Fine-Grained Brain Tumor Segmentation on BraTS 2020 Dataset Using an Attention-Enhanced U-Net Model

Thashmitha Bangalore Shekar
tbs@umass.edu

Rishitha Guttapalle Mohan
rguttapallem@umass.edu

Sreya Talasila
stalasila@umass.edu

Abstract

Precise and automatic segmentation of brain tumors is a significant difficulty in medical imaging. Automated methods are necessary to accurately delineate tumor boundaries in MRI scans, as manual delineation is time-consuming and prone to fluctuation across observers. The current study uses the BraTS 2020 dataset with multimodal MRI scans (FLAIR, T1, T1ce, T2) to build and examine an attention-based U-Net model for accurate tumor segmentation.The suggested architecture improves the classic U-Net framework through the addition of attention gates. This allows the model to dynamically prioritize relevant spatial aspects and ignoring less important information.

Data pre-processing pipelines optimize input quality by normalizing intensity, removing irrelevant regions, and addressing class imbalance. The model's performance was evaluated using standard indicators like the Dice coefficient, sensitivity, and specificity, resulting in competitive outcomes. Visualization of segmentation results efficiently delineates tumor locations in robust cases, including those with heterogeneously enhancing tumors. Incorporating attention mechanisms improves segmentation precision and robustness to noise compared to the basic U-Net model. [4]

This study suggests that attention-based architectures can effectively tackle challenging segmentation problems in medical imaging, with potential for real time medicinal purposes. Future initiatives include investigating lightweight designs to reduce computing cost and incorporating multi modal fusion techniques for better tumor characterization.

1. Introduction

Problem Statement: Brain tumors, especially gliomas, are one of the serious medical conditions that require prompt and accurate diagnosis for optimal treatment and positive patient outcomes. MRI is the preferred imaging technique for assessing brain tumors due to its higher spatial resolution and excellent contrast for soft tissues. Yet, accurately segmenting tumor regions from MRI scans is challenging due to varying tumor shape, heterogeneous tissue characteristics, and uneven enhancing patterns across sequences.

Objective: This study presents an attention-based U-Net model, which improves the standard design with the addition of attention gates. Integrating attention processes increases model performance by prioritizing critical areas for segmentation and minimizing the impact of irrelevant or noisy features.

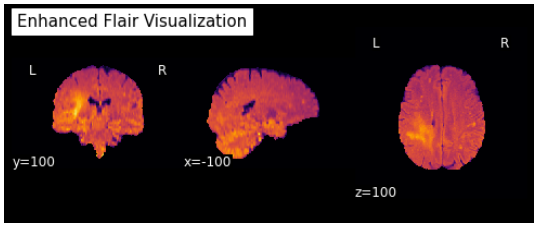

Figure 1. 3D view of a structural brain image

Dataset Used: The BraTS 2020 dataset, a popular dataset for brain tumor segmentation research, offers multimodal MRI images of gliomas. This dataset includes four MRI modalities (T1, T1ce, T2, and FLAIR), each offering unique anatomical and pathological information. The BraTS dataset comprises expert-annotated ground truth segmentations that can be used to evaluate segmentation method's performance. Annotations categorize tumors into three regions: enhancing tumor (ET), peritumoral edema (ED), and necrotic core (NC) determining each zone is

crucial for treatment planning. The Figure 1 illustrates the structural layout of a brain in 3D, serving as a reference for comprehending the spatial context of the regions studied during segmentation.

Expected Results and Evaluation: Our technique has been assessed on the BraTS 2020 dataset, and we do an in-depth examination of segmentation performance applying numerous parameters for evaluation, such as Dice coefficient, sensitivity, and specificity.

The key elements of this paper are as follows:

1. Constructing an attention-based U-Net model for segmenting brain tumors in MRI images.

2. Use the BraTS 2020 dataset for evaluating and validate the proposed model using real-world clinical data.

3. Improved segmentation performance by using attention processes to target tumor locations and suppress unimportant features.

This paper is structured as follows: Section 2 provides a review of related work in the field of brain tumor segmentation and attention-based U-Net architectures. Section 3 presents the methodology, detailing the dataset, model architecture, and training process. Section 4 discusses the results and performance evaluation of the model. Finally, Section 5 concludes the paper and outlines future research directions.

2. Literature Review

Convolutional Neural Networks (CNNs) have become prevalent in the area of medical imaging because of their ability to extract spatial data spontaneously. However, typical CNNs frequently struggle with exact positioning and border delineation, especially in complex applications such as tumor segmentation.[8] This approach, which can maintain high-resolution spatial information, outperforms typical convolutional neural networks (CNNs) at keeping fine-grained details, which are critical for segmenting complicated characteristics. The U-Net's symmetric skip connections preserve fine-grained features, making it more successful for medical segmentation of medical images than typical CNNs. [2]

Transformer-based models, such Vision Transformers (ViT), have recently been developed for tasks like segmentation and shown promising outputs in capturing long-range dependencies. Yet, the usage in medical imaging is limited since they often require vast data sets for training and are computationally costly. Research contrasting U-Net variations with ViT(Vision Transformers) based models has shown that U-Net is still reliable and computationally effective with limited data, which makes it the recommended option for real-time medical applications.[9]

Attention-based U-Net architectures have shown promising findings in segmenting brain tumors on the Brats dataset. Attention mechanisms are being added to traditional U-Net architectures, which are known for their efficient performance in medical image segmentation. This improves their potential to focus on relevant features, notably in cases of minor changes in tumor structures.[1]

Chen et al. (2023) found that adding attention mechanisms to U-Net improved its ability to segment brain tumors. This improvement allows the model to focus on crucial locations in MRI scans, especially when dealing with complicated tumor boundaries. Attention gates let the model learn spatial information and suppress irrelevant characteristics, improving segmentation accuracy.

Moreover, The latest advances involve studying hybrid models combining U-Net with deep learning-based fusion methodologies. These approaches improve segmentation precision and the model's capacity to generalize across different forms of tumors and MRI patterns. These advancements emphasize the significance of incorporating attention mechanisms into deep learning systems for medical image segmentation.[3] Combining U-Net and attention-based models has the potential to improve tumor identification and diagnostic accuracy particularly when dealing with challenging datasets such as Brats 2020.[6]

Gaps Addressed

1. The lack of capability of traditional U-Net models to dynamically focus on key tumor regions makes it difficult to segment complicated structures.[5]
2. The absence of specific loss functions or refinement methods in regular U-Net models frequently results in poor delineation of tumor boundaries, notably in compact or unclear regions.[7]
3. The practical implementation of classic and attention-based U-Nets is limited by their high computing costs, particularly when scaling to big 3D MRI datasets.[7] [5]
4. All features are processed similarly by standard CNNs, which may neutralize crucial region-specific information which degrades segmentation accuracy. [2]

Proposed Solution:

This work presents an improved attention-based U-Net model that uses attention gates to improve the standard architecture. By incorporating attention processes, the model

dynamically focuses on essential regions for tumor segmentation, decreasing the impact of irrelevant information. Furthermore, the approach makes use of multimodal MRI data, such as FLAIR, T1, T1ce, and T2 scans, via shared encoders and attention-guided fusion which allows for the extraction of complementing information from many modalities, increasing the accuracy and precision of tumor identification.

3. Technical approach

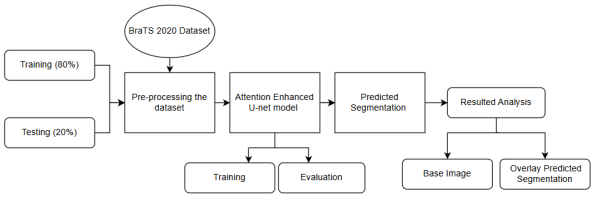

Figure 2. Flowchart presenting Workflow for Brain Tumor Segmentation

This study uses a modified U-Net architecture with attention mechanism to segregate brain tumors from the Brats 2020 dataset. The purpose is to enhance U-Net's encoder-decoder structure with an attention mechanism to focus on significant sections of the image, especially tumor regions. The flowchart of the sequential processes of the workflow for brain tumor segmentation, that includes various phases is shown in Figure 2.

3.1. Dataset

Data Pre-processing: The Brats 2020 dataset undergoes numerous pre-processing processes to prepare it for model training, specifically for brain tumor segmentation utilizing deep learning models such as U-Net with attention gates. The following is a full overview of the pre-processing pipeline using the provided code. The Figure 3 depicts Gray scale image, which shows a single MRI modality utilized as input to the model.

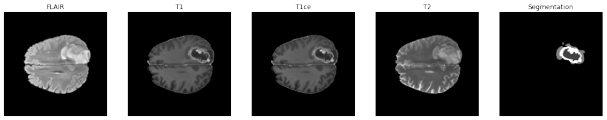

Figure 3. Gray scale image

1. Loading MRI Modalities: Preprocessing starts with loading various MRI modalities for each patient. The Brats dataset contains several MRI modalities: FLAIR, T1, T1c (contrast-enhanced), T2, and the segmentation mask (SG). The modalities are maintained in NIfTI (.nii) format.

The nibabel library loads NIfTI files, and data is transformed to numpy arrays via np.asarray(). Each modality is added to a list to create arrays that represent each patient's modalities.

2. Data Pre-processing Function: The Data Preprocessing function loads a list of modality file URLs and stores the data for each modality as a numpy array. Every patient in the dataset is subjected to this function.

The allmodalities array includes data for all modalities for each patient, indexed from 0 to 4. For instance, Modality 0: FLAIR, Modality 1: T1, Modality 2: T1c (contrast-enhanced), Modality 3: T2, Modality 4: Mask Segmentation

3. Concatenating Modalities for Each Patient: Data from the five modalities—FLAIR, T1, T1c, T2, and segmentation mask—is loaded and preprocessed for every patient.The concatenate_slices function creates 2D slices for each MRI modality by slicing the 3D volume data (i.e., each modality) along the depth axis (the z-axis). For tumor segmentation and prediction, the 2D slices are then appended to distinct lists for input (X_slices), output (Y_slices), and two more modalities (TR_slices, TRL_slices).

4. Preparing Data for Model Input:Once the 2D slices are extracted for each modality, the data is transformed into `numpy` arrays with a data type of `float32`, making them compatible with deep learning models. The processed arrays are then divided into four key components:

- **X_data**: This array contains the input images from various modalities (e.g., T1, T2) used for model training.
- **Y_data**: This array holds the ground truth segmentation masks corresponding to the input images.
- **TR_data and TRL_data**: These arrays store additional tumor-related information, such as the tumor region of interest (ROI) or other variables relevant to the segmentation task.

This preprocessing step ensures the data is properly structured and optimized for training and evaluating deep learning models.

5. Splitting the Dataset:Train_test_split is used to divide the dataset into training and testing sets. 15% of the data is in the testing set, while 85% of the data is in the training set.The X_data (input data) and Y_data (ground truth labels) arrays are used for the split, guaranteeing that the input data and matching labels are paired correctly.

6. Shape of Data: The final dimensions of the training and testing datasets were carefully verified to ensure consistency and compatibility with the deep learning model. The

datasets are structured as follows:

- **X_train and X_test**: These arrays contain the input data for training and testing, respectively, with dimensions of (3952,240,240)(3952, 240, 240)(3952,240,240) for each 2D slice.
- **Y_train and Y_test**: These arrays hold the corresponding ground truth segmentation masks for the training and testing datasets, also with dimensions of (3952,240,240)(3952, 240, 240)(3952,240,240).
- **TR_train and TRL_train**: These arrays store additional tumor-related data or variables, with dimensions of (4650,240,240)(4650, 240, 240)(4650,240,240) for each slice.

This dimensional verification step ensures the datasets are correctly prepared for the training and evaluation of the segmentation model.

3.2. Model Architecture

The proposed model is a U-Net-based architecture enhanced with **Attention Gates (AG)**, specifically designed for brain tumor segmentation. The integration of attention mechanisms enables the model to focus on the most relevant regions in the image, such as the tumor, while suppressing irrelevant background features. The overall architecture adopts an encoder-decoder structure, where the encoder extracts hierarchical features at multiple scales, and the decoder progressively reconstructs the segmentation mask by up-sampling the feature maps. The Figure 4 represents a generalized Traditional U-Net Architecture.

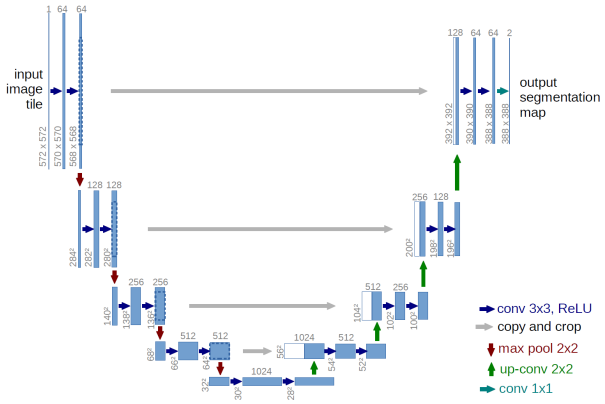

Figure 4. Traditional U-net architecture

The input layer takes a 240x240x1 grayscale image, representing an MRI scan, and processes it through a series of convolutional layers (using 32, 64, 128, 256, and 512 filters at different stages) followed by MaxPooling2D layers to down-sample the image. This down-sampling path helps reduce the spatial resolution while increasing the depth of the feature maps, allowing the network to capture abstract features of the input image. The convolutional layers are followed by BatchNormalization and ReLU activation functions to normalize the activations and introduce non-linearity. At the bottleneck, the model employs 512512512 filters to learn the most abstract and high-level features of the input image. This stage acts as a bridge between the encoder and the decoder paths, capturing key details necessary for accurate segmentation.

The decoder restores the spatial resolution of the image using **UpSampling2D** layers, progressively reducing the depth of the feature maps while reconstructing the segmentation mask. After each up-sampling step, convolutional layers refine the feature maps before they are passed to the subsequent layers. Skip connections are employed to concatenate feature maps from the encoder with those in the decoder, ensuring the retention of important spatial details lost during down-sampling.

A distinctive aspect of this model is the incorporation of **Attention Gates** between the encoder and decoder paths, particularly between the feature maps of the fourth and sixth convolutional layers. These gates apply convolutions to the feature maps from both the encoder and decoder, generating an attention map via a **sigmoid activation**. The attention map highlights the tumor regions and suppresses irrelevant areas, enhancing the segmentation accuracy. The resultant attention map is up-sampled and concatenated with the decoder feature maps, focusing the network's learning on critical image regions.

The final layers of the model involve concatenating up-sampled feature maps with their corresponding encoder feature maps through skip connections. A final convolutional layer reduces the depth of the output feature map to a single channel, followed by a **sigmoid activation function** to produce the binary segmentation mask. Each pixel in the output mask represents either the background (0) or the tumor (1). The Figure 5 shows the architecture that comprises a total of **3,218,434 parameters**, of which **3,214,978 are trainable**, reflecting the model's depth and sophistication. The remaining **3,456 non-trainable parameters** likely correspond to fixed components, such as batch normalization statistics.

3.2.1 Model Evaluation

The proposed model achieved the following average metrics. To assess the performance of the proposed model, the following evaluation metrics are used:

- **Dice Similarity Coefficient (DSC)**: This metric measures the overlap between the predicted segmentation and the ground truth annotation, with higher values indicating

better performance. It is defined as:

$$\text{DSC} = \frac{2|A \cap B|}{|A| + |B|}$$

where A and B are the predicted and true segmentation regions, respectively.

- **Intersection over Union (IoU)**: IoU is another commonly used metric for segmentation tasks. It calculates the ratio of the intersection of predicted and true tumor regions to the union of these regions:

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

The IoU curve shows a steady improvement for both training and validation sets over the epochs, with initial fluctuations observed in the validation set during early training. This instability stabilizes after approximately 10 epochs, indicating the model's convergence toward better predictions. The final IoU values for both training and validation demonstrate strong performance, with minimal deviation between the two, highlighting the model's generalization ability and effectiveness in segmenting the data.

- **Sensitivity and Specificity**: These metrics help evaluate the model's ability to detect true positives (tumor regions) and avoid false positives, respectively. Sensitivity is defined as:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

where TP is the number of true positives and FN is the number of false negatives. Specificity is defined as:

$$\text{Specificity} = \frac{TN}{TN + FP}$$

where TN is the number of true negatives and FP is the number of false positives.

Model: "functional_1"

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 240, 240, 1)	0	
conv2d (Conv2D)	(None, 240, 240, 32)	320	input_1[0][0]
batch_normalization (BatchNorma	(None, 240, 240, 32)	128	conv2d[0][0]
activation (Activation)	(None, 240, 240, 32)	0	batch_normalization[0][0]
max_pooling2d (MaxPooling2D)	(None, 120, 120, 32)	0	activation[0][0]
conv2d_1 (Conv2D)	(None, 120, 120, 64)	18496	max_pooling2d[0][0]
batch_normalization_1 (BatchNor	(None, 120, 120, 64)	256	conv2d_1[0][0]
activation_1 (Activation)	(None, 120, 120, 64)	0	batch_normalization_1[0][0]
max_pooling2d_1 (MaxPooling2D)	(None, 60, 60, 64)	0	activation_1[0][0]
conv2d_2 (Conv2D)	(None, 60, 60, 128)	73856	max_pooling2d_1[0][0]
batch_normalization_2 (BatchNor	(None, 60, 60, 128)	512	conv2d_2[0][0]
activation_2 (Activation)	(None, 60, 60, 128)	0	batch_normalization_2[0][0]
max_pooling2d_2 (MaxPooling2D)	(None, 30, 30, 128)	0	activation_2[0][0]
conv2d_3 (Conv2D)	(None, 30, 30, 256)	295168	max_pooling2d_2[0][0]
batch_normalization_3 (BatchNor	(None, 30, 30, 256)	1024	conv2d_3[0][0]
activation_3 (Activation)	(None, 30, 30, 256)	0	batch_normalization_3[0][0]
max_pooling2d_3 (MaxPooling2D)	(None, 15, 15, 256)	0	activation_3[0][0]
conv2d_4 (Conv2D)	(None, 15, 15, 512)	1180160	max_pooling2d_3[0][0]
batch_normalization_4 (BatchNor	(None, 15, 15, 512)	2048	conv2d_4[0][0]
activation_4 (Activation)	(None, 15, 15, 512)	0	batch_normalization_4[0][0]
up_sampling2d (UpSampling2D)	(None, 30, 30, 512)	0	activation_4[0][0]
conv2d_5 (Conv2D)	(None, 30, 30, 256)	1179904	up_sampling2d[0][0]
batch_normalization_5 (BatchNor	(None, 30, 30, 256)	1024	conv2d_5[0][0]
activation_5 (Activation)	(None, 30, 30, 256)	0	batch_normalization_5[0][0]
conv2d_6 (Conv2D)	(None, 30, 30, 128)	32896	activation_3[0][0]
conv2d_7 (Conv2D)	(None, 30, 30, 128)	32896	activation_5[0][0]
batch_normalization_6 (BatchNor	(None, 30, 30, 128)	512	conv2d_6[0][0]
batch_normalization_7 (BatchNor	(None, 30, 30, 128)	512	conv2d_7[0][0]
activation_6 (Activation)	(None, 30, 30, 128)	0	batch_normalization_6[0][0]
activation_7 (Activation)	(None, 30, 30, 128)	0	batch_normalization_7[0][0]
add (Add)	(None, 30, 30, 128)	0	activation_6[0][0] activation_7[0][0]
activation_8 (Activation)	(None, 30, 30, 128)	0	add[0][0]
conv2d_8 (Conv2D)	(None, 30, 30, 1)	129	activation_8[0][0]
up_sampling2d_1 (UpSampling2D)	(None, 60, 60, 256)	0	activation_5[0][0]
up_sampling2d_2 (UpSampling2D)	(None, 60, 60, 1)	0	conv2d_8[0][0]
concatenate (Concatenate)	(None, 60, 60, 257)	0	up_sampling2d_1[0][0] up_sampling2d_2[0][0]
conv2d_9 (Conv2D)	(None, 60, 60, 128)	296192	concatenate[0][0]
batch_normalization_8 (BatchNor	(None, 60, 60, 128)	512	conv2d_9[0][0]
activation_9 (Activation)	(None, 60, 60, 128)	0	batch_normalization_8[0][0]
up_sampling2d_3 (UpSampling2D)	(None, 120, 120, 128	0	activation_9[0][0]
conv2d_10 (Conv2D)	(None, 120, 120, 64)	73792	up_sampling2d_3[0][0]
batch_normalization_9 (BatchNor	(None, 120, 120, 64)	256	conv2d_10[0][0]
activation_10 (Activation)	(None, 120, 120, 64)	0	batch_normalization_9[0][0]
up_sampling2d_4 (UpSampling2D)	(None, 240, 240, 64)	0	activation_10[0][0]
concatenate_1 (Concatenate)	(None, 240, 240, 96)	0	up_sampling2d_4[0][0] activation[0][0]
conv2d_11 (Conv2D)	(None, 240, 240, 32)	27680	concatenate_1[0][0]
batch_normalization_10 (BatchNo	(None, 240, 240, 32)	128	conv2d_11[0][0]
activation_11 (Activation)	(None, 240, 240, 32)	0	batch_normalization_10[0][0]
conv2d_12 (Conv2D)	(None, 240, 240, 1)	33	activation_11[0][0]

Total params: 3,218,434
Trainable params: 3,214,978
Non-trainable params: 3,456

Figure 5. Summary of the model

3.2.2 Summary of Results

The evaluation results demonstrate that the proposed model performs effectively in both classification and segmentation tasks, with particularly strong results in terms of accuracy, sensitivity, and precision. While the model's overall performance is robust, further enhancements could be explored to improve segmentation boundary accuracy and to address more complex cases within the dataset.Figure

6 shows a comparative visualization of the original MRI scan and its segmented counterpart, emphasizing the ability of the framework to efficiently define tumor boundaries whereas the Figure 7 presents a magnified view of the tumor region predicted by the model.

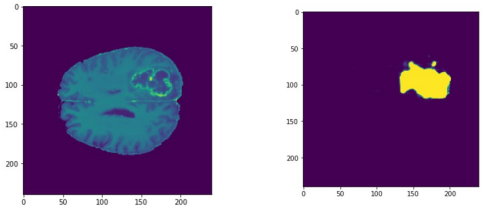

Figure 6. Base Image(left) and segmented Image(right)

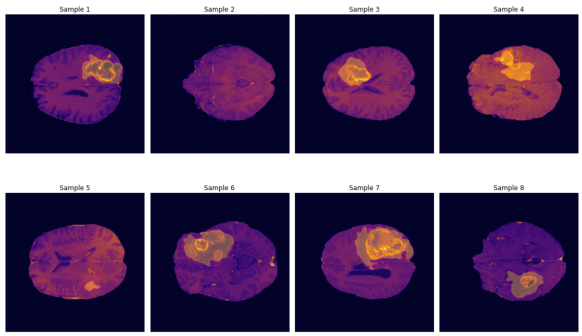

Figure 7. Image displaying the predicted tumor

4. Results

4.1. Evaluation Metrics

To assess the generalization performance of the proposed model, evaluation was conducted on a separate test dataset using a comprehensive set of metrics, including accuracy, precision, sensitivity, specificity, Dice Similarity Coefficient (DSC), and Intersection over Union (IoU). These metrics were selected to provide a well-rounded view of the model's segmentation and classification capabilities, particularly in identifying both positive and negative cases effectively.

1. Accuracy: The model achieved an accuracy of **98.56%** on the test dataset, signifying its ability to correctly classify the majority of cases. While high accuracy is indicative of strong overall performance, it must be considered alongside other metrics to ensure balanced results, especially in the context of imbalanced datasets.

2. Precision: With a precision score of **0.8318**, the model demonstrates a strong capability to minimize false positives. This indicates that the model reliably identifies true positive instances while effectively avoiding misclassifications.

3. Sensitivity: Sensitivity, or recall, reached **0.9237**, reflecting the model's success in detecting the majority of true positive cases. This high sensitivity indicates a low rate of false negatives, which is critical for accurate tumor detection.

4. Specificity: The model achieved near-perfect specificity, highlighting its proficiency in accurately classifying negative cases and minimizing false positives. This reinforces the model's robustness in distinguishing between tumor and non-tumor regions.

5. Dice Similarity Coefficient (DSC): The DSC, which measures the overlap between predicted and ground truth segmentation, achieved a value approximately of around **0.987** during training. Although there was a slight decline during testing, this metric underscores the model's ability to segment regions of interest with high accuracy. The reduction during testing suggests potential areas for improvement, such as refining predictions near segmentation boundaries.

6. Intersection over Union (IoU): An IoU score of **0.7784** indicates good alignment between the model's predictions and the ground truth. This metric demonstrates the model's ability to focus on relevant areas while maintaining a low rate of false positives and false negatives.

4.2. Qualitative Analysis

Visualization of predicted segmentation maps demonstrates the model's capability to accurately delineate tumor regions, even in challenging cases with irregular tumor boundaries.

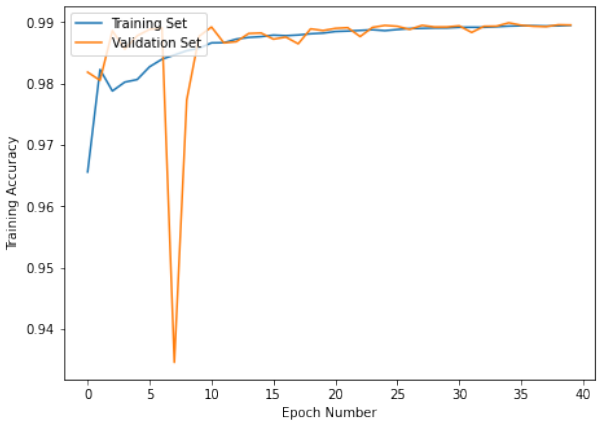

Figure 8. Accuracy vs epoch

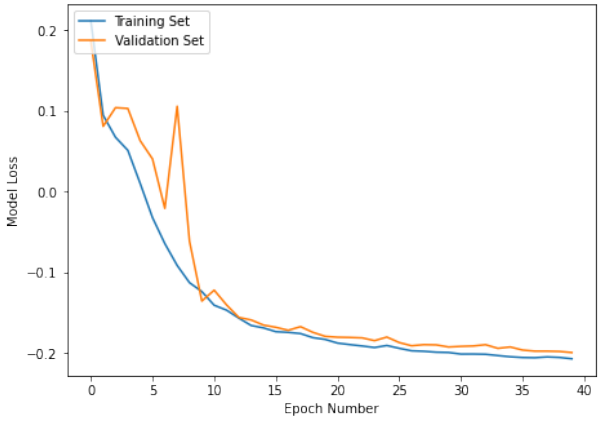

Figure 9. Loss vs epoch

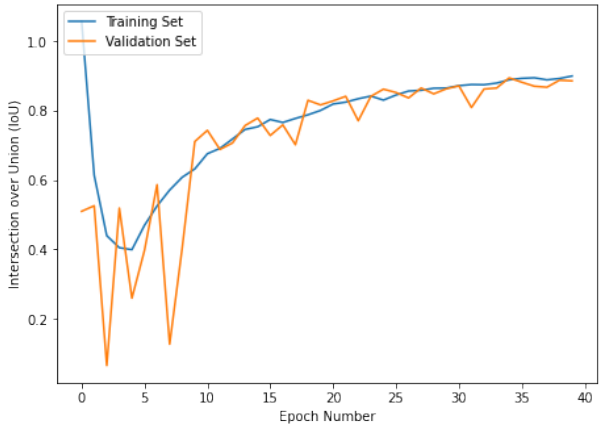

Figure 10. IoU vs epoch

The model's convergence and performance are demonstrated by the plots under Figures 8, 9, and 10 which illustrate the training metrics over epochs and display trends in accuracy, loss reduction, and IoU improvement respectively.

5. Conclusion

This study uses the BraTS 2020 dataset to investigate the use of a modified U-Net architecture improved with an attention mechanism for brain tumor segmentation. To enhance the model's capacity to concentrate on important sections of the MRI images, particularly the tumor regions, attention gates are being incorporated into the U-Net design. Multi-modal MRI data, including as FLAIR, T1, T1c, and T2, as well as matching tumor segmentation masks, are available in the BraTS 2020 dataset. These masks classify the tumor into necrotic core, peritumoral edema, and enhancing tumor. To import and arrange the multi-modal MRI data, slice the 3D volumes into 2D pictures, and guarantee correct alignment of the input and output data, a comprehensive preprocessing pipeline was put in place. This stage was essential for using to train the model with high-quality, well-structured data. The model was trained using a split dataset, with 85% of the data used for training and 15% for testing, allowing for effective evaluation of its segmentation capabilities.

The results of the model evaluation demonstrated promising performance, with improvements in both accuracy and intersection over union (IoU) metrics, showcasing its potential for medical image analysis.. The model successfully segmented several tumor locations by utilizing the attention mechanism and the U-Net architecture's characteristics, which makes it a useful tool for clinical applications. Nonetheless, there are still chances for optimization, including adjusting the model to perform even better. In order to improve tumor characterisation, future studies will look into lightweight architectures to lower computing overhead, optimize the model's performance, and incorporate multimodal fusion techniques. These developments may increase brain tumor segmentation's efficacy and precision, increasing its applicability in actual medical situations.

References

[1] Brain tumor segmentation using u-net. In *2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD)*, pages 214–218, 2024. 2

[2] Guantian Huang, Bixuan Xia, Haoming Zhuang, Bohan Yan, Cheng Wei, Shouliang Qi, Wei Qian, and Dianning He. A comparative analysis of u-net and vision transformer architectures in semi-supervised prostate zonal segmentation. *Bioengineering*, 11(9), 2024. 2

[3] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18:203 – 211, 2020. 2

[4] Bjoern H. Menze and Muhammad Yaqub. The multimodal brain tumor image segmentation benchmark (brats). In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2021. 1

[5] Ozan Oktay, Jo Schlemper, Loïc Le Folgoc, M. J. Lee, Mattias P. Heinrich, Kazunari Misawa, Kensaku Mori, Steven G. McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net: Learning where to look for the pancreas. *ArXiv*, abs/1804.03999, 2018. 2

[6] Vinod P, Kuppusamy P, and Manimaran A. Enhancing brain tumor segmentation using u-net and attention mechanism. In *2023 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, pages 1314–1321, 2023. 2

[7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *ArXiv*, abs/1505.04597, 2015. 2

[8] Nahian Siddique, Sidike Paheding, Colin P. Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, 9:82031–82057, 2021. 2

[9] Chen Zhang, Xiangyao Deng, and Sai Ho Ling. Next-gen medical imaging: U-net evolution and the rise of transformers. *Sensors*, 24(14), 2024. 2