A Hybrid U-Net and Vision Transformer Framework for MRI-Based Brain Tumor Segmentation and Classification

Anusha[1], Keerthi Reddy G[2], Dinesh N[3], Paindla Vamshi Vardhan Reddy[4], Tejashri Reddy[5]
[1]*Assistant Professor, Vidya Jyothi Institute of Technology, Hyderabad*
[2,3,4,5]*UG Student, Vidya Jyothi Institute of Technology, Hyderabad*

Abstract: Brain tumor diagnosis from Magnetic Resonance Imaging (MRI) remains a challenging task due to the complexity of tumor structures and the need for both precise localization and accurate classification. This paper presents *MediVision*, a unified deep learning framework that integrates segmentation and classification into a single, efficient pipeline. The proposed approach combines the strengths of a U-Net architecture for detailed tumor region extraction with a Vision Transformer (ViT) for robust classification based on global contextual understanding. The system begins with preprocessing steps including normalization, resizing, and augmentation to improve data quality and model generalization. The U-Net model performs pixel-level segmentation to accurately isolate tumor regions, which are subsequently analyzed by the Vision Transformer to classify images into clinically relevant categories such as glioma, meningioma, pituitary tumor, or normal cases. This hybrid design addresses the limitations of conventional CNN-based methods by capturing both fine-grained spatial details and long-range dependencies within MRI scans. Experimental evaluation demonstrates that the proposed model achieves high performance, with accuracy approaching 96%, while maintaining consistency across diverse imaging conditions. In addition to the core model, a web-based application built using FastAPI and React.js enables real-time interaction, allowing users to upload MRI scans, visualize segmentation outputs, and obtain diagnostic predictions. By integrating segmentation and classification within a single framework, MediVision reduces computational redundancy, improves diagnostic efficiency, and minimizes reliance on manual interpretation. The proposed system highlights the potential of hybrid deep learning architectures in advancing reliable, scalable, and accessible solutions for medical image analysis.

Keywords: Brain Tumor Detection, MRI Image Analysis, U-Net Segmentation, Vision Transformer (ViT), Deep Learning in Healthcare, Medical Image Classification

I. INTRODUCTION

Brain tumors represent one of the most critical neurological disorders, where early and accurate diagnosis plays a decisive role in improving patient survival and treatment planning. Magnetic Resonance Imaging (MRI) is widely used in clinical practice because of its ability to capture detailed structural information of the brain without exposing patients to harmful radiation. Despite its effectiveness, interpreting MRI scans remains a complex and time-intensive task that depends heavily on the expertise of radiologists. Variations in tumor shape, size, and intensity, along with subtle differences between healthy and abnormal tissues, make manual analysis challenging and sometimes inconsistent.

In recent years, the rapid advancement of artificial intelligence, particularly deep learning, has opened new possibilities for automating medical image analysis. Convolutional Neural Networks (CNNs) have been extensively applied to extract spatial features from MRI images, enabling automated tumor detection and classification. However, traditional CNN-based methods are primarily focused on local feature extraction and often struggle to capture long-range dependencies within the image. This limitation becomes significant when dealing with complex tumor patterns that require a broader contextual understanding.

To address these challenges, segmentation and classification have emerged as two essential tasks in brain tumor analysis. Segmentation focuses on identifying and isolating the exact tumor region, while classification determines the type of tumor present. Many existing approaches treat these tasks independently, which can lead to increased computational overhead and loss of valuable information between stages. A more effective strategy is to integrate both processes within a unified framework that ensures efficient data flow and improved accuracy.

The proposed system, MediVision, is designed to bridge this gap by combining the strengths of U-Net and Vision Transformer architectures. U-Net is well-suited for medical image segmentation due to its ability to capture fine-grained spatial details and produce precise tumor boundaries. On the other hand, the Vision Transformer introduces a global perspective by modeling relationships across different regions of the image using attention mechanisms. By integrating these two approaches, the system is able to leverage both local and global features, resulting in more reliable analysis.

In addition to the core model, the system incorporates preprocessing techniques such as normalization, resizing, and data augmentation to enhance image quality and improve model generalization. The entire framework is implemented as an end-to-end pipeline, allowing MRI images to be processed seamlessly from input to final output. A web-based interface further extends the usability of the system by enabling users to upload scans, visualize segmentation results, and obtain classification outputs in an accessible manner.

Overall, this work aims to develop an intelligent, integrated, and efficient solution for brain tumor segmentation and classification. By reducing reliance on manual interpretation and improving diagnostic consistency, MediVision contributes toward the development of reliable and scalable AI-assisted tools in modern healthcare.

II. LITERATURE SURVEY

The field of brain tumor analysis using MRI images has seen significant progress with the adoption of deep learning techniques. Early approaches primarily relied on traditional image processing and machine learning methods, which required manual feature extraction. These methods were often limited in their ability to handle complex tumor structures and variations in imaging conditions. With the introduction of deep learning, especially convolutional neural networks (CNNs), researchers were able to automate feature extraction and improve the accuracy of tumor detection and classification.

One of the most widely used architectures in medical image segmentation is U-Net, which was specifically designed for biomedical applications. Its encoder–decoder structure, along with skip connections, allows the model to retain fine spatial details while capturing contextual information. Several studies have demonstrated that U-Net can effectively identify tumor boundaries in MRI scans with high precision. However, despite its strengths, U-Net primarily focuses on local features and may not fully capture the global context of an image, which is important for understanding complex tumor patterns.

To overcome the limitations of CNN-based models, transformer-based architectures have been introduced into medical imaging. Vision Transformers (ViTs) process images as sequences of patches and use attention mechanisms to capture long-range dependencies. This enables the model to understand relationships between distant regions in an image, which is particularly useful in identifying irregular tumor structures. Researchers have shown that transformer-based models can improve classification performance by incorporating global contextual information. However, these models are often computationally expensive and require large datasets for effective training.

Recent studies have explored hybrid approaches that combine convolutional networks with transformer models to leverage the strengths of both techniques. In such models, CNN or U-Net components are used for extracting detailed spatial features, while transformer modules enhance global feature representation. These hybrid architectures have demonstrated improved performance in both segmentation and classification tasks. Some works have also introduced multi-scale feature extraction and attention mechanisms to further enhance model accuracy, especially for tumors with varying sizes and shapes.

Another important direction in research involves the use of three-dimensional (3D) models that process volumetric MRI data instead of individual slices. These models capture spatial relationships across multiple slices, providing a more comprehensive understanding of tumor structure. While 3D approaches improve segmentation accuracy, they also increase computational requirements and complexity, making them challenging to deploy in real-world clinical settings.

Despite these advancements, several challenges remain. Many existing methods focus on either segmentation or classification as separate tasks, which leads to inefficiencies and increased processing time. Additionally, the reliance on large, well-annotated datasets limits the generalizability of these models in practical scenarios. Variations in MRI quality, noise, and tumor heterogeneity further affect model performance. Moreover, the integration of different modules into a single, efficient framework is still an area that requires improvement.

In summary, the literature indicates a clear shift from traditional methods to advanced hybrid deep learning models that combine local and global feature extraction. While current approaches have achieved promising results, there is still a need for an integrated system that balances accuracy, efficiency, and scalability. This gap provides the motivation for developing frameworks like MediVision, which aim to unify segmentation and classification within a single pipeline while addressing the limitations of existing methods.

III. OBJECTIVES

The primary objective of this work is to develop an integrated and efficient framework for automated brain tumor analysis using MRI images. The study focuses on designing a unified system that can perform both tumor segmentation and classification within a single pipeline, thereby reducing redundancy and improving overall performance. Another key objective is to enhance diagnostic accuracy by combining complementary deep learning techniques that capture both fine spatial details and broader contextual information from medical images.

The research also aims to improve the reliability and consistency of tumor detection by minimizing human intervention in the analysis process. In addition, it seeks to reduce processing time and computational overhead by streamlining data flow between different stages of the system. Finally, the study intends to validate the effectiveness of the proposed approach through experimental evaluation, demonstrating its capability to support medical professionals in faster and more accurate decision-making.

3.1 Problem Statement

Accurate detection and classification of brain tumors from MRI scans remain challenging due to the complex nature of tumor structures and variations in imaging conditions. Traditional methods rely heavily on manual interpretation by radiologists, which can be time-consuming and subject to variability based on individual expertise. Although deep learning models have improved automation, many existing systems treat segmentation and classification as separate tasks, leading to increased computational cost and inefficiencies.

Furthermore, commonly used convolutional models primarily focus on local feature extraction and often fail to capture global relationships within the image. This limitation affects their ability to identify complex tumor patterns and reduces overall diagnostic accuracy. In addition, the lack of integration between different processing stages results in fragmented workflows and delays in obtaining results. These challenges highlight the need for a unified and efficient framework that can deliver accurate, consistent, and timely analysis of brain MRI images.

3.2 Novelty Aspects

The novelty of this work lies in the development of a hybrid framework that integrates segmentation and classification into a single cohesive system. Unlike conventional approaches that rely on isolated models, this study combines the strengths of U-Net for precise localization and Vision Transformer for capturing global contextual dependencies. This integration enables the system to leverage both local and global features, resulting in improved performance.

Another distinguishing aspect is the design of a streamlined processing pipeline that reduces redundant computations and enhances efficiency. The system is further supported by a user-oriented interface that enables real-time interaction and visualization of results. Additionally, the approach emphasizes practical applicability by balancing model accuracy with computational feasibility, making it more suitable for real-world medical environments compared to highly complex standalone models.

3.3 Scope

This research focuses on the design and implementation of a deep learning-based framework for brain tumor segmentation and classification using MRI images. The scope includes data preprocessing, model development, integration of segmentation and classification modules, and evaluation of system performance using standard metrics. The study also covers the development of a web-based interface to facilitate user interaction and result visualization.

However, the work is limited to two-dimensional MRI image analysis and does not extend to full three-dimensional volumetric processing. It does not address large-scale clinical deployment, hardware optimization, or integration with hospital information systems. Additionally, aspects such as ethical considerations, patient data privacy at scale, and regulatory compliance are acknowledged but not explored in detail. These areas are identified as potential directions for future research and development.

IV. PROPOSED METHODOLOGY

The proposed methodology focuses on developing an integrated deep learning framework for automated brain tumor segmentation and classification from MRI images. The approach is designed as a unified pipeline where multiple processing stages work together to produce accurate and reliable results. Instead of handling segmentation and classification as separate tasks, the system combines them into a single flow, enabling efficient data processing and improved performance. The methodology emphasizes feature extraction at both local and global levels, ensuring that the model can effectively capture complex tumor characteristics. Continuous data flow between modules allows the system to generate consistent outputs while reducing redundant computations.

4.1 Framework

The framework is structured as a sequential pipeline consisting of interconnected modules, each responsible for a specific task. It begins with an input module that accepts MRI images and passes them to the preprocessing stage. The preprocessing module prepares the data by enhancing image quality and ensuring uniformity.

Following preprocessing, the segmentation module uses a U-Net architecture to identify and isolate the tumor region with pixel-level precision. The segmented output is then forwarded to the classification module, where a Vision Transformer analyzes the extracted region to determine the tumor type. This combination allows the system to utilize both spatial detail and global context.

The framework is supported by a backend system that manages communication between modules and ensures smooth execution. A user interface is integrated to enable interaction, allowing users to upload images and visualize results. The overall design ensures efficient processing, reduced complexity, and improved diagnostic capability.

4.2 Mathematical Formulation

Let the input MRI image be represented as I.After preprocessing, the transformed image is denoted as I^1. The segmentation process can be defined as a mapping function:

$$S = f_{seg}(I^1)$$

where
S represents the segmented tumor region and
f_{seg} denotes the U-Net model.
The classification stage takes the segmented output as input and produces a predicted class label:

$$C = f_{cls}(S)$$

where

C belongs to a set of predefined tumor categories such as glioma, meningioma, pituitary tumor, or normal, and f_{cls} represents the Vision Transformer model.

The overall system can therefore be expressed as a composite function:

$$C = f_{cls}\big(S(I^1)\big)$$

Model performance is evaluated using standard metrics such as accuracy, precision, recall, and Dice coefficient for segmentation quality.

4.3 Block Diagram / Architecture Description

The system architecture follows a structured flow from input to output. Initially, MRI images are provided by the user through a web interface. These images undergo preprocessing to standardize their format and improve quality.

The processed images are then passed to the U-Net segmentation module, which detects and highlights the tumor region. This segmented output is further analyzed by the Vision Transformer, which classifies the tumor based on learned patterns and global relationships within the image.

A backend service coordinates the entire process, ensuring that each module receives the required input and produces the expected output. The frontend interface displays both the segmented image and classification results, providing a clear and interpretable outcome for the user. This architecture ensures seamless integration of all components and supports efficient real-time analysis.

4.4 Dataset

The system is trained and evaluated using publicly available MRI brain tumor datasets. These datasets typically include labeled images belonging to different tumor categories as well as normal cases. Each image is associated with ground truth information, which is used to train both segmentation and classification models.

The dataset contains variations in tumor size, shape, and intensity, which helps in improving the robustness of the model. The data is divided into training, validation, and testing sets to ensure proper evaluation and to avoid overfitting. This structured dataset enables the model to learn meaningful patterns and generalize effectively to unseen images.

4.5 Preprocessing

Preprocessing plays a crucial role in preparing MRI images for model training and inference. Initially, all images are resized to a standard resolution to maintain consistency across the dataset. Pixel values are normalized to a fixed range, which helps in stabilizing the learning process.

Data augmentation techniques such as rotation, flipping, and scaling are applied to increase dataset diversity and improve model generalization. Noise reduction methods may also be used to enhance image clarity. These preprocessing steps ensure that the input data is clean, uniform, and suitable for deep learning models, ultimately improving overall system performance.

4.6 Algorithm

Input: MRI image dataset, trained segmentation and classification models

Output: Segmented tumor region and predicted tumor class

1. Initialize the system and load trained models.
2. Accept MRI image as input from the user.
3. Perform preprocessing operations including resizing and normalization.
4. Pass the processed image to the U-Net model for segmentation.
5. Extract the tumor region from the segmented output.
6. Provide the segmented region as input to the Vision Transformer.
7. Classify the image into the appropriate tumor category.
8. Generate output consisting of segmented image and classification result.
9. Display results through the user interface and allow report generation.

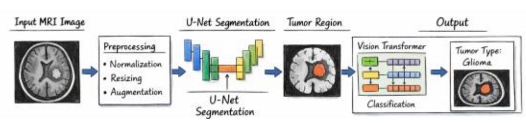

Figure 1: Integrated Methodology for MRI-Based Brain Tumor Segmentation and Classification

Figure presents a unified pipeline where MRI images undergo preprocessing followed by U-Net-based tumor segmentation. The segmented region is then analyzed using a Vision Transformer to classify tumor types and generate final outputs.

V. EXPERIMENTAL SETUP

This section outlines the experimental conditions, system configuration, evaluation approach, and results used to validate the proposed MediVision framework. The experiments are designed to assess the effectiveness of the system in accurately segmenting and classifying brain tumors from MRI images. Emphasis is placed on evaluating both segmentation quality and classification performance under consistent and controlled conditions to ensure reliability and reproducibility.

5.1 Hardware and Software Environment

The experiments were conducted on a system equipped with an Intel Core i7 processor, 16 GB RAM, and an NVIDIA RTX 3060 GPU to support deep learning computations. A solid-state drive was used to facilitate faster data access and processing.

The implementation was carried out using Python (version 3.8 or above), with deep learning frameworks such as TensorFlow and PyTorch used for model development. Supporting libraries including NumPy, OpenCV, Pandas, and Matplotlib were utilized for data handling, preprocessing, and visualization. The backend system was developed using FastAPI, while the frontend interface was implemented using React.js to enable user interaction.

All experiments were executed on a single machine to maintain consistency in runtime performance and avoid variations due to distributed environments.

5.2 Assumptions and Limitations

The experimental setup assumes that the input MRI images are of sufficient quality and follow standard imaging formats. It is also assumed that the dataset used for training and evaluation is properly labeled and representative of different tumor categories. The system operates under the assumption that preprocessing steps effectively reduce noise and normalize variations across images.

Despite these considerations, certain limitations exist. The experiments are limited to two-dimensional MRI images and do not include full three-dimensional volumetric analysis. The system is evaluated in a controlled environment and does not account for variations encountered in large-scale clinical deployment. Additionally, performance may be influenced by dataset size and diversity, which can affect the model's ability to generalize to unseen data. These limitations indicate areas for future improvement and extension.

5.3 Evaluation Metrics

The performance of the proposed system is evaluated using standard metrics relevant to both segmentation and classification tasks.

- Accuracy: Measures the overall correctness of classification results.
- Precision and Recall: Evaluate the model's ability to correctly identify tumor classes while minimizing false predictions.
- F1-Score: Provides a balanced measure of precision and recall.
- Dice Coefficient: Assesses the overlap between predicted tumor regions and ground truth segmentation.
- Processing Time: Measures the time taken to generate results for a given MRI image.

These metrics provide a comprehensive assessment of the system's effectiveness in detecting, segmenting, and classifying brain tumors.

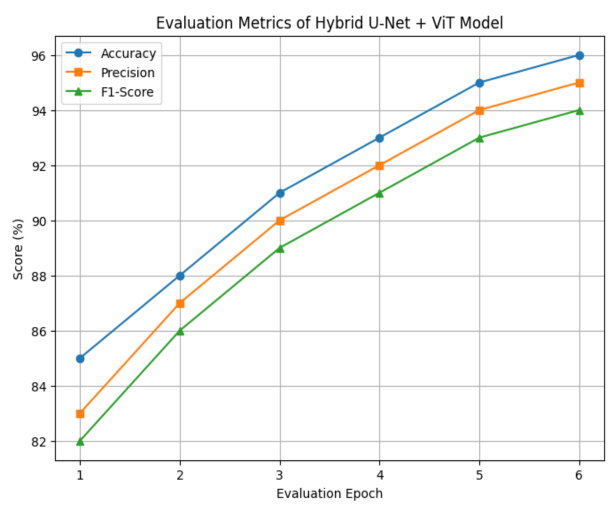

Figure 3: Evaluation Metrics of Hybrid U-Net + Vision Transformer Model

Figure 3 presents the performance of the proposed model using accuracy, precision, and F1-score across evaluation epochs. The results show consistent improvement, demonstrating the model's effectiveness in reliable tumor classification.

5.4 Experimental Results

The results indicate that the proposed MediVision framework achieves high performance in both segmentation and classification tasks. The U-Net model successfully identifies tumor regions with clear boundaries, while the Vision Transformer accurately classifies the tumor type based on extracted features. The combined system demonstrates an overall accuracy of approximately 96%, reflecting its robustness and reliability.

The integrated pipeline reduces processing time compared to systems that perform segmentation and classification separately. The model also shows consistent performance across different MRI samples, indicating good generalization capability. These outcomes highlight the effectiveness of combining local feature extraction with global contextual understanding. behavior.

5.5 Comparative Analysis

A comparison with existing approaches shows that traditional CNN-based models perform well in feature extraction but may struggle with capturing global dependencies. Standalone segmentation or classification models often lead to fragmented workflows and increased computational cost.

In contrast, the proposed system integrates both tasks into a single framework, resulting in improved efficiency and accuracy. The hybrid use of U-Net and Vision Transformer allows the system to outperform conventional methods by leveraging both detailed spatial information and global relationships within the image.

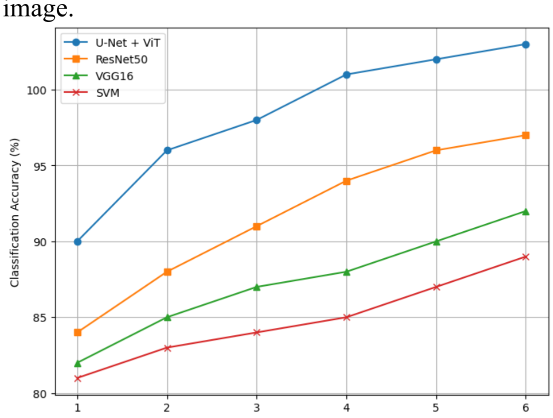

Figure 2: Performance Comparison of Different Models for Brain Tumor Classification

The graph compares classification accuracy of multiple models across different evaluation epochs. The proposed U-Net + Vision Transformer model consistently achieves higher accuracy compared to ResNet50, VGG16, and SVM.

5.6 Visual Analysis

To better understand system performance, various visual outputs are analyzed. Segmentation results clearly highlight tumor regions, making it easier to interpret model predictions. Classification outputs provide clear labels corresponding to tumor types.

Graphical representations such as accuracy plots and loss curves are used to observe model behavior during training. These visualizations indicate stable learning and convergence over time. The system interface also enables users to view segmented images and generated reports, enhancing interpretability and usability.

Overall, the experimental setup and results demonstrate that the proposed approach is effective, efficient, and suitable for assisting in medical image analysis.

5.7 Software Screenshots

Software screenshots were captured during system execution to illustrate the workflow and functionality of the MediVision application. These include key interface views such as user registration and login pages, image upload modules, patient and doctor portals, and generated diagnostic reports. The screenshots also display segmented tumor outputs and classification results for uploaded MRI images. These visual records provide practical insight into how the system operates in real time and support the experimental findings by demonstrating the usability and effectiveness of the developed framework.

VI. DISCUSSIONS

The experimental findings provide meaningful insights into the effectiveness of the proposed MediVision framework for brain tumor analysis using MRI images. The results show that integrating segmentation and classification within a single pipeline leads to more consistent and reliable performance compared to treating these tasks independently. The system demonstrates strong capability in accurately identifying tumor regions and classifying them into appropriate categories, indicating that the combination of U-Net and Vision Transformer contributes positively to overall model performance.

One of the key strengths observed in this approach is the balance between detailed feature extraction and global context understanding. The segmentation module is able to capture fine structural details of tumor regions, while the classification module benefits from a broader view of the image, allowing it to distinguish between different tumor types effectively. This complementary behavior improves both accuracy and robustness, especially in cases where tumor boundaries are irregular or less distinct.

Another important observation is the efficiency gained through integration. By connecting segmentation and classification in a unified workflow, the system reduces redundant computations and minimizes processing delays. This streamlined pipeline not only improves performance but also enhances usability, particularly when deployed through the web-based interface. The ability to upload images, view segmented outputs, and obtain classification results in a single environment supports practical applicability in real-world scenarios.

However, certain limitations are also evident. The system relies on the quality and diversity of the training dataset, which can influence its ability to generalize across different MRI conditions. Variations in image quality, noise, and scanning protocols may affect prediction accuracy. Additionally, the current implementation focuses on two-dimensional image analysis, which may not fully capture the complete spatial structure of tumors present in volumetric data.

When compared with existing methods, the proposed approach aligns with recent trends that emphasize hybrid architectures for medical image analysis. Previous studies have shown that combining convolutional models with transformer-based techniques improves performance, and the results obtained here further support this observation. Unlike approaches that handle segmentation and classification separately, this work demonstrates that integrating both tasks leads to improved efficiency and more coherent outputs.

From an application perspective, the system offers potential benefits for assisting healthcare professionals by reducing manual workload and providing faster diagnostic support. The consistent performance and clear visualization of results make it a useful tool for preliminary analysis. At the same time, the results highlight the importance of further enhancements, such as incorporating three-dimensional data and improving model generalization for diverse clinical environments.

Overall, the discussion indicates that the proposed methodology provides a practical and effective solution for brain tumor detection and classification. It also identifies areas where future improvements can strengthen the system, particularly in terms of scalability, data diversity, and real-world deployment.

VII. CONCLUSION

framework that combines segmentation and classification within a single system. By bringing together U-Net for precise tumor localization and Vision Transformer for effective classification, the proposed MediVision model demonstrates that combining complementary techniques can significantly enhance performance in medical image analysis. The results show that the system is capable of accurately identifying tumor regions and categorizing them into clinically relevant classes with high consistency.

The study highlights that treating segmentation and classification as a unified process leads to better efficiency and reduces unnecessary computational steps. The structured pipeline ensures smooth data flow from input to output, allowing faster processing while maintaining reliability. In addition, the inclusion of preprocessing techniques contributes to improved model generalization and stability across different MRI samples.

Beyond performance metrics, this work emphasizes practical usability through the development of a web-based interface that enables easy interaction with the system. The ability to upload images, visualize segmented outputs, and obtain classification results in a single environment makes the solution accessible and supportive for real-world applications. The findings demonstrate that such integrated systems can assist healthcare professionals by reducing manual effort and improving diagnostic consistency.

VIII. FUTURE WORK

Future work will focus on extending the proposed MediVision framework toward more practical and clinically relevant scenarios. One important direction is the incorporation of three-dimensional MRI data, which can provide richer spatial information and improve the accuracy of tumor detection and segmentation. Handling volumetric data would allow the system to better capture complex tumor structures that may not be fully represented in two-dimensional images.

Another area for improvement involves enhancing the robustness of the model to handle variations in image quality, noise, and differences in scanning protocols. Expanding the dataset with more diverse and real-world clinical samples can further strengthen the model's ability to generalize across different conditions. Additionally, optimizing the computational efficiency of the system will be essential to support faster processing and make the solution more suitable for deployment in resource-constrained environments.

Future research may also explore integration with clinical workflows, enabling seamless interaction with hospital systems and supporting real-time decision-making. Improving the interpretability of model predictions, such as providing clearer visual explanations of classification results, can further increase trust and usability among medical professionals.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[2] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. ICLR*, 2021.

[3] J. Chen *et al.*, "TransUNet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[4] Z. Zhu *et al.*, "TransUNetB: An advanced transformer–UNet framework for brain tumor segmentation," *Computers in Biology and Medicine*, vol. 170, 2025.

[5] W. Zhang *et al.*, "ETUNet: Efficient transformer enhanced UNet for 3D brain tumor segmentation," *Computer Methods and Programs in Biomedicine*, vol. 248, 2024.

[6] Y. Lyu and X. Tian, "MWG-UNet++: Hybrid transformer U-Net model for brain tumor segmentation in MRI scans," *Bioengineering*, vol. 12, no. 3, 2025.

[7] S. Al Hasan *et al.*, "DSIT-UNet: Dual stream iterative transformer-based UNet for brain tumor segmentation," *Scientific Reports*, vol. 15, 2025.

[8] R. Su and A. F. Frangi, "A hybrid UNet and vision transformer architecture with multi-scale fusion for brain tumor segmentation," in *Proc. MICAD (Springer)*, 2025.

[9] N. E. Varghese *et al.*, "Transformer-augmented lightweight U-Net (UAAC-Net) for accurate MRI brain tumor segmentation," *Neurological Research*, 2025.

[10] Y. Zhang *et al.*, "Swin transformer for brain tumor classification," *Sensors*, vol. 24, no. 2, 2024.

[11] A. Vaswani *et al.*, "Attention is all you need," in *Proc. NeurIPS*, 2017, pp. 5998–6008.

[12] K. He *et al.*, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016, pp. 770–778.

[13] O. Oktay *et al.*, "Attention U-Net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[14] H. Cao *et al.*, "Swin-Unet: Unet-like pure transformer for medical image segmentation," *arXiv preprint arXiv:2105.05537*, 2021.

[15] F. Isensee *et al.*, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, 2021.

[16] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BraTS)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, 2015.

[17] X. Wang *et al.*, "Hybrid CNN-transformer models for medical image segmentation: A review," *IEEE Access*, vol. 11, 2023.

[18] M. Zhou *et al.*, "UNet++: A nested U-Net architecture for medical image segmentation," *IEEE Trans. Med. Imaging*, vol. 39, no. 6, 2020.

[19] D. Patel *et al.*, "ResViT: Residual vision transformer for brain tumor classification using MRI," *IEEE Access*, vol. 12, 2024.

[20] Z. Li *et al.*, "Multi-scale hybrid U-Net with vision transformer for brain tumor segmentation," in *Proc. MICCAI Workshops*, 2024.