

MULTIMODAL MRI BRAIN TUMOR SEGMENTATION USING ATTENTION-ENHANCED INCEPTION U-NET

A. Bamila Rachel

Assistant Professor
Computer Science and Engineering,
Jayaraj Annapackiam CSI College of Engineering,
Nazareth, India.

M. Karpoora Jothi

PG Scholar
Computer Science and Engineering,
Jayaraj Annapackiam CSI College of Engineering,
Nazareth, India.

Abstract - Brain tumor segmentation using multimodal magnetic resonance imaging is an essential task in medical image analysis because precise identification of tumor regions helps clinicians during diagnosis, treatment planning, and patient follow-up. However, obtaining accurate segmentation remains difficult due to significant variations in tumor structure, boundary appearance, and signal intensity across different MRI modalities. To overcome these challenges, this work introduces an Attention-Enhanced Inception U-Net model designed for automated segmentation of brain tumors from multimodal MRI scans such as T1, T1c, T2, and FLAIR.

The proposed architecture combines inception-based feature extraction with an attention mechanism inside an encoder-decoder framework. Inception blocks enable the network to capture tumor characteristics at multiple spatial scales, while attention modules improve focus on relevant abnormal regions and reduce background interference. Before training, MRI volumes are processed through resizing and normalization to maintain input consistency. Soft Dice Loss together with Focal Loss is applied during optimization to improve segmentation quality in regions where class imbalance is significant.

Performance analysis on the BraTS dataset indicates that the proposed network achieves improved segmentation compared with conventional U-Net approaches. The predicted segmentation masks show stronger boundary localization and improved detection of irregular tumor structures. This framework can serve as an effective support tool for radiologists by providing more dependable tumor region identification for future clinical applications.

Keywords - Medical Image Processing, Brain Tumor Segmentation, Multimodal MRI, Attention-Enhanced Inception U-Net, BraTS, Deep Learning

I. INTRODUCTION

Brain tumors represent a critical medical condition in which abnormal tissue growth develops inside the brain and may seriously affect normal neurological function if not identified at an early stage. Brain Tumor Accurate determination of tumor boundaries, position, and extent is essential for effective diagnosis, treatment selection, and surgical planning. Among different medical imaging

modalities, Magnetic Resonance Imaging is commonly used because it offers superior soft-tissue contrast and clear visualization of intracranial structures without ionizing radiation.

Traditional manual tumor segmentation from MRI scans requires significant clinical expertise and often consumes considerable time, particularly when tumor regions have irregular shapes or unclear borders. These difficulties increase when different tumor tissues appear similar to surrounding healthy structures. Because of these challenges, automated segmentation methods based on Deep Learning have gained strong importance in medical image analysis, as they can learn complex image representations directly from large datasets and improve segmentation consistency.

Among modern segmentation networks, U-Net remains one of the most widely adopted architectures because its encoder-decoder design and skip connections preserve both high-level and fine-grained spatial features. However, conventional U-Net may still produce limited performance when tumor appearance varies significantly across multimodal MRI sequences. To address this limitation, inception modules are incorporated to capture features at multiple receptive fields, while attention mechanisms strengthen the network's ability to emphasize important tumor regions and reduce irrelevant background responses.

This study presents an Attention-Enhanced Inception U-Net framework for multimodal brain tumor segmentation using T1, T1c, T2, and FLAIR MRI modalities. Training and evaluation are performed using the BraTS dataset to generate accurate tumor segmentation masks. The overall aim is to improve boundary extraction performance and develop a dependable computer-assisted system that can assist radiologists in advanced clinical decision support.

II. LITERATURE SURVEY

A. Attention-Based Segmentation Models

Recent brain tumor segmentation research has shown that integrating Attention U-Net mechanisms into convolutional segmentation networks improves the identification of abnormal tumor regions in multimodal MRI scans. Attention gates help the network assign higher importance to clinically relevant areas while reducing interference from surrounding normal tissues.

This improves boundary detection when tumors appear with unclear edges or irregular structures. Studies indicate that attention-based models achieve stronger segmentation accuracy than standard U-Net architectures, particularly in small lesion regions. However, the additional attention layers increase computational requirements and training complexity.

B. Three-Dimensional MRI Segmentation Approaches

The use of 3D U-Net has become important in recent segmentation tasks because three-dimensional convolution preserves volumetric continuity across MRI slices. Instead of processing images individually, this method analyzes full MRI volumes, allowing better representation of tumor depth and neighboring tissue relationships. Multimodal inputs such as T1, T1c, T2, and FLAIR provide complementary information that improves segmentation reliability. Although segmentation quality improves significantly, these networks demand high memory consumption and advanced GPU resources during training.

C. Inception-Based Multi-Scale Feature Extraction

Deep learning methods built with Inception Network modules focus on extracting image features at different receptive fields simultaneously. Multiple kernel sizes inside a single block allow the model to detect both small-scale texture patterns and larger tumor structures. This is useful in brain tumor segmentation because tumor appearance differs greatly across patients and MRI modalities. Research findings suggest that inception-based models provide stronger feature representation than standard convolutional blocks, though model optimization becomes more sensitive to hyperparameter selection.

D. Transformer-Integrated Segmentation Networks

Recent developments in medical image segmentation have introduced Vision Transformer modules to improve global contextual understanding. Unlike conventional convolution layers that mainly focus on local spatial patterns, transformers capture long-distance feature relationships across the full MRI image. This helps improve segmentation in complex tumor regions where global structure matters. Hybrid CNN-transformer models show strong segmentation performance, but their training process generally requires larger datasets and increased computational power.

E. Hybrid Frameworks Evaluated on BraTS Dataset

Several recent studies using the BraTS dataset combine attention modules, multi-scale feature extraction, and specialized loss functions such as Soft Dice Loss and Focal Loss to improve tumor segmentation performance. These frameworks are designed to reduce class imbalance effects and improve segmentation precision across tumor subregions. Experimental results demonstrate better tumor mask generation than conventional segmentation architectures, although separating highly overlapping tumor components remains a continuing challenge.

III. PROBLEM STATEMENT

Accurate segmentation of Brain Tumor from Magnetic Resonance Imaging images is difficult because tumor regions vary in size, shape, and intensity across different MRI modalities such as T1, T1c, T2, and FLAIR. Conventional segmentation methods often fail to identify precise tumor boundaries when tumor tissues overlap with normal brain structures.

Although U-Net based deep learning models improve automation, they still face limitations in capturing multi-scale features and focusing accurately on complex tumor regions. Therefore, an improved segmentation model is required to enhance tumor boundary detection and produce reliable segmentation masks for clinical support.

IV. EXISTING SYSTEM

Existing brain tumor segmentation systems primarily rely on traditional image processing techniques and standard deep learning architectures such as U-Net for identifying tumor regions from Magnetic Resonance Imaging scans. In conventional approaches, preprocessing steps such as resizing, normalization, and noise removal are first applied to MRI images before feature extraction and segmentation are performed. U-Net based models use encoder and decoder layers with skip connections to generate segmentation masks by preserving spatial information from different feature levels.

Although these methods provide good segmentation performance in many cases, they still face important limitations when tumor structures are highly irregular or when MRI modalities show large intensity variations. Standard segmentation networks may not effectively capture both fine details and large tumor regions at the same time. In addition, small tumor boundaries and overlapping tissue patterns often lead to incomplete segmentation results. These limitations reduce segmentation reliability, especially in complex clinical MRI cases.

V. PROPOSED METHODOLOGY

A. Data Acquisition

The proposed segmentation system starts by obtaining multimodal Magnetic Resonance Imaging brain images from the BraTS dataset. The selected modalities include T1, T1c, T2, and FLAIR, because each modality contributes different diagnostic information related to tumor structure, edema, and surrounding tissues. Combining these modalities improves tumor visibility and supports more reliable segmentation.

B. Preprocessing of MRI Images

Before training, all MRI images are prepared through preprocessing operations to ensure consistent model input. Image resizing is performed to convert all scans into a fixed input dimension suitable for network processing. Normalization is then applied to standardize pixel intensity values, which helps reduce variation between MRI samples and improves training stability.

C. Deep Feature Encoding

The processed MRI images are passed into the encoder section of the Attention-Enhanced Inception U-Net network. In this stage, convolution operations progressively learn important visual features from the input images. Low-level and high-level image characteristics related to tumor appearance are extracted for further analysis.

D. Inception-Based Multi-Scale Learning

To improve feature extraction, inception blocks are incorporated into the segmentation architecture. Multiple convolution filters with different kernel sizes operate simultaneously, allowing the network to detect both fine tumor boundaries and broader abnormal tissue regions. This improves performance when tumor shapes are highly irregular.

E. Attention Mechanism for Tumor Localization

Attention modules are included to strengthen the focus of the network on important tumor regions. These modules help the model assign higher importance to relevant abnormal tissues while reducing the influence of non-tumor background areas. This improves boundary precision during segmentation.

F. Mask Generation and Optimization

In the final stage, decoder layers reconstruct the extracted features into a segmentation mask representing tumor regions. Soft Dice Loss and Focal Loss are used during training to improve segmentation quality and manage imbalance between tumor and background pixels. The final output provides a clear segmented tumor region for clinical interpretation.

VI. FLOWCHART

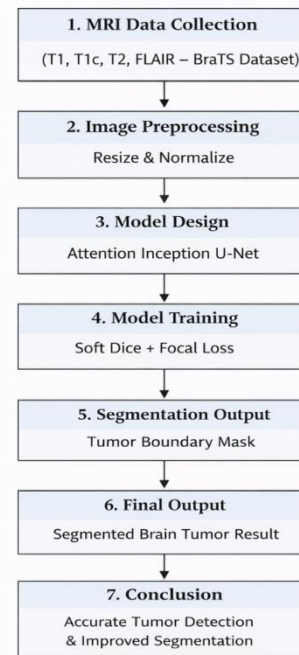
The flowchart of the proposed system begins with collecting multimodal Magnetic Resonance Imaging data, including T1, T1c, T2, and FLAIR images from the BraTS dataset. These multiple MRI modalities are used because each scan highlights different tissue characteristics, which helps improve tumor detection accuracy.

In the next stage, preprocessing is applied to the input MRI images. This includes resizing the images to a fixed dimension and normalizing pixel intensity values so that all input samples become consistent for model training. Proper preprocessing improves feature learning and reduces unwanted variation across MRI scans.

After preprocessing, the images enter the encoder section of the Attention-Enhanced Inception U-Net model, where inception blocks extract features at multiple scales. Different convolution filters help identify both small tumor boundaries and larger abnormal tissue regions.

The extracted features then pass through the attention mechanism, which allows the network to focus more strongly on important tumor regions while reducing

background interference. This improves localization of difficult tumor boundaries.



The extracted features then pass through the attention mechanism, which allows the network to focus more strongly on important tumor regions while reducing background interference. This improves localization of difficult tumor boundaries.

Finally, decoder layers reconstruct the learned features and generate the segmentation mask. The final output clearly highlights tumor regions, which can support medical diagnosis, treatment planning, and future clinical decision systems.

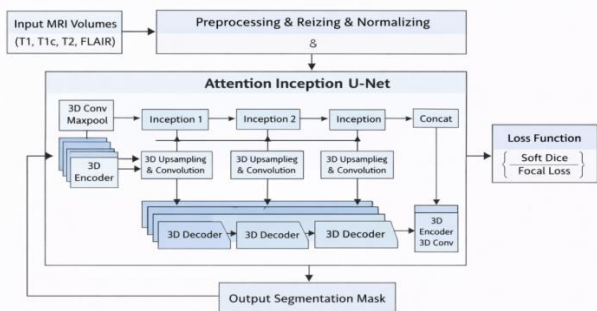
VII. ARCHITECTURE DIAGRAM

The proposed framework for brain tumor segmentation utilizes a simplified Attention-Enhanced Inception U-Net architecture to effectively learn discriminative multi-scale features from multimodal MRI data. The network receives four MRI modalities including T1, T1c, T2, and FLAIR as volumetric inputs. These inputs are processed through multiple encoder stages that progressively extract hierarchical feature representations. Each encoding layer performs convolutional operations followed by spatial downsampling, allowing the model to capture both local texture information and global contextual patterns related to tumor structures.

At the deepest level of the network, inception modules are employed to perform parallel convolutional operations with different receptive fields, enabling enhanced multi-scale feature extraction. Attention mechanisms are incorporated to guide the network in focusing on salient tumor regions while reducing the influence of non-tumorous background features. In the decoding path, feature maps are gradually upsampled to recover spatial resolution. Skip connections are used to merge encoder

and decoder features at corresponding levels, which helps preserve fine boundary details and improves segmentation precision.

The final layer generates a dense pixel-level tumor segmentation mask that delineates tumor boundaries accurately. To optimize model learning and handle class imbalance, a hybrid loss function combining Soft Dice Loss and Focal Loss is utilized during training. Overall, the proposed architecture improves segmentation robustness by jointly leveraging contextual information and detailed spatial features.



VIII. RESULTS AND DISCUSSION

The effectiveness of the proposed Attention-Enhanced Inception U-Net architecture was assessed using multimodal MRI data obtained from the BraTS benchmark dataset. The model demonstrated the ability to accurately segment tumor regions by learning discriminative features across multiple spatial scales. Experimental findings indicate that the network successfully generates detailed pixel-level segmentation maps that closely correspond to ground truth tumor annotations.

Performance analysis was conducted using widely accepted segmentation metrics including Dice Similarity Coefficient, Intersection over Union, and overall accuracy. The incorporation of attention mechanisms and inception-based feature extraction contributed to noticeable improvements in segmentation consistency and boundary precision. Furthermore, the use of a combined Soft Dice and Focal loss function facilitated stable model convergence and improved detection of comparatively smaller tumor regions.

Visual inspection of the predicted segmentation outputs revealed that the model effectively captures irregular tumor morphologies and intensity variations present in different MRI modalities. The encoder-decoder skip connections played a significant role in maintaining spatial continuity, thereby enhancing boundary refinement. In addition, attention modules supported selective feature enhancement, which reduced false positives in non-tumorous areas.

Despite achieving strong segmentation performance, slight prediction deviations were observed in challenging cases involving low contrast and diffuse tumor structures. These observations suggest that future improvements may

involve incorporating more advanced contextual learning strategies or expanding the training dataset. Overall, the proposed framework demonstrates reliable performance and shows promise for automated brain tumor segmentation tasks.

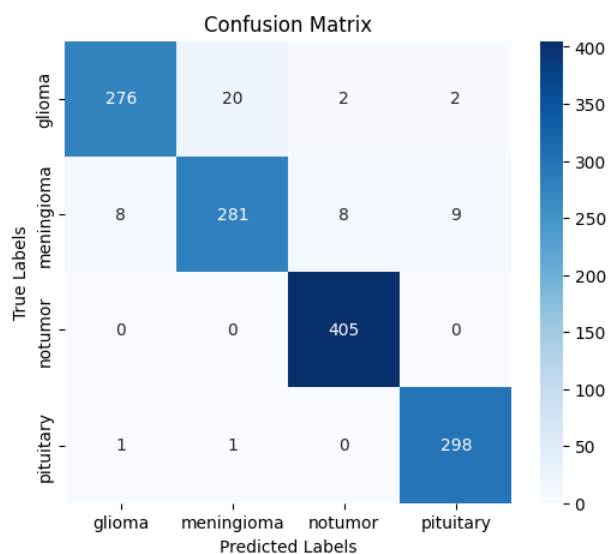
IX. CONCLUSION

In this work, an Attention-Enhanced Inception U-Net architecture was proposed for accurate brain tumor segmentation using multimodal MRI images. The integration of inception modules enabled effective multi-scale feature extraction, while attention mechanisms improved the network's ability to focus on relevant tumor regions. The encoder-decoder structure with skip connections helped preserve spatial information and enhanced boundary localization.

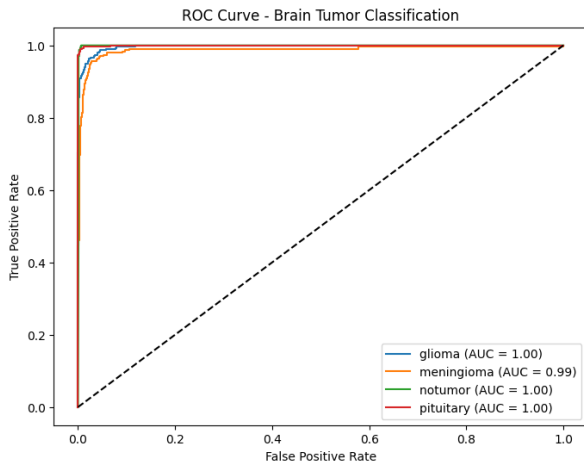
Experimental evaluation demonstrated that the proposed framework achieved reliable segmentation performance in terms of Dice Similarity Coefficient, Intersection over Union, and accuracy. The use of a hybrid loss function combining Soft Dice Loss and Focal Loss contributed to better handling of class imbalance and improved detection of small and irregular tumor regions. Both quantitative and qualitative results confirmed the robustness of the model in segmenting complex tumor structures across different MRI modalities.

Although the model produced promising results, certain limitations were observed in cases involving low-contrast tumor regions and highly heterogeneous tumor patterns. Future work can focus on incorporating advanced attention strategies, larger training datasets, and computational optimization techniques to further enhance segmentation accuracy and efficiency. Overall, the proposed approach demonstrates strong potential for supporting automated and reliable brain tumor diagnosis in clinical decision-making systems.

Confusion Matrix



Roc Curve



REFERENCES

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (MICCAI)*, 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[2] Oktay, O., Schlemper, J., Folgoc, L. L., et al. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*. <https://arxiv.org/abs/1804.03999>

[3] Szegedy, C., Liu, W., Jia, Y., et al. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>

[4] Bakas, S., Reyes, M., Jakab, A., et al. (2017). Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 4, 170117. <https://doi.org/10.1038/sdata.2017.117>

[5] Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proceedings of 3DV*, 565–571. <https://doi.org/10.1109/3DV.2016.79>

[6] Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*, 2980–2988. <https://doi.org/10.1109/ICCV.2017.324>

[7] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis*, 3–11. https://doi.org/10.1007/978-3-030-00889-5_1

[8] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2), 203–211. <https://doi.org/10.1038/s41592-020-01008-z>

[9] Abdusalomov, A. B., et al. (2023). Brain tumor detection based on deep learning approaches.

Sensors, 23(16), 7160. <https://doi.org/10.3390/s23167160>

[10] Liu, Z., et al. (2023). Deep learning-based brain tumor segmentation: A survey. *Complex & Intelligent Systems*, 9, 1–22. <https://doi.org/10.1007/s40747-022-00815-5>

[11] Zhang, Y., et al. (2023). Transformer-based multimodal brain tumor segmentation. *Medical Image Analysis*, 87, 102802. <https://doi.org/10.1016/j.media.2023.102802>

[12] Abidin, Z. U., et al. (2024). Recent deep learning-based multimodal brain tumor segmentation: A comprehensive survey. *Frontiers in Bioengineering and Biotechnology*, 12. <https://doi.org/10.3389/fbioe.2024.1392807>

[13] Yuan, J., et al. (2024). Hybrid attention-based Mask R-CNN for brain tumor segmentation. *Scientific Reports*, 14, 71250. <https://doi.org/10.1038/s41598-024-71250-4>

[14] Cao, Y., et al. (2024). A novel deep learning framework for multimodal brain tumor segmentation. *Applied Sciences*, 14(11), 4919. <https://doi.org/10.3390/app14114919>

[15] Correia de Verdier, M., et al. (2024). The BraTS 2024 challenge: Glioma segmentation benchmark. *arXiv preprint*. <https://arxiv.org/abs/2405.18368>