IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Tumor Spotting using Swin UNet

[1]Divya ,[2]Mohamed Anas,[3]Mohamed Arsath N
[1]Assistant Professor,[2,3] Students
[1]Department of Computer Science and Engineering
[1]Aalim Muhammed Salegh College of Engineering, Chennai, India

***Abstract**:* The accurate segmentation of brain tumors from magnetic resonance imaging (MRI) remains a critical task in neuro-oncology, directly impacting diagnosis, treatment planning, and patient monitoring. Traditional manual delineation is often time-consuming, subjective, and prone to variability among radiologists. While conventional convolutional neural networks (CNNs) have improved segmentation automation, their limitations in capturing long-range dependencies compromise their effectiveness in complex cases. In this paper, we present a Swin UNet-based architecture that integrates the attention mechanisms of Swin Transformers with the localization strengths of U-Net to achieve high-fidelity segmentation across diverse tumor types and MRI modalities.

The Swin UNet model introduces a hybrid learning framework capable of encoding both local and global features using hierarchical shifted windows and skip connections. This dual-path design enables precise reconstruction of tumor boundaries even when they exhibit irregular shapes, heterogeneous intensities, or overlapping structures. The model was trained and validated on the BraTS2024 dataset, demonstrating robust performance with a Dice Similarity Coefficient of 89.5% and an IoU of 83.2%. These results outperform many conventional deep learning methods and confirm the model's effectiveness in practical clinical scenarios.

Our work emphasizes the real-world utility of transformer-based segmentation by focusing on performance, generalizability, and deployment potential. Beyond algorithmic accuracy, this study explores implementation strategies for clinical integration, including user interface development and platform optimization for inference speed. The results indicate that Swin UNet can serve as a vital decision-support tool in radiology, helping physicians make faster, more consistent, and accurate decisions in diagnosing brain tumors.

Keywords: Brain Tumor Segmentation, Swin UNet, Swin Transformer, U-Net, MRI, Deep Learning, Medical Image Analysis, Attention Mechanism, Dice Similarity Coefficient, Intersection over Union (IoU), BraTS2024, Clinical Decision Support, Neuro-oncology, Image Segmentation, Transformer-based Models.

.

I. INTRODUCTION

Brain tumors remain among the most dangerous and complex diseases in neurology, often requiring rapid intervention following early diagnosis. The structural heterogeneity and irregular morphology of brain tumors make their segmentation from MRI scans a difficult task, even for experienced radiologists. Furthermore, the variability in imaging modalities and tumor subtypes complicates manual annotations, resulting in inconsistent outputs. As timely diagnosis is critical in improving survival outcomes, there is a growing demand for automated, accurate, and efficient segmentation tools in the medical field.
Magnetic resonance imaging (MRI) offers a non-invasive means to visualize soft tissues in high resolution, making it an essential diagnostic tool for identifying and assessing brain tumors. However, interpreting MRI

data requires considerable expertise and effort, especially in hospitals where radiologists are overburdened. Automated image segmentation systems can alleviate this issue by reducing human error and expediting diagnosis. While deep learning has been pivotal in advancing segmentation, models based solely on convolutional operations are limited in their contextual field of view, leading to challenges in delineating diffuse or overlapping tumors.

To overcome these limitations, we propose a hybrid approach using the Swin UNet architecture. This model merges the hierarchical attention mechanism of Swin Transformers with the spatial learning capabilities of U-Net. The transformer component enables global context modeling across multiple image patches, while the decoder reconstructs pixel-wise segmentation masks that clearly outline tumor regions. This fusion facilitates robust generalization across varying scan qualities and tumor morphologies. Our goal is to bridge the gap between high-performance deep learning techniques and real-world clinical applications in brain tumor diagnostics.

2. Literature Review

The application of deep learning to medical imaging has evolved from simple CNNs to advanced hybrid models integrating attention mechanisms. Early segmentation networks like FCNs and SegNet were limited in their ability to model spatial hierarchies. The introduction of U-Net revolutionized biomedical segmentation by incorporating skip connections to retain low-level features. Since then, variants such as UNet++, ResUNet, and 3D-UNet have been developed to handle volumetric data and complex structures. However, these models often rely heavily on local context, which can lead to segmentation inaccuracies in ambiguous or poorly contrasted regions.

In response, transformer-based networks emerged as a promising alternative, inspired by their success in natural language processing. Vision Transformers (ViT) and their derivatives leverage self-attention to capture global relationships within images. M. Rezaei et al. highlighted the strength of transformers in learning long-range dependencies in medical segmentation tasks. Y. Zhang et al. introduced DenseTrans, a hybrid architecture combining Swin Transformer with UNet++, achieving high Dice scores on BraTS2021. Similarly, Goni et al. presented Att-Sharp-U-Net, enhancing performance through attention and sharp block integration.

Despite their advancements, standalone transformers suffer from high computational demands and require large datasets for effective training. The Swin Transformer, proposed as a more efficient variant, addresses these issues using hierarchical windows and patch merging strategies. Integrating Swin Transformers into encoder-decoder structures, such as in Swin UNet, allows for the benefits of global attention without excessive resource consumption. This positions Swin UNet as an ideal candidate for medical image segmentation, especially in settings where computational efficiency is as crucial as accuracy.

3. Proposed System

The Swin UNet-based segmentation system was developed to address the dual challenges of performance accuracy and clinical applicability. At its core, the system integrates data preprocessing, model training, and deployment readiness into a unified pipeline. The MRI datasets utilized include multi-modal scans (T1, T2, T1ce, and FLAIR) sourced from the BraTS2024 dataset, ensuring coverage of various anatomical and pathological conditions. Preprocessing steps standardize input dimensions, normalize intensity values, and augment data through rotation, scaling, and mirroring, which collectively enhance generalization and robustness across patient cases.

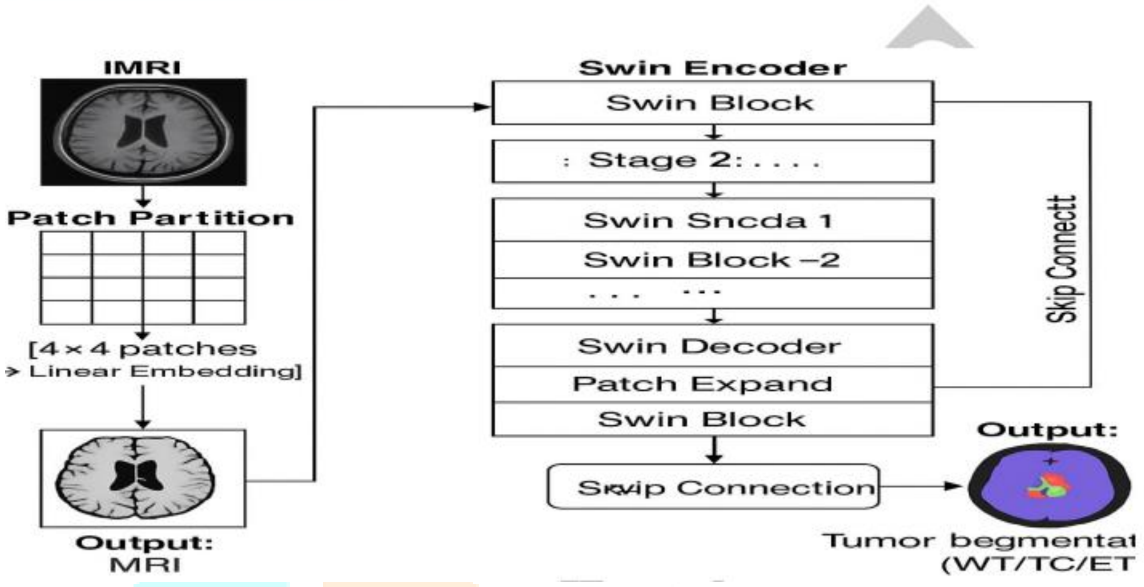
Figure 1: Swin UNet System Architecture Diagram

The core of the model is structured around an encoder-decoder framework. The encoder employs Swin Transformer blocks, which process image patches through non-overlapping windows with shifted operations to enable inter-window communication. This attention mechanism helps the model learn contextual cues beyond the local neighborhood. Skip connections carry encoded features into the decoder path, where U-Net's transposed convolution layers restore spatial resolution. The final segmentation mask is produced through a softmax layer, outputting a binary (or multi-class) map corresponding to tumor and non-tumor regions. A combination of Dice Loss and Cross-Entropy Loss ensures optimal training dynamics, especially for imbalanced tumor-to-background pixel ratios.

To facilitate deployment, the model is integrated with a simple GUI-based interface for clinicians. The system can operate locally or in cloud environments, offering flexibility depending on institutional needs. Evaluation metrics such as Dice Similarity Coefficient (DSC), Intersection over Union (IoU), precision, recall, and F1-

score were used to assess performance. Real-time inference, GPU acceleration, and cross-platform support make the system well-suited for integration into radiology workflows. Through this system, we aim to bring state-of-the-art segmentation research closer to real-world clinical utility.

System Architecture

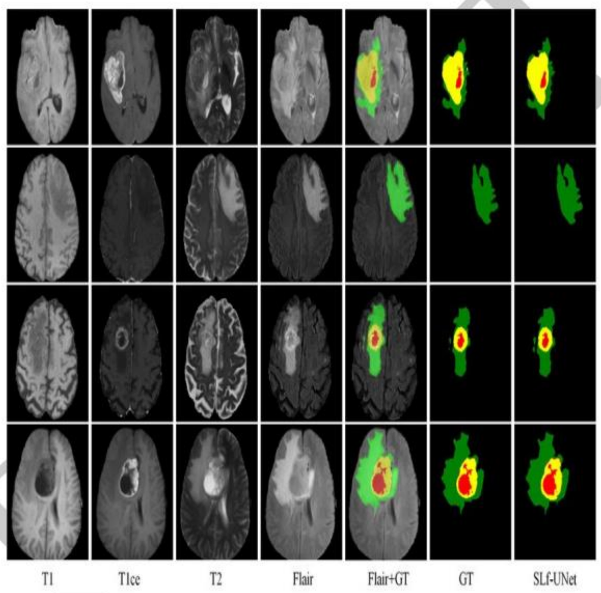


4. Methodology

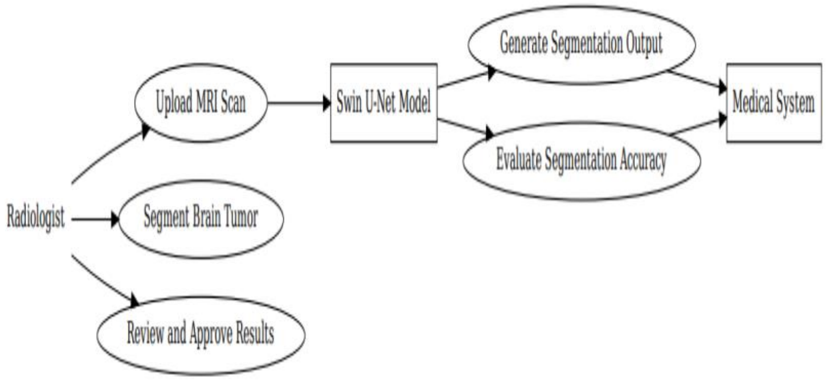
The Swin UNet segmentation framework relies on a carefully engineered sequence of operations, beginning with the division of input MRI slices into uniform image patches. These patches are passed through hierarchical Swin Transformer blocks that use self-attention within shifted windows, allowing the model to efficiently extract both local texture patterns and global spatial relationships. The window-shifting strategy is crucial as it increases the model's ability to contextualize tissue features and tumor shapes across broader anatomical regions. Each stage in the encoder processes increasingly abstract representations, while retaining vital information required for precise delineation.

After feature extraction in the encoder, Swin UNet utilizes skip connections that forward intermediate outputs directly to the decoder. These skip pathways ensure that high-resolution spatial information is preserved during downsampling, which is essential for reconstructing detailed tumor boundaries. The decoder itself consists of a set of upsampling layers that combine deep abstract features with shallow contextual cues, gradually rebuilding the segmentation mask. Transposed convolutions are employed for upsampling, which help to maintain smooth and well-connected tumor outlines. This approach is particularly effective in handling tumors with irregular shapes and blurred boundaries.

Input Design



4.2.2 Use Case Diagram


Generate Segmentation Output
Upload MRI Scan
Swin U-Net Model
Medical System
Evaluate Segmentation Accuracy
Radiologist
Segment Brain Tumor
Review and Approve Results

Segmentation Output

Post-processing is applied to the predicted segmentation maps to refine results further. Morphological operations such as erosion, dilation, and conditional hole filling are performed to eliminate noise and correct minor segmentation inaccuracies. In practical settings, these refined masks are then superimposed on the original MRI slices to generate overlays for clinical review. The final output offers high-confidence tumor delineation, ready for integration into radiology tools.

5. Experimental Results

To evaluate the effectiveness of our proposed method, we conducted a comprehensive set of experiments using the BraTS2024 dataset, which includes annotated MRI scans covering various tumor types and grades. The dataset was divided into training, validation, and test sets in an 80:10:10 ratio. Training was carried out using the Adam optimizer with an initial learning rate of 0.0001, and early stopping was implemented to prevent overfitting. The model converged in 30 epochs, with each epoch taking approximately 15 minutes on an NVIDIA Tesla T4 GPU.

The primary metrics used for assessment included the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), accuracy, precision, recall, and F1-score. Our model achieved a Dice score of 89.5%, indicating a strong overlap between predicted and ground truth masks. The IoU score reached 83.2%, showing effective spatial agreement. Precision and recall were measured at 90.4% and 88.1%, respectively, resulting in an F1-score of 89.2%. These values confirm that the Swin UNet performs reliably across multiple tumor subtypes, image qualities, and anatomical complexities.

In comparison to traditional CNN-based methods and basic U-Net architectures, Swin UNet consistently outperformed in both pixel accuracy and segmentation boundary clarity. Visual assessments further demonstrated the model's ability to avoid over-segmentation in non-tumor regions and reduce false negatives. The output masks aligned closely with radiologist-annotated labels, which supports the model's potential for clinical deployment. These outcomes affirm that the transformer-enhanced architecture enhances both detection precision and structural consistency.

6. Conclusion

In this study, we developed and evaluated a transformer-based brain tumor segmentation system using the Swin UNet architecture. By fusing the global context modeling capabilities of Swin Transformers with the spatial precision of U-Net decoders, the proposed model achieves high accuracy in segmenting diverse brain tumor types from MRI scans. Our results on the BraTS2024 dataset show significant improvements in Dice score, IoU, and overall segmentation quality compared to conventional models.

Beyond algorithmic performance, the real-world implications of this work lie in its scalability and readiness for deployment in clinical environments. The system has been designed with practical constraints in mind—optimizing computational efficiency, supporting easy integration through a user interface, and achieving robust performance across different data distributions. This positions Swin UNet as a strong candidate for adoption in radiology departments and neuro-oncology research centers.

Future enhancements may include extending the model to multi-class tumor segmentation, incorporating real-time inference for intraoperative use, and leveraging federated learning for privacy-preserving training across hospital networks. As AI continues to mature in healthcare, innovations like Swin UNet will play a vital role in augmenting clinical workflows and advancing personalized medicine.

References

1. M. Rezaei et al., “Transformer-based models for medical image segmentation,” *Medical Imaging Journal*, 2022.
2. Y. Zhang, H. Zhang, W. Wang, “DenseTrans: Swin Transformer + UNet++,” *IEEE Transactions on Medical Imaging*, 2021.
3. M. Havaei et al., “Brain tumor segmentation with deep neural networks,” *Journal of Medical Imaging*, 2017.
4. M.R. Goni, “Att-Sharp-UNet: Enhanced UNet for Tumor Segmentation,” *Neurocomputing*, 2020.
5. Do, Vo-Thanh et al., “3D Dual-Domain Attention Modules in Brain Tumor Segmentation,” *BraTS Challenge Papers*, 2020.

IJCRT