Semantic Segmentation of Gliomas with 3D U-Net in Multimodal MRI

Eduardo Henrique da Silva Santana
undergraduate in computer science
CIn - *Federal university of Pernambuco*
Recife-PE, Brazil
ehss@cin.ufpe.br

Matheus Ayres dos Santos
undergraduate in computer engineering
CIn - *Federal university of Pernambuco*
Recife-PE, Brazil
mas11@cin.ufpe.br

Abstract*— The diagnosis and treatment of brain tumors, such as gliomas, critically depend on the precise delineation of affected areas in medical imaging. Manual segmentation, performed by specialists, is a time-consuming, subjective, and variable process. This paper presents the development and evaluation of a Deep Learning model, based on an optimized 3D U-Net architecture, for the automatic segmentation of gliomas from multimodal magnetic resonance imaging (MRI). Using the BraTS 2020 dataset, the model was trained to identify and delineate tumor sub-regions, including the necrotic core, peritumoral edema, and the enhancing tumor. The methodology involved a pre-processing stage focused on data normalization and resizing, the 3D U-Net architecture with specific optimizations, and a combined loss function (Dice Loss and Cross-Entropy) to handle class imbalance. The model achieved a Dice Score of 0.811 for whole tumor (WT) segmentation, 0.644 for the tumor core (TC), and 0.563 for the enhancing tumor (ET), demonstrating the feasibility of the approach to accelerate and standardize the diagnostic process.

I. *INTRODUCTION*

Gliomas are among the most common and aggressive primary brain tumors. The complexity of these lesions, which exhibit histological heterogeneity and different sub-regions (necrotic core, edema, active tumor), makes diagnosis and therapeutic planning challenging tasks [2]. Magnetic Resonance Imaging (MRI) is the primary imaging modality for visualizing these tumors, providing detailed information about the brain's anatomy.

However, the manual analysis and segmentation of MRI images is a time-intensive process, highly dependent on the specialist's experience, and subject to inter- and intra-observer variability. This "bottleneck" can delay diagnoses and the development of treatment plans, such as surgeries and radiotherapies, where precise tumor delineation is fundamental.

To overcome these limitations, Deep Learning techniques have emerged as a powerful solution. Architectures like the U-Net [3] have revolutionized biomedical image segmentation, offering the ability to learn complex features and perform automatic segmentations with high accuracy.

This work aims to develop and evaluate a model based on the 3D U-Net architecture for the automatic and precise segmentation of gliomas in multimodal MRI images, with the goal of creating a tool that can assist healthcare professionals by making the diagnosis faster, more reliable, and reproducible.

II. METHODOLOGY

The methodology was structured into four main stages: dataset selection and preparation, model architecture definition, training, and finally, evaluation.

A. Dataset: BraTS 2020

The project utilized the dataset from the *Multimodal Brain Tumor Segmentation Challenge* (BraTS) 2020 [4]. This is a public and widely recognized dataset containing MRI images of patients with gliomas. Each patient has four MRI modalities:

- **T1:** T1-weighted, useful for visualizing general anatomy.
- **T1ce:** T1-weighted with contrast (gadolinium), which enhances areas with a disrupted blood-brain barrier, typical of active tumors.
- **T2:** T2-weighted, sensitive to areas with high water content, such as edema.
- **T2-FLAIR:** T2 with suppression of cerebrospinal fluid signal, highlighting edema and periventricular lesions.

The provided ground truth classifies voxels into four categories: 0 (background), 1 (necrotic and non-enhancing tumor core - NCR/NET), 2 (peritumoral edema), and 4 (enhancing tumor - ET). For training, label 4 was remapped to

B. Data Split

The full dataset, comprising 369 patients, was divided into three subsets. The training set contained 250 patients (67.8%), the validation set had 74 patients (20.1%), and the test set consisted of 45 patients (12.2%). This split ensures that the model is trained on a large portion of the data, validated on an

independent set to tune hyperparameters, and finally evaluated on a completely unseen test set to assess its generalization performance.

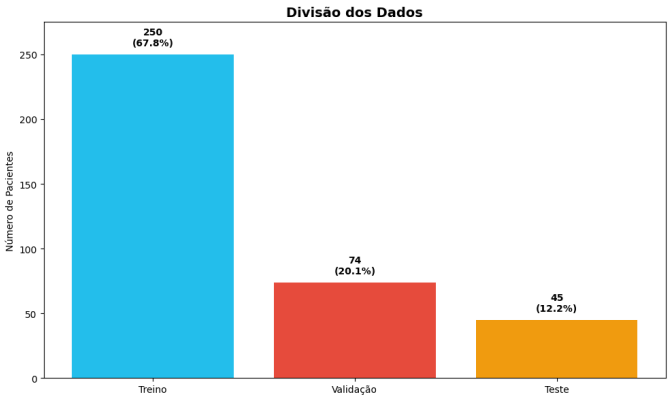

Fig. 1. Data distribution into training, validation, and test sets.

C. Data Pre-processing

To optimize training and standardize the data, the following pre-processing steps were applied:

1. **Modality Selection:** The T1 modality was disregarded due to its low contrast for the regions of interest, keeping the T1ce, T2, and T2-FLAIR modalities as input channels for the model.
2. **Normalization:** Min-Max normalization was applied to scale the voxel intensities to the [0, 1] range.
3. **Resizing:** The volume of each image was resized from (240, 240, 155) to (128, 128, 64) using tri-linear interpolation to reduce the computational load while maintaining spatial proportions.

D. Model Architecture: 3D U-Net

The implemented architecture is an adaptation of the 3D U-Net model [1], [3], which consists of an encoding path (encoder) and a decoding path (decoder) connected by skip connections.

- **Encoder:** Responsible for capturing the image context and extracting hierarchical features. It is composed of sequential blocks of 3D convolution (3x3x3 kernel), Group Normalization, and LeakyReLU activation. In each block, the spatial resolution is halved (MaxPool3D), while the channel depth is doubled.
- **Decoder:** Performs precise tumor localization by reconstructing the segmentation mask to its original resolution. It uses Upsample operations to increase spatial resolution, combining contextual information from the encoder via skip connections.
- **Optimization:** A distinctive feature of the model is the use of different slopes for the LeakyReLU activation function: 0.01 in the encoder to maintain a stable gradient flow, and 0.2 in the decoder to allow for greater distinction in the reconstruction of fine features.

The final architecture has a total of 5,650,684 trainable parameters.

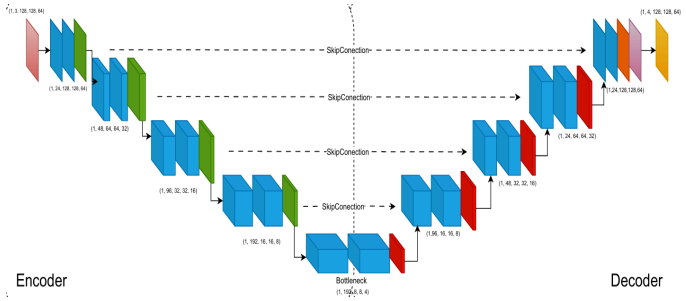

Fig. 2. Diagram of the implemented 3D U-Net architecture.

E. Training

The model was trained with the following parameters:

Optimizer: Adam, with an initial learning rate of 1e-3 and a *ReduceLROnPlateau* scheduler for dynamic adjustment.

Loss Function: A weighted combination of Cross-Entropy Loss and Dice Loss was used to mitigate the imbalance between classes (tumor vs. background). The total loss was calculated as: Specific weights for each tumor class ([0.5, 2.0, 1.0, 2.0]) were applied in the loss function to give more importance to minority classes.

Training Parameters: The model was trained for 20 epochs with a batch size of 16. Early Stopping with a patience of 5 epochs was used to prevent overfitting.

III. RESULTS AND DISCUSSION

The model was evaluated both quantitatively and qualitatively on the validation data.

A. Quantitative Evaluation

The model's performance was measured using the Dice Score, an overlap metric ranging from 0 (no overlap) to 1 (perfect overlap). The evaluation was performed for three clinically relevant regions, which are combinations of the original classes:

- **Whole Tumor (WT):** The complete tumor (NCR/NET + Edema + ET).
- **Tumor Core (TC):** The core of the tumor (NCR/NET + ET).

- **Enhancing Tumor (ET):** Only the active part of the tumor.

After 20 epochs of training, the model achieved the following average Dice Scores on the validation set:

- **Dice Score - WT:** 0.811
- **Dice Score - TC:** 0.644
- **Dice Score - ET:** 0.563

The evolution of these scores during training can be seen in Fig. 3, which shows a consistent upward trend for all three metrics, indicating that the model progressively learned to better delineate the tumor regions. This complements the loss history (Fig. 4), which shows a stable convergence.

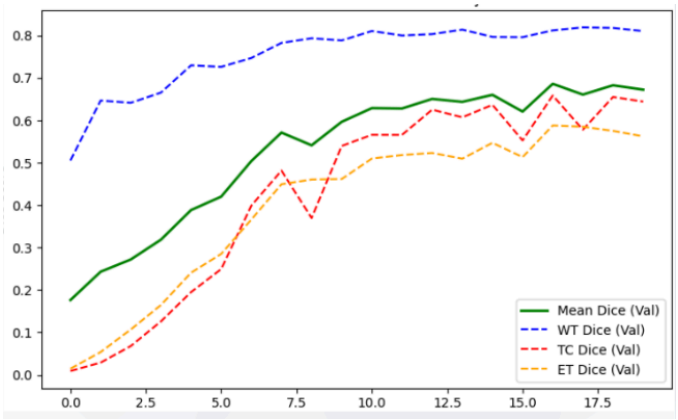

Fig. 3. Dice Score history on the validation set for Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET).

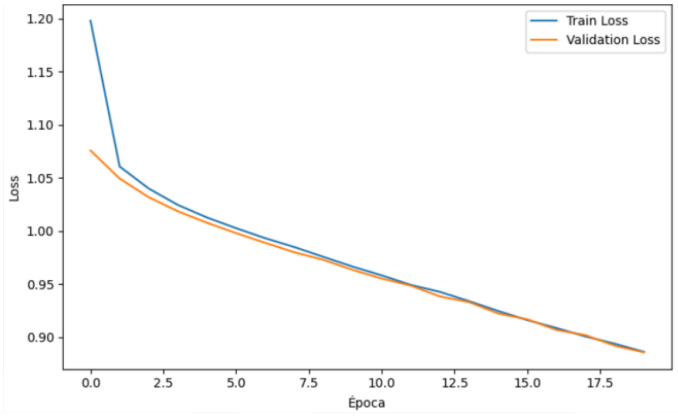

Fig. 4. Loss graph per epoch.

B. Qualitative Evaluation

Visual analysis of the predictions confirms the quantitative results and provides deeper insights into the model's behavior. We highlight two representative cases.

Figure 5 shows the result for patient BraTS20_Training_090. In this case, the model demonstrates strong performance, accurately identifying all three sub-regions. The prediction closely matches the ground truth, validating its effectiveness in complex scenarios where multiple tumor components are present. The specific metrics for this patient were:

- ***Whole Tumor:*** *Dice: 0.7384, Precision: 0.5946*
- ***Tumor Core:*** *Dice: 0.6259, Precision: 0.5734*
- ***Enhancing Tumor:*** *Dice: 0.4904, Precision: 0.3833*

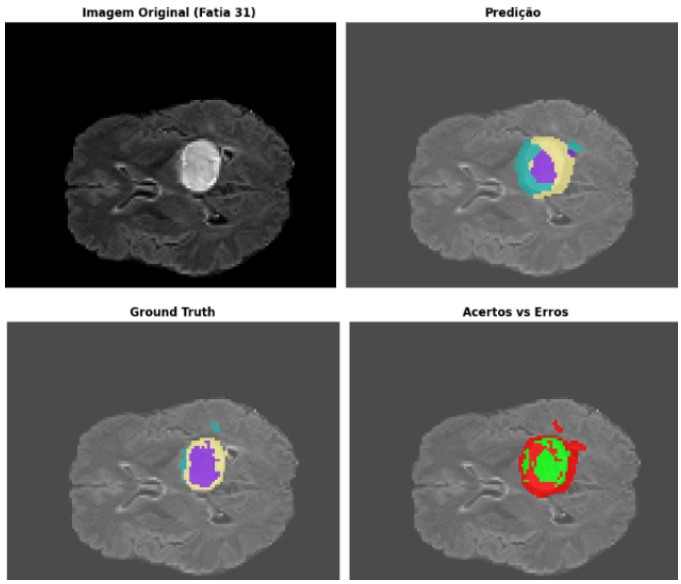

Fig. 5. Example of a successful segmentation for patient BraTS20_Training_090. : Original Image (T2-FLAIR), Model Prediction, and Ground Truth.

Conversely, Figure 6 illustrates the case of patient BraTS20_Training_278. Here, the ground truth indicates the absence of an enhancing tumor (ET). The model correctly predicts no such region, showcasing its high specificity. Although this leads to a Dice Score of 0.0 for the ET class, it represents a correct negative prediction. This demonstrates the model's ability to not only identify present features but also to correctly recognize when a specific tumor sub-region is absent. The metrics for this case were:

- ***Whole Tumor:*** *Dice: 0.7710, Precision: 0.6479*
- ***Tumor Core:*** *Dice: 0.4926, Precision: 0.8471*
- ***Enhancing Tumor:*** *Dice: 0.0000, Precision: 0.0000*

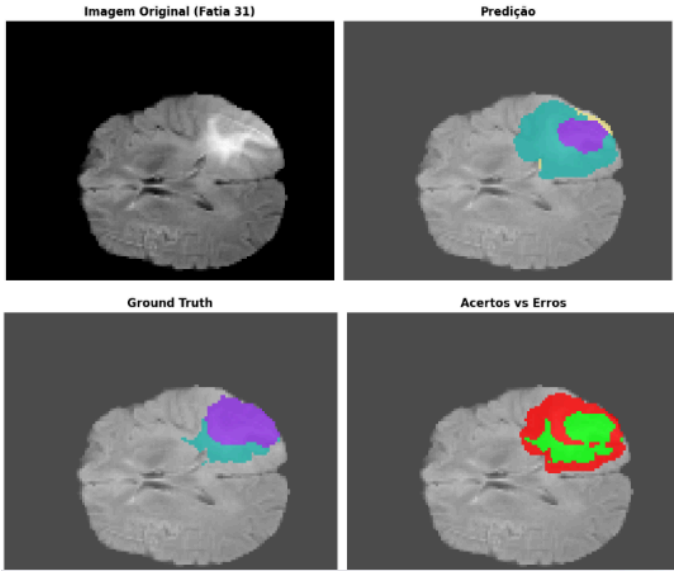

Fig. 6. Example of segmentation for patient BraTS20_Training_278, where the enhancing tumor is absent. The model correctly identifies its absence. Original Image, Model Prediction, and Ground Truth.

IV. CONCLUSION

This work successfully demonstrated the implementation of a 3D U-Net Deep Learning model for the automatic segmentation of gliomas in MRI images. The model achieved promising results, especially in delineating the whole tumor (WT) with a Dice Score of 0.811, indicating its potential as a tool to aid in diagnosis.

The results also highlighted the inherent challenges of the task, such as the difficulty in accurately segmenting smaller and more heterogeneous sub-regions like the enhancing tumor (ET).

For future work, we suggest exploring more advanced architectures (e.g., with attention mechanisms), using more sophisticated data augmentation techniques to increase the variability of the training data, and implementing ensemble methods by combining predictions from multiple models to improve the overall robustness and accuracy of the segmentation.

V. REFERENCES

[1] Futrega, M., Milesi, A., Marcinkiewicz, M., & Ribalta, P. (2021). *Optimized U-Net for Brain Tumor Segmentation*. arXiv:2110.03352.

[2] Havaei, M., et al. (2017). *Brain tumor segmentation with Deep Neural Networks*. Medical Image Analysis, 35, 18-31.

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (pp. 234-241).

[4] Menze, B. H., et al. (2014). *The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)*. IEEE Transactions on Medical Imaging, 34(10), 1993-2024.

[5] Isensee, F., et al. (2021). *nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation*. Nature Methods, 18(2), 203-211.