Deep Learning based Brain Tissue Segmentation from MRI: A 2D CNN Approach

Mahdi Islam[1] and Musarrat Tabassum[1]

[1] *University of Girona, Erasmus Mundus Joint Master's Program in Medical Imaging and Applications (MAIA).*

January 13, 2025

Abstract

Brain tissue segmentation plays a pivotal role in medical imaging by facilitating the precise differentiation of anatomical structures, which is crucial for diagnosing neurological conditions and conducting advanced research. Brain tissue segmentation has demonstrated great utility in quantifying MRI data by serving as a precursor to further post-processing analysis. In this project, we propose a deep learning-based framework to segment white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) in magnetic resonance imaging (MRI). Leveraging 2D convolutional neural networks (CNNs), our approach integrates advanced preprocessing and model optimization techniques to enhance segmentation accuracy. We demonstrate the prformance of our model on the IBSR18 dataset by training and validating our models, focusing on the segmentation of gray matter, white matter, and cerebrospinal fluid. We utilize the Dice similarity coefficient as the primary evaluation metric to assess segmentation performance. The proposed methodology demonstrates the potential for robust and efficient brain tissue segmentation, paving the way for improved clinical diagnostics and neuroscientific investigations.

Keywords—Brain Tissue Segmentation, MRI, Convolutional Neural Networks, White Matter, Gray Matter, Cerebrospinal Fluid, Deep Learning, Dice Similarity, Hausdorff Distance.

1 Introduction

Segmentation of medical images has long been a cornerstone in advancing diagnostic accuracy and therapeutic interventions in healthcare. By delineating distinct regions of interest based on intensity, texture, or structure, segmentation enables in-depth analysis of anatomical and pathological features. In brain imaging, the segmentation of magnetic resonance (MR) images into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) is of paramount importance for quantitative analysis. Such segmentation not only aids in the diagnosis and management of neurological disorders, including Alzheimer's disease and multiple sclerosis, but also serves as a critical tool for understanding brain anatomy and function.

With the advent of deep learning, convolutional neural networks (CNNs) have emerged as a powerful solution for medical image segmentation, offering unparalleled accuracy and automation. Building on these advancements, this project introduces a deep learning-based framework for brain tissue segmentation. Our methodology incorporates advanced preprocessing steps to normalize the data and address inherent class imbalances, enabling precise and consistent segmentation across varying MR images.

The focus of this work is on segmenting WM, GM, and CSF using a CNN-based architecture, with the evaluation centered around the Dice similarity coefficient and Hausdorff Distance. By addressing key challenges in brain tissue segmentation, such as boundary delineation and class disparity, this project aims to contribute to the development of reliable and interpretable solutions for clinical and research applications.

2 Dataset

The dataset contains 18 T1-weighted scan and ground truth of segmentation masks for each cases that corresponds to white matter (WM), grey matter (GM), and cerebrospinal fluid (CSF). The dataset contains scans of normal subject from the Internet Brain Segmentation Repository (IBSR) [1], known as IBSR 18. The dataset has been split for training, validation and testing the model. Case 1,3,4,5,6,7,8,9,16 and 18 are used for training, case 11,12,13,14 and 17 for validation and finally case 2,15,10 for test set evaluating the model performance.
Figure 1 shows the images of the dataset.

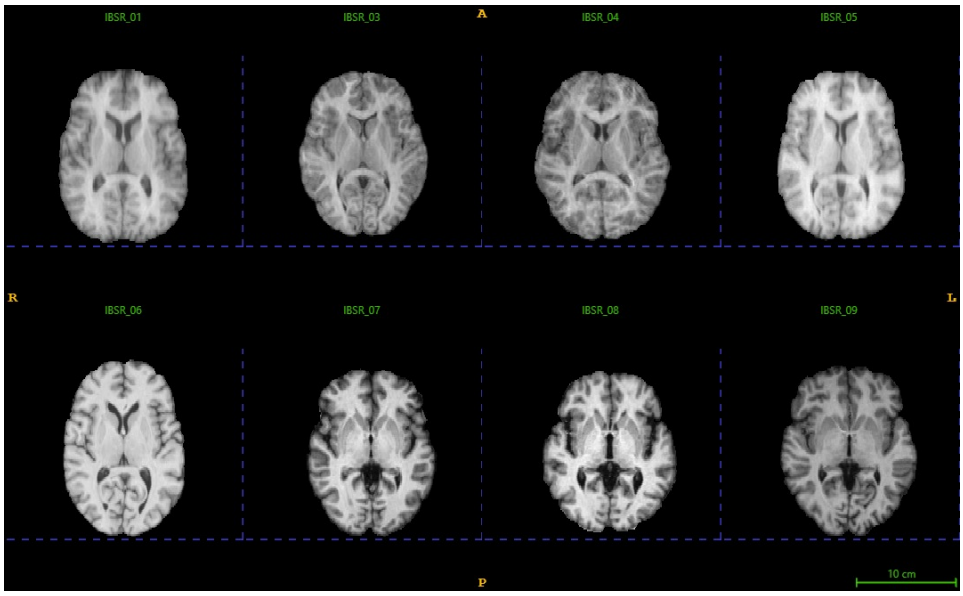


Figure 1: IBSR Dataset

3 Methodology

The methodology is divided into three main subsections; Preprocessing, Models Architectures and Model Training.

3.1 Preprocessing

3.1.1 Normalization

During the preprocessing of brain MRI images, Z-normalization was employed to standardize the intensity values of the images. This technique ensures that the intensity values have a mean of 0 and a standard deviation of 1, calculated using the pixel intensity distribution of the entire image. By transforming the data into this standardized scale, Z-normalization minimizes intensity variations across images caused by differences in acquisition protocols or scanners. Since Group 2's cases has a presence of outliers towards higher values observed from its non-Gaussian distribution. Equation 1 shows the normalization technique used.

$$X_{\text{normalized}} = \frac{X - \text{mean}_Q(X)}{\text{std}_Q(X)}; \quad \forall X > 0 \tag{1}$$

Where:

- X are the intensity values.
- $mean_Q(X)$ is the mean computed using the 25th and 75th quantiles.

- $std_Q(X)$ is the standard deviation computed over the same quantile range.

3.1.2 Data Augmentation

Random flips and rotations were performed on the original images as a part of the augmentation process to improve the algorithm's reliability.

3.1.3 Slice Selection

For training purposes, only the axial slices were utilized. Initially, it was considered to prioritize slices containing cerebrospinal fluid (CSF) intensities. However, after completing the baseline experiment, which included all axial slices from each MRI volume in the training set, the results demonstrated excellent performance. Consequently, this approach was retained for all subsequent trials.

3.2 Models Architectures

Multiple model architectures have been developed from scratch, fine-tuned, and evaluated for this project. These models are as follows:

1. **2D U-Net**:
   The implemented U-Net is a convolutional neural network designed for semantic segmentation tasks, particularly in medical imaging, as originally proposed by Ronneberger et al. [2]. It follows an encoder-decoder architecture with batch normalization and ReLU activation in convolutional blocks. Skip connections are used to merge features between the downsampling (encoder) and upsampling (decoder) paths, preserving spatial information for accurate segmentation. The encoder reduces spatial dimensions through max-pooling while increasing feature depth, and the decoder restores resolution using transposed convolutions, concatenating them with encoder features. A bottleneck captures high-level features, and the network outputs segmentation maps via a final 1x1 convolution followed by a sigmoid activation. The architecture is efficient and versatile, suited for various segmentation challenges due to its computationally efficient design and parameter scalability.

2. **2D Sep-Res U-Net**:
   The 2D Sep-Res U-Net enhances the baseline U-Net by introducing separable convolutions and residual blocks, as inspired by research advancements [3], [4]. The encoder consists of convolutional blocks with batch normalization and ReLU activation, followed by max-pooling. The decoder employs transposed convolutions for upsampling and incorporates skip connections that can either be concatenated or added. These features improve gradient flow and enhance feature representation, particularly for tasks involving complex structures. The addition of separable convolutions reduces computational overhead, making the model more efficient for large-scale medical datasets. A 1x1 convolution in the output layer produces segmentation maps with a sigmoid activation for binary segmentation tasks.

3. **2D Attention U-Net**:
   The Attention U-Net builds upon the U-Net architecture by integrating attention gates, as introduced by Oktay et al. [5]. These gates are applied at skip connections, allowing the network to focus on relevant features while suppressing irrelevant ones. The encoder extracts hierarchical features using convolutional blocks with batch normalization, followed by max-pooling. The decoder path reconstructs spatial resolution via transposed convolutions and additive skip connections with attention gating. The attention mechanism uses gating signals from the decoder and spatial information from the encoder to generate attention maps, improving segmentation performance in low-contrast and noisy datasets. A bottleneck further refines features before decoding, and a 1x1 convolution with sigmoid activation produces the final output. This architecture is particularly effective for medical imaging tasks requiring precise localization and segmentation.

Table 1: Comparison of 2D U-Net, 2D Sep-Res U-Net, and 2D Attention U-Net Architectures

Feature	**2D U-Net (Baseline)**	**2D Sep-Res U-Net**	**2D Attention U-Net**
Batch Normalization	✓	✓	✓
Skip Connections		✓	✓
Separable Convolution		✓	✓
Inverse Bottleneck		✓	✓
Attention Gate			✓

3.3 Model Training

- Loss Function: In this challenge, several loss functions were explored, including Dice loss and cross-entropy loss. However, to effectively handle the issue of class imbalance, a weighted categorical cross-entropy loss was ultimately chosen. This approach assigns greater importance, or weight, to the minority class of cerebrospinal fluid, ensuring a more balanced optimization process. The formula for this loss is presented in Equation 2

$$L = -\sum w \cdot (y \log(\hat{y})) \tag{2}$$

  Here,

  - L is the weighted categorical cross-entropy loss.
  - w are the weights given to each class.
  - y is the ground truth.
  - $\hat{y}$ are the predicted labels.

  The weights for each class were empirically chosen based on the distribution and difficulty of segmentation. The weights can be found in Table 2.

Table 2: The weights assigned for each class for the weighted categorical cross-entropy loss

Class	**Weight** w
Background	1
Cerebrospinal Fluid	10
Gray Matter	3
White Matter	3

- Learning Rate:

  To prevent early convergence on a plateau, a learning rate scheduler was employed. If the loss does not decrease after ten epochs, the learning rate is reduced by a factor of ten. This adjustment enables a more precise gradient descent towards the minimum. The initial learning rate was set at 5×10^{-4}.

- Early Stopping:

  Early stopping was applied with a patience of twenty epochs, based on the validation loss. This means training is halted if the validation loss does not improve for twenty consecutive epochs, helping to minimize unnecessary computational overhead.

- Best Model: The selection of the best model is determined using the validation loss. The model corresponding to the epoch with the lowest validation loss is chosen, regardless of whether early stopping is triggered or not.

Training and validation performances were monitored using Weights & Biases [6]

4 Results

4.1 Metrics

To evaluate the effectiveness of our approach and analyse the laboratory experiments, we assessed the various cases in our dataset using two key metrics: the Dice Score (DSC), as outlined in Equation 3, and and the Hausdorff Distance (HD), which is presented in Equation 4.:

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3}$$

$$d_H(X, Y) = \max\{d_{XY}, d_{YX}\} = \max\left\{\max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(x, y)\right\} \tag{4}$$

5 Results & Discussion

5.1 Quantitative Results

The results reveal some critical insights into the performance of the three models; Attention U-Net, Sep U-Net, and U-Net. The analysis focuses on two metrics, Dice score and Hausdorff distance, which provide complementary views of the segmentation quality. The summary of the results is presented in Table 3, 4 which shows that the Attention U-Net achieves the highest Dice scores across all validation cases.

Table 3: Mean Dice Scores By Model And Validation Case

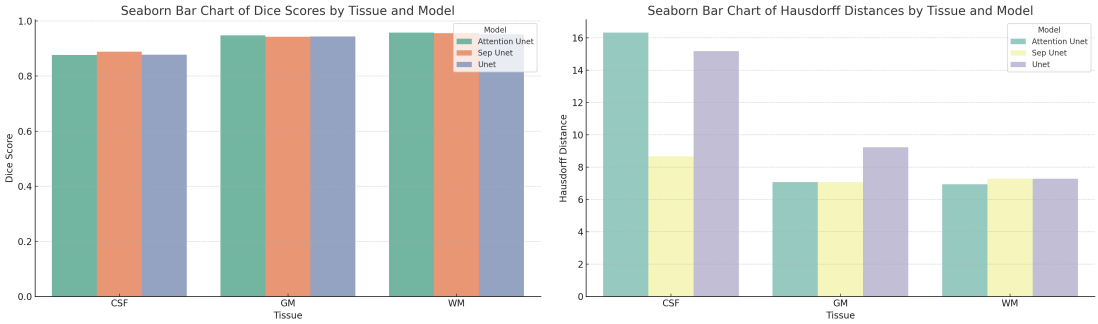
Validation Case	Attention U-Net	Sep U-Net	U-Net
IBSR_11	0.9275	0.8887	0.8778
IBSR_12	0.9304	0.899	0.8954
IBSR_13	0.9265	0.8939	0.8868
IBSR_14	0.9284	0.8987	0.8893
IBSR_17	0.9292	0.9037	0.8923

Table 4: Hausdorff Distances By Model And Validation Case

Validation Case	Attention U-Net	Sep U-Net	U-Net
IBSR_11	16.31	8.66	15.17
IBSR_12	12.48	16.63	24.83
IBSR_13	15.88	18.7	23.33
IBSR_14	15.63	20.4	23.67
IBSR_17	15.29	22.86	23.17

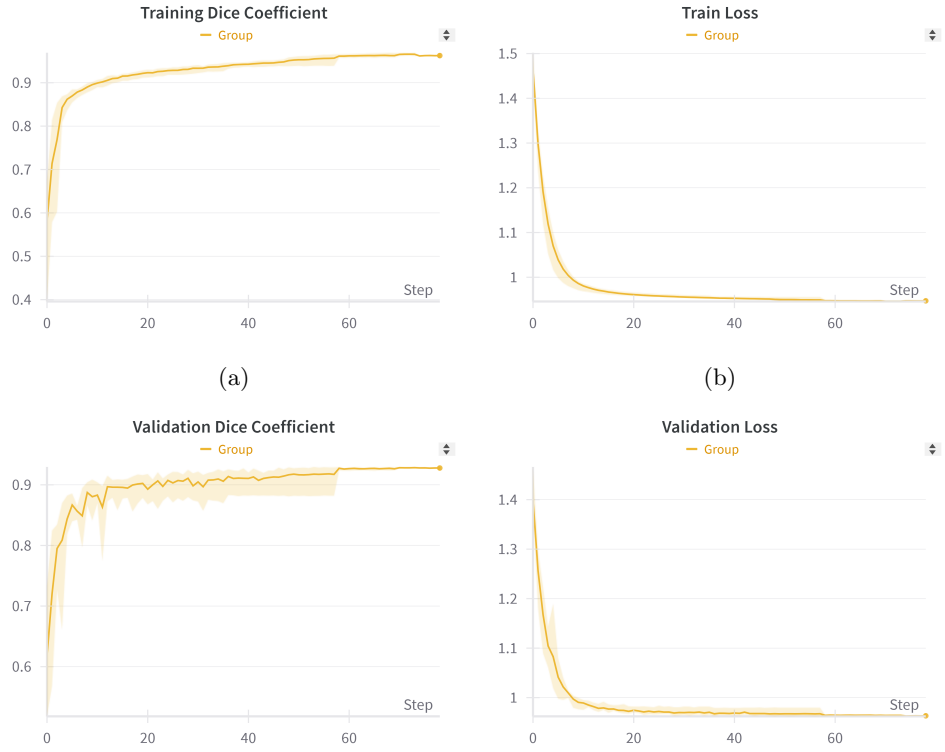
For example, in the IBSR_17 validation case, the Attention U-Net attained a Dice score of 0.9292, surpassing the Sep U-Net (0.9037) and U-Net (0.8923). These results, visualized in the Dice score bar plot (Figure 2a), highlight the superior ability of the Attention U-Net to segment all three tissues—gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). This model's architecture likely enables it to capture relevant spatial features more effectively, leading to high segmentation accuracy. The Hausdorff distance, which measures the alignment between predicted and ground truth segmentation boundaries, is summarized in the same table and further visualized in Figure 2b. Here, the Sep U-Net demonstrates its strength, achieving the lowest Hausdorff distances in most cases. For instance, in IBSR_14, the Sep U-Net achieved a distance of 20.4, which is better than the Attention U-Net (15.63) and

U-Net (23.67). This suggests that the Sep U-Net excels in refining segmentation boundaries, making it particularly effective for applications that require precise delineation of tissue edges.



(a) (b)

Figure 2: Bar Plot. (a) Dice Score (b) Hausdorff Distance.



(a) (b)

(c) (d)

Figure 3: Training and validation performance throughout each epoch ; (a) Train Dice Coefficient (b) Train Loss (c) Validation Dice Coefficient (d) Validation Loss.

The learning curves for validation loss and validation Dice coefficient, shown in Figure 3, provide additional insights into the models' training processes. The validation loss graph reveals a rapid decrease during the initial training steps, followed by stabilization, suggesting effective learning across all models. Correspondingly, the Dice coefficient graph shows a steady increase, with the Attention U-Net achieving the highest validation Dice scores overall. This indicates that the Attention U-Net not only performs well in terms of segmentation accuracy but also generalizes effectively to unseen validation data. The learning curves for validation loss and validation Dice coefficient, shown in Figure 3, provide additional insights into the models' training processes. The validation loss graph reveals a rapid decrease during the initial training steps, followed by stabilization, suggesting effective learning across all models. Correspondingly, the

Dice coefficient graph shows a steady increase, with the Attention U-Net achieving the highest validation Dice scores overall. This indicates that the Attention U-Net not only performs well in terms of segmentation accuracy but also generalizes effectively to unseen validation data. In conclusion, the Attention U-Net stands out as the most reliable and robust model for brain image segmentation, combining high Dice scores with reasonable boundary accuracy. It is particularly suited for general segmentation applications, as visualized in the bar plots and graphs. The Sep U-Net, however, shows promise for tasks where minimizing boundary errors is a priority, as evidenced by its lower Hausdorff distances.

5.2 Qualitative Results

Qualitative Result in Figure 4

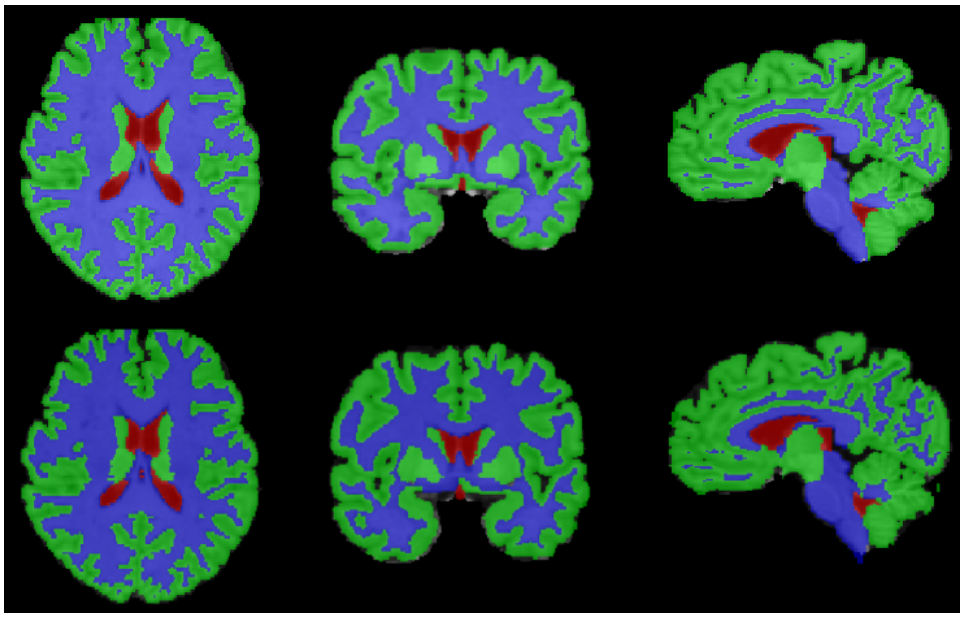


Figure 4: Comparison between the prediction and the validation labels. First row represents the predicted labels and the second row represent the validation levels for axial, coronal and sagittal respectively.

6 Conclusion

This study presents a deep learning-based approach for brain tissue segmentation in MRI images, focusing on segmenting white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). Among the three evaluated models—U-Net, Sep U-Net, and Attention U-Net—the Attention U-Net demonstrated superior performance, achieving the highest Dice scores across all validation cases, making it the most reliable for accurate segmentation tasks. The Sep U-Net, on the other hand, excelled in minimizing boundary errors, as evidenced by its lower Hausdorff distances, making it particularly effective for tasks requiring precise boundary delineation. The integration of advanced preprocessing techniques, architecture optimization, and a robust evaluation framework ensures the reliability and applicability of the proposed methodology. These findings emphasize the importance of tailoring model selection to specific clinical or research needs, paving the way for improved diagnostic and analytical applications in neuroscience and medical imaging.

References

[1] NITRC: IBSR: Tool/Resource Info. https://www.nitrc.org/projects/ibsr. Accessed: [date of access].

[2] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. https://arxiv.org/abs/1505.04597

[3] Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://arxiv.org/abs/1610.02357.

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://arxiv.org/abs/1512.03385.

[5] Oktay, O., Schlemper, J., Folgoc, L. L., et al. (2018). Attention U-Net: Learning Where to Look for the Pancreas. *arXiv preprint arXiv:1804.03999*. https://arxiv.org/abs/1804.03999.

[6] Weights & Biases.https://wandb.ai/site/