Advantages of brain parcellation in Multiple Sclerosis Lesion Segmentation

1st Dario Samuele Pishvai
Dept. of Math and Computer Science
University of Catania
Catania, Italy
dario.pishvai@phd.unict.it

2nd Alessia Rondinella
Dept. of Math and Computer Science
University of Catania
Catania, Italy
alessia.rondinella@unict.it

3rd Francesco Guarnera
Dept. of Math and Computer Science
University of Catania
Catania, Italy
francesco.guarnera@unict.it

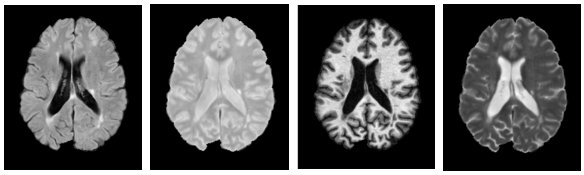
4th Sebastiano Battiato
Dept. of Math and Computer Science
University of Catania
Catania, Italy
sebastiano.battiato@unict.it

***Abstract*—Segmentation of multiple sclerosis lesions plays an important role in understanding disease status. In this work, we focus on the effectiveness of brain parcellation in enhancing the performance of segmentation for multiple sclerosis lesions in Magnetic Resonance Imaging. Brain parcellation does not improve the segmentation performance, but make the results more robust in terms of overall variability (e.g. standard deviation), by dividing the brain into physically significant sub-regions that the model can concentrate on. Our approach combines parcellation with the existing diffusion-based model to increase sensitivity, particularly in regions with small anomalies. We conducted a thorough evaluation of a reference dataset on the field using all available modalities. Our results show how the parcellation of the brain when integrated into a diffusion-based pipeline, makes the segmentation of MS more stable, lowering deviations from the average, and improving some of the results w.r.t. state-of-the-art. This method achieves good segmentation capabilities even with small datasets, providing promising indications for further research.**

***Index Terms*—Multiple Sclerosis, Brain parcellation, Diffusion Models, Medical image segmentation, MRI.**

I. INTRODUCTION

Multiple Sclerosis (MS) is a chronic autoimmune disease that primarily affects the Central Nervous System (CNS) affecting the brain and spinal cord [1]. MS is characterized by focal regions of inflammation together with myelin and axonal degradation, causing a wide spectrum of neurological symptoms and impairments. For treatment and clinical assessments, precise identification of MS lesions could play a crucial role. MS lesions can be seen with Magnetic Resonance Imaging (MRI) appearing in different parts of the brain and spinal cord; moreover the location of them often reflects the severity of the disease [2]. Differentiating lesions according to their specific locations, such as periventricular, cortical/iuxtacortical, brain stem/cerebellar, and spinal cord, is crucial for accurate diagnosis, tracking the course of the disease, and determining the effectiveness of treatment [3]. Although necessary, manually annotating MS lesions on MRI scans is a labor-intensive process that calls for specific knowledge. Furthermore, the intrinsic subjectivity introduces heterogeneity within and across operators, which may affect the repeatability and precision of lesion segmentation [4]. In order to eliminate human-induced biases and provide consistent and reliable clinical evaluations, automated technologies must be developed. The technique for longitudinal brain MRI includes a range of sequences, each providing distinct contrasts for distinguishing between different brain tissues. Specifically, T1-weighted (T1-w), T2-weighted (T2-w), PD-weighted (PD-w) and with Fluid Attenuated Inversion Recovery (FLAIR) images, have proven indispensable in the identification of MS lesions. Of them, FLAIR images are particularly notable because they offer a sharp and contrasted view of the lesions, making it possible to clearly distinguish them from the surrounding tissues (see Fig. 1 for illustrative examples).



(a) (b) (c) (d)

Fig. 1: Examples of axial brain MRI images in different modality of acquisition showing MS lesions: FLAIR (a), PD-w (b), T1-w (c) and T2-w (d).

In this work, we investigate the application of a novel technique that enhances MS lesion segmentation by utilizing brain parcellation [40]. With the term parcellation we usually mean dividing an item into meaningful subregions as opposed to dividing a picture into random segments [37] [39]. Partitioning the brain into discrete areas according to connectivity, structural, or functional based criteria is known as parcellation in the context of brain imaging [40]. This

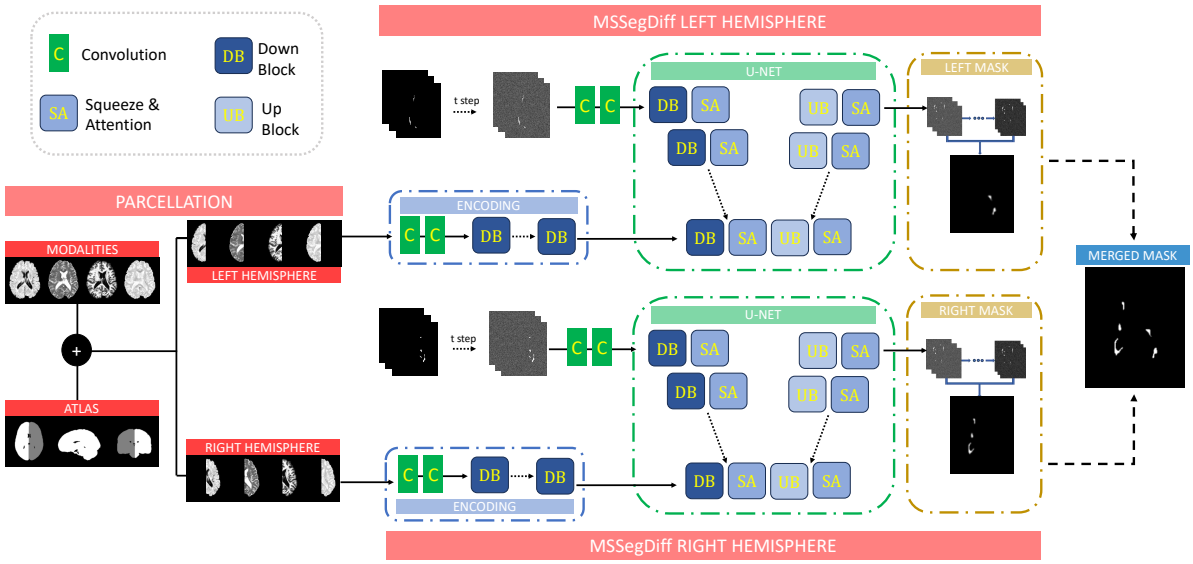


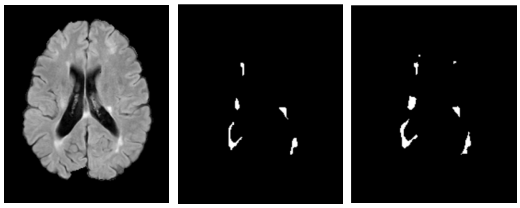
Fig. 2: Schematic representation of proposed pipeline with parcellation step, use of MSSegDiff [36] for both hemispheres and merging for final mask generation.

approach lessens the possibility of misunderstanding arising from the brain's complex structure by breaking the brain down into smaller, physically appropriate components and enables the model to concentrate on focused areas [38]. To further improve segmentation performance, we combine the parcellation with a Diffusion Model (DM) based architecture called MSSegDiff proposed by [36] as depicted in Fig. 2. Our analysis demonstrates how the parcellation combined with the DM, greatly enhances sensitivity, particularly in areas with mild irregularities, improving the segmentation algorithm's overall stability. Our technique performs well, as evidenced by extensive trials on the ISBI2015 [5] dataset's magnetic resonance volumes, which yield a significant mean Dice Score (DSC) on the test set. Therefore, international competitions such as ICPR 2024 ([49]) or ISBI, offer a benchmark to assess progress in automated segmentation methods and encourage the development of more robust techniques. The main goal of this read, remains to build a model capable of detecting lesions due to MS within the patient's brain. The novelty concerns the work of parcellation and thus the identification and volumetric extraction of specific areas of the brain. The optimization of the input through the parcellation optimise feature extraction in the training phase, resulting in more stable segmentation. Brain atlases are crucial tools in neuroimaging that help in brain parcellation, functional mapping, and the study of anatomical variability [45]. They provide consistent identification of specific regions, which facilitates comparative analysis of anatomical areas between subjects in clinical and research contexts. Utilizing atlases can help with tasks like matching pathological deviations from the norm, segmenting anatomical components, and aligning individual brain scans to a common area. Our experiments make use of DM's capabilities to provide reliable 3D medical image segmentation. With their revolutionary strategy of infusing controlled noise at the input and iteratively improving the segmentation label map to improve prediction stability, DMs mark a new frontier in Deep Learning methodology. The architecture outlined in [18] was built upon, and the framework was expanded to produce more potent and customized solutions [36]. Those designs are prepared to handle medical imaging data effectively. The rest of the paper is structured as follows. The employment of the parcellation tool with different application areas in medical picture segmentation is highlighted in Section II, which gives an overview of the state-of-the-art; Section III explores the dataset used, the employed approach, and the specifics of the suggested architecture. Section IV reports on the experimental findings and ablation experiments, whereas Section V contains the paper's closing thoughts.

II. STATE OF THE ART

A. *Medical image segmentation*

Recent advances in deep learning have facilitated its application in medical imaging for neurological disorders like Alzheimer's, MS, and stroke. Brain scan analysis using multimodal data presents challenges, especially due to variability in lesion shape, intensity, and imaging procedures. Deep learning is increasingly used for classification [6] and segmentation tasks, especially with CT and MRI data ([50], [51]). A comprehensive review [7] outlines deep learning methods for MRI and suggests future directions, while another study [8] summarizes MS lesion detection research. CNNs are frequently employed for accurate MS segmentation ([54]), with some works incorporating dual-path designs [19] for patch-wise lesion extraction [9]. Multiple sclerosis lesion segmen-



(a) (b) (c)

Fig. 3: Examples of FLAIR image (a) with the corresponding masks annotated by Rater 1 (b) and Rater 2 (c), respectively. To note that, different Raters have different decision about pixel location in the mask.

tation often uses FLAIR MRI, as noted in studies ([52], [53]), and attention mechanisms are introduced to enhance temporal data interactions. Neural network methods for 3D brain MRI lesion detection are explored in [34] and [35]. Additionally, [36] investigates DMs in medical segmentation, providing a baseline with their MSSegDiff model (used in our experiments). Brain parcellation, dividing the brain into distinct anatomical and functional areas, enhances segmentation accuracy, crucial for diagnosis and treatment in conditions like epilepsy, MS, and Alzheimer's. Parcellation reduces inter-subject variability and enhances anatomical fidelity, as seen in Alzheimer's research where atrophic regions like the hippocampus and entorhinal cortex are isolated [48]. Parcellation-based segmentation improves diagnostic accuracy, aids disease progression tracking [46], and is advantageous for resting-state fMRI [47]. This approach enables researchers to study intrinsic connectivity patterns, aiding in the identification of biomarkers for psychiatric, neurodevelopmental, and cognitive disorders.

III. METHODOLOGY

Brain parcellation divides the brain into regions with distinct physical or functional characteristics, based on the understanding that different brain areas have unique architectures and roles. For instance, the division into left and right hemispheres reflects functional and anatomical asymmetries:

- **Functional specialization:** The left hemisphere typically supports language and analytical thinking, while the right focuses on spatial awareness and creativity.
- **Disorder asymmetry:** Neurological conditions (e.g., stroke, epilepsy) often affect one hemisphere more significantly.

Based on resting-state fMRI data, [38] introduced a widely recognized parcellation of the cerebral cortex into 17 functional networks, revealing brain segmentation into discrete cognitive networks. Alternatively, the Human Connectome Project's multimodal approach [39] divides the cortex into 180 regions using functional, anatomical, and connectivity data, supporting a specialized brain region approach over a monolithic one. Research suggests MS may affect brain hemispheres differently, with altered lateralization potentially linked to cognitive changes. Studies reveal that MS-induced lateralization shifts may impact processes like visual-spatial memory, possibly due to changes in white matter microstructure [42]. Advanced imaging has shown asymmetric atrophy in specific brain regions affected by MS, potentially influencing hemispheric dominance and symptom manifestation [41]. This highlights a need for further studies on MS's lateralized effects, potentially aiding personalized rehabilitation approaches [43]. Deep learning research supports network training based on region-specific brain volumes. For example, DeepMedic [44] demonstrated that multiscale input data improves segmentation. Expanding on this, DM-based generative models learn a diffusion process to produce refined, noise-free segmentation maps. The architecture tested for parcellation, MSSegDiff [36], incorporates a standalone Encoder and Denoising Attention U-Net for extracting key information from the input volume.

TABLE I: Configurations employed in the study, based on 20 distinct folds. A leave-one-subject-out cross-validation was used, with 3 patients for training, 1 for validation, and 1 for testing. The patient numbers refer to all of their time-point scans.

# Fold	Training set	Validation set	Test set
1	[1, 2, 3]	4	5
2	[1, 2, 4]	3	5
3	[1, 3, 4]	2	5
4	[2, 3, 4]	1	5
5	[1, 2, 3]	5	4
6	[1, 2, 5]	3	4
7	[1, 3, 5]	2	4
8	[2, 3, 5]	1	4
9	[1, 2, 4]	5	3
10	[1, 2, 5]	4	3
11	[1, 4, 5]	2	3
12	[2, 4, 5]	1	3
13	[1, 3, 4]	5	2
14	[1, 3, 5]	4	2
15	[1, 4, 5]	3	2
16	[3, 4, 5]	1	2
17	[2, 3, 4]	5	1
18	[2, 3, 5]	4	1
19	[2, 4, 5]	3	1
20	[3, 4, 5]	2	1

A. Dataset

The ISBI2015 dataset [5] was used to evaluate the proposed designs via leave-one-subject-out cross-validation on patients with lesions in both baseline and follow-up scans. This dataset, from the ISBI 2015 Longitudinal MS Lesion Segmentation Challenge, includes 21 MRI scans from 5 patients acquired with a 3.0 T scanner over multiple time points and labeled by two expert raters. Figure 3 illustrates the variability between rater masks, highlighting the difficulty of MS lesion segmentation even for experts. One patient has five time points, while the remaining four have four, with each scan taken a year apart. Each scan includes raw and preprocessed MRI data (e.g., brain extraction, co-registration, and non-uniformity correction) and masks by both raters, with modalities T1-w, T2-w, PD-w,

and FLAIR, all measuring $181 \times 217 \times 181$. The ISBI 2015 dataset is commonly used in MS lesion segmentation studies ([11], [20], [21], [34], [35], [36]). In this study, we used all available modalities for training and testing, utilizing only Rater 1's annotations to assess model stability. Table I details the 20 fold combinations, as applied in prior works ([34], [35], [36]), ensuring consistent training and testing across all configurations.

B. Parcellation

In order to parcel the brain into multiple regions by recording on the patient at a given time point of an atlas, we use the one proposed by Neuroparc[2] [55]. Neuroparc is a public registration algorithm available on GitHub, that gives the opportunity to choose from several MNI-registered atlases[3] for different volumes, while also providing the corresponding csv file, containing the names of the anatomical regions of the brain and the labels assigned to each. The atlas chosen among the atlas proposed by the Neuroparc, for the next recording phase is the HEMISPHERIC[1], with the following partitioning of brain areas:

- **left hemisphere;**
- **right hemisphere;**
- **background.**

The partition of the brain into physically or functionally different areas is referred to as brain parcellation. The brain's divide into the right and left hemispheres, for instance, can be a helpful strategy due to the functional and anatomical asymmetries as suggested in numerous studies [38] [39] by focusing on relevant features of a particular brain area. Therefore, in accordance with what has been reported in the scientific literature in the area of parcellation, an atlas, i.e., a mapping of the various areas of the brain, was needed that was recorded and then matched for each patient at a specific time point. The Neuroparc pipeline is only one part of the parcellation algorithm that we use. In fact, was decided to divide the process of parcellation of each patient's brain into three basic steps:

1) **Atlas Registration**: on the patient's scanned brain at a specific time point. In this step, thanks to Neuroparc's pipeline we are able to create an atlas that fit the patient's MRI, avoiding sharp cuts and volume losses;
2) **Atlases Creation**: at this stage, from the atlas recorded on the patient's scan we derive the left hemisphere atlas and the right hemisphere atlas;
3) **Extracting Volumes**: in this last part, having the atlas of each hemisphere, we extract the volumes of each available mri modalities (T1-w, T2-w, PD-w, FLAIR) and mask (MASK by Rater 1).

[2]https://github.com/neurodata/neuroparc/blob/master/scripts/json_generation.py

[3]https://github.com/NeuroDataDesign/the-ents/blob/explore-atlases/atlases/Results/brainAtlases_color_v2.png

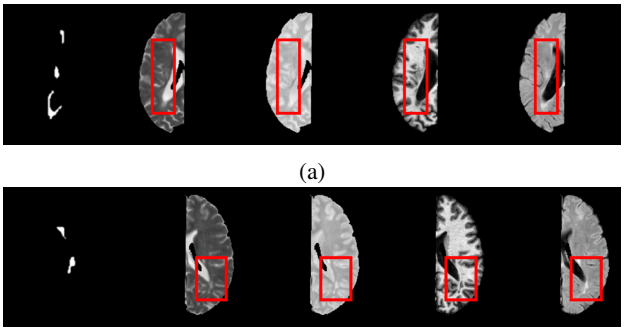
[1]defined by Neuroparc: https://github.com/neurodata/neuroparc/blob/master/scripts/json_generation.py

Fig. 4: Visualization of the registered Atlas of an example patient. The atlas is composed of left hemisphere colored white, while right hemisphere colored gray.

The Neuroparc pipeline is designed to facilitate the processing of MRI parcellation files by performing tasks such as resampling, registration, and label generation, with a specific focus on neuroimaging data. The pipeline it is capable of handling brain images, applying transformations such as resampling and registration, and generating corresponding output files in standardized formats, as depicted in Fig. 4. The class resamples a given brain image to a desired resolution and registers it to a reference brain. This is achieved using affine transformation methods, producing a resampled and registered brain image. After registration, the processor assigns labels to different regions of interest (ROIs) based on a pre-defined labeling scheme or by analyzing the image itself. It rounds and constrains image voxel values to map to valid labels. The Neuroparc pipeline computes the coordinates of centroids for each identified brain region. These are extracted from the processed brain image and are useful for further connectome analysis. The processing pipeline culminates in the generation of labeled brain image files and corresponding metadata files, providing a comprehensive workflow for brain image parcellation and analysis.

After the Atlas Registration phase, we find the Atlas Creation phase. This phase is designed to process a registered and labeled brain atlas by separating its regions into distinct hemispheric components. This function enables the generation of individual brain atlases for both the left and right hemispheres from a labeled brain image. It identifies the unique regions present in the atlas, excluding background or null values. These unique values represent different brain areas that can be assigned to either the left or right hemisphere. Then the function assigns regions corresponding to the left hemisphere and right hemisphere into two separate arrays. The output consists of two separate brain atlas files, each corresponding to one of the hemispheres, facilitating independent analysis or visualization of the hemispheric brain regions. The last step in the process involves the extraction by macro areas of the patient's brain volume at a given time point. The volume extraction is done not only by single time-point, but for each of the 4 modalities (FLAIR, T1-w, T2-w, PD-w), and also for the MASK annotated by Rater 1, how described in Fig. 5.

After the various steps of the process, we performed a volume check on each patient for each modality. As a matter of fact, the volume of the patient's brain after compartmentalisation is on average 5.29% less than the original volume (calculated on FLAIR, MPRAGE, T2 and PD), whereas, in



(a)

(b)

Fig. 5: Visualization of the left(a) and right(b) hemisphere volumes of an example patient after the extraction phase. In the first column we have the lesion masks centered and zoomed for both hemispheres, while for the other modalities, a red boundig box is used to indicate the lesion area.

masks, we have negligible volume loss. This is due to the composition of the atlas, which in fact, as already shown in the Neuroparc preview, omits some anatomical parts. The dataset after parcellation turns out to be divided into Train, Validation and Test sets. Each of these three datasets in turn is subdivided into: right hemisphere and left hemisphere. In each of these divisions we have patient scans divided into time points and modality, so for example a patient may have several time points and for each of them include the 4 modalities (FLAIR, PD-w, T1-w, T2-w) and the MASK. The split into train, validation and was done according the leave-one-out-cross-validation method described in the Table I. After an initial training phase of two models, trained on complementary brain regions obtained by the parcellation, and a subsequent phase of merging the outputs generated by each model for a specific patient, we have the final phase of comparison of the results obtained in article [36] and ours. As explained in the preceding lines, one network will be trained on the volume corresponding to left hemisphere (Fig. 5a), the other complementary network will be trained on the volumes representing right hemisphere (Fig. 5b). Using this method, the model can be trained to identify the unique features inherent in the two parcelled areas, potentially improving segmentation accuracy. The objective is to demonstrate that in terms of performance on the test set is better for the region-specific networks than for the model described in [36]. For a better visualization and a fairer evaluation of ours method, the comparison is made by merging for the same patient, the mask predicted by the Left Hemisphere Network with the mask predicted by the Right Hemisphere Network, using the weights that resulted in the highest dice score on validation. In this way it is possible to create a merged mask formed by the predicted mask on the right-hand side and the predicted mask on the left-hand side. Therefore, the metrics described below were calculated with respect to the ground truth of the entire original volume. In particular, the dice score, true positive rate and positive predictive rate were a way to compare our results with those

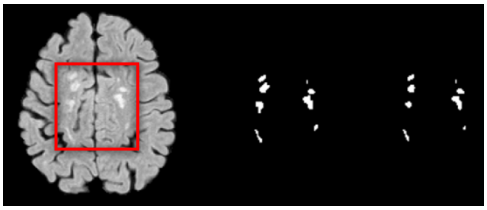


Fig. 6: Example of a segmentation mask obtained with our approach for Fold-15. The image shows: the original slice with a bounding box (red) to indicate the lesion area, the ground-truth mask and the predicted mask.

of the article [36], as illustrated in Fig. 6. This made it possible to compare the results and see if there really is an improvement in stability brought about by the parcellation.

C. Architecture

To evaluate Diffusion Models (DM) for MS lesion segmentation in MRI data, the architecture in [36] was used. The denoising process adds noise to multimodal MRI data and progressively removes it to produce segmentation maps. MSSegDiff, based on [18] and employing BasicUNet [31] as its backbone, integrates attention mechanisms to improve lesion detection. In the forward process, noise is incrementally added to ground-truth masks over ”t” steps, while an Encoder Module (EM) extracts discriminative features from the MRI volume [18]. These features combine with noisy labels in the downsampling path of the Denoising Attention U-Net (DAUNet), followed by 3D Squeeze-and-Attention blocks [32] to focus on important pixel clusters. Feature fusion is achieved by maintaining consistent feature sizes in both encoders. The Denoising Diffusion Implicit Model (DDIM) [27] generates segmentation masks, with accuracy and certainty increasing as steps progress. During testing, the StepUncertainty-based Fusion (SUF) module [18] combines these masks for robust segmentation results.

D. Evaluation metrics

The segmentation masks predicted for each hemisphere are merged and compared to the original challenge masks for evaluation. For fair comparison with [36], the metrics used include: Dice Score (DSC) [30], indicating overlap between predicted and true masks; True Positive Rate (TPR), or recall, measuring correctly identified positives; and Positive Predictive Value (PPV), or precision, assessing the correctness of positive predictions. Additional metrics are included for comprehensive evaluation: False Positive Rate (FPR), the proportion of false positives; Hausdorff Distance (HD), capturing the maximum distance between the predicted and true surfaces; Lesion-wise True Positive Rate (LTPR) and Lesion-wise False Positive Rate (LFPR), representing the count of positive and negative labels in the ground truth, respectively; Absolute Volume Difference (AVD), quantifying volume discrepancy between prediction and ground truth; and Average Symmetric Surface Distance (ASSD), the mean symmetric distance between predicted and actual surfaces.

TABLE II: Results obtained by the proposed method evaluated across all 20 folds.

Fold	DSC↑	TPR↑	PPV↑	FPR↓	HD↑	LTPR↑	LFPR↓	AVD↓	ASSD↓
Fold1	0.7116	0.6938	0.735	$1.75e^{-4}$	41.36	0.7392	0.2805	10.82	1.79
Fold2	0.7450	0.7807	0.7168	$1.75e^{-4}$	22.22	0.7689	0.2598	12.41	0.83
Fold3	0.7258	0.7133	0.7405	$1.75e^{-4}$	36.68	0.7693	0.2110	6.36	1.11
Fold4	0.7149	0.7815	0.6646	$2.5e^{-4}$	36.48	0.7789	0.2257	17.00	1.33
Fold5	0.7444	0.7564	0.7353	$1.0e^{-4}$	38.46	0.6099	0.3196	5.98	2.47
Fold6	0.6813	0.7850	0.6068	$1.2e^{-4}$	31.83	0.6372	0.5606	22.39	3.68
Fold7	0.5975	0.8062	0.4764	$2.5e^{-4}$	45.32	0.7605	0.6566	40.75	0.50
Fold8	0.7185	0.8135	0.6521	$1.2e^{-4}$	31.52	0.6928	0.4223	19.81	2.90
Fold9	0.7535	0.7373	0.7709	$1.8e^{-4}$	30.13	0.8136	0.2376	4.57	1.55
Fold10	0.7332	0.6516	0.8391	$1.0e^{-4}$	47.12	0.7109	0.1582	28.97	2.49
Fold11	0.7010	0.7876	0.6325	$3.4e^{-4}$	36.98	0.7417	0.3425	19.56	1.50
Fold12	0.7457	0.7296	0.7627	$2.0e^{-4}$	29.08	0.7719	0.2674	4.55	1.28
Fold13	0.7872	0.6689	0.9580	$1.25e^{-4}$	28.03	0.3021	0.0754	43.63	0.97
Fold14	0.7807	0.6595	0.9602	$1.0e^{-4}$	23.15	0.3592	0.0410	46.61	0.87
Fold15	0.8739	0.8887	0.8600	$6.5e^{-4}$	27.64	0.4659	0.1895	3.19	0.55
Fold16	0.7844	0.7391	0.8512	$6.25e^{-4}$	31.22	0.3293	0.2458	13.27	1.03
Fold17	0.8469	0.8506	0.8450	$3.5e^{-4}$	31.79	0.62	0.2187	5.50	0.65
Fold18	0.8506	0.8652	0.8381	$3.75e^{-4}$	34.99	0.6154	0.2111	5.71	0.68
Fold19	0.7220	0.6640	0.8095	$4.0e^{-4}$	33.23	0.54	0.3226	29.54	1.55
Fold20	0.6094	0.5272	0.7352	$4.5e^{-4}$	39.70	0.5493	0.4730	41.21	2.88
Mean	0.7414	0.7450	0.7595	$2.63e^{-4}$	33.84	0.6288	0.2859	19.09	1.53

TABLE III: Comparison between proposed method and MSSegDiff [36] across 20 folds.

	Ours			MSSegDiff [36]		
Fold	**DSC↑**	**TPR↑**	**PPV↑**	**DSC↑**	**TPR↑**	**PPV↑**
Fold1	0.7116	0.6938	**0.7350**	**0.7200**	**0.7621**	0.6891
Fold2	0.7450	**0.7807**	0.7168	**0.7520**	0.7371	**0.7721**
Fold3	**0.7258**	0.7133	**0.7405**	0.5854	**0.9155**	0.4312
Fold4	**0.7149**	**0.7815**	0.6646	0.7011	0.7294	**0.6782**
Fold5	0.7444	0.7564	0.7353	**0.7875**	**0.7608**	**0.8193**
Fold6	0.6813	0.7850	0.6068	**0.7671**	**0.7908**	**0.7490**
Fold7	0.5975	0.8062	0.4764	**0.6575**	**0.8560**	**0.5352**
Fold8	0.7185	**0.8135**	0.6521	**0.7629**	0.7709	**0.7574**
Fold9	0.7535	0.7373	0.7709	**0.8045**	**0.8156**	**0.7950**
Fold10	0.7332	0.6516	**0.8391**	**0.8002**	**0.8066**	0.7952
Fold11	0.7010	0.7876	0.6325	**0.7498**	**0.8716**	**0.6590**
Fold12	**0.7457**	0.7296	**0.7627**	0.7419	**0.8640**	0.6507
Fold13	0.7872	0.6689	0.9580	**0.8191**	**0.7121**	**0.9640**
Fold14	0.7807	0.6595	**0.9602**	**0.8076**	**0.6986**	0.9581
Fold15	**0.8739**	**0.8887**	0.8600	0.8162	0.7142	**0.9530**
Fold16	0.7844	0.7391	0.8512	**0.8337**	**0.7603**	**0.9231**
Fold17	**0.8469**	**0.8506**	0.8450	0.7233	0.6431	**0.8479**
Fold18	**0.8506**	**0.8652**	**0.8381**	0.7675	0.7248	0.8247
Fold19	0.7220	0.6640	0.8095	**0.7770**	**0.7393**	**0.8251**
Fold20	**0.6094**	**0.5272**	**0.7352**	0.4574	0.5699	0.3831
Mean	0.7414	0.7450	**0.7595**	**0.7416**	**0.7622**	0.7505

TABLE IV: Volumes and lesion's number of each acquisition Ground Truth mask denoted by the couple patient/timestep

Patient	Timestep	Volume (mm³)	Number of Lesions
01	01	17456.0	30
01	02	14981.0	45
01	03	17826.0	22
01	04	16426.0	20
02	01	26892.0	46
02	02	31405.0	35
02	03	31494.0	36
02	04	32333.0	36
03	01	5990.0	27
03	02	5264.0	29
03	03	5281.0	32
03	04	5176.0	29
03	05	5335.0	25
04	01	2298.0	21
04	02	2196.0	25
04	03	1950.0	22
04	04	2281.0	19
05	01	4331.0	23
05	02	4957.0	26
05	03	4540.0	23
05	04	4383.0	20

E. Implementation details

During preprocessing, MONAI adjustments [29], such as padding to $96 \times 96 \times 96$ and input intensity normalization, were applied. Data augmentation included random flips, scaling, intensity changes, and clipping, as in [36], excluding crop foreground transformation to allow subsequent mask merging. The architecture was developed using PyTorch [28] and MONAI [29] and trained on an NVIDIA A100 GPU. Training used the Adam optimizer with a learning rate of 1×10^{-4}, weight decay of 1×10^{-5}, and batch size of 2. A Cosine Annealing schedule [33] was applied: after a warm-up, the learning rate increased linearly, then decreased with a cosine function, over 1200 cycles. For data augmentation, random $96 \times 96 \times 96$ patches were sampled per epoch to ensure diverse batches from the same patient. During inference, overlapping patches of 0.6 were used to enhance data volume and reduce memory requirements while maintaining detail.

IV. EXPERIMENT AND RESULTS

A leave-one-subject-out cross-validation (LOSO-CV) was performed on annotated subjects as shown in Table I, where three patients are used for training, one for validation, and one

for testing per fold. This approach ensures a subject-wise split, preventing the network from seeing any test patient or their previous time points during training. Unlike other methods ([11], [20], [34]) that use entire patient data for training and a single time point for testing—posing an overfitting risk—this setup avoids patient-wise data leakage. All available modalities were used, similar to state-of-the-art architectures. The "best" model iteration was selected based on the highest Dice score in validation, with results also reported for the final model (after 1200 epochs) to assess overall performance. This LOSO-CV process was applied separately to each hemisphere, and for a fair comparison, predictions from the right and left hemispheres were merged and compared with the ISBI challenge masks [5]. The table III presents a comparison of the proposed method against the state-of-the-art method (MSSegDiff [36]) in terms of segmentation performance across 20 cross-validation folds. The evaluation metrics include DSC, TPR, PPV, with arrows indicating the direction of desired improvement. Overall, our method demonstrates the same performance across all metrics. Specifically, the average DSC of our approach is **0.7414**, compared to **0.7416** for MSSegDiff; while this difference in DSC might appear minimal, it is consistent across most folds, highlighting the robustness of our approach in achieving stable results. Furthermore, our method achieves a higher average PPV (**0.7595** vs. **0.7505**), indicating a little improvement in terms of precision. In fact, the proportion of positive predictions that are correct is a little bit higher across different validation folds. In addition to achieving higher average values, our method exhibits lower standard deviations in DSC, TPR, and PPV across folds, indicating greater robustness and consistency. The reduced variability is particularly evident in challenging cases such as folds 3, 17, and 20, where MSSegDiff shows larger performance fluctuations. Lower standard deviation is critical in medical image segmentation, as it implies that our method performs reliably across diverse data subsets, reducing the risk of underperformance in certain cases. Finally, it is important to note the scenario in which the proposed method work better w.r.t. [36]. Table IV describes number of lesions and volumes of the Ground Truths where observing patient 01, although the number of lesion is high the volume is low. It is known that the described scenario (small sparse lesions) is very challenging, which is the same on which our method reach better results.

V. CONCLUSIONS REMARKS

In this work, we demonstrated the efficacy of parcellation within a segmentation task, and affirmed its effectiveness by applying DM for the segmentation of MS lesions, a task of primary importance for clinical diagnosis and therapy planning. Our results highlight the difficulties posed by the small size of the dataset, which had an impact on the performance of the diffusion models. However, our studies with different choices of atlases have shown several encouraging directions for further development of this method. We are confident that the use of parcellation will be the main view to obtain better results in lesion segmentation, and for this reason a deep analysis of different parcellation atlases will be carried out, with the objective to exceed the state-of-the-art results.

VI. ACKNOWLEDGMENT

Alessia Rondinella is a PhD candidate enrolled in the National PhD in Artificial Intelligence, XXXVII cycle, course on Health and life sciences, organized by Università Campus Bio-Medico di Roma. Francesco Guarnera is funded by the PNRR MUR project PE0000013-FAIR.

REFERENCES

[1] H. Lassmann, "Multiple sclerosis pathology," *Cold Spring Harbor Perspectives in Medicine*, vol. 8, no. 3, 2018.

[2] R. Dobson and G. Giovannoni, "Multiple sclerosis–a review," *European Journal of Neurology*, vol. 26, no. 1, pp. 27–40, 2019.

[3] M. J. Hohol, E. J. Orav, and H. L. Weiner, "Disease steps in multiple sclerosis: a simple approach to evaluate disease progression," *Neurology*, vol. 45, no. 2, pp. 251–255, 1995.

[4] M. Filippi, M. A. Horsfield, S. Bressi, V. Martinelli, C. Baratti, P. Reganati, A. Campi, D. H. Miller, and G. Comi, "Intra-and inter-observer agreement of brain MRI lesion volume measurements in multiple sclerosis: a comparison of techniques," *Brain*, vol. 118, no. 6, pp. 1593–1600, 1995.

[5] A. Carass, S. Roy, A. Jog, J. L. Cuzzocreo, E. Magrath, A. Gherman, J. Button, J. Nguyen, F. Prados, C. H. Sudre, and others, "Longitudinal multiple sclerosis lesion segmentation: resource and challenge," *NeuroImage*, vol. 148, pp. 77–102, 2017.

[6] F. Guarnera, A. Rondinella, O. Giudice, A. Ortis, S. Battiato, F. Rundo, G. Fallica, F. Traina, and S. Conoci, "Early Detection of Hip Periprosthetic Joint Infections Through CNN on Computed Tomography Images," in *Image Analysis and Processing – ICIAP 2023*. Lecture Notes in Computer Science, vol. 14234, Springer, Cham, 2023. [Online]. Available: https://doi.org/10.1007/978-3-031-43153-1_12.

[7] A. Mec¸a, "Applications of Deep Learning to Magnetic Resonance Imaging (MRI)," *2023 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*, Swansea, United Kingdom, 2023, pp. 113–120. doi: 10.1109/iCCECE59400.2023.10238598.

[8] A. Shoeibi, M. Khodatars, M. Jafari, P. Moridian, M. Rezaei, R. Alizadehsani, F. Khozeimeh, J. M. Gorriz, J. Heras, M. Panahiazar, and others, "Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review," *Computers in Biology and Medicine*, vol. 136, p. 104697, 2021.

[9] A. Kaur, L. Kaur, and A. Singh, "DeepCONN: patch-wise deep convolutional neural networks for the segmentation of multiple sclerosis brain lesions," *Multimedia Tools and Applications*, pp. 1–33, 2023.

[10] R. Gamal, H. Barka, and M. Hadhoud, "GAU U-Net for multiple sclerosis segmentation," *Alexandria Engineering Journal*, vol. 73, pp. 625–634, 2023.

[11] A. Rondinella, E. Crispino, F. Guarnera, O. Giudice, A. Ortis, G. Russo, C. Di Lorenzo, D. Maimone, F. Pappalardo, and S. Battiato, "Boosting multiple sclerosis lesion segmentation through attention mechanism," *Computers in Biology and Medicine*, vol. 161, p. 107021, 2023.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, and others, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint*, arXiv:2010.11929v2, 2020.

[14] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, "Transbts: Multimodal brain tumor segmentation using transformer," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*, Springer International Publishing, 2021.

[15] Q. Jia and H. Shu, "Bitr-unet: a cnn-transformer combined network for mri brain tumor segmentation," in *International MICCAI Brainlesion Workshop*, Springer International Publishing, 2021.

[16] U. Baid, S. Ghodasara, S. Mohan, M. Bilello, E. Calabrese, E. Colak, K. Farahani, J. Kalpathy-Cramer, F. C. Kitamura, S. Pati, and others, "The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification," *arXiv preprint*, arXiv:2107.02314, 2021.

[17] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *International MICCAI Brainlesion Workshop*, Springer International Publishing, 2021.

[18] Z. Xing, L. Wan, H. Fu, G. Yang, and L. Zhu, "DiffUNet: A Diffusion Embedded Network for Volumetric Segmentation," *arXiv preprint*, arXiv:2303.10326, 2023.

[19] N. Gessert, J. Kruger, R. Opfer, A.-C. Ostwaldt, P. Manogaran, H. H. Kitzler, S. Schippling, and A. Schlaefer, "Multiple sclerosis lesion activity segmentation with attention-guided two-path CNNs," *Computerized Medical Imaging and Graphics*, vol. 84, p. 101772, 2020.

[20] M. Hashemi, M. Akhbari, and C. Jutten, "Delve into multiple sclerosis (MS) lesion exploration: a modified attention U-net for MS lesion segmentation in brain MRI," *Computers in Biology and Medicine*, vol. 145, p. 105402, 2022.

[21] B. Sarica, D. Z. Seker, and B. Bayram, "A dense residual U-net for multiple sclerosis lesions segmentation from multi-sequence 3D MR images," *International Journal of Medical Informatics*, vol. 170, p. 104965, 2023.

[22] Wu, J., Fu, R., Fang, H., Zhang, Y., Yang, Y., Xiong, H., ... & Xu, Y. (2024, January). Medsegdiff: Medical image segmentation with diffusion probabilistic model. In Medical Imaging with Deep Learning (pp. 1623-1639). PMLR.

[23] Wu, J., Ji, W., Fu, H., Xu, M., Jin, Y., & Xu, Y. (2024, March). Medsegdiff-v2: Diffusion-based medical image segmentation with transformer. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 6, pp. 6030-6038).

[24] J. Wolleb, R. Sandkuhler, F. Bieder, P. Valmaggia, and P. C. Cattin, "Diffusion models for implicit image segmentation ensembles," in *International Conference on Medical Imaging with Deep Learning*, PMLR, 2022.

[25] Chen, T., Wang, C., & Shan, H. (2023, October). Berdiff: Conditional bernoulli diffusion model for medical image segmentation. In International conference on medical image computing and computer-assisted intervention (pp. 491-501). Cham: Springer Nature Switzerland.

[26] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[27] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint*, arXiv:2010.02502, 2020.

[28] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, and others, "Pytorch: An imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[29] M. J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murrey, A. Myronenko, C. Zhao, D. Yang, and others, "Monai: An open-source framework for deep learning in healthcare," *arXiv preprint*, arXiv:2211.02701, 2022.

[30] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[31] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Bohm, J. Deubner, Z. Jäckel, K. Seiwald, and others, "U-Net: deep learning for cell counting, detection, and morphometry," *Nature Methods*, vol. 16, no. 1, pp. 67–70, 2019.

[32] Z. Zhong, Z. Q. Lin, R. Bidart, X. Hu, I. Ben Daya, Z. Li, W.-S. Zheng, J. Li, and A. Wong, "Squeeze-and-attention networks for semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[33] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint*, arXiv:1608.03983, 2016.

[34] F. Raab, S. Wein, M. Greenlee, W. Malloni, and E. Lang, "A multimodal 2D Convolutional Neural Network for Multiple Sclerosis Lesion Detection," 2022.

[35] S. Aslani, M. Dayan, L. Storelli, M. Filippi, V. Murino, M. A. Rocca, and D. Sona, "Multi-branch convolutional neural network for multiple sclerosis lesion segmentation," *NeuroImage*, vol. 196, pp. 1–15, 2019.

[36] A. Rondinella, F. Guarnera, O. Giudice, A. Ortis, G. Russo, E. Crispino, F. Pappalardo, and S. Battiato, "Enhancing Multiple Sclerosis Lesion Segmentation in Multimodal MRI Scans with Diffusion Models," in *Proc. 2023 IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM)*, pp. 3733–3740, 2023, doi: 10.1109/BIBM58861.2023.10385334.

[37] S. B. Eickhoff, B. T. T. Yeo, and S. Genon, "Brain parcellation: A perspective," *NeuroImage*, vol. 170, pp. 396–400, 2018.

[38] B. T. Yeo, F. M. Krienen, J. Sepulcre, et al., "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *Journal of Neurophysiology*, vol. 106, no. 3, pp. 1125–1165, 2011, doi: 10.1152/jn.00338.2011. [Online]. Available: https://journals.physiology.org/doi/full/10.1152/jn.00338.2011.

[39] M. F. Glasser, T. S. Coalson, E. C. Robinson, et al., "A multimodal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, pp. 171–178, 2016, doi: 10.1038/nature18933. [Online]. Available: https://www.nature.com/articles/nature18933.

[40] S. B. Eickhoff, B. T. T. Yeo, and S. Genon, "Brain parcellation: A perspective," NeuroImage, vol. 170, pp. 396–400, 2018.

[41] Zhang, F., Luo, C., Xu, J., Luo, Y., & Zheng, F. C. (2022). Deep learning based automatic modulation recognition: Models, datasets, and challenges. Digital Signal Processing, 129, 103650.

[42] K. Hanken, P. Eling, and H. Hilde, "Functional connectivity lateralization shift of resting state networks is linked to visuospatial memory and white matter microstructure in relapsing–remitting multiple sclerosis", Brain Topography, vol. 27, 301–311, 2014, issn: 1573-6803. doi:10.1007/s10548-014-0405-7. Available: https://link.springer.com/article/10.1007/s10548-014-0405-7

[43] M. A. Rocca, P. Preziosa, and M. Filippi, "Advanced neuroimaging techniques to explore the effects of motor and cognitive rehabilitation in multiple sclerosis", Journal of Neurology, vol. 266, 276–287, 2019, issn: 0340-5354. doi: 10 . 1007 / s00415 - 018 - 9157 - 7. [Online]. Available: https://link.springer.com/article/10.1007/s00415-018-9157-7

[44] K. Kamnitsas, C. Ledig, V. F. Newcombe, et al., "Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation", Medical Image Analysis, vol. 36, pp. 61–78, 2017, issn: 1361-8415. doi: 10.1016/j.media.2016.10.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841516301697.

[45] A. C. Evans, D. L. Collins, and B. Milner, "MRI-based stereotactic atlas of the human brain: Mapping the human brain using three-dimensional proportional system: An approach to cerebral imaging," Neurosurgery, vol. 34, pp. 777–788, 1993.

[46] Bron, E. E., Smits, M., van der Flier, W. M., et al. (2014). Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge. *NeuroImage*, 31(2), 455-469.

[47] Nunes, P. M., Zareba, M., Dierckx, R. A., & Van Laere, K. (2018). Multimodal cardiac segmentation combining atlas-based and region growing techniques for myocardium delineation. *IEEE Transactions on Biomedical Engineering*, 65(5), 2312-2321.

[48] MEGA, M. S. (2001). The entorhinal cortex in Alzheimer's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 71(4), 431-432. doi:10.1136/jnnp.71.4.431a.

[49] Rondinella, A., Guarnera, F., Crispino, E., Russo, G., Di Lorenzo, C., Maimone, D., ... & Battiato, S. (2024). ICPR 2024 Competition on Multiple Sclerosis Lesion Segmentation–Methods and Results. arXiv preprint arXiv:2410.07924.

[50] Shang, L. (2024). Deep Learning Models for Stroke Lesion Segmentation and Brain Aging Clocks (Doctoral dissertation, The University of Wisconsin-Madison).

[51] - Spagnolo, F., Depeursinge, A., Schädelin, S., Akbulut, A., Müller, H., Barakovic, M., ... & Granziera, C. (2023). Automated MS lesion detection and segmentation in clinical workflow: a systematic review. NeuroImage: Clinical, 103491.

[52] Naeeni Davarani, M., Arian Darestani, A., Guillen Cañas, V., Azimi, H., Havadaragh, S. H., Hashemi, H., & Harirchian, M. H. (2024). Efficient segmentation of active and inactive plaques in FLAIR-images using DeepLabV3Plus SE with efficientnetb0 backbone in multiple sclerosis. Scientific Reports, 14(1), 16304.

[53] Alpar, O., Soukup, O., Ryska, P., Paluska, P., Valis, M., & Krejcar, O. (2024). Automated multiple sclerosis progression rate computation of a patient from 2D FLAIR images with Rayleigh-Weibull-Fuzzy imaging and augmented morphing method. Knowledge-Based Systems, 112580.

[54] Ghosal, P., Roy, A., Agarwal, R., Purkayastha, K., Sharma, A. L., & Kumar, A. (2024). Compound attention embedded dual channel encoder-decoder for ms lesion segmentation from brain MRI. Multimedia Tools and Applications, 1-33.

[55] Lawrence, R. M., Bridgeford, E. W., Myers, P. E., Arvapalli, G. C., Ramachandran, S. C., Pisner, D. A., ... & Vogelstein, J. T. (2021). Standardizing human brain parcellations. Scientific data, 8(1), 78.