Hybrid Atlas and Cascade U-NET Brain Tissue Segmentation

Agustin Cartaya and Micaela Rivas

Introduction and problem definition

Image segmentation plays a fundamental role in medical image processing, particularly in the field of brain studies, where it becomes essential for the delimitation of different regions. This project focuses on the development of a segmentation method for the three main tissues of the brain: white matter (WM), grey matter (GM), and cerebrospinal fluid (CSF), from magnetic resonance imaging (MRI).

The dataset used is IBSR18 composed of 18 MRI volumes distributed in train (10), validation (5), and test (3). These images come from three different scanners, presenting variations in resolutions (pixel spacing) and intensities, which adds complexity to the segmentation task.

The evaluation of segmentation is based on the use of standard metrics commonly used in similar challenges, such as the Dice Score Coefficient (DSC), the Hausdorff Distance (HD), and the Average Volumetric Difference (AVD). For the calculation of these metrics, the validation set was used.

To achieve the objective of this project, an algorithm based on probabilistic atlas passed as prior spatial information to a Deep learning model was implemented.

Proposal analysis / Design and implementation of the proposed solution

The proposed algorithm consists of three key phases: preprocessing for intensity homogenization, obtaining spatial information from a probabilistic atlas, and computing the final masks with Deep Learning using intensity information and the previous spatial information. Fig. 1 shows the main pipeline of the proposed algorithm.

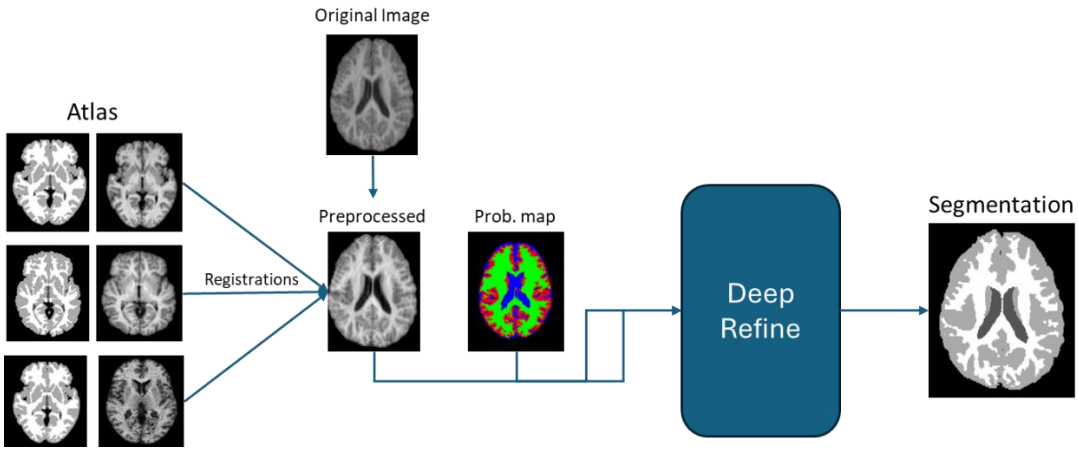

Fig. 1: Main pipeline of the algorithm

Preprocessing

Given the diversity of scanners and variations in image intensities, preprocessing becomes essential for homogenizing the data. The similarity between resulting images is critical for effective generalization and achieving precise and consistent segmentation. The preprocessing comprises two main stages: first, normalizing images considering only the brain region, and second, applying histogram matching with respect to an average histogram.

Normalization

The normalization step was implemented with the purpose of all images being within the range of 0 to 1. Considering that the minimum value of each voxel is 0, normalization was carried out by dividing by the maximum value of each image. This maximum value was searched only in the brain region with the help of a mask calculated as follows:

- For train images: The mask provided in the training set (groundtruth) was used directly.
- For the test images: An Otsu threshold was applied to obtain a brain mask with the help of the SimpleITK library.

Histogram matching

This step was carried out to improve the similarity between the images, contributing to a more uniform base. The process is based on matching or fitting the histogram of an image to another reference histogram. In this case, an average histogram was created using all the images in the training set, and then all the images were adjusted with respect to this histogram.

To reduce noise and artifacts that could affect the average histogram, a Gaussian filter with sigma=20 was applied. Subsequently, the histogram was normalized using the L1 standard.

Probabilistic atlas segmentation

A probabilistic atlas was generated for each image by registering the preprocessed training images to the preprocessed target image, using an affine registration followed by a non-rigid one. The same transformation was then applied to the training labels. By combining all these masks in the patient space, three images were created with probabilities of belonging to each tissue.

Computing atlas with deep learning using prior spatial information

After having obtained the spatial information from the probabilistic masks of each tissue. A deep learning model was trained with this information and the intensity image. This approach helped to combine special information and intensity information to obtain final masks.

Model structure

The model was based on a cascaded multipath U-Net optimized with a multi-objective loss similar to the one proposed in [1]. First, the T1w image and its probabilistic atlas are passed as input into three networks arranged in parallel. Each one of these networks accounts separately for changes in CSF, WM, and GM. Second, the resulting individual

latent space of these networks is appended and passed to another U-shaped network which merges them effectively to produce the final output. The architecture of each U-Net module is similar to the one presented in [1] with the difference that the activation function on the output of each model was changed to sigmoid and that 4 output channels are obtained instead of 1 in the latest U-Net module. These changes made it possible to obtain disjoint probabilities for each tissue and thus to be able to calculate an individual and a set for the 4 possible classes, 3 corresponding to the main tissues (CSF, WM, GM) and an extra one for the background (BG). Fig. 2 shows the model structure.

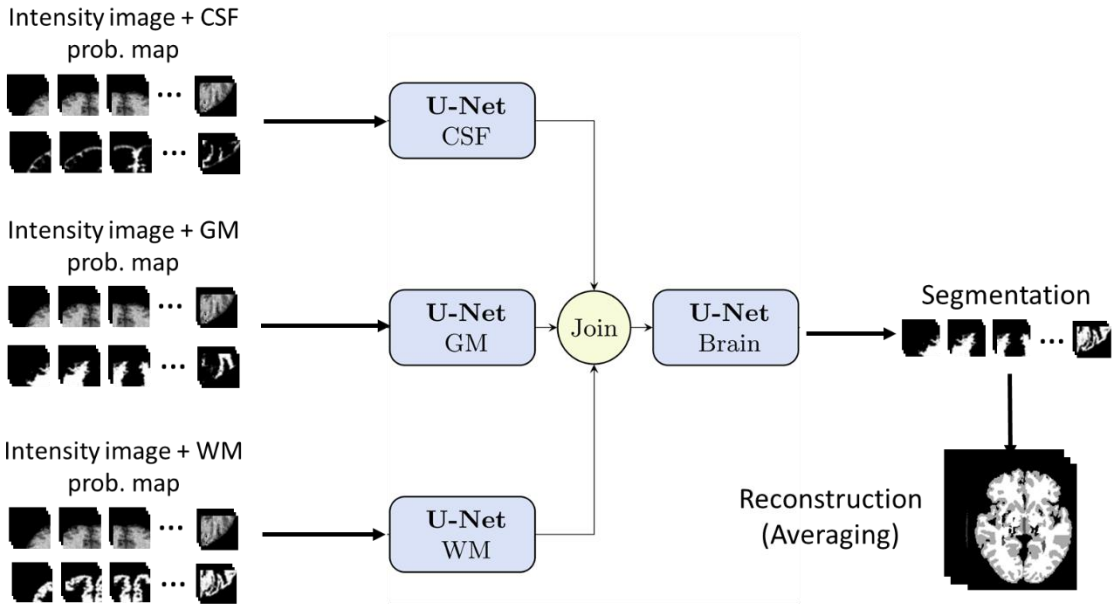

Fig. 2: Model structure

Input

The model takes as input 4 patches of 32x32x32 extracted from the T1w image and the probabilistic atlas in the same way as [1]. For both training and testing, overlapping blocks were extracted to gather more samples, reduce block boundary artifacts, and enforce spatial consistency (Bernal et al. 2019b). Additionally, empty, or partially empty training patches were discarded to prevent background biased predictions. the minimum content rate and overlap extent were set to 20% and 50%, respectively. Both values were favorable experimentally.

Data Augmentation

During the training phase data augmentation was used to avoid overfitting. The implemented augmentation was applied to the complete image and not to the individual patches to maintain spatial consistence. The applied augmentation was based on random 3D elastic deformation applied to the intensity image and to the atlas, and gaussian noise only applied to the intensity image. This was performed using MONAI library.

Loss function

The loss function of the model is based on a composite loss function with penalties to the most difficult tissues to classify. Given a real segmentation of the image, y, the corresponding tissue probability map for each tissue, sCSF, sGM, and sWM, the approximation obtained with the model, $\tilde{y}$, and the early approximation obtained by

parallel U-Net modules $\widetilde{ye}_{csf}$, $\widetilde{ye}_{wm}$, $\widetilde{ye}_{gm}$, the compose loss function is defined as follows:

$$L(y,\tilde{y}) = pred_{loss}(y,\tilde{y}) + w_{csf} * bce\left(y_{csf},\widetilde{ye}_{csf}\right) + w_{wm} * bce(y_{wm},\widetilde{ye}_{wm}) + w_{gm} * bce\left(y_{gm},\widetilde{ye}_{gm}\right)$$

$$pred_{loss}(y,\tilde{y}) = w_{csf} * bce\left(y_{csf},\tilde{y}_{csf}\right) + w_{wm} * bce(y_{wm},\tilde{y}_{wm}) + w_{gm} * bce(y_{gm},\tilde{y}_{gm})$$

where bce is the binary cross entropy loss function.

Segmentation reconstruction for overlapping patches

Once the segmentation by patches was obtained, the reconstruction of the complete image was obtained by reorganizing the patches in the same way in which they were extracted and calculating the average for the overlays, this is done for the 4 channels obtained from the network. Then, to obtain the discrete segmentation, the channel with the highest probability was obtained for each voxel: BG=0, CSF=1, WM=2, GM=3.

Experimental section and results analysis

Preprocessing

All experiments were performed using the preprocessed images to ensure a uniform database. Preprocessing included normalization and histogram matching.

An attempt was made to implement a bias field correction; however, this step was discarded because it generated very similar results and significantly increased computation time.

Atlas-based segmentation

1. Multi-atlas without registration to MNI

Initially, a multi-atlas segmentation without registration to the MNI space was done. This involved registering preprocessed training images to a target image to create a multi-atlas. The registration process consisted of an initial affine registration followed by a non-rigid one. The same parameters were applied to transform labels, which were subsequently used to generate the multi-atlas.

1.1. Majority voting

The first attempt involved combining the atlases using majority voting. Tables 1 and 2 present the resulting metrics and overall means, respectively.

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.7756	0.8249	0.7874	21.3776	7.6811	9.4340	18956	936920	431835
12	0.8122	0.8285	0.8004	19.6469	6.7082	8.6603	20091	955968	410093
13	0.7639	0.8688	0.7855	12.0830	10.0499	10.2956	17067	921148	402204
14	0.8103	0.8659	0.8045	20.2731	6.3246	9.8995	21771	961632	443693
17	0.8729	0.8683	0.7814	22.4499	11.5758	10.9545	25593	1055267	477191
Means	**0.8070**	**0.8513**	**0.7919**	**19.1661**	**8.4679**	**9.8488**	**20696**	**966187**	**433003**

Table 1: Resulting metrics of the multi-atlas segmentation with majority voting combination

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.7960	12.8309	462570
12	0.8137	11.6718	462051
13	0.8061	10.8095	446806
14	0.8269	12.1657	475699
17	0.8409	14.9934	519350
Means	**0.8167**	**12.4943**	**473295**

Table 2: Overall means of the multi-atlas segmentation with majority voting combination

This approach yielded Dice scores within an acceptable range but showed variability across patients. In pursuit of optimization, a weighted voting strategy was employed to enhance the segmentation accuracy.

1.2. Weighted voting

To improve upon the initial results, atlas combination was performed using a weighted voting approach, assigning greater weight to images more similar to the target image. The weighting of votes for each image involved the calculation of similarity metrics, which were subsequently treated to be the weights for combining the atlases, influencing the final segmentation outcome.

1.2.1. Metric Squared differences (SD)

This metric was chosen as the initial similarity measure due to its simplicity and ease of implementation. This metric assesses the dissimilarity between an atlas image and the target image by computing the squared sum of pixel-wise differences. The formula is defined as follows:

$$SD = \frac{\sum_{i=1}^{n}(I_i - T_i)^2}{n}$$

where I_i is the pixel intensity in the atlas image, T_i refers to the corresponding pixel intensity in the target image, and n is the total number of pixels.

The weight is then calculated as the reciprocal of the SD, giving more importance to the images with lower SD.

Tables 3 and 4 show the resulting metrics and overall means, respectively.

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.7780	0.8235	0.8000	21.3776	8.4853	8.1240	18378	911120	467820
12	0.8092	0.8214	0.8018	19.6469	6.5574	7.8740	19614	929910	444924
13	0.7669	0.8642	0.7821	12.7279	9.4868	10.2956	15801	913139	427892
14	0.8127	0.8638	0.8079	20.2731	6.3246	8.6023	20581	947771	473440
17	0.8645	0.8682	0.7832	22.4499	11.7898	10.9545	24102	1047705	505028
Means	**0.8063**	**0.8482**	**0.7950**	**19.2951**	**8.5288**	**9.1701**	**19695**	**949929**	**463821**

Table 3: Resulting metrics of the multi-atlas segmentation with weighted voting combination using metric SD

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.8005	12.6623	465773
12	0.8108	11.3594	464816
13	0.8044	10.8368	452277
14	0.8281	11.7333	480597
17	0.8387	15.0647	525612
Means	**0.8165**	**12.3313**	**477815**

Table 4: Overall means of the multi-atlas segmentation with weighted voting combination using metric SD

The results of this combination show less variability than the previous one; nevertheless, the mean values of the metrics generally worsen.

1.2.2. Metric Normalized Cross correlation (NCC)

The metric NCC was used trying to improve the segmentation results. This decision was prompted by the aim to assess whether employing a different similarity measure could contribute to improved mean metric values. The NCC metric evaluates the similarity between an atlas image and the target image by measuring the normalized cross-correlation of their pixel intensities. The formula of the NCC is:

$$NCC = \frac{\sum_{i=1}^{n}(I_i - \bar{I}) * (T_i - \bar{T})}{\sqrt{\sum_{i=1}^{n}(I_i - \bar{I})^2 * \sum_{i=1}^{n}(T_i - \bar{T})^2}}$$

where I_i and T_i are the pixel intensities of the atlas and target images, respectively. $\bar{I}$ and $\bar{T}$ are the mean intensities of the atlas and target images, and n is the total number of pixels.

The resulting similarity score, calculated using the normalized cross-correlation, aims to provide insights into the spatial correspondence of pixel intensities, potentially leading to refined segmentation outcomes.

Tables 5 and 6 show the resulting metrics and overall means, respectively.

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.7780	0.8234	0.8000	21.3776	8.4853	8.1240	18338	912419	467552
12	0.8073	0.8213	0.8018	19.6469	6.5574	7.8740	19472	932309	444456
13	0.7666	0.8641	0.7820	12.7279	9.4868	10.2956	15785	913156	427906
14	0.8135	0.8639	0.8077	20.2731	6.3246	9.0000	20540	949892	472195
17	0.8645	0.8682	0.7833	22.4499	11.7898	10.9545	24097	1047703	505032
Means	**0.8060**	**0.8482**	**0.7950**	**19.2951**	**8.5288**	**9.2496**	**19646**	**951096**	**463428**

Table 5: Resulting metrics of the multi-atlas segmentation with weighted voting combination using metric NCC

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.8005	12.6623	466103
12	0.8101	11.3594	465412
13	0.8042	10.8368	452283
14	0.8284	11.8659	480876
17	0.8387	15.0647	525611
Means	**0.8164**	**12.3578**	**478057**

Table 6: Overall means of the multi-atlas segmentation with weighted voting combination using metric NCC

The results of this combination still show less variability than the majority voting approach, but the mean values of the metrics worsen with respect to the previous one.

1.2.3. Metric Mutual information (MI)

The last metric tried to improve the combination method was the MI. This choice was motivated by the exploration of diverse metrics to assess whether they could contribute to improved mean metric values. The MI metric quantifies the amount of information shared between the pixel intensities of an atlas image and a target image. It can be calculated as:

$$MI = H(I) + H(T) - H(I,T)$$

where $H(I)$ represents the histogram of the atlas image, $H(T)$, the histogram of the target image and $H(I,T)$ the joint histogram.

The resulting similarity score, were normalized and aimed to capture the information content shared between the images, providing valuable insights for refining segmentation outcomes.

Tables 7 and 8 show the resulting metrics and overall means, respectively

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.7785	0.8235	0.8000	21.3776	8.4853	8.1240	18279	913367	467367
12	0.8018	0.8209	0.8015	19.6469	6.5574	7.8740	19188	934797	444264
13	0.7665	0.8642	0.7823	12.7279	9.4868	10.2956	15772	913212	428058
14	0.8130	0.8638	0.8073	20.2731	6.3246	9.0000	20475	950915	471965
17	0.8640	0.8680	0.7828	22.4499	11.7898	10.9545	24084	1047732	505279
Means	**0.8048**	**0.8481**	**0.7948**	**19.2951**	**8.5288**	**9.2496**	**19560**	**952005**	**463387**

Table 7: Resulting metrics of the multi-atlas segmentation with weighted voting combination using metric MI

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.8007	12.6623	466338
12	0.8081	11.3594	466083
13	0.8043	10.8368	452347
14	0.8280	11.8659	481119
17	0.8383	15.0647	525699
Means	**0.8159**	**12.3578**	**478317**

Table 8: Overall means of the multi-atlas segmentation with weighted voting combination using metric MI

The results of this combination gave the worst results among all the metrics, but the variability was decreased with respect to the others.

The results obtained with multi-atlas segmentation without registration to MNI are shown in Table 9.

Metric	Majority Voting	Weighted SD	Weighted NCC	Weighted MI
DSC	0.8167	0.8165	0.8164	0.8159
HD	12.4943	12.3313	12.3578	12.3578
AVD	473295	477815	478057	478317

Table 9: Overall means of the multi-atlas segmentation without registration to MNI

These results show that the best combination approach for multi-atlas segmentation is the one using the majority voting combination, although it gave the most variability among all the approaches. Fig. 3 shows the results obtained with multi-atlas segmentation without registration to MNI.

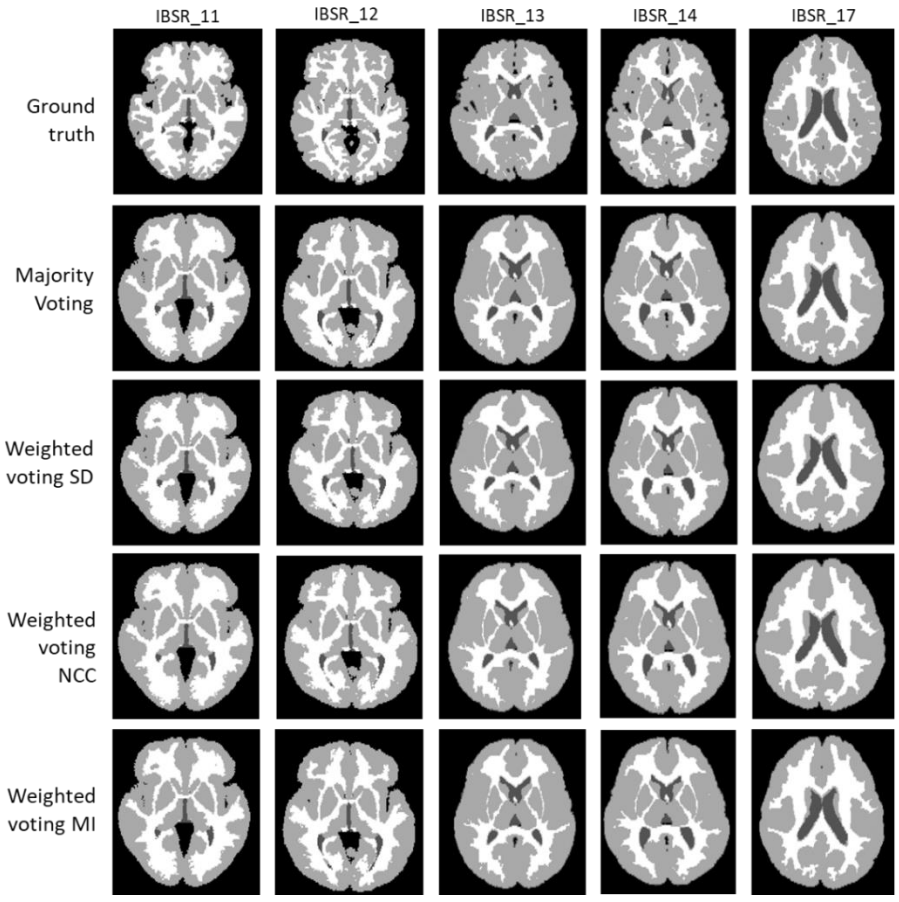

Fig. 3: results for all the patients of the validation set obtained with multi-atlas segmentation without registration to MNI.

2. Multi-atlas with previous registration to MNI

The initial attempt at multi-atlas segmentation without registration to the MNI space exhibited challenges in achieving accurate segmentations due to the inherent diversity

among the images. In response to this, an exploration was conducted to evaluate the impact of a preliminary registration to the MNI atlas on the preprocessed images. The rationale behind this adjustment was to assess whether aligning the images to the MNI space beforehand could enhance segmentation results.

Following the same procedural steps as in the previous multi-atlas approach, the difference lay in the choice of the base image. Instead of utilizing the preprocessed image as the foundation for atlas creation, we employed the image registered to the MNI atlas. This adjustment aimed to account for the variability among the images and assess whether aligning them to a common reference space would lead to improved segmentation outcomes.

The subsequent sections detail the outcomes of the multi-atlas segmentation with both majority voting and weighted voting combinations, utilizing three distinct similarity metrics: Squared Differences (SD), Normalized Cross Correlation (NCC), and Mutual Information (MI).

2.1. Majority voting

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.7986	0.8341	0.8081	30.6757	9.0554	9.9499	3361	172157	134667
12	0.8003	0.8463	0.8327	27.3861	7.5498	9.6954	3359	134180	53902
13	0.7692	0.8675	0.8191	11.0454	13.1529	13.4536	3016	85647	44454
14	0.8102	0.8690	0.8225	30.8221	7.4833	10.0499	3693	29088	39269
17	0.8427	0.8674	0.8137	29.3087	13.7477	12.8452	327	27428	18393
Means	**0.8042**	**0.8569**	**0.8193**	**25.8476**	**10.1978**	**11.1988**	**2751**	**89700**	**58137**

Table 10: Resulting metrics of the multi-atlas segmentation with majority voting combination

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.8136	16.5603	103395
12	0.8264	14.8771	63814
13	0.8186	12.5506	44372
14	0.8339	16.1184	24017
17	0.8413	18.6339	15383
Means	**0.8268**	**15.7481**	**50196**

Table 11: Overall means of the multi-atlas segmentation with majority voting combination

2.2. Weighted voting

2.2.1. Metric Squared differences (SD)

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.8007	0.8345	0.8205	30.6757	9.4340	8.7750	2622	136360	87098
12	0.7947	0.8420	0.8358	27.3861	7.8102	9.0554	4249	105780	10985
13	0.7704	0.8640	0.8148	11.0454	12.7279	13.4536	1401	95487	78279
14	0.8105	0.8666	0.8233	30.8221	7.4833	9.4340	2349	12978	1999
17	0.8409	0.8688	0.8154	29.3087	13.9642	12.8452	1239	34677	13729
Means	**0.8034**	**0.8552**	**0.8220**	**25.8476**	**10.2839**	**10.7126**	**2372**	**77056**	**38418**

Table 12: Resulting metrics of the multi-atlas segmentation with weighted voting combination using metric SD

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.8186	16.2949	75360
12	0.8242	14.7506	40338
13	0.8164	12.4090	58389
14	0.8335	15.9131	5775
17	0.8417	18.7061	16548
Means	**0.8269**	**15.6147**	**39282**

Table 13: Overall means of the multi-atlas segmentation with weighted voting combination using metric SD

2.2.2. Metric Normalized Cross correlation (NCC)

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.8007	0.8344	0.8202	30.6757	9.4340	8.7750	2481	140301	88363
12	0.7944	0.8420	0.8358	27.3861	7.8102	9.0554	4295	107171	11284
13	0.7698	0.8642	0.8150	11.0454	12.7279	13.4536	1385	94160	77473
14	0.8105	0.8668	0.8233	30.8221	7.4833	9.4340	2264	15967	3384
17	0.8405	0.8688	0.8154	29.3087	13.9642	12.8452	1240	34394	13511
Means	**0.8032**	**0.8553**	**0.8219**	**25.8476**	**10.2839**	**10.7126**	**2333**	**78399**	**38803**

Table 14: Resulting metrics of the multi-atlas segmentation with weighted voting combination using metric NCC

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.8184	16.2949	77048
12	0.8240	14.7506	40917
13	0.8163	12.4090	57673
14	0.8336	15.9131	7205
17	0.8416	18.7061	16382
Means	**0.8268**	**15.6147**	**39845**

Table 15: Overall means of the multi-atlas segmentation with weighted voting combination using metric NCC

2.2.3. Metric Mutual information (MI)

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.8009	0.8343	0.8201	30.6757	9.4340	8.7750	2495	139608	87855
12	0.7914	0.8419	0.8359	27.3861	7.8102	9.0554	4508	108773	10771
13	0.7698	0.8641	0.8150	11.0454	12.7279	13.4536	1378	94675	77874
14	0.8104	0.8668	0.8231	30.8221	7.4833	9.4340	2215	16353	3356
17	0.8405	0.8686	0.8152	29.3087	13.9642	12.8452	1239	35038	14005
Means	**0.8026**	**0.8552**	**0.8219**	**25.8476**	**10.2839**	**10.7126**	**2367**	**78889**	**38772**

Table 16: Resulting metrics of the multi-atlas segmentation with weighted voting combination using metric MI

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.8184	16.2949	76653
12	0.8231	14.7506	41351
13	0.8163	12.4090	57976
14	0.8334	15.9131	7308
17	0.8414	18.7061	16761
Means	**0.8265**	**15.6147**	**40010**

Table 17: Overall means of the multi-atlas segmentation with weighted voting combination using metric MI

The results obtained with multi-atlas segmentation with previous registration to MNI are shown in Table 18.

Metric	Majority Voting	Weighted SD	Weighted NCC	Weighted MI
DSC	0.8268	0.8269	0.8268	0.8265
HD	15.7481	15.6147	15.6147	15.6147
AVD	50196	39282	39845	40010

Table 18: Overall means of the multi-atlas segmentation without registration to MNI

As seen in the previous tables, the results improved compared to the previous experiments. However, it is worth noting that the metrics were calculated in the MNI space, and upon re-registering to the patient space, there was a loss of approximately 0.008 in the Dice coefficient. Fig. 4 shows the results obtained with multi-atlas segmentation with previous registration to MNI

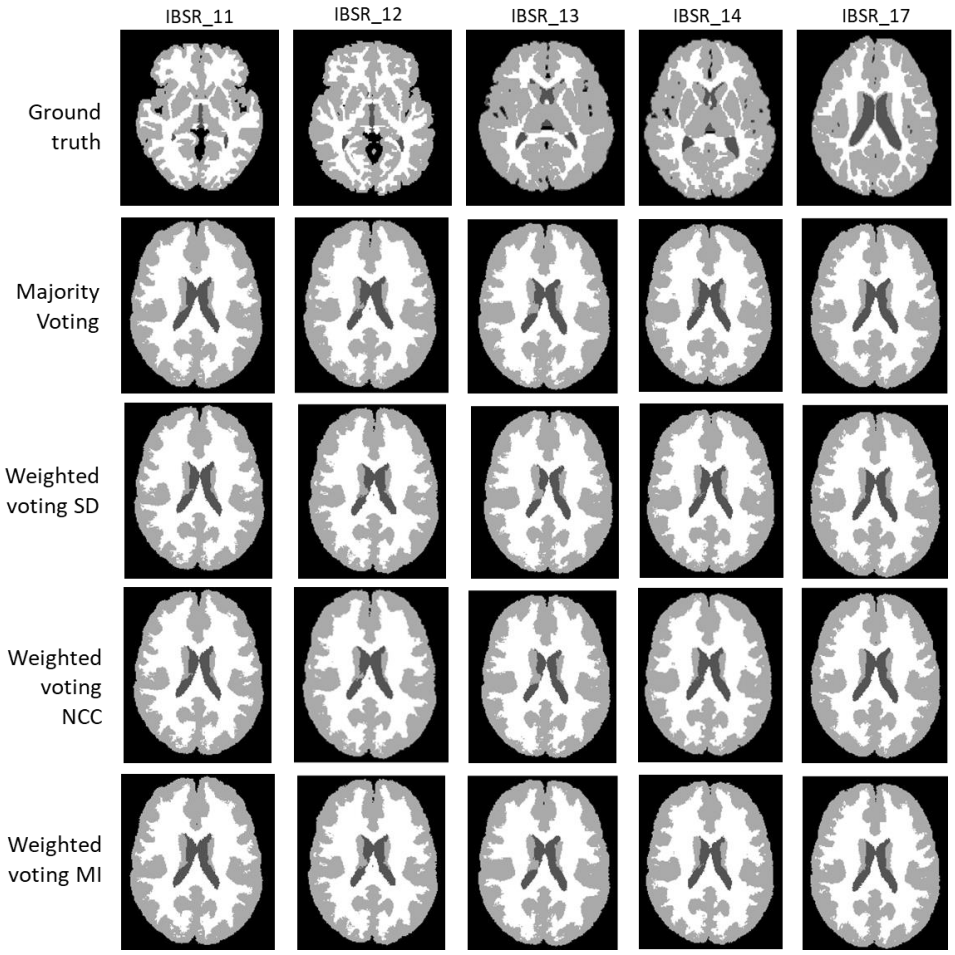

Fig. 4: Results for all the patients of the validation set obtained with multi-atlas segmentation with previous registration to MNI.

Refining with Deep learning

After obtaining the segmentations with the multi-atlas, experiments were conducted to refine them using Deep Learning.

To compare the different experiments, some common settings were established to train the models. It was trained for a maximum of 50 epochs on the pre-established preprocessed training set with a batch size of 8. At the end of each epoch, the training set was augmented, and the performance on the validation set was evaluated. The training phase stops after 10 consecutive epochs without improvement. The model leading to the higher mean Dice score is retained. The models were optimized using the Adam optimization method with an initial learning rate of $1\ x\ 10^{-2}$. The learning rate decreases every 2 epochs by a factor of 0.5 if there is no improvement in the validation loss until it reaches 1×10^{-4}. The loss was computed with the following tissue

weights: $w_{csf} = 15$, $w_{wm} = 0.1$, $w_{gm} = 6$, $w_{bg} = 2$, which were obtained from some tests in early experiments.

The following sections show the results of the 4 main experiments:

1. Unregistered multi-atlas (majority voting)

In this experiment, an attempt was made to refine the results obtained from the segmentation performed using the multi-atlas (combined using majority voting) in the original patient space.

Tables 19 and 20 show the resulting metrics and overall means, respectively.

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.8396	0.8822	0.8298	28.9137	12.8062	10.0995	15560	615396	613146
12	0.8069	0.7508	0.9038	18.0278	8.0623	14.0712	24482	861972	264030
13	0.9329	0.8871	0.8674	14.2478	10.4403	11.8743	13623	1007169	320318
14	0.9215	0.8683	0.9079	8.1240	10.0995	11.0000	18330	1040684	382298
17	0.9297	0.8803	0.9192	28.0891	13.6382	9.1652	23239	1160934	418355
Means	**0.8861**	**0.8537**	**0.8856**	**19.4805**	**11.0093**	**11.2420**	**19047**	**937231**	**399629**

Table 19: Resulting metrics of the refining with deep learning after a multi-atlas segmentation with majority voting

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.8505	17.2731	414701
12	0.8205	13.3871	383495
13	0.8958	12.1875	447036
14	0.8992	9.7412	480437
17	0.9097	16.9642	534176
Means	**0.8751**	**13.9106**	**451969**

Table 20: Overall means of the refining with deep learning after a multi-atlas segmentation with majority voting

2. MNI registered multi-atlas (majority voting)

In this experiment, an attempt was made to refine the results obtained from the segmentation performed using the multi-atlas (combined using majority voting) in the MNI space.

Tables 21 and 22 show the resulting metrics and overall means, respectively.

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.9248	0.9230	0.9076	40.4722	8.3666	7.8102	514	101969	83414
12	0.9326	0.9455	0.8872	35.0143	7.1414	10.8628	4230	63381	5887
13	0.9232	0.9148	0.8903	9.2195	13.0384	11.4018	217	73568	68802
14	0.9506	0.9473	0.9308	26.1725	6.4031	8.8318	2136	24465	22661
17	0.9425	0.9329	0.9156	28.9482	7.1414	11.3578	142	30096	25414
Means	0.9347	0.9327	0.9063	27.9654	8.4182	10.0529	1448	58696	41236

Table 21: Resulting metrics of the refining with deep learning after a multi-atlas segmentation with majority voting in the MNI space

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.9185	18.8830	61966
12	0.9218	17.6728	24499
13	0.9095	11.2199	47529
14	0.9396	13.8025	16421
17	0.9303	15.8158	18551
Means	**0.9239**	**15.4788**	**33793**

Table 22: Overall means of the refining with deep learning after a multi-atlas segmentation with majority voting in the MNI space

Note that while the overall results improve, these metrics decrease when transforming the segmentation back to the patient space.

3. Probabilistic atlas

After analyzing the obtained results, further experiments were conducted by replacing the multi-atlas segmentation with a probabilistic approach as a prior for deep learning. This was done with the aim of achieving smoother transitions between tissues by incorporating spatial information to enhance the edges of the final segmentation.

3.1. MNI registered probabilistic atlas

In this experiment, an attempt was made to obtain segmentation from the results obtained from the probabilistic atlas registered in the patient's MNI space.

Tables 23 and 24 show the resulting metrics and overall means, respectively.

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.9126	0.9126	0.8917	13.1529	8.1240	8.6603	1876	153112	96185
12	0.9159	0.9365	0.8538	22.6716	7.8740	10.8628	6195	128515	21520
13	0.9164	0.9134	0.8696	10.0499	12.8062	12.5698	1770	25515	53371
14	0.9446	0.9392	0.9183	6.9282	8.8318	10.8167	512	70786	37309
17	0.9413	0.9331	0.9065	27.3861	9.8489	11.3578	1515	18999	10482
Means	**0.9262**	**0.9270**	**0.8880**	**16.0377**	**9.4970**	**10.8535**	**2374**	**79385**	**43773**

Table 23: Resulting metrics of the results obtained with deep learning using spatial prior information with probabilistic atlas in the MNI space

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.9056	9.9791	83724
12	0.9021	13.8028	52077
13	0.8998	11.8086	26885
14	0.9340	8.8589	36202
17	0.9270	16.1976	10332
Means	**0.9137**	**12.1294**	**41844**

Table 24: Overall means of results obtained with deep learning using spatial prior information with probabilistic atlas in the MNI space

3.1. Unregistered probabilistic atlas

In this experiment, an attempt was made to obtain segmentation from the results obtained from the probabilistic atlas registered in the original patient space.

Tables 25 and 26 show the resulting metrics and overall means, respectively.

Patient	Dices			Hausdorff distances			AVD		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
11	0.88806	0.92681	0.91756	8.4853	7.0711	6.7082	16019	861921	468042
12	0.90123	0.93653	0.94268	8.9443	6.3246	5.3852	19841	884640	423656
13	089791	0.93616	0.91793	11.4455	10.0499	9.1104	14516	919748	414960
14	0.91574	0.95503	0.94301	8.7750	7.6811	6.0000	20497	947740	455645
17	0.94036	0.95042	0.93353	19.0263	7.8740	9.2736	26508	1049859	500366
Means	**0.90866**	**0.94099**	**0.93094**	**11.3353**	**7.8001**	**7.2955**	**19476**	**932782**	**452534**

Table 25: Resulting metrics of the results obtained with deep learning using spatial prior information with probabilistic atlas in the patient space

Patient	Mean dices	Mean hausdorff distances	Mean AVD
11	0.91081	7.4215	448661
12	0.92681	6.8847	442712
13	0.91733	10.2019	449742
14	0.93793	7.4854	474627
17	0.94144	12.0580	525578
Means	**0.92686**	**8.8103**	**468264**

Table 26: Overall means of results obtained with deep learning using spatial prior information with probabilistic atlas in the patient space

Metric	Unregistered MA	Registered MA	Registered PA	Unregistered PA
DSC	0.8751	0.9239	0.9137	0.9268
HD	13.9106	15.4788	12.1294	8.8103
AVD	451969	33793	41844	468264

Table 27: Overall means of the results obtained with deep learning segmentations

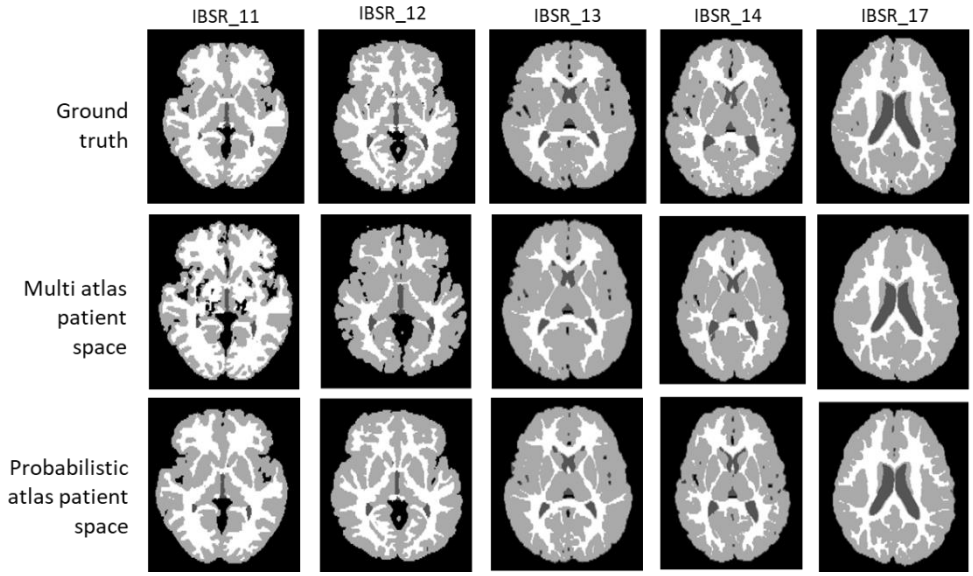

Fig. 5: esults for all the patients of the validation set obtained with deep refining after multi-atlas segmentation and probabilistic-atlas segmentation unregistered.

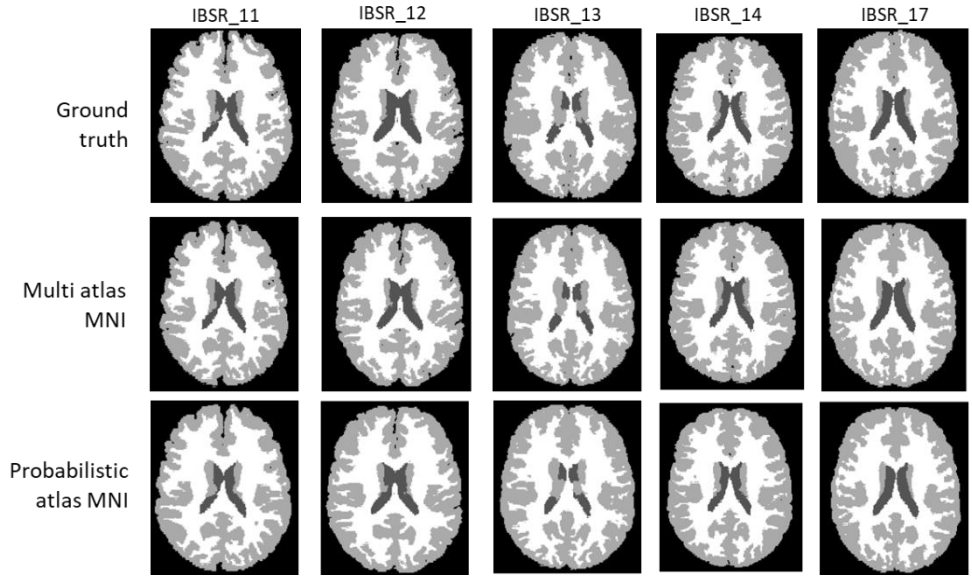

Fig. 6: results for all the patients of the validation set obtained with deep refining after multi-atlas segmentation and probabilistic-atlas segmentation registered to MNI.

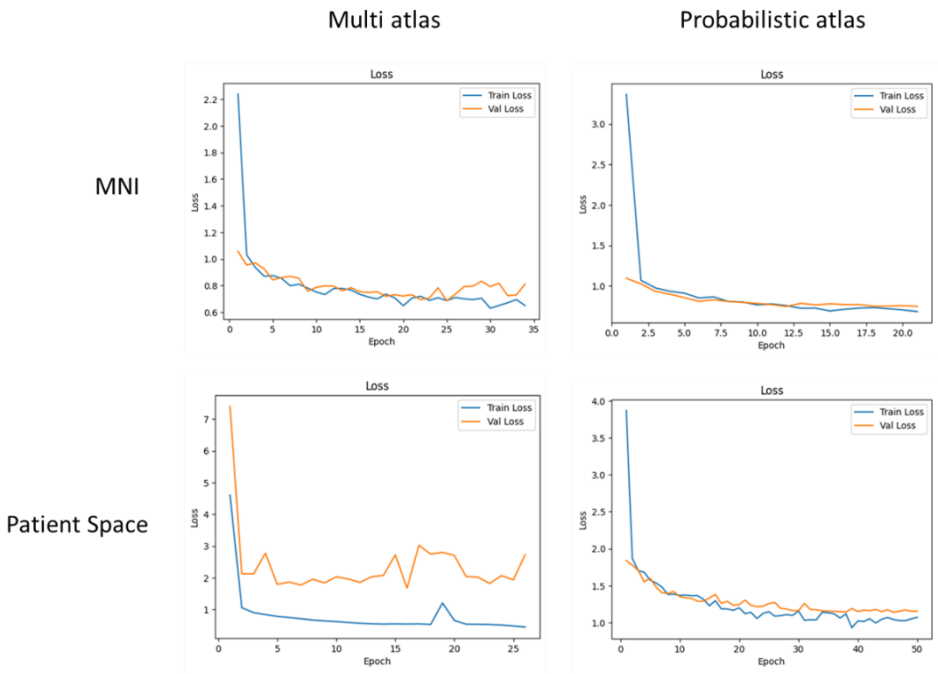

Fig. 7: Graphs of the loss in the refining of probabilistic and multi atlas in the patient space and in the MNI space

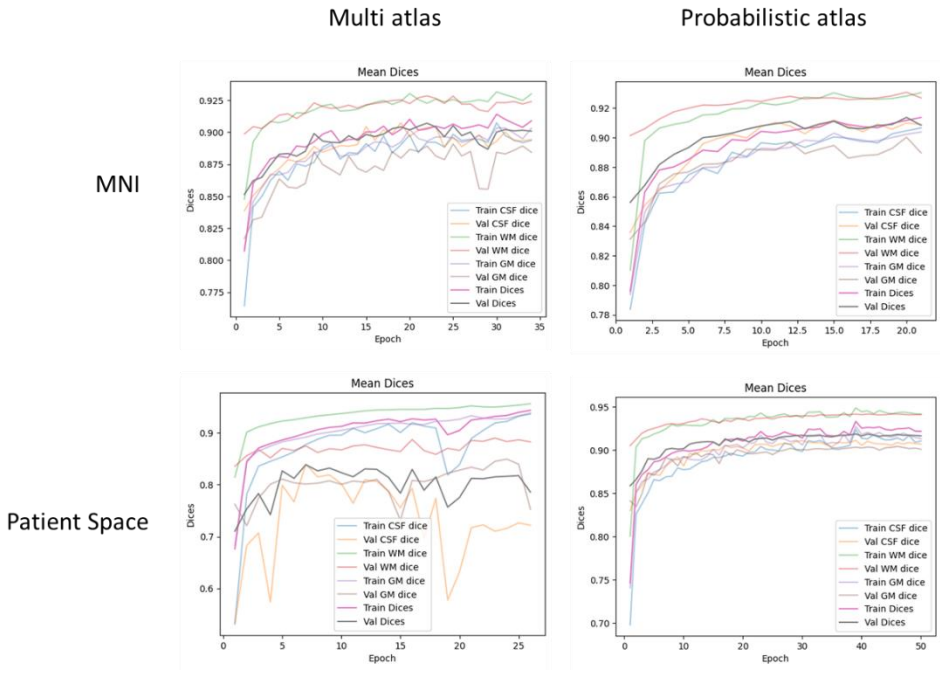

Fig. 8: Graphs of the dice of all the tissues in the refining of probabilistic and multi atlas in the patient space and in the MNI space

As observed in the graphs from Fig. 7 and 8, the outcomes demonstrate more promise when registering to the MNI space, showcasing alignment among all images and a faster convergence of the model to an optimal solution. However, it is crucial to note that these results are preliminary, as the segmentation requires registration back to its original space. Due to the application of non-rigid transformations using B-splines during registration, the inverse registration may lack exact precision. The attempt to

register back to the patient space resulted in a marginal decrease of approximately 0.01 in the mean Dice. Considering these findings, the probabilistic atlas as a prior and without performing a registration to the MNI space was chosen.

Conclusions

In this project, the goal was to develop a segmentation method for the three main tissues of the brain from MRI. The proposed algorithm combines a probabilistic atlas approach with deep learning to achieve accurate and robust segmentation.

The three-step algorithm involves initial preprocessing, ensuring image uniformity through intensity normalization and histogram matching. Subsequently, probabilistic atlas segmentation is executed, generating probability maps for each tissue by registering preprocessed training images to a target image. The key innovation lies in the third step, where deep learning refinement occurs using a cascaded multipath U-Net, incorporating both spatial and intensity information for enhanced segmentation masks. Emphasizing the importance of both intensity and spatial information in brain segmentation, the method underscores their complementary roles in achieving accurate results.

A notable feature of the method is its ability to impose additional penalization on the chosen tissue, which proves advantageous when segmenting one tissue is more challenging than the others.

Registration before deep learning was unnecessary since the model was trained with patches, accommodating variations in resolutions.

In conclusion, the proposed algorithm, uniting probabilistic atlas segmentation with deep learning refinement, holds substantial promise for achieving precise and robust brain tissue segmentation in MRI.

Bibliography

[1] Jose B. et al. “Generating Longitudinal Atrophy Evaluation Datasets on Brain Magnetic Resonance Images Using Convolutional Neural Networks and Segmentation Priors” 2021 Jul;19(3):477-492. doi: 10.1007/s12021-020-09499-z.