AI IN HEALTHCARE – FINAL PROJECT

COVER PAGE

Link to Code:

https://colab.research.google.com/drive/1kTVzyTWDBhSiNPJM8tr3IOX_Nvi4dNZp?usp=sharing

Link to Google Drive (all files):

https://drive.google.com/drive/folders/1XvoTqJ3mhiyFbeK6oFlOANn_lJhcIo9Q?usp=sharing

Link to GitHub (all files):

https://github.com/VinhNguyen8889/MS-AI-program/tree/main/AIH/Final%20Project

Link to Video:

https://utexas.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=3b2b8cf1-14c1-4cef-ae35-b23d01782130

DECEMBER 4, 2024

VINH NGUYEN
UNIVERSITY OF TEXAS AT AUSTIN

Brain Tumor Segmentation Using the UNET Model: An Approach for Clinical Diagnosis

Vinh Nguyen
vinh.nguyen@my.utexas.edu
University of Texas at Austin
Austin, Texas, USA

ABSTRACT

Brain tumor segmentation is crucial in medical image analysis, assisting clinicians in diagnosing and planning patient treatments. This paper presents an efficient approach for brain tumor segmentation using the UNET model, a Convolutional Neural Network (CNN) architecture originally developed for pixel-wise classification in biomedical image segmentation. The BraTS2020 dataset [1], containing multi-modal MRI scans with annotated brain tumor regions, was used to train, validate, and test the model.

The work emphasizes preprocessing techniques, including normalization and data augmentation, to enhance model performance and address dataset variability. The different modalities of MRI are used as input for the model, however, due to computational constraints, the scope of the work is focused on 2D slices extracted from the middle slice of each modality due to computational constraints. Evaluation using Intersection over Union (IoU) metrics demonstrates that the proposed approach achieves high segmentation accuracy, showcasing its potential in clinical settings for precise and automated tumor identification.

This research highlights the adaptability and efficacy of the UNET model with ResNet encoder for brain tumor segmentation and its potential for integration into clinical workflows. By automating the tumor identification process, the proposed approach seeks to enhance diagnostic accuracy and reduce the time required for manual segmentation, thereby improving overall treatment effectiveness for patients.

KEYWORDS

Brain Tumor Segmentation, Convolutional Neural Network, UNET, ResNet, Intersection over Union, Transfer Learning, Deep Learning, Skip Connection, MRI, Medical Imaging, and Medical Diagnosis.

I. INTRODUCTION

Brain tumors present significant challenges in healthcare, where early and accurate detection is critical for effective treatment and improved prognosis. MRI is the best and most commonly used non-invasive method for brain tumor detection. However, manual segmentation of brain tumors from MRI scans is labor-intensive, prone to inter-observer variability, and further constrained by the quality of images produced by different scanning machines or the expertise of radiologists.

To address these challenges, recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have transformed medical image analysis [2]. The UNET model, introduced by Ronneberger et al., has become a leading solution for biomedical image segmentation, thanks to its encoder-decoder architecture and skip connections that effectively preserve spatial details [3]. By leveraging such models, it is possible to analyze MRI scans with high accuracy and consistency across different machines and scanning techniques, reducing reliance on manual segmentation.

This paper explores the application of the UNET model for brain tumor segmentation using the BraTS2020 dataset, a benchmark dataset for brain tumor analysis. The study focuses on preprocessing techniques, model optimization, and evaluation to create an efficient solution for clinical use. The key objectives include implementing a UNET-based segmentation model optimized for 2D MRI slices, evaluating its performance using IoU, and emphasizing the clinical implications of integrating automated systems into healthcare workflows. These efforts aim to enhance diagnostic precision and streamline workflows, providing consistent and reliable results in brain tumor analysis.

II. RELATED WORK

Brain tumor segmentation has evolved significantly, marked by milestones in traditional and deep learning-based methods. Early approaches relied on rule-based systems and probabilistic models, such as Adaptive Templates (1999) and Graph Cuts (2008), which often suffered from limitations in generalization and scalability. The advent of deep learning revolutionized the field, with AlexNet (2012), and Autoencoder Regularization (2018) paving the way for CNN-based segmentation. [2]

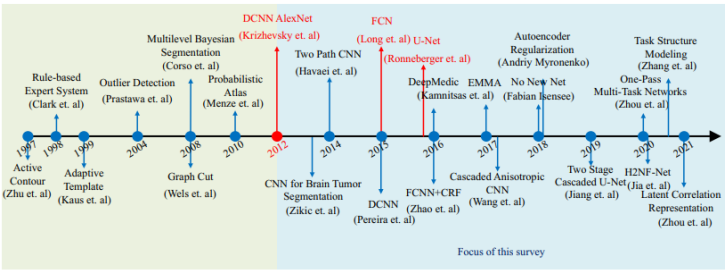


Figure 1: The evolution of brain tumor segmentation [2]

III. METHODOLOGY

A. Dataset

The BraTS2020 dataset was used for this study. It includes multi-modal MRI scans (T1, T1ce, T2, and FLAIR) and corresponding ground truth masks with pixel-wise tumor annotations, all stored in .nii format. The dataset consists of 369 cases with ground truth and 125 cases without ground truth, each containing 155 slices. Due to the computing constraints, the median slice (77th slice) is extracted from each set of images for training and testing the model.

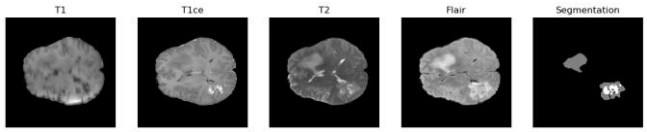


Figure 2: MRI Modalities (T1, T1ce, T2, FLAIR) and Segmentation Mask

B. *Preprocessing*

The raw MRI scans from the BraTS2020 dataset need a series of preprocessing steps that were carefully designed to ensure uniformity across all samples and prepare the dataset for training, validation, and testing.

- ToTensor: All the MRI modalities (T1, T2ce, T2, and FLAIR) and the masked scans were loaded using nib library as numpy arrays, and then converted into tensors to optimize performance on the GPU.
- Resizing: each image was resized from 250 x 250 pixels to 256 x 256 pixels to align with the input size required by the pre-trained UNET model used as a starting point for training.

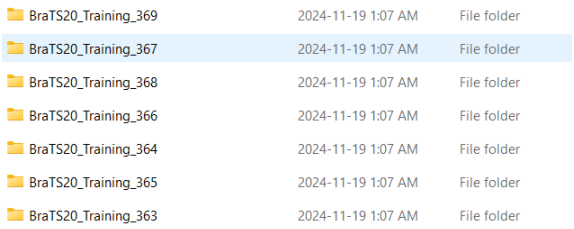
Then, data augmentation techniques were applied to the training data to increase dataset diversity and improve the model's ability to generalize. These augmentation techniques were designed to simulate variations that could occur in real-world medical imaging scenarios while preserving the integrity of the data. The following transformation techniques were used:

- Random Flip: each image had a 50% probability of being flipped horizontally and a 50% probability of being flipped vertically, simulating variations in scan orientations.
- Random Rotation: images were rotated randomly within a range of -5 to +5 degrees to account for slight misalignments during the scanning process.
- Color Jitter: the brightness, contrast, saturation of the images were adjusted randomly within a range of ±60%, while the hue was adjusted randomly within a range of ±20%. This transformation ensures variations in image-capturing conditions or different machine models, such as lighting and contrast adjustments.

C. *Dataset Preparation:*

A custom dataset class, BraTSDataset, was implemented to prepare the dataset for training. This class handles loading, preprocessing, and formatting the data into a suitable structure for the designed model architecture. Below is a detailed explanation of its functionality:

- Sample Retrieval: the training dataset directory is scanned to identify individual cases (folders), each containing multi-modal MRI scans (T1, T2ce, T2, and FLAIR) and corresponding segmentation masks. The testing dataset directory follows the same procedure, except it does not have the segmentation masks.



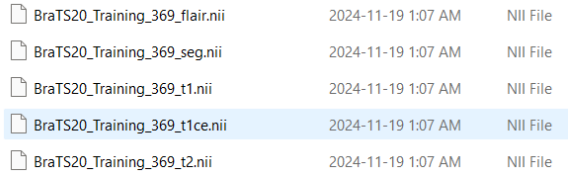


Figure 3: BraTS 2020 Dataset Directory Structure for Samples and MRI Modalities

- Slice Extraction: due to computational constraints, only the median slice (77th slice) from each modality and mask is extracted. This slice is chosen because it typically provides a representative cross-section of the tumor.
- Input Modalities: Four MRI modalities (T1, T2ce, T2, and FLAIR) are loaded as 2D NumPy arrays from the .nii files.
- Segmentation Mask: For training, the corresponding ground truth segmentation mask is also loaded, while a placeholder is assigned for testing. Each pixel of the ground truth segmentation mask is then converted to a binary class format, aligning with the expected target output of the UNET model.
- Transformation: Each loaded input and segmentation mask goes through a transformation pipeline defined previously.
- Stacking Modalities: The four MRI modalities are stacked along the channel dimension to create a multi-channel input tensor (4 x 256 x 256), representing the model input. The segmentation mask is a single-channel tensor with dimensions (1 x 256 x 256), serving as the output.

The dataset was divided into three subsets to train, validate, and test the model effectively:

- Training Set: 85% of the cases from the training folder were used for training the model.
- Validation Set: The remaining 15% of the cases from the training folder were set aside for validation to monitor the model's performance during training.
- Test Set: A separate folder containing unseen cases was used for testing the utility of the model. The test set does not have the ground truth segmentation mask.

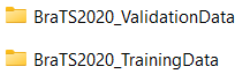


Figure 4: BraTS 2020 Dataset Directory Structure

D. *Model Architecture*

The UNET architecture is used as the core model for this task. The UNET architecture consists of a contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a convolutional network, consisting of the repeated application of two 3x3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for down-sampling. At each down-sampling step, the number of feature channels was doubled. Each step in the expansive path consists of an up-sampling of the feature map followed by a 2x2 convolution that halves the number of feature channels, a concatenation with the corresponding feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU [3].

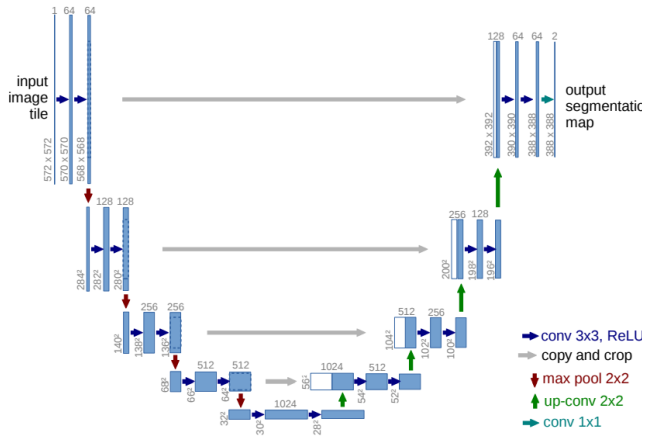


Figure 5: U-Net Architecture for Biomedical Image Segmentation [4]

However, deeper neural networks are more difficult to train; hence, a residual learning framework is often used to ease the training of substantially deep networks. The ResNet architecture was developed to address this issue, thereby enabling the construction of deep networks. In this project, the contracting path (left side) of the U-Net model is substituted with a ResNet-34 architecture to enhance feature extraction capabilities and leverage the powerful residual learning framework.

Pre-trained weights from the ImageNet dataset are used to improve performance further and accelerate convergence to initialize the ResNet-34 model. This transfer learning approach allows the network to benefit from previously learned features, which are transferable across various domains [4].

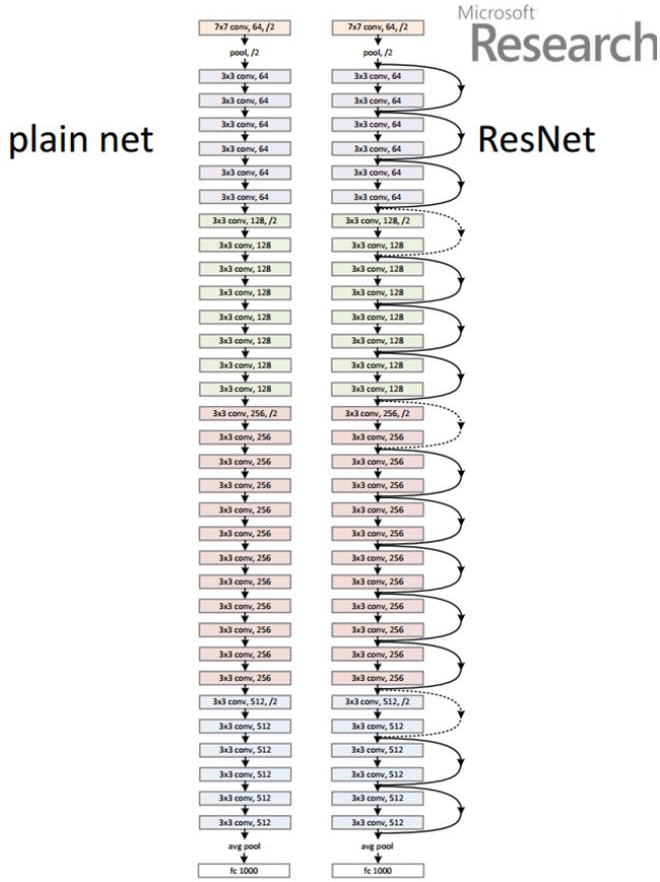


Figure 6: Example network architectures for ImageNet [4]

In this project, the final architecture is a combination of ResNet-34, utilizing pre-trained weights from ImageNet to enhance feature extraction capabilities, and U-Net, to improve the segmentation performance of the deep learning model.

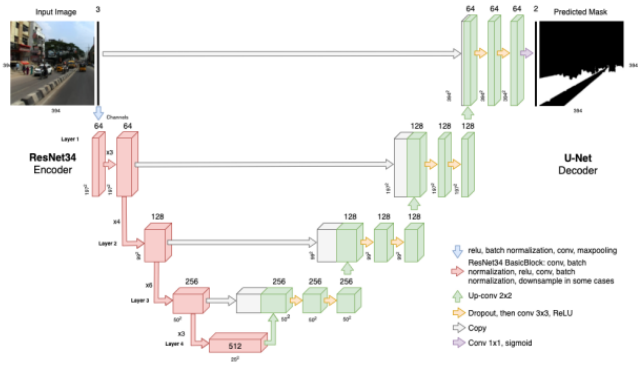


Figure 7: Example network architectures for ImageNet [5]

E. Training Procedure

The model was trained on 85% of the dataset and validated on the remaining 15%. The training process employed the AdamW optimizer with a learning rate of 1 x 10^{-4}, conducted over 200 epochs with a batch size of 64, and an early stopping based on validation's IoU loss. To enhance computational efficiency, the model and all inputs were processed on a GPU using CUDA.

Laptop Configuration: the model is trained on a laptop equipped with a 13th Gen Intel Core i9-13900HX processor, 32GB RAM, and and NVIDIA GeForce RTX 4060 GPU with 8GB of VRAM.

Training time: the training takes place in approximately 250 minutes with 200 epochs. The model is still learning at this epoch, and the early stopping setup is not applied.

The combination of BCELoss (Binary Cross Entropy) and IoU loss is used as the custom loss function guiding the segmentation training for this model.

BCELoss is a loss function specifically designed for pixel-wise prediction tasks. This loss function is well-suited for binary classification at the pixel level, allowing the model to learn the probability of each pixel belonging to the target class. The loss function calculates the mean negative log loss for the sigmoid output values and the ground truth values. The formula for BCELoss is defined as below:

$$\text{Log loss} = \frac{1}{N}\sum_{i=1}^{N} -\left(y_i * \log(p_i) + (1-y_i) * \log(1-p_i)\right)$$

Figure 8: Formula for Binary Cross-Entropy Loss

Intersection over Union (IoU), also known as the Jaccard Index, is a widely used metric in segmentation, object detection, and tracking tasks. It measures the overlap between the predicted segmentation and the ground truth, providing a robust evaluation of the model's ability to identify target regions accurately. By incorporating IoU as part of the optimization process, the model was encouraged to focus on improving the quality of the segmentation, especially in terms of boundary precision and minimizing false positives or negatives.

$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{|I|}{|U|}$$

Figure 9: Intersection over Union (IoU) Mathematical Definition

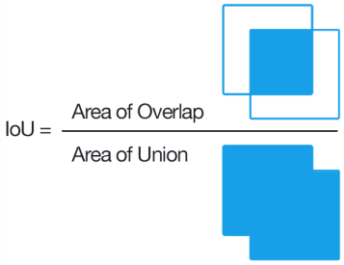


Figure 10: Intersection over Union (IoU) Visual Representation

F. Evaluation Metrics

The model's performance was evaluated using the metrics Intersection over Union (IoU) only, which is particularly important for the segmentation task.

The below image demonstrates the overlap between the predicted and ground truth boundaries to explain the IoU score.

Figure 11: IoU Score Performance Visual Representation

IV. RESULTS

Quantitative and qualitative analyses were conducted to assess the model's accuracy and its ability to delineate tumor regions effectively.

Quantitative Results

IoU score: The model obtained an average IoU score of 0.65 on the validation set, demonstrating its ability to accurately segment tumor regions while minimizing false positives and negatives. According to the pictorial explanation above, the model performance is categorized as good.

Qualitative Results

Qualitative analyses were performed on both the validation and test sets. For the validation set, visual inspection of segmentation results revealed that the UNET model with a ResNet-34 encoder successfully delineated tumor regions in MRI slices, even in challenging cases with irregular tumor boundaries.

Validation set: Ground truth segmentation masks were utilized to compare and validate the predicted outputs. Figure 12 presents examples of segmentation results on the validation set, demonstrating a close match between the predicted masks and ground truth annotations.

Test set: Qualitative evaluation was performed without ground truth masks. Instead, the predicted segmentation masks were overlaid on the MRI images to visually inspect their alignment with the apparent tumor regions. Figure 13 further highlights the model's ability to accurately detect potential tumor regions in the scan files. However, the lack of ground truth annotations limited the ability to compute objective metrics like IoU for the test set.

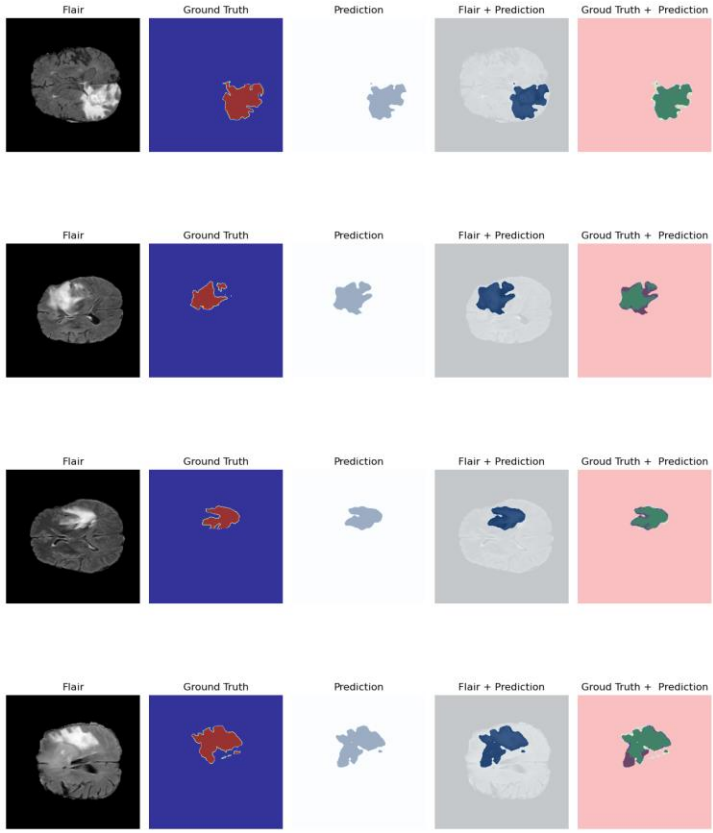


Figure 12: Visualization of Ground Truth and Predictions for validation dataset

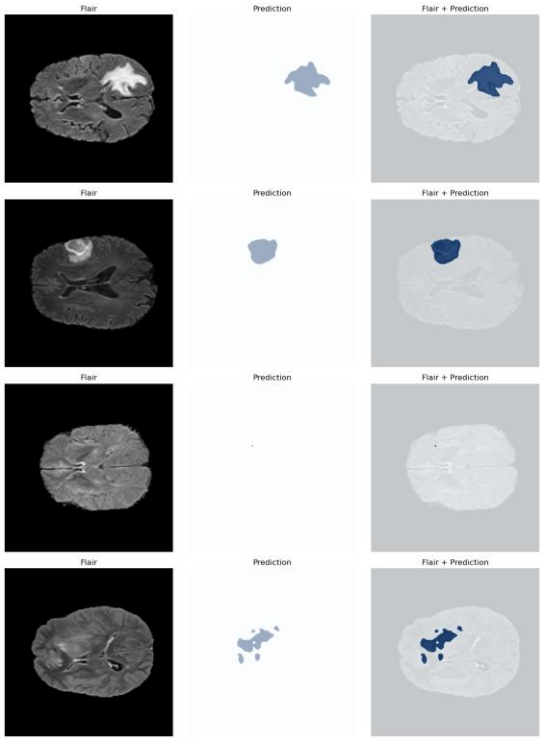


Figure 13: Visualization of Predictions for test dataset

Training and Validation Performance

The model demonstrates effective learning, with the Combined loss, BCE loss, and IoU loss decreasing consistently for training and validation. The training losses follow a smooth downward trend, while validation losses stabilize towards the end, with validation loss slightly higher than training loss, suggesting mild overfitting.

The IoU scores highlight strong segmentation performance on the training data, reaching close to 0.9 for training. However, the validation IoU stabilizes around 0.65, reflecting a noticeable performance gap, likely due to overfitting or challenges in generalization to unseen data.

Figure 14: Training and Validation Combined Loss Across Epochs

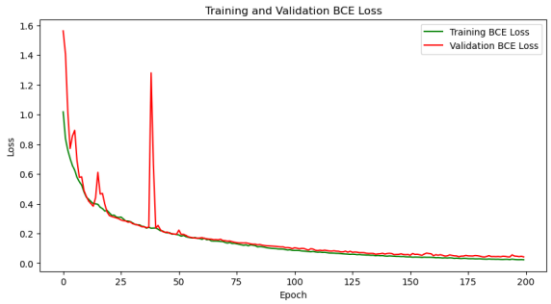


Figure 15: Training and Validation BCE Loss Across Epochs

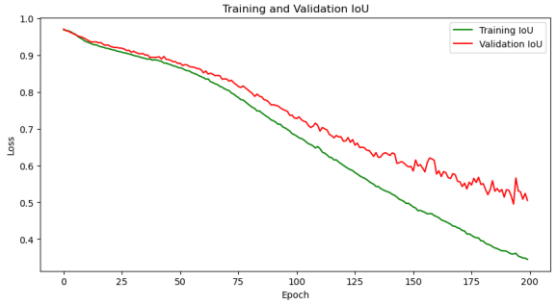


Figure 16: Training and Validation IoU Loss Across Epochs

Figure 17: Training and Validation IoU Scores Across Epochs

V. CONCLUSION

Limitations

While the model achieved high accuracy, its performance was slightly affected by cases with extremely small tumors and curve edges in MRI modalities. It also shows some slight issues with generalization on the validation set.

Moreover, the lack of ground truth annotations in the test set limited objective evaluation of unseen data, making qualitative analysis the only assessment method for the test set.

Enhanced Opportunity

To improve the model's performance and address the gap between training and validation, several enhancements can be implemented:

- Increase Dataset Size: The current train dataset includes 300 samples. While augmentation adds variability, increasing the dataset size with additional real samples will improve generalization.
- Hyperparameter Fine-tuning: Hyperparameter optimization can further refine the training process. Due to the computing constraints, the model is not fine-tuned with different hyperparameters such as optimizers, learning rates, batch size.
- Incorporate Dice Loss: Dice Loss is particularly effective for addressing class imbalance, which is common in medical imaging datasets where the tumor region (positive class) is significantly smaller than the background (negative class). By incorporating the Dice coefficient, the model can focus more on overlapping areas between the predicted and actual masks, thereby achieving better segmentation results.
- Incorporate Classification Task: the model can be extended to perform a dual task: tumor segmentation and classification (benign vs. malignant, or different tumor types), thereby providing comprehensive diagnostic insights in a single step.

REFERENCES

[1] Awsaf. (2020). Brats20 Dataset Training Validation [Data set]. Kaggle. https://www.kaggle.com/datasets/awsaf49/brats20-dataset-training-validation/data

[2] Liu, Z., Tong, L., Chen, L., Jiang, Z., Zhou, F., Zhang, Q., ... & Zhou, H. (2022). Deep learning-based brain tumor segmentation: a survey. Complex & Intelligent Systems, 9, 1001-1026. https://doi.org/10.1007/s40747-022-00815-5

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv preprint arXiv:1505.04597*. https://arxiv.org/pdf/1505.04597

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *arXiv preprint arXiv:1512.03385*. Retrieved from https://arxiv.org/pdf/1512.03385

[5] Salcedo, E.; Jaber, M.; Requena Carrión, J. A Novel Road Maintenance Prioritisation System Based on Computer Vision and Crowdsourced Reporting. J. Sens. Actuator Netw. 2022, 11, 15. https://doi.org/10.3390/jsan11010015