Brain Tumor Segmentation using 3D-Unet with FPN and Attention mechanism

Neha Dyaberi[1], Bhagyashree Nyamagoudar[2], Prateek Chittoor[3], Swadeep Sawant[4], and Channabasappa Muttal[5]

School of Computer Science and Engineering
KLE Technological University, Hubballi, India
01fe22bcs300@kletech.ac.in, 01fe22bcs299@kletech.ac.in,
01fe22bcs292@kletech.ac.in, 01fe22bcs308@kletech.ac.in,
Channabasappa.muttal@kletech.ac.in

Abstract. Brain tumor segmentation is a critical task in medical imaging, aiding in diagnosis and treatment planning for brain tumor patients. This study presents an enhanced 3D U-Net architecture integrated with Feature Pyramid Networks (FPN) and attention mechanisms to improve segmentation accuracy. Utilizing the BraTS 2020 dataset, which comprises multimodal MRI scans (T1, T2,T1ce and FLAIR sequences), the data preprocessing involved intensity normalization, alignment of segmentation masks, and organization of 3D MRI scans. The architecture incorporates an encoder, bottleneck, and decoder path, leveraging FPN for multi-resolution feature extraction and attention mechanisms for tumor region focus. A combination of categorical cross-entropy and Dice loss optimized with the Adam optimizer was employed during training. The model achieved a training accuracy of 96.16% ,Dice loss of 0.0086 ,a test accuracy of 96.14% and F1 score of 98.03% , outperforming other baselines, including U-Net and its variants. Practical applications include automating tumor segmentation to assist in diagnosis, surgical planning, and treatment monitoring, improving radiological workflows' speed and precision. However, challenges such as high computational demands and limited dataset diversity remain, requiring further optimization for resource-constrained environments and validation on larger, multi-institutional datasets. This study highlights the potential of combining FPN and attention mechanisms to advance clinical tumor segmentation workflows, with scope for scalability and integration into real-world medical imaging systems.

Keywords: MRI, UNet, FPN, Attention Mechanism, Image segmentation, Deep learning, Brain tumor

1 INTRODUCTION

Magnetic Resonance Imaging (MRI)[1] plays a crucial role in modern clinical diagnostics, providing detailed insights into the structure and function of the human body, especially the brain. Its exceptional ability to clearly visualize soft

tissues makes it invaluable for detecting and diagnosing conditions such as tumors, strokes, and neurodegenerative diseases[2]. However, manually segmenting brain MRI scans is a challenging and time-consuming task. It is prone to errors and highly dependent on the expertise of clinicians, further complicated by the significant variability[3] in brain MRI data. These challenges highlight the need for automated methods to ensure more accurate and efficient segmentation.

Building on the challenges of manual brain MRI segmentation, deep learning[4] has brought transformative advancements to medical imaging by enabling automated analysis and interpretation of complex datasets. Among these, Convolutional Neural Networks (CNNs)[1,5] have emerged as powerful tools, capable of extracting intricate patterns and features from MRI scans. Architectures like UNet[1,6], with its encoder-decoder structure, have proven particularly effective in biomedical segmentation due to their ability to preserve spatial precision while capturing valuable contextual information. The Feature Pyramid Network (FPN)[7] further enhances UNet by improving multiscale feature extraction, enabling the model to handle variability in tumor sizes and detect fine-grained details. Additionally, attention mechanisms[8] refine this process by dynamically focusing on essential features and reducing the influence of noise and irrelevant regions in the data.

This study introduces a novel framework that integrates UNet, Feature Pyramid Networks (FPN), and attention mechanisms to improve brain tumor segmentation in MRI datasets. This combined approach leverages the strengths of each component: UNet's ability to preserve spatial resolution, FPN's capacity for multi-scale feature fusion, and attention mechanism's precision in focusing on critical features. Together, they not only automate the segmentation process but also provide robustness against the inherent variability and complexity of brain MRI data.

Building on the limitations of manual segmentation and enhancing existing deep learning architectures[9], this framework aims to deliver more accurate and reliable brain tumor segmentation. The proposed model contributes to more efficient clinical workflows and supports personalized treatment planning, paving the way for improved patient care and advancing medical imaging practices.

The paper explores advancements in brain tumor segmentation through deep learning2 techniques. It begins with a review of related works, focusing on the evolution of U-Net2.1 and its significant impact on medical imaging, particularly for brain tumor segmentation. The transition to 3D U-Net2.2 is highlighted, emphasizing its ability to process volumetric data and effectively capture spatial dependencies. Further enhancements, such as the integration of Feature Pyramid Networks (FPN) with 3D U-Net2.3, are discussed, showcasing their role in improving multi-scale feature extraction and segmentation accuracy.

Building on these advancements, the proposed method3 leverages the BraTS 2020 dataset for training and validation. The preprocessing stage3.1 ensures consistency by normalizing intensity values and aligning segmentation masks with input images, optimizing the model's performance. Implementation details3.2 cover the training process, which utilized a combination of Categorical cross-

entropy and Dice loss, the architecture integrating UNet, FPN, and attention mechanisms, and the testing phases.

The results4 demonstrate the effectiveness of the proposed framework, achieving an accuracy of 96.16% on the training set and 96.14% on the test set. Compared with existing models like U-Net, 3D U-Net, and 3D U-Net with FPN, the proposed method outperforms these baselines, offering higher accuracy and better generalization. The integration of FPN and attention mechanisms significantly enhances the model's ability to detect fine-grained details and handle variability in tumor sizes, addressing limitations seen in previous approaches.

The paper concludes5 by summarizing the key findings, emphasizing the potential of the proposed method in advancing automated brain tumor segmentation, and suggesting future research directions, including further optimization of the architecture and its application in clinical settings for personalized treatment planning.

2 Related work

Deep learning has brought major changes to medical imaging, especially in tasks like brain tumor segmentation. This shift is mainly due to its ability to automatically learn complex features from data, reducing the need for manual feature extraction.Convolutional Neural Networks (CNNs)[5] is widely used for image-related tasks because they are great at processing visual data. Over time, deep learning models have evolved, leading to specialized architectures like U-Net[10], which was designed specifically for biomedical image segmentation. These advancements have significantly improved both accuracy and efficiency, making deep learning an essential tool in modern medical imaging[11].

2.1 Role of Unet in Brain Tumor Segmentation

U-Net is a highly effective model for brain tumor segmentation due to its encoder-decoder architecture. It combines a contracting path for context extraction and an expansive path for precise localization, making it ideal for complex medical images[12]. The contracting path reduces spatial dimensions and increases feature depth, while the expansive path uses up-convolutions and skip connections to restore spatial resolution. This U-shaped architecture, shown in Figure 1, is simple yet effective for segmentation tasks[12].

One of U-Net's key advantages is its ability to perform well with limited annotated data, which is common in medical imaging. Through data augmentation techniques like elastic deformations[12], U-Net improves robustness and reduces the need for large datasets. Additionally, its overlap-tile strategy allows it to process high-resolution MRI scans by dividing them into smaller, overlapping patches, maintaining accuracy even at the edges[12].

U-Net has demonstrated its versatility by achieving state-of-the-art results in various biomedical segmentation challenges, further proving its suitability for brain tumor segmentation[12]. Its efficient use of data and ability to handle

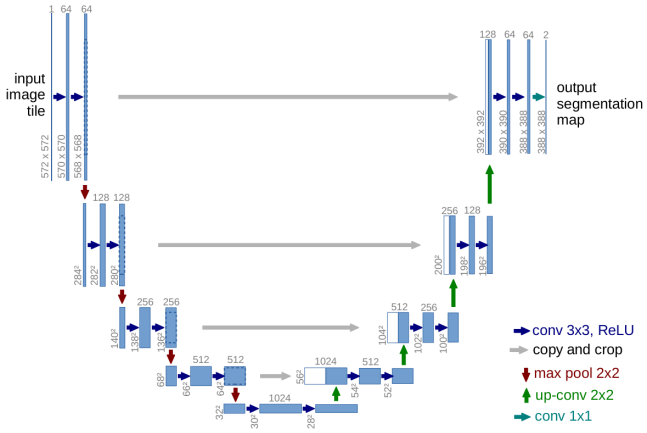

Fig. 1. Unet architecture(example for 32x32 pixels in the lowest resolution).Each blue box corresponds to a multi-channel featuremap.The number of channels is denoted on top of the box.The x-y-size is provided at the lower left edge of the box.White boxes represent copied featuremaps . The arrows denote the different operations.[12]

complex medical images make it a foundational tool for brain tumor segmentation, with further improvements possible through the integration of advanced techniques like Feature Pyramid Networks and attention mechanisms.

However, U-Net has limitations, such as struggling with accurately segmenting tumors with varying shapes and sizes. Additionally, its performance can be affected by noise or imbalanced data. These issues highlight the need for upgrades, like incorporating multi-scale feature extraction and focusing mechanisms, to improve its segmentation accuracy and robustness.

2.2 Role of 3D U-Net in Brain Tumor Segmentation

The 3D U-Net architecture is a significant advancement in biomedical image segmentation, especially for volumetric data like MRI scans[13]. This extension of the original U-Net framework captures both spatial and volumetric features, addressing challenges like alignment errors and complex spatial relationships[14]. The encoder extracts hierarchical features through 3D convolutions and max-pooling, while the decoder uses up-sampling layers and skip connections to integrate high-resolution information, preserving fine details essential for brain tumor segmentation[12].

Unlike 2D models, 3D U-Net processes volumetric data directly, improving segmentation accuracy for 3D MRI scans. The use of soft dice loss helps address class imbalance in medical datasets, while data augmentation techniques like random rotations and elastic deformations improve model generalization[15]. This is especially valuable in medical imaging, where annotated data is often scarce.

For brain tumor segmentation, 3D U-Net processes MRI volumes in overlapping sub-volumes, overcoming GPU memory limitations and ensuring accurate segmentation at patch edges[14]. Additionally, integrating methods like feature pyramids enhances segmentation performance, especially for tumors with low contrast or irregular shapes[16].

In summary, 3D U-Net's ability to integrate spatial and volumetric features, along with data augmentation and advanced segmentation techniques, makes it a cornerstone in brain tumor segmentation and a foundation for future innovations in biomedical imaging.

2.3 Working of FPN-Feature Pyramid Networks

The 3D U-Net model, a widely recognized architecture for medical image segmentation, serves as the foundation for many approaches[6]. Its encoder-decoder structure, with skip connections to retain spatial details, is particularly effective for pixel-wise segmentation tasks[12].

To enhance U-Net's performance, Feature Pyramid Network (FPN) structures have been integrated to address the limitations in multi-scale feature extraction[17]. FPNs combine semantic information across scales using a combination of bottom-up pathways, top-down pathways, and lateral connections. This structure significantly improves the model's ability to capture fine-grained details and contextual information, especially for boundary refinement[18].

The performance of segmentation models is evaluated using standard metrics such as the Dice Similarity Coefficient (DSC)1, Jaccard Index (IoU)2, and Accuracy (ACC)3[19].

The evaluation metrics are defined as follows:

$$\mathrm{DSC} = \frac{2|A \cap B|}{|A| + |B|} \tag{1}$$

$$\mathrm{IoU} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{2}$$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{3}$$

where A is Predicted segmentation, B is Ground truth segmentation TP, TN, FP, FN are True positive, true negative, false positive and false negative values, respectively.

An improved model of U-Net integrated with FPN enhances U-Net's ability to capture multi-scale features and refines edge segmentation, addressing the limitations of traditional approaches[20]. The improved model demonstrated superior performance with a Dice score of 92%, Jaccard index of 86%, and precision of 99.1%, highlighting its effectiveness in segmenting brain tumor regions[19].

Existing approaches to brain tumor segmentation, such as UNet, 3D UNet, and UNet with FPN, have their limitations. UNet and 3D UNet, which are effective at capturing spatial and volumetric features, often struggle to focus on

the critical regions needed for precise segmentation of tumor subregions, such as the enhancing tumor. Similarly, UNet with FPN improves multi-scale feature extraction but does not fully address issues with hierarchical feature imbalance or prioritize important areas effectively. 3D UNet models also come with significant computational overhead, which limits their applicability in practical settings. A common observation across these methods is the lack of attention mechanisms to dynamically emphasize relevant features, making it difficult to handle small or irregular tumor regions. Our proposed approach of combining 3D UNet with FPN and an attention mechanism directly addresses these gaps by improving contextual understanding, enhancing feature fusion, and reducing computational complexity, making it more practical and effective for clinical use.

3 Methodology

This structured approach for developing a robust brain tumor segmentation model using multimodal MRI scans from the BraTS 2020 dataset. Key steps include data preprocessing, feature selection, and the implementation of a 3D U-Net architecture enhanced with a Feature Pyramid Network (FPN) and attention mechanisms. These techniques ensure precise segmentation by leveraging multi-resolution features and focusing on tumor-specific regions. The workflow integrates advanced preprocessing techniques and carefully selected MRI modalities to optimize model training and testing, enabling accurate segmentation of brain tumor subregions for clinical and computational applications.

3.1 Preprocessing of data

Preprocessing is a crucial phase that ensures the data is clean, consistent, and ready for effective model learning. This process involves organizing 3D MRI scans for each patient, integrating various imaging modalities such as T1, T1ce, T2, and FLAIR. These modalities provide complementary views of the brain and its abnormalities. By combining these different *modalities into a single 3D volume with multiple channels*, the model can simultaneously analyze all the modalities, leading to a comprehensive understanding of the brain's structure and any present abnormalities.

Given the substantial variations in intensity across MRI scans due to different scanners and patient conditions, intensity normalization is essential. This step scales all values into the range [0, 1],using the formula4 ensuring the model focuses on tumor-related variations rather than inconsistencies caused by the scanning process. By standardizing the intensity values, the normalization process enhances the reliability and accuracy of the segmentation model.

The normalization formula used is:

$$I_{\text{normalized}} = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \tag{4}$$

where $I_{\text{normalized}}$ represents the normalized intensity value, I is the original intensity value, $I_{\min}$ is the minimum intensity value in the scan, and $I_{\max}$ is the

maximum intensity value in the scan. This formula effectively scales the intensity values to a standardized range, which is crucial for consistent and accurate model performance.

The *segmentation masks* delineating tumor regions (necrotic core, edema, and enhancing tumor) undergo meticulous processing. Each voxel in the mask is categorized into one of four labels, including the background. These masks are meticulously aligned with the input images, ensuring the model accurately learns the spatial relationships between the images and their corresponding tumor regions.

Feature selection, a key part of preprocessing, involves identifying the most relevant features for model input. In this study, this step is carried out by selecting three MRI modalities: T1, T2, and FLAIR. These modalities were chosen based on domain expertise[1], as they provide complementary and meaningful information for tumor segmentation.

Specifically: *T1 offers detailed structural information,T2 highlights fluid-based contrasts, FLAIR emphasizes lesion areas by suppressing signals from normal fluids.*

The modality T1ce (contrast-enhanced T1) is excluded because it provides structural details similar to T1 but with additional contrast in areas affected by a contrast agent. Including both could introduce redundancy, as T1 already captures the necessary structural features. By focusing on these three modalities, the model gains the most informative inputs while avoiding unnecessary complexity.

The data is organized patient by patient using a naming convention that helps identify and group related files. Each patient's normalized image and mask are stored in arrays, making it easy to feed them into the training process. These preprocessing steps simplify the data while preserving the important details needed for tumor segmentation, and are integrated into the model training workflow to ensure clean and standardized data. The entire process is designed to be efficient and scalable, ensuring robust and accurate tumor segmentation.

3.2 Implementation Approach:

The 3D U-Net architecture is specifically designed for 3D volumetric data like MRI scans and consists of three main components: the Encoder Path, Bottleneck, and Decoder Path. The Encoder Path extracts hierarchical features from input images through successive 3D convolutions and pooling layers, progressively reducing spatial dimensions to capture high-level semantic features. The Bottleneck acts as a bridge, providing compact representations of the learned features. The Decoder Path reconstructs the segmentation output by progressively upsampling the features and using skip connections to merge them with corresponding encoder features, ensuring better spatial localization of tumor regions.

To enhance performance, a Feature Pyramid Network (FPN) is integrated into the model. As shown in Figure 2, FPN captures multi-resolution features and improves tumor localization across different scales. It employs a top-down

approach to upsample higher-level features and merges them with lower-level features through skip connections. This feature fusion process, illustrated in Eq. 5, allows the model to combine fine details with high-level contextual information for accurate segmentation.

$$F_{\text{output}} = \text{Conv3D}(\text{Upsample}(F_{\text{high}}) + F_{\text{low}}) \tag{5}$$

Here, F_{output} represents the fused feature map, F_{high} and F_{low} are higher and lower-resolution feature maps, respectively.

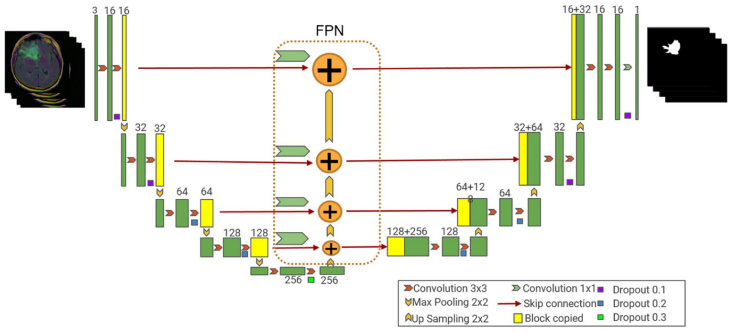

Fig. 2. The figure shows a UNet-based architecture with an FPN for image segmentation. The encoder consists of convolutional layers (3×3), max-pooling (2×2), and dropout blocks (0.1, 0.2, 0.3). The FPN aggregates multi-scale features using addition operations. The decoder uses up-sampling (2×2) and skip connections to restore spatial resolution. The output is a single-channel segmentation map. Legends denote operations and dropout rates.[19]

An attention mechanism is applied to focus on the most relevant areas of the MRI scans, particularly the tumor regions. This mechanism assigns higher importance to features corresponding to tumors while reducing the influence of irrelevant areas like background noise. The attention block, as described in Eqs. 6 and 7, ensures that only the most relevant features are emphasized.

$$A = \sigma(\text{Conv3D}(F_{\text{encoder}} \cdot F_{\text{decoder}})) \tag{6}$$

$$F_{\text{refined}} = A \cdot F_{\text{decoder}} \tag{7}$$

Here, A is the attention weight map, F_{encoder} and F_{decoder} are feature maps from the encoder and decoder layers, respectively. F_{refined} is the refined feature map obtained after applying attention.

By combining the FPN (Eq. 5) and attention mechanisms (Eqs. 6 and 7), the model achieves effective segmentation of tumors, leveraging multi-scale feature representations and focused attention on relevant regions.

The workflow for brain tumor segmentation, depicted in Figure 3, involves preprocessing multiple MRI modalities such as T1, T2, and FLAIR, feature selection, and stacking of images. This ensures that the segmented output highlights specific brain regions for improved diagnosis and treatment planning.

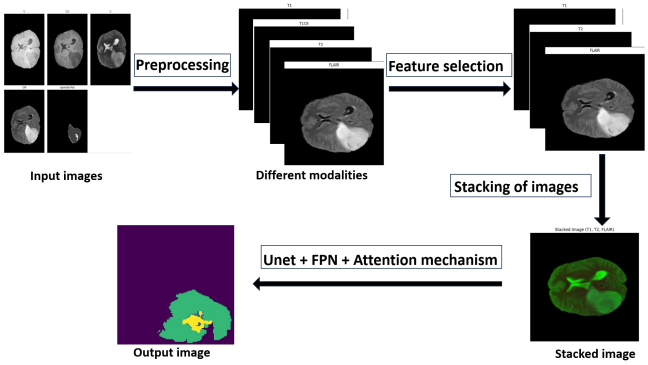

Fig. 3. This image showcases the workflow for brain tumor image segmentation using multiple MRI modalities (T1, T2, FLAIR). It includes steps such as preprocessing, feature selection, and stacking of images, ultimately resulting in a segmented image that highlights specific brain regions for improved diagnosis and treatment planning.

The model was trained using data from 295 patients, comprising 80.1% of the dataset. Training spanned 50 epochs with a batch size of 2, using an Adam optimizer with a learning rate of 0.001. A combination of categorical cross-entropy and Dice loss was employed to address class imbalance and improve multi-class segmentation performance. Post-training, the model was evaluated using MRI scans from patients 296 to 369, excluding patient 355 due to data issues. Preprocessing and normalization steps were applied consistently to both training and testing data, ensuring reliable evaluation and effective generalization.

4 Result and Discussion

The model achieved a training accuracy of **96.16%**, highlighting its strong ability to learn patterns from the dataset. Training was conducted over 50 epochs using a small batch size of 2, with data from 295 patients (approximately 80.1% of the dataset). The combination of categorical cross-entropy and Dice loss as the loss function, optimized using the Adam optimizer with a learning rate of 0.001, contributed to this performance. The approximate Dice loss for the model was **0.0086**, reflecting its high segmentation accuracy.

During testing, the model reached an accuracy of **96.14%**, which is very close to the training accuracy. The approximate F1 score during testing was **98.03%**, further confirming the model's ability to achieve precise segmentation.

This minimal difference between training and testing accuracy indicates that the model generalizes well to unseen data and is not overfitting. These results demonstrate the robustness of the model across both training and testing phases, showcasing its reliability for real-world applications. Additionally, the architecture effectively balances feature extraction and spatial resolution, crucial for accurate tumor delineation.

The dataset used in this study is the BraTS 2020 dataset [21], a standard benchmark for brain tumor segmentation tasks. It provides multimodal MRI scans, including T1, T1ce, T2, and FLAIR sequences (Figure 4), which capture essential tissue characteristics. Tumors are segmented into key subregions: enhancing tumor (ET), tumor core (TC), and whole tumor (WT). This detailed segmentation framework supports effective model training by providing comprehensive insights into tumor morphology. The dataset's diversity in tumor subregions allows the model to learn intricate patterns, further enhancing segmentation performance.

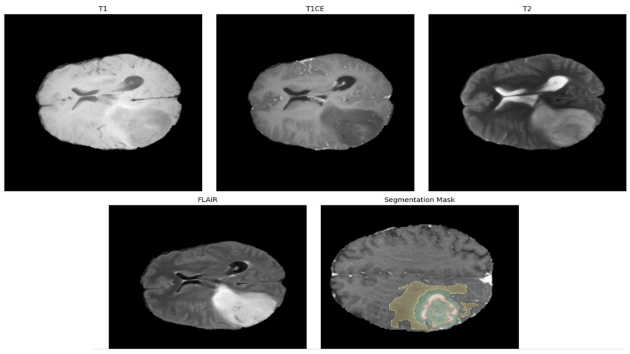

Fig. 4. Example of a brain tumor from the BraTS 2020 training dataset. Red: Enhancing Tumor (ET), Green: Non-enhancing Tumor/Necrotic Tumor (NET/NCR), Yellow: Peritumoral Edema (ED). T1-weighted sequence, T1-weighted contrast-enhanced sequence, T2-weighted sequence, FLAIR sequence, and T1-weighted contrast-enhanced sequence with labelmap overlay.

The integration of advanced techniques such as *Feature Pyramid Networks (FPN) and Attention Mechanisms* significantly contributed to the model's success. FPN enabled the model to capture features at multiple scales, improving its ability to localize both small and large tumors. Simultaneously, the attention mechanism focused on the most relevant tumor areas, reducing the influence of irrelevant regions like background noise. These combined enhancements made the model robust and effective for real-world applications in medical imaging.

The performance of the proposed model was compared with other brain tumor segmentation models (Table 1). It achieved the highest accuracy of **96.16%**,

outperforming other approaches. This result underscores the effectiveness of combining FPN and attention mechanisms for accurate tumor segmentation.

Table 1. Comparison of Brain Tumor Segmentation Accuracy with Existing Models on BraTS 2020 Dataset.

Model	Accuracy (%)
U-Net (Standard)	91.8
Attention U-Net	93.2
Dense U-Net	92.6
U-Net with FPN	91.0
ResUNet	94.3
U-Net with FPN and Attention	**96.16**

5 Conclusion

The proposed study successfully developed a novel brain tumor segmentation framework by integrating 3D U-Net, Feature Pyramid Networks (FPN), and an attention mechanism, achieving superior performance compared to traditional methods, with an accuracy value of 96.16%. This approach effectively addresses the challenges of accurate tumor delineation, especially in complex cases involving enhancing tumors and peritumoral edema. The use of the attention mechanism enabled the model to focus on key tumor regions, enhancing segmentation precision, whereas the FPN allowed for improved multi-scale feature fusion.

However, this study also highlights certain limitations. The computational cost associated with training and deploying the model is significant, posing challenges for resource-constrained environments. Additionally, the generalizability of the framework to other datasets remains uncertain due to the reliance on the BraTS 2020 dataset. Future work should focus on optimizing the model to reduce computational demands and validating it on larger, more diverse, and multi-institutional datasets to enhance robustness.

Incorporating advanced imaging modalities such as PET or CT alongside MRI could further improve segmentation accuracy and provide a more comprehensive diagnostic perspective. Expanding the model's applications to related tasks, such as tumor progression tracking or treatment response evaluation, would also broaden its clinical utility. By addressing these challenges and exploring potential optimizations, the framework could evolve into a more versatile and practical tool for real-time clinical use, ultimately advancing diagnostic accuracy and efficiency in healthcare.

References

1. Derrick Yeboah, Li Dequan, and George K Agordzo. Enhancing brain mri data visualization accuracy with unet and fpn networks. *Biomedical Signal Processing and Control*, 96:106418, 2024.

2. Natalia Shusharina, Denis Yukhnenko, Stepan Botman, Viktor Sapunov, Vladimir Savinov, Gleb Kamyshov, Dmitry Sayapin, and Igor Voznyuk. Modern methods of diagnostics and treatment of neurodegenerative diseases and depression. *Diagnostics*, 13(3):573, 2023.
3. Zahra Mirikharaji, Kumar Abhishek, Alceu Bissoto, Catarina Barata, Sandra Avila, Eduardo Valle, M Emre Celebi, and Ghassan Hamarneh. A survey on deep learning for skin lesion segmentation. *Medical Image Analysis*, 88:102863, 2023.
4. Mengfang Li, Yuanyuan Jiang, Yanzhou Zhang, and Haisheng Zhu. Medical image analysis using deep learning algorithms. *Frontiers in Public Health*, 11:1273253, 2023.
5. Leon O Chua. Cnn: A vision of complexity. *International Journal of Bifurcation and Chaos*, 7(10):2219–2425, 1997.
6. Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval, Afshin Bozorgpour, Sanaz Karimijafarbigloo, Joseph Paul Cohen, Ehsan Adeli, and Dorit Merhof. Medical image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
7. Wei Hu, Shouyi Yang, Weifeng Guo, Na Xiao, Xiaopeng Yang, and Xiangyang Ren. Stc-unet: renal tumor segmentation based on enhanced feature extraction at different network levels. *BMC Medical Imaging*, 24(1):179, 2024.
8. Xiang Ye, Zihang He, Wang Heng, and Yong Li. Toward understanding the effectiveness of attention mechanism. *AIP Advances*, 13(3), 2023.
9. Arkapravo Chattopadhyay and Mausumi Maitra. Mri-based brain tumour image detection using cnn based deep learning method. *Neuroscience informatics*, 2(4):100060, 2022.
10. Nahian Siddique, Sidike Paheding, Colin P Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE access*, 9:82031–82057, 2021.
11. Kenji Suzuki. Overview of deep learning in medical imaging. *Radiological physics and technology*, 10(3):257–273, 2017.
12. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.
13. Li-Ming Hsu, Shuai Wang, Lindsay Walton, Tzu-Wen Winnie Wang, Sung-Ho Lee, and Yen-Yu Ian Shih. 3d u-net improves automatic brain extraction for isotropic rat brain magnetic resonance imaging data. *Frontiers in Neuroscience*, 15:801008, 2021.
14. Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.
15. Fabian Isensee, Paul F Jäger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. Automated design of deep learning methods for biomedical image segmentation. *arXiv preprint arXiv:1904.08128*, 2019.
16. Mengxian Chi, Hong An, Xu Jin, and Zhenguo Nie. An n-shaped lightweight network with a feature pyramid and hybrid attention for brain tumor segmentation. *Entropy*, 26(2):166, 2024.

17. Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.
18. Rui Li, Shunyi Zheng, and Chenxi Duan. Feature pyramid network with multi-head attention for se-mantic segmentation of fine-resolution remotely sensed im-ages. *arXiv e-prints*, pages arXiv–2102, 2021.
19. Haitao Sun, Shuai Yang, Lijuan Chen, Pingyan Liao, Xiangping Liu, Ying Liu, and Ning Wang. Brain tumor image segmentation based on improved fpn. *BMC Medical Imaging*, 23(1):172, 2023.
20. Abdul Haseeb Nizamani, Zhigang Chen, Ahsan Ahmed Nizamani, and Kashif Shaheed. Feature-enhanced fusion of u-net-based improved brain tumor images segmentation. *Journal of Cloud Computing*, 12(1):170, 2023.
21. AWSAF49. Brats20 dataset - training and validation. `https://www.kaggle.com/datasets/awsaf49/brats20-dataset-training-validation`, 2020. Accessed: 2024-12-27.