Universitat de Girona
Escola Politècnica Superior

MAIA

Medical image segmentation and applications (MISA)

Final Report

# "Brain tissue segmentation using Multi-atlas and Deep learning approaches"

Melika Pooyan and Hadeel Awwad
EPS, University of Girona, Spain.

*u1984923@campus.udg.edu, u1985573@campus.udg.edu*

January 12, 2024

## 1 Abstract

In this study, we focus on the segmentation of the three primary brain tissues - cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM) - using the brain MRI data from the IBSR 18 dataset. Our approach integrates two key methodologies: traditional techniques involving multi-atlas and probabilistic atlas models, and advanced methods utilizing deep learning, specifically the U-net model. The objective was to identify the most effective segmentation strategy for our specific problem. Through attentive experimentation, we achieved Dice scores of 0.81, 0.89, and 0.80 for CSF, GM, and WM respectively using a single probabilistic atlas. Notably, the U-net model yielded even higher scores of 0.89, 0.94, and 0.93 for each tissue type in sequence.

### 1.1 Introduction and problem definition

Medical image segmentation is a key component in computer-assisted diagnosis and intelligent healthcare due to its marked enhancement in diagnostic precision and efficiency. This process aims to highlight structural changes, whether anatomical or pathological, in medical images. Common applications of medical image segmentation include segmenting breast tumors, brain and brain tumors, and the optic disc. Our project specifically addresses the segmentation of brain tissue.

Segmentation of brain tissue represents a highly relevant field within medical image analysis. It offers in-depth, quantitative assessments of the brain, crucial for the accurate diagnosis of diseases, as well as the identification and categorization of irregularities. Magnetic resonance imaging (MRI) is predominantly used for brain tissue segmentation because of its widespread application in both clinical and research settings.

The automated segmentation of brain tissue presents significant challenges due to factors such as uneven intensity distribution, noise, and the intricate structure of the targeted tissues. The objective of our project is to develop an automated method to segment the three principal brain tissues: cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM).

## 1.2 Dataset

Our project employed the IBSR 18 MRI dataset, which includes 18 T1-weighted MRI images from healthy individuals, all of whom have undergone skull-stripping. This dataset was divided as follows for our project purposes:

- Training set: 10 images, all with ground truth labels
- Validation set: 5 images, each with ground truth labels
- Test set: 3 images, without ground truth labels

The dataset is characterized by its variability in spatial resolution and intensity distribution across the images. Each image consists of the same number of slices (256, 128, 256), yet they differ in pixel size across different dimensions. Specifically, the dataset encompasses three distinct pixel size categories. The following table details the pixel size distribution across the training and validation sets:

- **Training Set:**
- ID: 01, 03, 04, 06 - Pixel Size: (0.9375, 1.5, 0.9375) - Type A
- ID: 07, 08, 09 - Pixel Size: (1.0, 1.5, 1.0) - Type B
- ID: 16, 18 - Pixel Size: (0.8371, 1.5, 0.8371) - Type C
- **Validation Set:**
- ID: 11, 12 - Pixel Size: (1.0, 1.5, 1.0) - Type B
- ID: 13, 14 - Pixel Size: (0.9375, 1.5, 0.9375) - Type A
- ID: 17 - Pixel Size: (0.8371, 1.5, 0.8371) - Type C
- **Test Set:**
- ID: 02 - Pixel Size: (0.9375, 1.5, 0.9375) - Type A
- ID: 10 - Pixel Size: (1.0, 1.5, 1.0) - Type B
- ID: 15 - Pixel Size: (0.8371, 1.5, 0.8371) - Type C

Additionally, the grayscale intensities vary among the images. For instance, Figure 1 displays the intensity distribution within the validation set. These two principal characteristics – diverse intensities and varying pixel sizes – must be considered in the preprocessing phase before the design and training of any models, as they could significantly impact the effectiveness of the segmentation algorithms and the final segmentation results.

# 2 Proposal Analysis

We chose to divide the primary workflow into two key sections: preprocessing and segmentation methodologies:

## 2.1 Preprocessing

The main preprocessing pipeline to enhance the quality and consistency of images is as follows:

- Bias-field correction: for rectifying intensity inhomogeneities typical in MRI image.
- Intensity normalization: for stabilizing the range of intensity values, making the image more analytically consistent.
  - Individual normalization methods (each image individually)
- Gaussian smoothing: for reducing noise and detail, thereby enhancing the image's key features and suppressing minor imperfections.

The conditional preprocessing pipeline is as follows:

- Histogram matching: for maintaining visual consistency across medical imaging data.

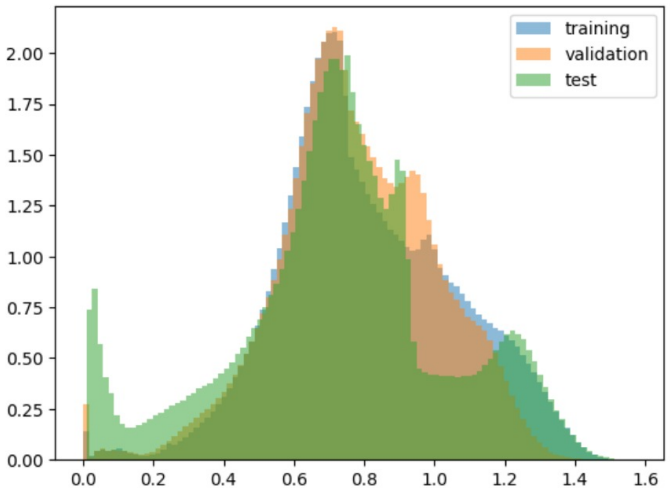


Figure 1: Results of histogram matching to IBSR-09

## 2.2 Segmentation algorithms

In this section, we explore various segmentation algorithms, focusing on traditional methods and their adaptations for effective brain tissue segmentation. These techniques range from intensity-based to atlas-based approaches, customized for optimal performance in our MRI dataset analysis.

### 2.2.1 Traditional approaches

Given our familiarity with traditional brain tissue segmentation techniques from this course, we naturally started with these methods. We decided to revisit and apply what we had learned from previous models used in our lab sessions, using the best parameters and optimization techniques we had picked up. This covered a range of segmentation methods, including those based on intensity and atlas-based approaches. Over time, we've noticed that atlas-based techniques, especially when combined with intensity information, often produce promising results. This realization prompted us to dive deeper into these methods, focusing on multi-atlas algorithms. Our subsequent goals were as follows:

- Multi-atlas
  - Most similar atlas
  - Majority voting atlas
- Probabilistic atlas
- Multi Probabilistic atlas (using probabilistic atlas for each scanner set of images)

**Image registration** A crucial component of the atlas-based methods revolves around image registration. During several classes this semester, we were fortunate to work with Elastix, a robust registration software known for its user-friendly Python integration through the subprocess module. Specifically, we employed Elastix with Parameter 9, first using the Affine transformation and then transitioning to the Bspline transformation, demonstrating its versatility and adaptability for our research needs.

## 2.3 Deep Learning approach

In the final stage of our study, recognizing the lack of exploration in Deep Learning-based segmentation methods, we seized the opportunity to investigate similar projects. Our objective was to adopt their methodologies and evaluate how they perform in comparison to the classic algorithms we are familiar with.

### 2.3.1 Unet architecture

In this context, we decided to focus on implementing and assessing the U-net segmentation model, a prominent approach in the field of Deep Learning for medical image segmentation. This exploration is aimed at broadening our understanding and capabilities in applying advanced machine-learning techniques to our dataset.

# 3 Design and implementation of the proposed solutions

## 3.1 Preprocessing

**Bias field correction** Given that skull stripping was already accomplished in our dataset, our focus shifted exclusively to addressing bias field correction. We successfully implemented this by employing the *SimpleITK N4 Bias Field Correction Filter*. This filter operates directly on the

SimpleITK image format, outputting an image that has been adjusted to rectify the bias field anomalies.

**Normalization** Normalization was achieved using a Python library called *intensity-normalization* specifically designed for brain MRI imaging normalization. This library contains several normalization methods which we used: Individual image-based methods: normalize images independently, rather than in comparison or to a set of images. Utilizing *Fuzzy C-Means (FCM)*-based tissue-based mean normalization focusing on white matter tissues.

The normalization using FCM effectively standardizes the brightness and contrast specifically in white matter tissue regions, which ensures that these areas have consistent visual properties across various scans.

## 3.2 Classic approaches

- **Multi-atlas:** This technique involves leveraging spatial information derived from other similar images and their associated segmentation labels (referred to as individual atlases) and combining these labels through a label fusion process. The effectiveness of this approach hinges on two primary factors: the accuracy of segmentation labels within the individual atlases and the quality of the image registration. To initiate the multi-atlas methodology, a crucial preliminary step was the individual registration of all ten training images to the target (validation) images. This preparatory registration process was essential to ensure the subsequent application of the multi-atlas method.

- **Most similar atlas:**One straightforward way to perform segmentation when multiple atlases are registered to the same target image (validation image) is to choose the atlas with the highest similarity metric. Essentially, we rely on the label information from the atlas that exhibits the closest resemblance to our target image. Various similarity metrics are available, but in our approach, we utilized the Mean Squared (MS) metric. For each validation image, we selected the atlas with the most favorable MS value, and the labels derived from this selected atlas were employed as the segmentation masks. Table 1 presents a case study of the atlas most closely aligned with image 11 from the validation set. Here, image 5 stands out with the lowest mean square (MS) error and the highest similarity, resulting in the most favorable Dice coefficients. However, despite these relative merits, the Dice coefficients achieved were not sufficiently high when benchmarked against other established methods. Consequently, this led to the decision to discontinue further exploration of this particular approach.

- **Majority voting (mean atlas):**In our project, we have also implemented a majority voting strategy, an alternative approach to using the most similar atlas for segmentation. This method involves aggregating multiple atlas labels that have been registered to the same target image. We first load the registered label images using NIfTI files. These labels are then converted to integer types and stacked together. Majority voting is performed by selecting the most frequent label at each voxel across all stacked labels. This process is systematically applied to each case in the validation dataset. The final segmented image is determined by the majority vote at each voxel, providing a consensus-based segmentation result. This approach is advantageous in handling variations and uncertainties in the labels, aiming to produce a more reliable and accurate segmentation by considering the collective information from multiple registered atlases.

- **Mulit-Probabilistic atlas:**The concept behind the Multi-Probabilistic Atlas approach entails generating a probabilistic atlas for each distinct set of scanner images. This process

Table 1: Example of a most similar atlas table, with Dice score per tissue and the MS metric for validation image 11, where image 5 shows the lowest MS and the most similar atlas

| Image ID | Dice Class 1 | Dice Class 2 | Dice Class 3 | Similarity Score |
|---|---|---|---|---|
| 01 | 0.7424 | 0.8034 | 0.8054 | 0.0610 |
| 03 | 0.7038 | 0.7988 | 0.7817 | 0.0576 |
| 04 | 0.7251 | 0.7823 | 0.7753 | 0.0642 |
| 05 | 0.7635 | 0.8073 | 0.8128 | 0.0535 |
| 06 | 0.7454 | 0.7815 | 0.7938 | 0.0617 |
| 07 | 0.7272 | 0.7897 | 0.8083 | 0.0556 |
| 08 | 0.6818 | 0.7759 | 0.7954 | 0.0660 |
| 09 | 0.7376 | 0.7798 | 0.7993 | 0.0579 |
| 16 | 0.6847 | 0.7779 | 0.7567 | 0.0630 |
| 18 | 0.7329 | 0.7738 | 0.7654 | 0.0630 |

involves creating a separate probabilistic atlas for each scanner category using images specific to that category. Consequently, three distinct probabilistic atlases are produced for each scanner type, with each atlas tailored for the corresponding validation image sourced from the same scanner group. It's worth noting that, for the creation of a reliable probabilistic atlas, a minimum of three images is typically required. However, in the case of scanner type 3, there were only two images available in the training dataset, which led to concerns about the reliability of the atlas. Despite this limitation, an attempt was made to assess the outcomes, and label propagation was employed to derive the segmentation results.

- **Single Probabilistic atlas:**Following the execution of the primary preprocessing phase, the next step involved the selection of a reference image for histogram matching. This reference image was chosen based on two primary criteria: it exhibited the highest average correlation with all other images, signifying similarity, and it possessed high contrast characteristics. Subsequently, Image 09 was identified as the optimal reference image due to its pronounced resemblance to the histograms of the remaining images in the dataset. Histogram matching was then applied to align all images with the intensity profile of Image 09. Next, a registration process was carried out to align the final preprocessed images with Image 09, facilitating the creation of a probabilistic atlas. This atlas, in turn, served as the foundation for segmentation, accomplished through label propagation techniques, yielding the final segmented results for the validation images.

## 3.3 Deep Learning approach

In our project, since the dataset contains labels for both the training and validation sets and the primary objective was to assign categorical labels to individual pixels, it naturally falls under the category of image semantic segmentation with a supervised learning approach. In this context, we specifically focused on implementing and evaluating the U-Net segmentation model. This approach was chosen based on its prominence in the field of Deep Learning for medical image segmentation. It's important to note that we adopted Jose Bernal's notebook, which employed a 2D patch-wise approach for implementing the U-Net model. This endeavor was undertaken to enhance our understanding and proficiency in applying state-of-the-art machine-learning techniques to our dataset.

### 3.3.1 Data preparation

In the data preparation phase for our deep learning model, our input data undergoes initial processing through the primary preprocessing pipeline. Subsequently, we perform histogram matching on all processed images, aligning them with Image 09 as the reference. This choice is driven by the reference image's exceptional characteristics, including a notably high average correlation and pronounced contrast, as previously elucidated.

The utilization of data sampling via patches proves to be an efficient strategy for model training. This approach involves extracting a subset of smaller patches from the data, enabling more streamlined and effective model training while still encapsulating crucial data features and variations. In our dataset preparation process, we have specifically extracted patches with dimensions of 64x64 pixels. While we did explore different patch sizes in our experiments, the 64x64 size emerged as the optimal choice for our specific application. The patches are filtered out based on the proportion of foreground (non-zero) pixels in the segmentation labels.

### 3.3.2 Model parameters

Within this section, our focus will be directed toward an in-depth exploration of both the parameters utilized and the hyperparameter tuning for our U-Net models.

For our optimizer selection, we opted for ADAM, a widely acclaimed choice in deep learning. ADAM offers several advantages, including efficient convergence, adaptive learning rates for individual parameters, and the ability to handle noisy gradients. This adaptive optimization method enhances the training process, making it more stable and effective.

Our choice for the loss function was Categorical Cross-Entropy, a popular and effective choice for multi-class classification tasks. Categorical Cross-Entropy provides advantages such as sensitivity to prediction errors and a clear measure of model performance. It encourages the model to make confident and accurate predictions, which is crucial for accurate segmentation in our task.

We employed the Hausdorff Distance (HD) and Dice Similarity Coefficient (DSC) as our evaluation metrics. HD provides a measure of the maximum discrepancy between predicted and ground truth boundaries, offering insights into boundary alignment. DSC, on the other hand, quantifies the overlap between segmented and ground truth regions. Together, these metrics offer a comprehensive assessment, enabling us to evaluate both boundary accuracy and region overlap, which are critical for effective segmentation model evaluation.

Utilizing a patch size of (64, 64) and a patch stride of (32, 32), we applied a *CONTENT-THRESHOLD* of 0.1 for patch filtering. Our model underwent training for 50 epochs, with early stopping patience set at 20, and a batch size of 64. Additionally, we implemented a model checkpoint mechanism based on validation loss to ensure the preservation of the best-performing model during training.

# 4 Experimental section and results analysis

## 4.1 Traditional approaches results

**majority voting results:** The table provides a detailed quantitative evaluation of the majority voting atlas method applied to various cases, specifically IBSR-11 through IBSR-17, except for IBSR-16. The evaluation metrics used are the Dice scores and Hausdorff distance for the cerebrospinal fluid (CSF), white matter (WM), and gray matter (GM).

The Dice scores, which measure the overlap between the segmented tissue and the ground truth, are fairly consistent across the cases, with the total average scores being 0.84 for CSF, 0.87 for WM, and 0.84 for GM. These scores indicate a high level of accuracy in the segmentation process, with white matter segmentation being slightly more accurate on average than the other tissues.

On the other hand, the Hausdorff distance (HD) provides insight into the maximum distance between the surface of the ground truth and the segmented results. For CSF, the distances vary, with a total average of 24.42, suggesting some variability in the segmentation boundary accuracy for this tissue type. The WM and GM have lower total average distances, 9.87 and 9.33 respectively, indicating tighter conformity to the true tissue boundaries and, hence, more precise segmentations.

Overall, the majority voting approach yielded robust segmentations with good agreement to the ground truth, as evidenced by the high Dice scores and relatively low Hausdorff distances, especially in the white and gray matter segmentations.

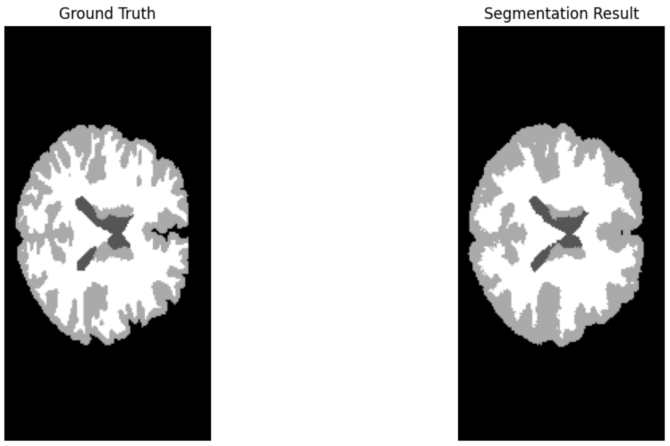


Figure 2: Result of IBSR-11 for majority voting

Table 2: Dice scores and Hausdorff distance for majority voting atlas

| Cases | CSF | WM | GM | CSF HD | WM HD | GM HD |
|---|---|---|---|---|---|---|
| IBSR-11 | 0.84 | 0.85 | 0.85 | 31.71 | 10.55 | 10.55 |
| IBSR-12 | 0.82 | 0.85 | 0.85 | 24.88 | 9.00 | 7.76 |
| IBSR-13 | 0.80 | 0.88 | 0.82 | 11.25 | 11.78 | 10.25 |
| IBSR-14 | 0.85 | 0.89 | 0.85 | 26.07 | 9.74 | 6.83 |
| IBSR-17 | 0.89 | 0.89 | 0.83 | 28.18 | 8.28 | 11.27 |
| Total | 0.84 | 0.87 | 0.84 | 24.42 | 9.87 | 9.33 |

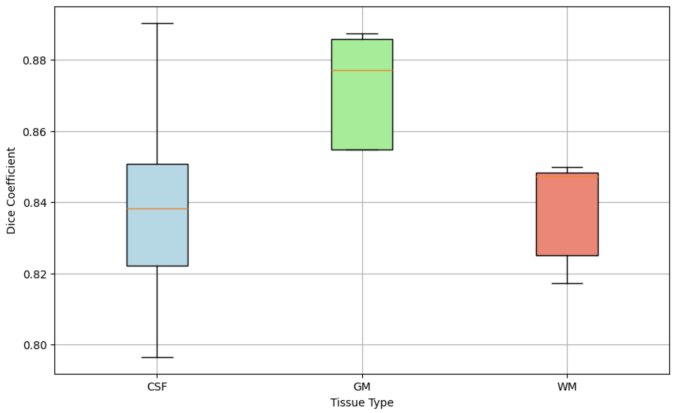


Figure 3: Result of Dice Coefficient Distribution by Tissue Type for majority voting

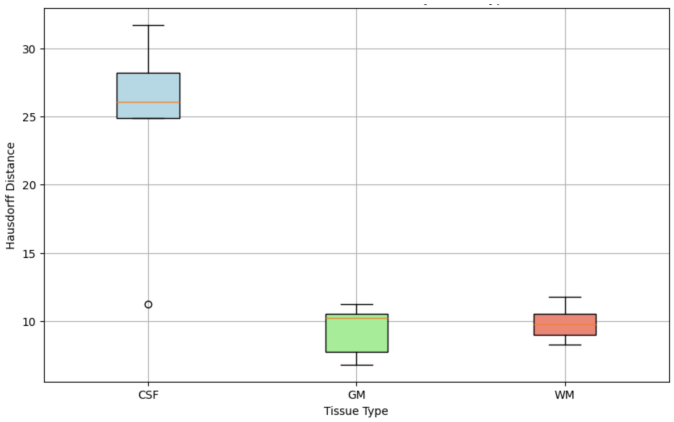


Figure 4: Result of Hausdorff Distance Distribution by Tissue Type for majority voting

**Single probabilistic atlas:** The results shown in the table and figures indicate the performance of a single probabilistic atlas segmentation method applied to case IBSR-09. The Dice scores for this particular case demonstrate proficient segmentation across the cerebrospinal fluid (CSF), white matter (WM), and gray matter (GM), with gray matter achieving the highest level of accuracy as indicated by the highest Dice score. The Hausdorff distance for each tissue type in case IBSR-9 reflects a congruent trend with the Dice scores, where the GM shows the smallest distance, suggesting a precise match to the ground truth boundary compared to CSF and WM. This is particularly important for clinical and research applications that require exact delineation of gray matter. In the graphical representations, the boxplots for Dice scores and Hausdorff distances provide a visual summary of the segmentation performance. The gray matter stands out with a tight distribution of high Dice scores and low Hausdorff distances, emphasizing the effectiveness of the single probabilistic atlas method for this specific case. The results for CSF and WM also show a good level of accuracy, though with a slightly broader range in the Hausdorff distances, indicating some variations in the boundary precision for these tissue types. Overall, the single probabilistic atlas has demonstrated a commendable segmentation capability for case IBSR-9, with the quantitative metrics suggesting that it could be particularly reliable for tasks requiring meticulous gray matter segmentation.

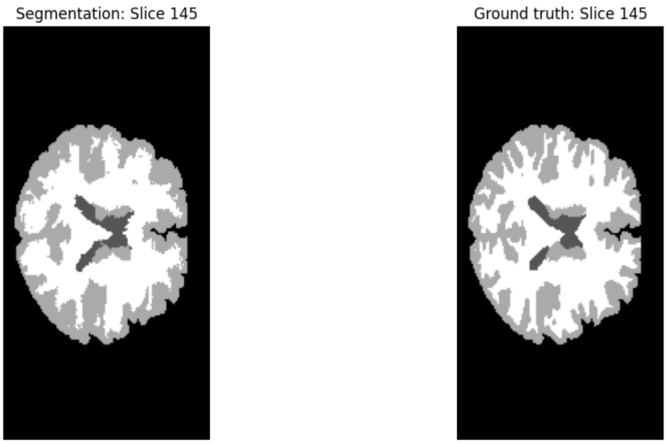


Figure 5: Result of IBSR-11 for single probabilistic atlas of Case IBSR-09

Table 3: Dice scores and Hausdorff distance for Single probabilistic atlas

| Cases | CSF | WM | GM | CSF HD | WM HD | GM HD |
|---|---|---|---|---|---|---|
| IBSR-11 | 0.79 | 0.82 | 0.89 | 30.99 | 10.26 | 9.70 |
| IBSR-12 | 0.82 | 0.82 | 0.91 | 26.52 | 12.41 | 6.02 |
| IBSR-13 | 0.77 | 0.79 | 0.90 | 10.71 | 12.47 | 7.30 |
| IBSR-14 | 0.83 | 0.82 | 0.90 | 26.55 | 9.34 | 6.73 |
| IBSR-17 | 0.87 | 0.79 | 0.90 | 26.67 | 9.05 | 9.50 |
| Total | 0.82 | 0.81 | 0.90 | 24.29 | 10.71 | 7.85 |

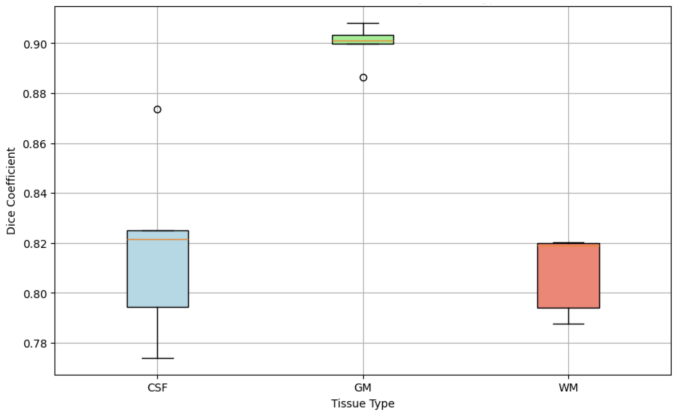


Figure 6: Dice score Distribution by Tissue Type for Single probabilistic atlas

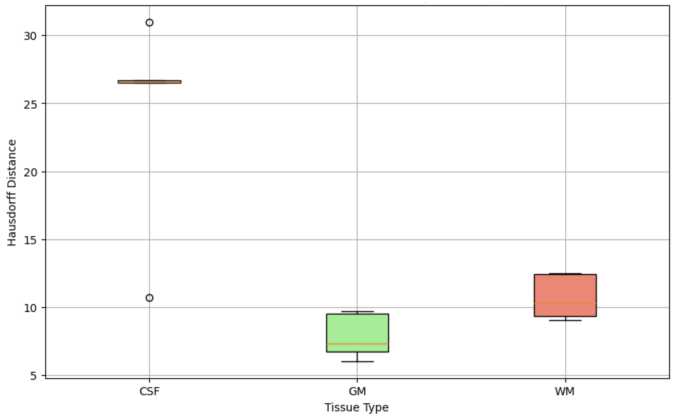


Figure 7: Hausdorff Distance Distribution by Tissue Type for Single Probabilistic Atlas

## 4.2 Deep learning approaches results

The outcomes from the application of a 2D patch-wise U-net architecture demonstrate exceptional segmentation performance, as evidenced by robust Dice coefficients and Hausdorff Distance metrics. Table 4 details the Dice coefficients and Hausdorff distances for the segmentation of cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM) across multiple IBSR cases. Notably, the Dice scores for GM and WM are consistently above 0.9, indicating a superior level of segmentation accuracy which is indicative of the U-net's proficiency in delineating these particular tissues with high fidelity to the ground truth.

Correspondingly, the Hausdorff distances align with the high Dice scores, particularly for GM and WM, where the measurements are markedly low. This suggests that the segmented boundaries produced by the U-net closely mirror the actual boundaries of the tissues, underscoring the precision of the model. Although the Hausdorff distances for CSF are higher, they are still within a reasonable range, denoting an acceptable accuracy in segmentation.

The distribution of these metrics is visually encapsulated in Figures 8 and 9. The Dice coefficient distribution by tissue type (Figure 8) exhibits a narrow interquartile range for GM and WM, reflecting the model's consistent segmentation accuracy across these tissues. Similarly, the Hausdorff distance distribution (Figure 9) corroborates the model's precision, with lower ranges for GM and WM, suggesting a compact boundary definition.

Table 4: Dice scores and Hausdorff distance for U-net

| Image Name | Dice_class_1 | HD_class_1 | Dice_class_2 | HD_class_2 | Dice_class_3 | HD_class_3 |
|---|---|---|---|---|---|---|
| IBSR-11 | 0.8816 | 22.3830 | 0.9336 | 6.8023 | 0.9488 | 0.9488 |
| IBSR-12 | 0.6780 | 29.0867 | 0.9227 | 7.8102 | 0.9265 | 0.9265 |
| IBSR-13 | 0.8713 | 16.9153 | 0.9379 | 12.5698 | 0.9163 | 0.9163 |
| IBSR-14 | 0.9138 | 22.9128 | 0.9542 | 13.3840 | 0.9499 | 0.9499 |
| IBSR-17 | 0.9169 | 32.6497 | 0.9477 | 13.6747 | 0.9260 | 0.9259 |

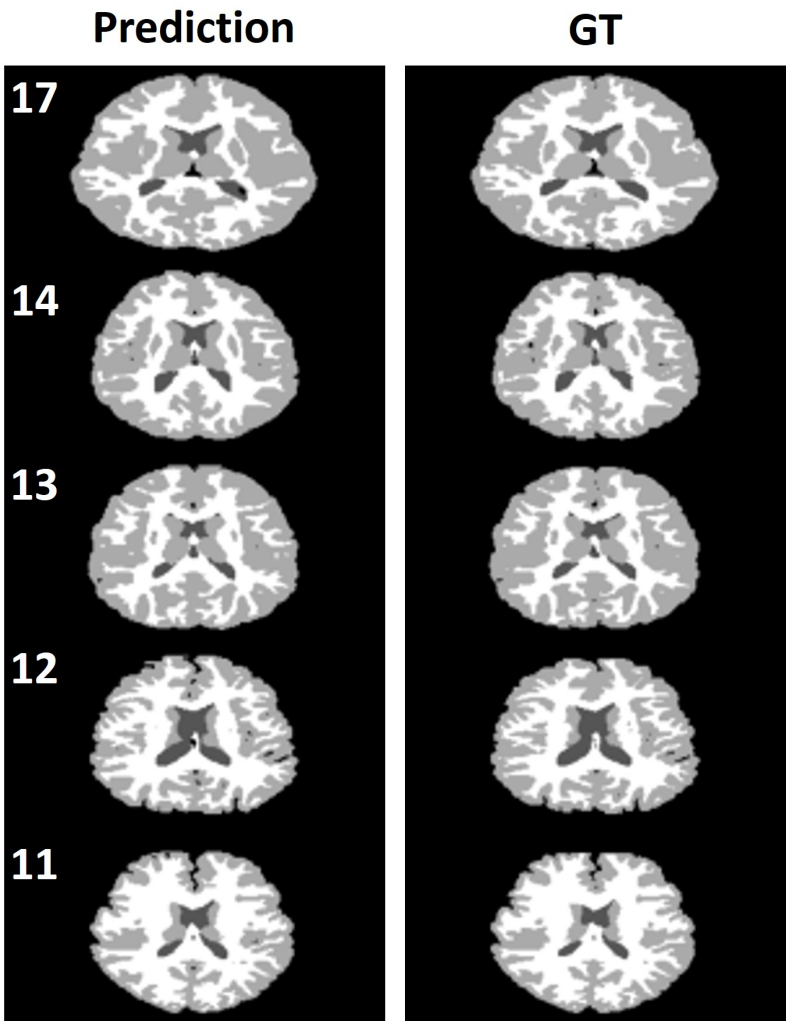


Figure 8: The ground truth and the predictions of the Validation set using U-net

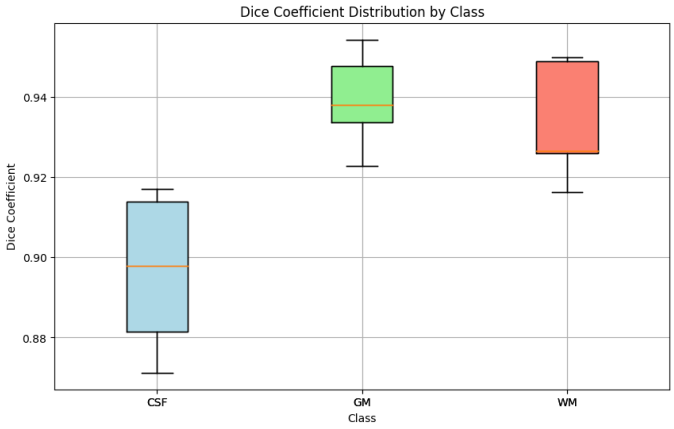


Figure 9: Result of Dice Coefficient Distribution by Tissue Type for U-net

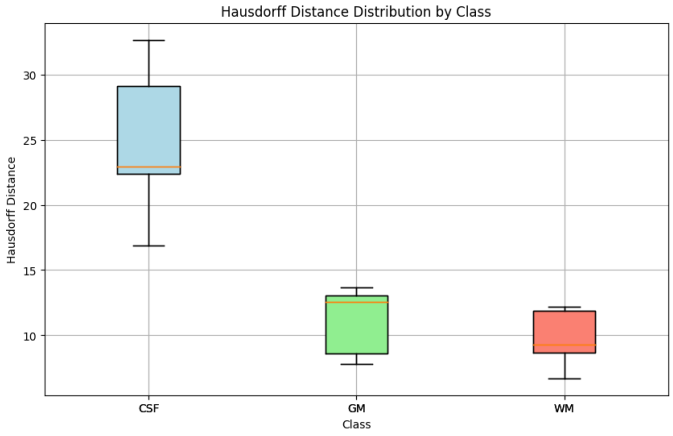


Figure 10: Result of Hausdorff Distance Distribution by Tissue Type for U-net

## 4.3 Comparing the best classic and DL approaches

In assessing the comparative effectiveness of traditional and deep learning-based segmentation methods, we find that each possesses unique advantages. The traditional techniques, including multi-atlas, single probabilistic atlas, most similar atlas, and majority voting, rely on the integration of pre-existing anatomical knowledge, which can sometimes limit their adaptability but also provide a degree of interpretability and transparency that some clinical applications may favor.

Among these traditional methods, majority voting distinguished itself as the most effective, achieving reliable segmentation with average Dice scores of around 0.84 for CSF, 0.87 for WM, and 0.84 for GM. The corresponding Hausdorff distances suggested a reasonable match to the ground truth, indicating that majority voting was particularly adept at handling gray matter segmentation, likely due to its consensus-based decision-making process.

The single probabilistic atlas, while it performed slightly less effectively for CSF and WM compared to majority voting, demonstrated a high Dice score average for GM, comparable to majority voting. Despite its simplicity, the single probabilistic atlas showed promise in specific applications.

However, when these traditional methods are juxtaposed with the U-net, a deep learning model, there is a notable shift in performance. The U-net model delivered higher Dice scores, especially for GM and WM, with values consistently above 0.9, denoting highly precise segmentation. The Hausdorff distances for the U-net model were also notably lower for GM and WM, indicating a segmentation that more closely aligns with the actual tissue boundaries.

While traditional methods have their place, particularly for their simplicity and clear interpretability, the U-net model outstripped them in terms of accuracy and precision, essential for clinical diagnosis and research applications. The U-net's robustness to variations in data and its ability to learn complex patterns from the data itself makes it a superior tool for medical image segmentation tasks. It benefits from automation, minimizing the need for manual intervention, which is a significant advantage over traditional methods that often require more hands-on tuning.

In conclusion, the U-net model emerges as the leading approach for segmentation in our analysis, offering superior accuracy and precision, critical for the exact delineation of brain tissues. Its performance in the segmentation of gray and white matter demonstrates the potential of deep learning to not only complement but also enhance and perhaps eventually surpass the capabilities of traditional image segmentation methods in medical imaging.

# 5 Organization and development of the coursework

The project's organizational framework was meticulously devised during the final lab sessions, where group members engaged in collaborative brainstorming sessions. The insights and ideas generated from this collective brainstorming were methodically documented in a shared file. Subsequently, in the ensuing weeks, extensive literature review and research were diligently undertaken to lay the foundation for the project's direction.

As we transitioned into the early weeks of December, a prudent reassessment led to the elimination of certain ideas due to resource constraints and project time limitations. This pivotal decision paved the way for the adoption of a dual-pronged approach encompassing both classical and contemporary (deep learning) methodologies. This strategic choice not only allowed us to apply the knowledge imparted during lectures but also facilitated exploration into modern techniques.

The project's experimental phase unfolded in December, where we extremely conducted various experiments. To ensure the integrity and completeness of our work, we carried out the most recent revisions and accurately documented all aspects of the project, culminating in the finalization of documentation in early January.

# 6 Conclusions

In conclusion, medical image segmentation presents a formidable challenge, given the inherent complexities of medical images replete with artifacts and noise. In this project, our primary objective was to perform segmentation on the three key brain tissues within MRI images. Despite the formidable obstacles encountered, such as varying voxel sizes and intensity irregularities, we attained commendable results through a combination of classical and deep learning approaches.

Within the realm of classical approaches, we conducted a series of experiments encompassing multi-atlas methodologies, including the utilization of the most similar atlas, majority voting (mean atlas), and single and multi-probability approaches. Notably, the single probability atlas model yielded noteworthy mean dice scores of 0.81, 0.89, and 0.80 in the validation test for CSF, GM, and WM, respectively.

In the domain of deep learning, we explored state-of-the-art U-Net-based algorithms, recognizing the need for a patch-based approach due to dataset size constraints. To address the nuances of semantic segmentation, we experimented with various loss functions, including focal loss, dice loss, and categorical cross-entropy loss.

Ultimately, our most promising results were achieved using the U-net model with a categorical cross-entropy loss function, yielding remarkable mean dice scores of 0.89, 0.94, and 0.93 for CSF, GM, and WM, respectively, in the validation set. The outstanding performance of this model led to its selection for deployment in our test dataset for the challenge, showcasing its potential for real-world applications.

# References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. cite arxiv:1505.04597Comment: conditionally accepted at MICCAI 2015

[2] Dora, L., Agrawal, S., Panda, R., Abraham, A. (2017). State-of-the-art methods for brain tissue segmentation: A review. IEEE reviews in biomedical engineering, 10, 235-249.