 10.48550/arXiv.2109.12271.

[15] Zhang J., Qin Q., Ye Q., Ruan T., "ST-Unet: Swin Transformer boosted U-Net with Cross-Layer Feature Enhancement for medical image segmentation," *Computers in Biology and Medicine*, vol. 153, p. 106516, Feb. 2023. DOI: 10.1016/j.compbiomed.2022.106516.

[16] Peiris, H., Hayat, M., Chen, Z., Egan, G., Harandi, M. (2021). A Robust Volumetric Transformer for Accurate 3D Tumor Segmentation. *arXiv preprint arXiv:2111.13300*. https://doi.org/10.48550/arXiv.2111.13300.

[17] Çetiner H., Metlek S., "DenseUNet+: A novel hybrid segmentation approach based on multi-modality images for brain tumor segmentation," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 8, p. 101663, 2023. DOI: 10.1016/j.jksuci.2023.101663.

[18] A Survey of MRI-Based Brain Tumor Segmentation Methods, *Tsinghua Science Technology*, vol. 19, no. 6, pp. 578-595, 2014. DOI: https://doi.org/10.1109/TST.2014.6961028.

[19] Abidin Z. U., Naqvi R. A., Haider A., Kim H. S., Jeong D., Lee S. W., "Recent deep learning-based brain tumor segmentation models using multi-modality magnetic resonance imaging: a prospective survey," *Frontiers in Bioengineering and Biotechnology*, vol. 12, 2024, DOI: 10.3389/fbioe.2024.1392807.

[20] "Published in: IEEE Transactions on Medical Imaging (Volume: 31, Issue: 3, March 2012)," pp. 790–804, 2012. DOI: 10.1109/TMI.2011.2181857. Published by IEEE, PubMed ID: 22207638.

[21] Schlemper J., Oktay O., Schaap M., Heinrich M., Kainz B., Glocker B., Rueckert D., "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, pp. 197–207, Apr. 2019. DOI: 10.1016/j.media.2019.01.012. PMID: 30802813, PMCID: PMC7610718.

[22] "MA-Net: A Multi-Scale Attention Network for Liver and Tumor Segmentation," *IEEE Access*, vol. 8, pp. 179656–179665, Sep. 2020. DOI: 10.1109/ACCESS.2020.3025372. Published by IEEE.

[23] "Computer-assisted segmentation of brain tumor lesions from multi-sequence Magnetic Resonance Imaging using the Mumford-Shah model," presented at the *2010 25th International Conference of Image and Vision Computing New Zealand*, Queenstown, New Zealand, Nov. 8–9, 2010. DOI: 10.1109/IVCNZ.2010.6148803. Published by IEEE.

[24] Zhou T., Canu S., Vera P., Ruan S., "Brain Tumor Segmentation with Missing Modalities via Latent Multi-source Correlation Representation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV*, pp. 533–541. DOI: 10.1007/978-3-030-59719-1_52. Published Oct. 4, 2020.

[25] The MNIST Database of Handwritten Digit Images for Machine Learning Research, IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 141-142, Nov. 2012. doi: 10.1109/MSP.2012.2211477.

[26] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-Institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation," *Brainlesion*, vol. 11383, pp. 92-104, 2019. doi: 10.1007/978-3-030-11723-8_9. Epub 2019 Jan 26. PMID: 31231720, PMCID: PMC6589345.

[27] A. Hamamci, N. Kucuk, K. Karaman, K. Engin, and G. Unal, "Tumor-Cut: segmentation of brain tumors on contrast enhanced MR images for radiosurgery applications," *IEEE Trans. Med. Imaging*, vol. 31, no. 3, pp. 790-804, Mar. 2012.

doi: 10.1109/TMI.2011.2181857. Epub 2011 Dec 26. PMID: 22207638.

[28] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 10781-10790.

[29] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, and L. Lanczi, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993-2024, Oct. 2014. doi: 10.1109/TMI.2014.2377694. PMID: 25494501, PMCID: PMC4833122.

[30] K.-L. Tseng, Y.-L. Lin, W. Hsu, and C.-Y. Huang, "Joint Sequence Learning and Cross-Modality Convolution for 3D Biomedical Segmentation," in *Proc. CVPR*, 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1704.07754.

[31] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," 2018.

[32] A. Dosovitskiy, et al., "ICLR camera-ready version: Fine-tuning code and pre-trained models available," arXiv, vol. 2010.11929, pp. 2021.

[33] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 7794–7803, 2018.

[34] S. Bakas *et al.*, "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge," *arXiv preprint arXiv:1811.02629*, 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1811.02629.

[35] A. Vaswani *et al.*, "Attention Is All You Need," *arXiv preprint arXiv:1706.03762*, 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1706.03762.

[36] C. Chen, Q. Dou, Y. Jin, H. Chen, J. Qin, and P.-A. Heng, "Robust Multimodal Brain Tumor Segmentation via Feature Disentanglement and Gated Fusion," *arXiv preprint arXiv:2002.09708*, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.2002.09708.

[37] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L.u., Polosukhin, I.: "Attention is all you need," in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc. (2017).

[38] D. Li, G. Chen, X. Wu, Z. Yu, and M. Tan, "Face anti-spoofing with cross-stage relation enhancement and spoof material perception," *Neural Networks*, 2024,

https://doi.org/10.1016/j.neunet.2024.106275.

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv preprint arXiv:1505.04597*, 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1505.04597.

[40] B. H. Menze *et al.*, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993-2024, Oct. 2015. [Online]. Available: https://doi.org/10.1109/TMI.2014.2377694.

[41] P. Zhou *et al.*, "HCCANet: Histopathological image grading of colorectal cancer using CNN based on multichannel fusion attention mechanism," *Sci. Rep.*, vol. 12, no. 1, p. 22242, Dec. 2022. [Online]. Available: https://doi.org/10.1038/s41598-022-18879-1.

[42] J. Zhang *et al.*, "SEM-RCNN: A Squeeze-and-Excitation-Based Mask Region Convolutional Neural Network for Multi-Class Environmental Microorganism Detection," *Appl. Sci.*, vol. 12, no. 19, p. 9902, Oct. 2022. [Online]. Available: https://doi.org/10.3390/app12199902.

[43] M. F. Ahamed *et al.*, "A review on brain tumor segmentation based on deep learning methods with federated learning techniques," *Comput. Med. Imaging Graph.*, vol. 99, p. 102313, 2023. [Online]. Available: https://doi.org/10.1016/j.compmedimag.2023.102313.

[44] S. Banerjee, S. Mitra, and B. U. Shankar, "Multi-planar spatial-convnet for segmentation and survival prediction in brain cancer," in *Int. MICCAI Brainlesion Workshop*, 2018, pp. 94–104.

[45] G. Chen, Q. Li, F. Shi, I. Rekik, and Z. Pan, "RFDCR: Automated brain lesion segmentation using cascaded random forests with dense conditional random fields," *Neuroimage*, vol. 211, p. 116620, 2020. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2020.116620.

[46] L. Sun and S. Zhang, "Multi-view 3D CNN with dense CRF for brain tumor segmentation and survival prediction," in *Proc. Pre-Conf. 7th MICCAI BraTS Challenge*, 2018, pp. 448–456.

[47] B. Wang *et al.*, "Deeply supervised 3D fully convolutional networks with group dilated convolution for automatic MRI prostate segmentation," *Med. Phys.*, vol. 46, no. 4, pp. 1707–1718, 2019. [Online]. Available: https://doi.org/10.1002/mp.13416.

[48] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 7132–7141, 2018.

[49] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. Roth, and D. Xu, "UNETR: Transformers for 3D Medical Image Segmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2022. https://doi.org/10.48550/arXiv.2103.10504

[50] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," *arXiv preprint arXiv:2103.14030*, 2021.

[51] F. Martínez, E. Romero, G. Dréan, A. Simon, P. Haigron, R. De Crevoisier, and O. Acosta, "Segmentation of pelvic structures for planning CT using a geometrical shape model tuned by a multi-scale edge detector," *Med. Phys.*, vol. 41, no. 3, pp. 1–12, 2014. PMID: 24594798, PMCID: PMC5103036.

[52] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[53] Z. Ullah, M. Usman, M. Jeon, and J. Gwak, "Cascade multiscale residual attention CNNs with adaptive ROI for automatic brain tumor segmentation," *Inf. Sci. (Ny)*, vol. 608, pp. 1541–1556, 2022. [Online]. Available: https://doi.org/10.1016/j.ins.2022.07.044.