Hybrid V-Net and Swin Transformer–Based Deep Learning Model for Brain Tumor Segmentation in Low-Quality MRI

Fajar Astuti Hermawati[1], Andre Pramudya[2]
[1,2] Departement of Informatics, Universitas 17 Agustus 1945 Surabaya, Surabaya, Indonesia
[1] fajarastuti@untag-sby.ac.id, [2]andrepramudya3@gmail.com

Accepted 04 January 2026
Approved 23 January 2026

***Abstract*— Brain tumor segmentation from low-quality magnetic resonance imaging (MRI) remains a challenging task due to noise, resolution variation, and low contrast between tumor and healthy tissues. Improving segmentation accuracy is essential to support more precise diagnosis and treatment planning. This study proposes a hybrid deep learning model that integrates V-Net and Swin Transformer–based architecture (Swin UNETR) for automatic brain tumor segmentation in multimodal MRI images. The MICCAI BraTS 2020 dataset was used, consisting of T1, T1c, T2, and FLAIR sequences with corresponding segmentation labels. The preprocessing pipeline includes resampling, skull stripping, intensity normalization, and data augmentation. V-Net is employed to extract local spatial features from 3D volumetric data, while the Swin UNETR captures global spatial relationships through a self-attention mechanism. Postprocessing procedures such as thresholding, morphological refinement, and false-positive removal are applied to enhance segmentation quality. The proposed hybrid model achieves Dice scores of 0.8635 for Whole Tumor (WT), 0.7179 for Tumor Core (TC), and 0.8073 for Enhancing Tumor (ET), with additional evaluation using precision, recall, and IoU further confirming its effectiveness. These results highlight the model's potential to improve automated brain tumor segmentation in low-quality MRI images and support its applicability as an efficient AI-assisted diagnostic tool in clinical practice.**

***Index Terms*— Brain Neoplasms; MRI; Deep Learning; Segmentation; V-Net, Transformer.**

I. Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive medical imaging technology that is crucial for detecting and diagnosing various diseases, particularly brain tumors. MRI's advantage lies in its ability to produce high-resolution images with good contrast against soft brain tissue. Imaging modalities such as T1-weighted, T2-weighted, and FLAIR can provide comprehensive information about brain structure [1]. The multimodal MRI approach has proven effective in improving diagnostic accuracy because each modality provides distinct information about the structure and morphology of brain tissue [2].

However, segmenting brain tumors from MRI images is a significant challenge. This is due to the complexity of tumor shape and size, irregular boundaries, differences in intensity between tissues, and the presence of noise and imaging artifacts [3]. Therefore, automated methods based on artificial intelligence, particularly deep learning, are needed to improve segmentation efficiency and accuracy.

Recent advancements in brain tumor segmentation and classification from MRI scans highlight the shift toward sophisticated deep learning and hybrid models. Early methods, like the one proposed by [4], utilized a classical approach combining a Modified Region Growing (MRG) algorithm for segmentation with Adaptive Support Vector Machine (ASVM) and Grasshopper Optimization Algorithm (GOA) feature selection to manage computational complexity. However, the field has rapidly moved toward Convolutional Neural Networks (CNNs) and Transformers. Key developments include 3D U-Net models for accurate volumetric segmentation [5], and advanced U-Net variants like the Trans U-Net [6] and UNETR [2], which leverage the Transformer's self-attention mechanism to capture long-range spatial dependencies. Further innovation includes hybrid approaches such as the 3D U-Net with Contextual Transformer and Double Attention [1], multi-pathway 3D FCNs for multimodal data fusion [7], and the introduction of computational efficiency techniques like QuantSR [8] for high-resolution medical imaging. These studies collectively demonstrate a trend of integrating advanced architectures and multi-modal data processing to achieve superior segmentation and classification accuracy for clinical application.

One deep learning architecture that has proven effective is the U-Net, which uses a symmetric encoder-decoder approach with skip connections. The U-Net performs well in medical image segmentation, but is less than optimal when handling images with high noise [9]. On the other hand, Swin UNETR is capable of capturing global and local relationships in medical images, but requires significant computational resources [10]. Other approaches such as 3D U-Net and Modified Region Growing (MRG) have also been explored. 3D U-Net can process volumetric images, but still faces challenges when intensity is non-uniform [11]. V-Net, which uses a 3D convolutional neural network, is specifically designed for volumetric data such as MRI. V-Net is effective in understanding spatial context between layers, but has limitations in comprehensively capturing global features [12].

Combining V-Net with Swin UNETR in a hybrid approach is expected to overcome the weaknesses of each method. This combination allows for the integration of the local strengths of V-Net and the global strengths of Swin UNETR, thereby improving the accuracy of brain tumor segmentation in low-quality MRI images. This system uses multimodal image input (T1, T1c, T2, FLAIR) from the BraTS dataset that has undergone preprocessing stages, including noise removal, intensity normalization, and data augmentation. The system outputs a label map (mask) that clearly shows the brain tumor area. The segmentation results were then compared with ground truth to evaluate performance. With this approach, the research is expected to significantly contribute to the development of more accurate and efficient automated segmentation tools, accelerate medical diagnosis, and enrich the academic literature in the field of deep learning-based medical image segmentation.

II. METHOD

A. Data

The dataset used in this study is the MICCAI Brain Tumor Segmentation Challenge (BraTS) 2020 dataset, which provides multimodal MRI scans and expert-annotated ground truth labels for brain tumor segmentation [13]. The data consists of four main imaging modalities: T1, T1 with contrast (T1c), T2, and FLAIR, as shown in Fig.1. Segmentation labels are provided in three categories: Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET).

The dataset described in Table I consists of two main parts: Training Data, used to train the segmentation model, and Validation Data, used to evaluate the model's performance. The dataset can be used for the development of deep learning-based segmentation methods because it provides tumor segmentation labels that include Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET).

TABLE I. DATASET SPECIFICATIONS

Specifications	Description
Amount of MRI Images	369 total (295 for training and 74 for validation)
Amount of Image slices	2,349 total images (1,847 for training and 502 for validation)
Resolution	240×240×155 voxel.
Modalities	T1, T1c, T2, FLAIR.
Color Depth	16-bit per channel.
Format	NIfTI (.nii).

Computations were performed using a laptop with the following specifications: an Intel Core i5-13000 processor, an NVIDIA RTX4050 GPU, 24 GB of RAM, and Windows OS. Programming was performed using Python via the Google Colab and Jupyter Notebook platforms, with libraries such as PyTorch, MONAI, and Scikit-image for medical image processing and deep learning model implementation.

B. The Proposed Methods

This research was conducted through several main stages systematically arranged to ensure optimal segmentation results, as shown in Fig. 2. These stages include data preprocessing, segmentation using a hybrid model, post-processing to refine the results, and testing scenarios to evaluate model performance. Each stage is interconnected and designed to address common issues encountered in segmenting low-quality MRI images, such as noise, intensity variations, and low contrast between tissues.

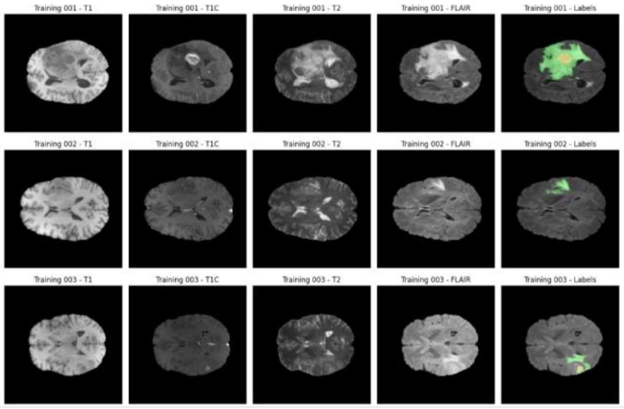

Fig. 1. Some Examples of Images from the BraTS 2020 MRI Dataset

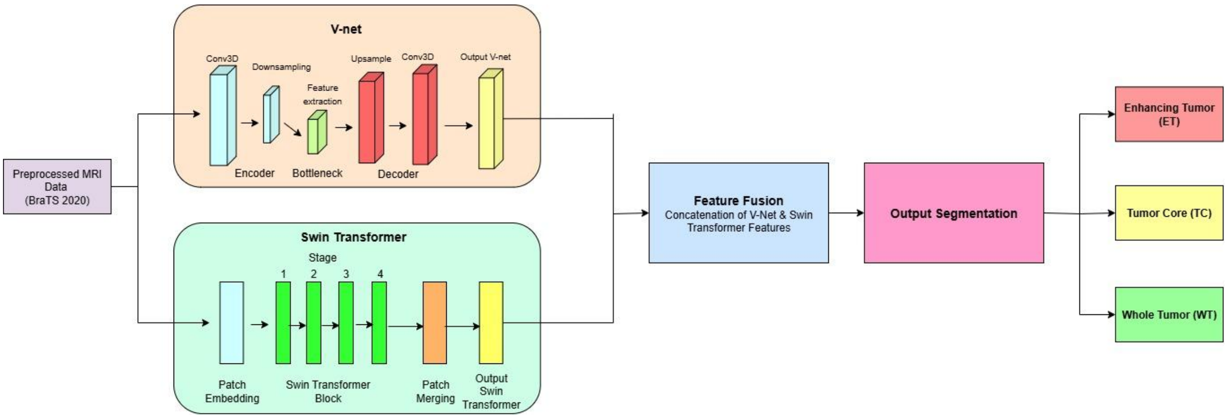

Fig. 2. Architecture Diagram of Hybrid Model of V-Net and Swin UNETR for Brain Tumor Segmentation

1). *Preprocessing*

Preprocessing is key to producing consistent and optimal input images. First, resampling is performed to standardize voxel resolution, given that the MRI data originate from different institutions. Next, skull stripping is performed using threshold-based and connected components algorithms to remove non-brain tissue [14]. The third step is intensity normalization, applying *z*-score normalization to each modality. Normalized intensity value, I_{norm}, is calculated based on (1) as follow:

$$I_{norm} = \frac{I-\mu}{\sigma} \quad (1)$$

Where *I* is an original voxel intensity value in the MRI image, μ is a mean intensity value within an MRI volume, σ is a standard deviation of intensity within an MRI volume.

The modalities are then combined into a 3D tensor consisting of four channels. The dataset is then converted to tensor format for compatibility with deep learning architectures [15]. The final step is converting the image into a 3D tensor format ready for processing by the model. Once these steps are complete, the data is ready to be fed into the Hybrid V-Net and Swin UNETR, where V-Net handles local spatial features, while Swin UNETR focuses on broader spatial relationships. With proper preprocessing, the model is expected to perform more accurately in brain tumor segmentation.

2). *Segmentation Approach*

The hybrid V-Net and Swin UNETR model was designed using a dual-path approach. V-Net, as a 3D convolutional network, focuses on local spatial features through an encoder-decoder with skip connections [12]. Swin UNETR with a hierarchical architecture based on local self-attention, is used to extract global spatial context [10]. After feature extraction, feature fusion is performed through concatenation and convolution layers to combine the representations from both models. To combine the feature outputs from V-Net (local) and Swin UNETR (global), a concatenation operation is used followed by a 3D convolution to reduce the dimensionality and fuse the features, as shown in (2).

$$F_{fusion} = \mathrm{Conv3D}([F_v||F_t] \quad (2)$$

F_v is features extracted from the V-Net pipeline, which captures local spatial information from volumetric MRI (e.g., shape, texture around the tumor). F_t is the extracted features from the Swin UNETR pipeline, which brings global context through a self-attention mechanism (long-range relations).

The training process uses *n* epochs, with a loss function based on a combination of Dice Loss and Cross Entropy Loss [16]. Equation (3) is formula of the Dice Loss as:

$$\mathcal{L}_{Dice} = 1 - Dice \quad (3)$$

where Dice is as presented in follow equation:

$$Dice = \frac{\sum_i p_i g_i + \epsilon}{\sum_i p_i + \sum_i g_i + \epsilon} \quad (4)$$

with p_i = model prediction at the *i*-th voxel (0 or 1, or probabilistic 0–1), g_i = ground truth at the *i*-th voxel (0 or 1), $\sum_i p_i g_i$= number of correctly detected voxels (intersection) and ϵ = small value to prevent division by zero.

Equation (5) is formula of Cross Entropy Loss:

$$\mathcal{L}_{CE} = -\sum_i \sum_{c=1}^{C} g_{i,c} \log(p_{i,c}) \quad (5)$$

with *C* = number of classes, $g_{i,c}$= 1 if the *i*-th voxel belongs to class *c* and 0 otherwise, $p_{i,c}$= predicted probability that the *i*-th voxel belongs to class *c*.

The final segmentation results are grouped into three sections, namely: Enhancing Tumor (ET) that is the active tissue after contrast, Tumor Core (TC) that is the

UMN

interior of the tumor without edema and Whole Tumor (WT) that is the entire tumor mass.

3). *Postprocessing*

The postprocessing stage begins with thresholding and probability masking, where a threshold value is set (usually between 0.3 and 0.7) to filter predictions based on the probabilities generated by a model, such as the Swin UNETR. Only voxels with probabilities above the threshold are considered valid, thus reducing noise and preventing minor misclassifications at the tumor edge [17]. Next, morphological refinement is performed using closing and dilation techniques to address contour roughness, fill small holes, and strengthen segmentation boundaries to align with the original anatomical structure [9]. This refinement is crucial because initial segmentation results are often discrete and not perfectly connected.

The third step is the removal of false positives, which are areas of the image incorrectly identified as tumors. This process uses Connected Component Analysis (CCA) to eliminate small predicted regions that are not spatially related to the main tumor structure, thereby increasing the model's specificity [13]. To refine the final results, smoothing using Gaussian or median filtering is applied, which is useful for smoothing segmentation edges and reducing unnatural intensity variations due to noise or unstable predictions [6]. This stage also improves the accuracy of volume measurements and facilitates 3D visualization.

As a final step, the segmentation results are converted into standard medical formats, namely NIfTI (.nii) and DICOM (.dcm). The NIfTI format is very commonly used in neuroimaging research because it is compatible with software such as FSL and SPM, while DICOM is a universal format in clinical medical practice and supports integration with hospital PACS systems [18]. This conversion makes the segmentation results ready for further analysis and clinical applications, bridging research findings with real-world applications.

4). *Testing Scenario*

The test scenario in this study was designed to evaluate the performance and reliability of a hybrid V-Net and Swin UNETR model in brain tumor segmentation based on the MICCAI BraTS 2020 dataset. The dataset includes various MRI imaging modalities such as T1, T1c, T2, and FLAIR, equipped with ground truth labels, allowing for objective evaluation of prediction accuracy. Initial testing was conducted by applying the trained model to validation data to measure the model's ability to identify and separate tumor structures from healthy brain tissue. Next, model performance was analyzed using evaluation metrics such as the Dice Score, Jaccard Index, sensitivity, and specificity. The Dice Score measures the similarity between the predicted segmentation and the reference label, while the Jaccard Index measures the degree of overlap between the two. Sensitivity assesses the model's ability to correctly detect tumors, while specificity assesses its accuracy in avoiding misclassification of healthy tissue. In addition to the Dice Score, other commonly used segmentation metrics, including Intersection over Union (IoU), Precision, Recall (Sensitivity), and Specificity, were employed to ensure broader comparability with existing brain tumor segmentation studies.

To assess the model's robustness to variations in image quality, testing was conducted on noisy or low-resolution MRI data. This testing is crucial for assessing the model's resilience under less-than-ideal imaging conditions. Furthermore, the resulting segmentations were also analyzed post-processing, using techniques such as morphological refinement and removal of false positives to ensure the final results were more accurate and freer from false predictions.

III. Result and Discussions

A. *Training Result*

The training process was carried out using a stepwise approach. Initially, the model was trained for 5 epochs to test the stability of the architecture and data pipeline. The results of this initial testing showed that the loss value was still relatively high and the segmentation performance was not optimal. The segmentation produced at this stage appeared coarse, with a very low Dice Score (WT = 0.159, TC = 0.0, ET = 0.0), and was not able to differentiate well between Whole Tumor (WT), Enhancing Tumor (ET), and Tumor Core (TC). After increasing the number of epochs from 5 to 40, there was an improvement in both the Loss and Dice score for predicting brain tumors, as seen in Fig. 3 and 4.

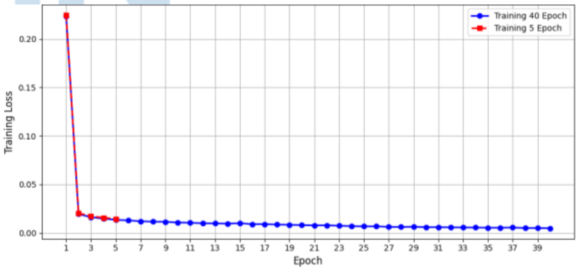

Fig. 3. Loss Comparison Graph between 5 epochs and 40 epochs.

Fig. 3 shows a comparison of the loss graphs between the model training for 5 epochs and 40 epochs. It can be seen that in the initial training with 5 epochs, the loss value decreased quite drastically but stopped before reaching stable convergence. Conversely, in the training for 40 epochs, the loss decrease was more consistent and sustained, reaching a value approaching 0.0048 at the end of the training. This graph shows that increasing the number of epochs provides a longer learning period for the model, allowing it to better adjust its weights and resulting in more accurate and stable segmentation performance.

B. Segmentation Result

Segmentation visualization for a representative MRI slice is shown in Fig. 4. For this particular sample, the model achieved Dice Scores of 0.8932 for Whole Tumor (WT), 0.9327 for Tumor Core (TC), and 0.8304 for Enhancing Tumor (ET). These values represent the segmentation quality on a single example image and are intended to illustrate the model's behavior visually. Table II summarizes the quantitative performance of the proposed model across multiple evaluation metrics, including Dice Similarity, Precision, and Recall, aggregated over the entire validation set.

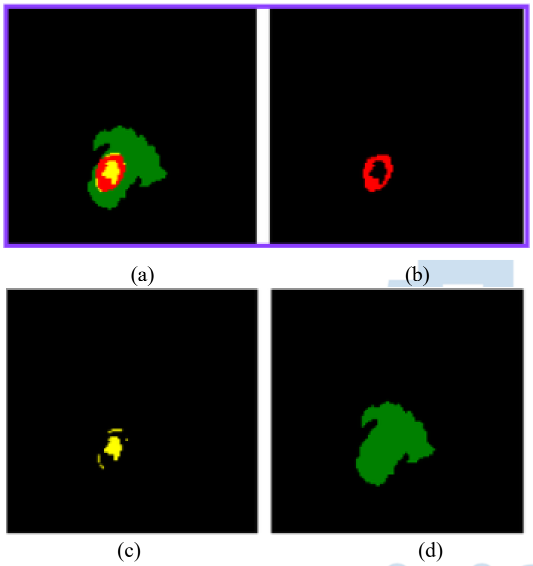

Fig. 4. Prediction results example for 40 epochs: (a) Prediction of the entire tumor, (b) Enhancing tumor, (c) Tumor core, (d) Whole tumor.

TABLE II. SEGMENTATION PERFORMANCE EVALUATION RESULTS FOR EACH BRAIN TUMOR SUBREGION MAP

Brain Tumor Subregions	Dice Similarity	Precision	Recall
Tumor Core (TC)	0.7179	0.8346	0.6675
Whole Tumor (WT)	0.8635	0.8622	0.8677
Enhancing Tumor (ET)	0.8073	0.7904	0.8392

Based on the test results, Whole Tumor (WT) achieved the highest scores across almost all evaluation metrics, with a Dice Score of 0.8635, Precision 0.8622, and Recall 0.8677. This indicates that the model is capable of identifying the entire tumor area with good accuracy and sensitivity.

Meanwhile, Enhancing Tumor (ET) also demonstrated quite solid performance with a Dice Score of 0.8073, indicating the model's ability to detect active tumor regions or those experiencing contrast enhancement following contrast agent administration in MRI. However, the relatively small variation in shape and size of ET compared to WT makes it more difficult to fully segment.

For the Tumor Core (TC), the Dice Score of 0.7179 indicates that the model still faces challenges in precisely detecting the tumor core. This may be due to the similarity in intensity between the TC and the surrounding tissue, as well as the more limited distribution of TC data compared to WT.

Overall, this evaluation results indicate that the hybrid V-Net and Swin UNETR approach is capable of providing competitive segmentation performance on low-quality MRI images. However, accuracy improvements, particularly for the TC segment, can still be achieved through strategies such as adding various data augmentations, adjusting the loss function (e.g., a combination of Dice Loss and Focal Loss), and implementing more adaptive post-processing techniques to reduce segmentation errors in small areas.

C. Qualitative Evaluation Results

Visual evaluation was performed by displaying axial MRI image slices along with predicted segmentation results and ground truth labels. This visualization demonstrates that the model is able to map tumor areas with relatively accurate shapes, although there are minor inaccuracies at the edges of small tumors.

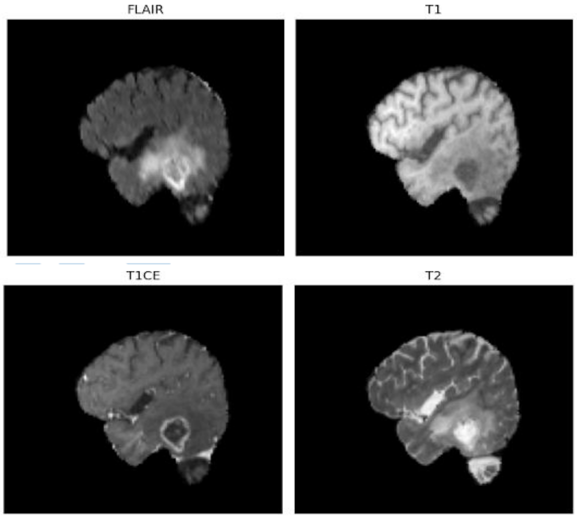

Fig. 5. MRI Modality Output: FLAIR, T1, T1CE, T2

Fig. 5 shows the four main MRI modalities that have undergone preprocessing and postprocessing, used as input for the segmentation process: FLAIR, T1, T1CE, and T2. Each modality provides different information about brain tissue structures, such as edema, active tumor contrast, and anatomical brain boundaries. The combination of these four modalities is crucial in providing a complete representation of various types of brain tumor tissue.

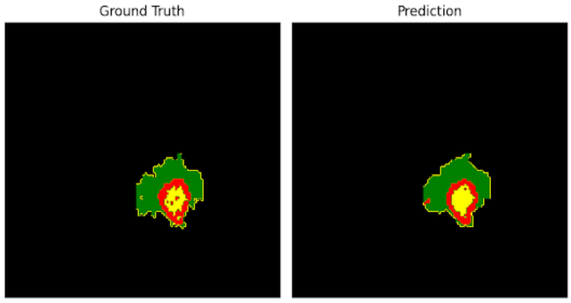

Fig. 6. Ground Truth and Segmentation

Fig. 6 displays a comparison between the segmentation results generated by the model and the ground truth labels. It can be seen that the model's predictions successfully follow the tumor shape and area quite accurately. Despite slight differences in tumor edges, the model was generally able to identify relevant tumor locations and shapes, including their internal structures, such as Tumor Core (TC) and Whole Tumor (WT).

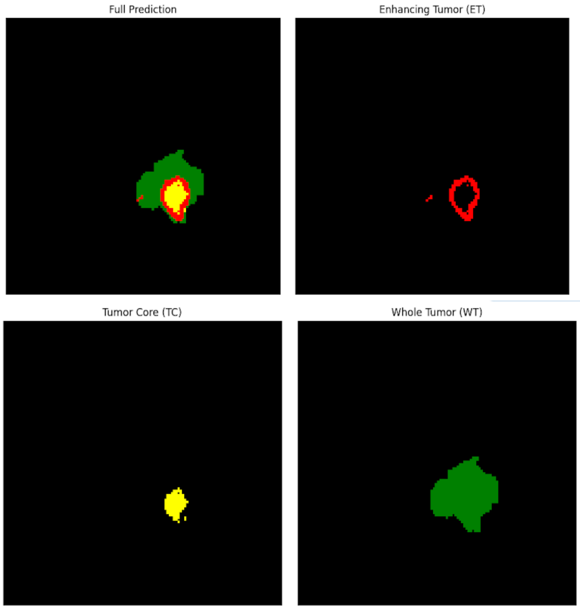

Fig. 7. Tumor Class Mask: ET (red), TC (yellow), WT (green)

Fig. 7 clarifies the classification of tumor classes predicted by the model. Red indicates the Enhancing Tumor (ET) region, yellow represents the Tumor Core (TC), and green indicates the Whole Tumor (WT). This mask helps assess how well the model can spatially distinguish and characterize each tumor subregion and highlights the model's ability to detect complex tumor structures with precise segmentation.

D. 3D Visualization

As part of the qualitative evaluation, a three-dimensional visualization of the brain tumor segmentation results was performed using the *Plotly* library. The purpose of this visualization was to provide a comprehensive understanding of the spatial structure of the tumor predicted by the *FusionModel* model, while also more intuitively evaluating the accuracy, integrity, and distribution of each tumor component. The visualization was performed by mapping each voxel classified as tumor into 3D space based on the (*x*, *y*, *z*) coordinates of the *pred_mask*, which is the final segmentation prediction result. Each voxel is displayed as a point scatter in 3D space and assigned a different color to distinguish the tumor components: yellow for Tumor Core, green for Whole Tumor, and red for Enhancing Tumor.

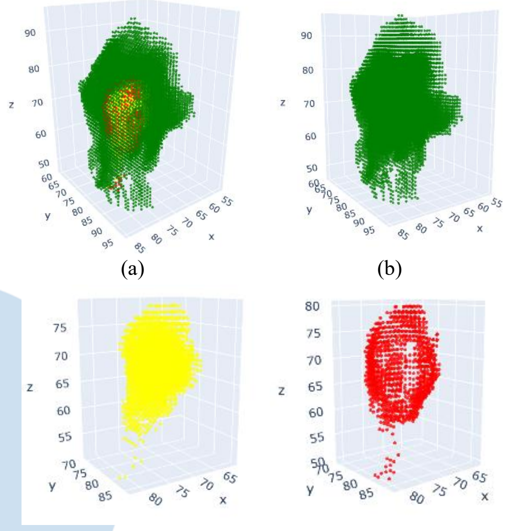

Fig. 8. 3D visualization of (a) the entire tumor (b) the Whole Tumor (c) the Tumor Core and (d) Enhancing Tumor (ET)

This visualization consists of four main sections. First, the full tumor prediction visualization displays all voxels classified as part of the tumor, colored based on their respective labels. This visualization illustrates the overall structure and distribution of the tumor, including irregular borders, asymmetric distribution, and areas of high density that may indicate tumor dominance. The clarity of the color labels allows identification of spatial relationships between tumor components and helps assess whether the model's predictions logically follow the biological pattern of the brain tumor.

The Whole Tumor (WT) visualization focuses on the entire tumor mass regardless of label type. In this stage, all voxels with a label greater than zero are displayed in green, reflecting the total size and shape of the tumor. This visualization is useful for evaluating whether the model covers the entire tumor volume as intended, or whether it is missing important areas (under-segmentation) or over-segmenting.

The third visualization displays only the Tumor Core (TC), the deepest area of the tumor, typically composed of dense tissue and crucial for diagnosis. The TC is displayed in yellow to highlight whether the model accurately and consistently identifies the tumor core, consistent with the general pattern of tumor growth. The distinctive shape and location of the TC can help assess the potential for compression of vital brain structures.

The Enhancing Tumor (ET) was visualized, which is the area of the tumor that shows increased signal on T1-weighted contrast (T1c) imaging. This area is often associated with high levels of biological activity and increased vascularization, making it important for diagnosis and therapy planning. Voxels in the ET are visualized in red, which helps observe the distribution and potential aggressiveness of the tumor in 3D space.

Overall, this 3D visualization not only strengthens the quantitative evaluation results but also provides a more realistic spatial representation of model predictions, making it very useful for medical practitioners in understanding and analyzing brain tumor development more comprehensively.

Discussion

A. *Findings*

This study demonstrates that combining the Swin UNETR and V-Net architectures into a single hybrid model (Fusion model) can improve the accuracy of brain tumor segmentation in volumetric MRI images. Quantitative evaluation results demonstrate that the average Dice Score for three tumor types, Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET), reaches 0.8635, 0.7179, and 0.8073, respectively. These values are considered high, indicating that the model successfully recognizes and maps tumor areas accurately, even in low-resolution MRI images.

Although several recent studies report higher Dice scores, particularly for Whole Tumor segmentation, these methods often rely on extensive architecture tuning, large-scale computational resources, or ideal imaging conditions. In contrast, the proposed hybrid V-Net and Swin UNETR model demonstrates balanced performance across Dice, Precision, and Recall metrics, especially under low-quality MRI conditions. This indicates that the proposed approach prioritizes robustness and generalizability rather than solely optimizing a single metric.

When compared conceptually to non-hybrid baselines, a standalone V-Net effectively captures local volumetric features but lacks global contextual awareness, often leading to fragmented boundaries. Conversely, Swin UNETR models emphasize global spatial relationships but may miss fine-grained local details critical for small tumor regions. The proposed hybrid architecture integrates both strengths, resulting in improved segmentation consistency across WT, TC, and ET regions.

This achievement aligns with a previous study by [2] which demonstrated that the use of a transformer-based architecture like Swin UNETR is able to capture global spatial context better than conventional CNN models, especially for 3D segmentation tasks. Furthermore, the V-Net-based encoder-decoder approach proved effective in extracting local spatial features from medical volumes, as also demonstrated by [12] in their original study on V-Net for internal organ segmentation.

The success of the fusion model in this study also demonstrates that an architectural ensemble approach can mitigate the weaknesses of each model when used alone. This is reinforced by findings [16] in nnU-Net, which suggest that appropriate architecture and pipeline adaptation, including fusion strategies and post-processing, significantly impact segmentation quality. Furthermore, qualitative evaluation through 3D visualization demonstrated that the model's predictions were not only numerically accurate but also morphologically and spatially consistent. The Tumor Core (TC) and Enhancing Tumor (ET) areas were successfully mapped with shapes and distributions consistent with the general biological structure of brain tumors.

However, several challenges and potential improvements remain. One is the reliance on training data, due to the lack of publicly available labels in the official BraTS validation set, evaluation was conducted using an internal validation split. This opens up the possibility of evaluation bias. Furthermore, some samples exhibited lower Dice scores in the Enhancing Tumor (ET) class, indicating that the model still has limitations in capturing small, mixed, low-contrast tumor areas. A similar finding was also reported by [19], who emphasized that ET segmentation is a major challenge because its intensity contrast often overlaps with normal tissue.

The implications of these findings are important in the context of developing AI-based clinical decision support systems. Fusion models such as those proposed in this study have the potential to be used as non-invasive and efficient early diagnostic tools, particularly for brain tumor screening in 3D images. However, further validation against external datasets and integration with feedback from healthcare professionals are needed for the system to be adopted clinically.

B. *Limitation*

Although the results are promising, this study still has several limitations that should be considered for future development. While multiple evaluation metrics, including Dice, Precision, and Recall, were employed, the inclusion of additional distance-based metrics such as the Hausdorff Distance could provide deeper insight into boundary accuracy. Furthermore, clinical evaluation involving radiologists or specialist physicians would offer more comprehensive validation of the segmentation results.

A second limitation lies in the lack of segmentation labels in the validation data, so the evaluation was conducted only on the training data. This results in a lack of objective testing of the model's performance on data the model has never encountered before. Therefore, future research is recommended to perform manual splitting or add an external validation dataset for more comprehensive and accurate model evaluation.

Furthermore, the current loss formulation can be further improved by incorporating advanced loss combinations, such as Dice Loss with Focal Loss, to enhance sensitivity for small tumor regions. Combining *DiceLoss* with *CrossEntropyLoss* is recommended to improve segmentation performance, especially for minority classes. Finally, future research is expected to explore hyperparameter optimization in more depth, such as variations in learning rate and batch size, as well as the application of spatial data augmentation such as rotation, flipping, and elastic deformation. These steps have the potential to improve the model's generalization and robustness to variations in tumor shape and size in MRI images.

By considering these various suggestions and improvements, it is hoped that future research will produce a more reliable, accurate, and applicable brain tumor segmentation system in real-world clinical settings.

IV. Conclusions

The hybrid V-Net-Swin UNETR model successfully improves brain tumor segmentation performance on the BraTS 2020 MRI dataset. By combining comprehensive preprocessing, a dual-path feature extraction strategy, and adaptive postprocessing, the model achieves Dice scores of 0.8635 for Whole Tumor (WT), 0.8073 for Enhancing Tumor (ET), and 0.7179 for Tumor Core (TC), demonstrating its ability to integrate local volumetric features with global contextual information effectively. These results highlight the model’s potential as a reliable AI-based diagnostic support tool in clinical workflows. For future development, further validation on external datasets is needed to assess generalization across imaging protocols, along with enhancements in detecting small tumor regions through improved loss functions and augmentation strategies. Incorporating uncertainty estimation, developing lightweight versions for real-time or resource-limited settings, and enabling interactive or semi-supervised segmentation could also enhance clinical usability. Additionally, integrating imaging data with clinical or molecular information offers opportunities for more comprehensive tumor characterization.

Acknowledgment

The authors would like to thank all parties who supported this research, particularly the Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) of Universitas 17 Agustus 1945 Surabaya

References

[1] T. B. Nguyen-Tat, N. H. Nghia, and V. M. Ngo, “Enhancing Brain Tumor Segmentation in MRI Images: A Hybrid Approach Using UNet, Attention Mechanisms, and Transformers,” Egyptian Informatics Journal, vol. 7, 2024, doi: 10.13140/RG.2.2.18164.36485.

[2] A. H. Nvidia et al., “UNETR: Transformers for 3D Medical Image Segmentation,” arXiv:2003.10504v3, Oct. 2021, [Online]. Available: https://monai.io/research/unetr

[3] E. Sami, H. Ebied, S. Amin, and M. Hassaan, “Brain Tumor Segmentation Using Modified U-Net,” Res Sq, no. preprint, May 2022, doi: 10.21203/rs.3.rs-1653006/v2.

[4] A. Srinivasa Reddy and P. Chenna Reddy, “MRI brain tumor segmentation and prediction using modified region growing and adaptive SVM,” Soft comput, vol. 25, no. 5, pp. 4135–4148, Mar. 2021, doi: 10.1007/s00500-020-05493-4.

[5] P. Agrawal, N. Katal, and N. Hooda, “Segmentation and classification of brain tumor using 3D-UNet deep neural networks,” International Journal of Cognitive Computing in Engineering, vol. 3, pp. 199–210, Jun. 2022, doi: 10.1016/j.ijcce.2022.11.001.

[6] J. Chen et al., “3D TransUNet: Advancing Medical Image Segmentation through Vision Transformers,” arXiv:2310.07781v1, Oct. 2023, [Online]. doi: 10.48550/arXiv.2310.07781

[7] Y. Ding, L. Gong, M. Zhang, C. Li, and Z. Qin, “A multi-path adaptive fusion network for multi-modal brain tumor segmentation,” Neurocomputing, vol. 412, pp. 19–30, Oct. 2020, doi: 10.1016/j.neucom.2020.06.078.

[8] H. Qin et al., “QuantSR: Accurate Low-bit Quantization for Efficient Image Super-Resolution,” in 37th Conference on Neural Information Processing Systems (NeurIPS 2023), 2023. [Online]. Available: https://github.com/htqin/QuantSR.

[9] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015., Springer, Cham, 2015, ch. Lecture No, pp. 1–8.

[10] Z. Liu et al., “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021. doi: https://doi.org/10.1109/ICCV48922.2021.00986.

[11] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” in International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, Jun. 2016. [Online]. Available: http://arxiv.org/abs/1606.06650

[12] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” arXiv:1606.04797v1, Jun. 2016, [Online]. Available: http://arxiv.org/abs/1606.04797

[13] 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, doi: 10.1109/TMI.2014.2377694.

[14] A. Hoopes, J. S. Mora, A. V. Dalca, B. Fischl, and M. Hoffmann, “SynthStrip: skull-stripping for any brain image,” Neuroimage, vol. 260, Oct. 2022, doi: 10.1016/j.neuroimage.2022.119474.

[15] M. Havaei et al., “Brain tumor segmentation with Deep Neural Networks,” Med Image Anal, vol. 35, pp. 18–31, Jan. 2017, doi: 10.1016/j.media.2016.05.004.

[16] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioin-formatics), Springer Verlag, 2017, pp. 240–248. doi: 10.1007/978-3-319-67558-9_28.

[17] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,” Nat Methods, vol. 18, no. 2, pp. 203–211, Feb. 2021, doi: 10.1038/s41592-020-01008-z.

[18] A. A. Taha and A. Hanbury, “Metrics for evaluating 3D medical image segmentation: Analysis, se-lection, and tool,” BMC Med Imaging, vol. 15, no. 1, Aug. 2015, doi: 10.1186/s12880-015-0068-x.

[19] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation,” IEEE Trans Med Imaging, vol. 39, no. 6, pp. 1856–1867, Jun. 2020, doi: 10.1109/TMI.2019.2959609

UMN