

École polytechnique de Louvain

Towards improving multiple sclerosis lesion segmentation performance with synthetic lesions

Author: **Juliette DROCHMANS**

Supervisor: **Benoit MACQ**

Readers: **Laurence DRICOT, Colin VANDEN BULCKE, Maxence WYNEN**

Academic year 2024–2025

Master [120] in Biomedical Engineering

Table of contents

Acknowledgements	iii
Abstract	iv
List of Figures	vii
List of Tables	viii
Abbreviations	ix
Introduction	1
1 State of the art	3
1.1 Multiple sclerosis	3
1.1.1 Lesions phenotypes and MRI vizualisation	5
1.2 Medical image segmentation	12
1.2.1 MS lesion segmentation	12
1.2.2 Performance evaluation of automated segmentation methods	14
1.3 Synthetic lesion generation	16
1.3.1 Deep learning based methods	17
1.3.2 Non deep-learning based methods	21
1.4 Texture Analysis/Generation	26
1.4.1 Medical image texture analysis	26
1.4.2 MS lesion texture analysis	26
1.4.3 Texture generation with an High Order Singular Value De- composition	26
1.4.4 Texture generation with THGOP	27
2 Methodology and implementation	29
2.1 Proposed method to generate synthetic lesion	29
2.1.1 Lesion placement	31
2.1.2 Lesion shape	32

2.1.3	Lesion texture and intensity	33
2.1.4	Generation of a synthetic lesion	34
2.2	Experiments	35
2.2.1	Statistical analysis	37
2.2.2	Material	38
3	Results	41
3.1	Assessing the best synthetic lesion's parameters	41
3.1.1	Lesion's blurring	41
3.1.2	Lesion's number	42
3.1.3	Healthy images augmentation	43
3.2	Results on the test set	44
3.2.1	Results on the test set in domain	44
3.2.2	Results on the test set out-of-domain	47
4	Discussion	54
	Conclusion	58
	Références et Bibliographie	60

Acknowledgements

First, I would like to express my gratitude to Prof. Benoît Macq for allowing me to work on this very interesting topic: Multiple Sclerosis.

I would also like to thank Maxence Wynen and Colin Vandebulcke for all the time they spent helping me with this work and for their valuable advice.

Finally, I would like to thank my parents and friends, who have always supported me throughout my academic journey.

Abstract

Multiple sclerosis (MS) is the most prevalent neurological disease in the young adult population, affecting around 2.5 million people worldwide. It is a chronic neuroinflammatory condition that causes lesions in various regions of the central nervous system (CNS). Magnetic resonance imaging (MRI) is widely used for diagnosis, as it enables the visualization of MS lesions. Automatic segmentation of these lesions can provide neurologists with valuable insights into disease progression and support more accurate diagnosis. However, the limited availability of clinical data can decrease the performance of segmentation algorithms. Data augmentation, particularly the generation of synthetic MS lesions on MRI scans, has the potential to improve the performance of deep learning models, leading to more reliable diagnostic tools. In this master's thesis, a non-deep learning method for synthetic lesion generation is presented and then evaluated with a deep learning segmentation algorithm, namely nnU-Net. Synthetic lesions generated by our pipeline were inserted into pathological MRI scans and added to the in-domain training set. The resulting model was then compared to a baseline model trained without synthetic data, by evaluating both on an in-domain test set and two out-of-domain datasets. A slight, non-significant performance improvement (+1.2% in DSC when using synthetic lesions in the training set compared to the baseline model) was observed on the in-domain test set. Results on the out-of-domain datasets were more variable: on the first out-of-domain dataset, a significant improvement was observed (+1.1% in DSC), while on the second one, a significant performance drop occurred (-4.4% in DSC). This suggests that the use of synthetic lesions alone is not sufficient to overcome the model's limitations in terms of generalizability outside of domain. However, the increased segmentation accuracy in-domain suggests better generalization within the domain, which is encouraging for future work. Github link for all codes used in this master thesis

List of Figures

1.1	Typical periventricular MS lesions, adapted from the article written by Filippi and al. (2019) [7]	8
1.2	"Juxtacortical MS lesion located in the U-fiber" taken from the website [12]	9
1.3	"Axial 7T T2*w images showing cortical lesions (arrows) in different patients with MS : a) 41 years old man with SPMS and b) a 31 years old woman with RRMS." Image adapted from the paper by Treaba and al. (2021) [22]	9
1.4	Examples of infratentorial MS lesions, adapted from the paper written by Filippi and al. (2019) [7]	10
1.5	"Proton-density-image of the spinal cord in a patient with MS", adapted from the website[12]	11
1.6	U-Net architecture, with the blue boxes representing a multi channel feature map and white boxes the copied feature maps. Arrows represents different operations. Adapted from Ronneberger and al. (2015) [41]	13
1.7	Pipeline of the method proposed by Mostafa Salem and al. (2019) [57], an U-Net based method to generate synthetic MS lesions in healthy brain MRI	18
1.8	Image adapted from Basaran and al. (2022) [54], this figure illustrates the pipeline for generating both healthy and pathological images with a GAN	20
1.9	Image adapted from LesionMix (Basaran et al. 2023) [56] representing the generating and the inpainting branch present in the method to create synthetic lesions or inpaint real lesions following the lesion's load targeted.	22
1.10	Adapted from Tahghighi and al.(2024) [70], (a) brain MRI without synthetic lesion, (b) WM mask, (c) generated lesion, (d) brain MRI scan with the synthetic lesion.	23
1.11	Adapted from [78]: pipeline of TGHOP, S_n being the reduced space, E and G respectively the forward and backward transformations.	28

2.1	Intuitive idea.	29
2.2	Synthetic lesion generation.	31
2.3	Pipeline for the likelihood map creation.	31
2.4	Axial view of the likelihood map at two different level.	32
2.5	Lesion volume separation, with the first mask coming from MS-MIST[72] database.	33
2.6	Avoiding masks.	34
2.7	Comparison between high blur and low blur lesion on the MRI scan.	36
2.8	Comparison between high blur and low blur lesion.	36
2.9	Comparison between a low load and an high load of lesion.	37
3.1	DSC and nDSC results as a function of the lesion's number.	43
3.2	Comparison of DSC (left) and nDSC (right) for in-domain test set segmentation with or without synthetic lesions in the training phase.	45
3.3	Boxplots for the DSC, nDSC and lesion-wise metrics associated to the in-domain test set segmentation.	45
3.4	(a) DSC obtained for the segmentation of our test set in-domain as a function of lesion load, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of our test set in-domain as a function of lesion load, with and without synthetic lesions added to the training set. (c) Difference in DSC obtained on the test set in-domain between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC obtained on the test set in-domain between the model trained with synthetic lesions and the model trained without.	46
3.5	(a) DSC obtained for the segmentation of our test set in-domain as a function of the lesion's number, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of the in-domain test set as a function of lesion's number, with and without synthetic lesions added to the training set. (c) Difference in DSC obtained on the test set in domain between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC obtained on the test set in domain between the model trained with synthetic lesions and the model trained without.	47
3.6	Boxplots for the DSC, nDSC and lesion-wise metrics associated to the MSLesSeg database segmentation	48

3.7	(a) DSC obtained for the segmentation of MSLesSeg as a function of lesion load, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of MSLesSeg as a function of lesion load, with and without synthetic lesions added to the training set. (c) Difference in DSC associated to the MSLesSeg database segmentation between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC associated to the MSLesSeg database segmentation between the model trained with synthetic lesions and the model trained without.	49
3.8	(a) DSC obtained for the segmentation of MSLesSeg as a function of the lesion's number, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of MSLesSeg as a function of lesion's number, with and without synthetic lesions added to the training set. (c) Difference in DSC associated to the MSLesSeg database segmentation between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC associated to the MSLesSeg database segmentation between the model trained with synthetic lesions and the model trained without.	50
3.9	Boxplots for the DSC, nDSC and lesion-wise metrics associated to the MSShift database segmentation	51
3.10	(a) DSC obtained for the segmentation of MSShift as a function of lesion load, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of MSShift as a function of lesion load, with and without synthetic lesions added to the training set. (c) Difference in DSC associated to the MSShift database segmentation, between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC associated to the MSShift database segmentation, between the model trained with synthetic lesions and the model trained without.	52
3.11	(a) DSC obtained for the segmentation of MSShift as a function of the lesion's number, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of MSShift as a function of lesion's number, with and without synthetic lesions added to the training set. (c) Difference in DSC associated to the MSShift database segmentation, between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC associated to the MSShift database segmentation, between the model trained with synthetic lesions and the model trained without.	53

List of Tables

3.1	Mean DSC and nDSC for the 5 validation sets, when trained with synthetic lesions (low blur and high blur) or without synthetic lesions.	42
3.2	Mean DSC and nDSC for the 5 validation sets, when trained with synthetic lesions added on pathological or healthy images or without synthetic lesions.	43
3.3	Results obtained when testing both model (<i>Baseline</i> is the model without synthetic lesions and <i>SynLes</i> the model trained with synthetic lesions) on the test set in domain, MSLesSeg and MSShift. (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$	44
3.4	DSC and nDSC associated to the in-domain test set segmentation, with and without synthetic lesions (<i>Baseline</i> is the model without synthetic lesions and <i>SynLes</i> the model trained with synthetic lesions) added in the training set.	44
3.5	DSC and nDSC associated to MSLesSeg database segmentation, with and without synthetic lesions added in the training set (<i>Baseline</i> is the model without synthetic lesions and <i>SynLes</i> the model trained with synthetic lesions)	48
3.6	DSC and nDSC associated to the test set segmentation, with and without synthetic lesions added in the training set (<i>Baseline</i> is the model without synthetic lesions and <i>SynLes</i> the model trained with synthetic lesions).	51

Abbreviations

CIS	Clinically Isolated Syndrome
CNS	Central Nervous System
CSF	Cerebro-Spinal Fluid
DIS	Dissemination In Space
DIT	Dissemination In Time
DSC	Dice Similarity Coefficient
FLAIR	Fluid Attenuated Inversion Recovery
GAN	Generative Adversarial Network
GLCM	Grey Level Co-occurrence Matrix
GM	Grey Matter
HOSVD	High Order Singular Value Decomposition
LFPR	Lesion False Positive Rate
LTPR	Lesion True Positive Rate
MPRAGE	Magnetization Prepared Rapid Gradient Echo Imaging
MS	Multiple Sclerosis
MRI	Magnetic Resonance Imaging
nDSC	Normalized Dice Similarity Coefficient
PPMS	Primary Progressive Multiple Sclerosis

RIS	Radiologically Isolated Syndrome
RRMS	Relapsing Remitting Multiple Sclerosis
SPMS	Secondary Progressive Multiple Sclerosis
TDA	Traditional Data Augmentation
WM	White Matter

Introduction

Multiple sclerosis (MS) is a chronic neuroinflammatory disease affecting over 2.5 million people worldwide[1] and occurs three times more often in women than men.[1] It is the most prevalent neurological disorder among young adults[2], and is characterized by demyelination of axons, resulting in focal lesions in the brain and spinal cord.[1] Diagnosis of MS is primarily based on MRI scans, which allow visualization of MS-specific lesions that vary widely in shape, intensity, and location. These imaging findings, combined with others fluid biomarkers, allow neurologists to propose treatment strategies tailored to the disease phenotype observed on MRI. [1]

To enhance MS diagnosis and support healthcare professionals, automatic MS lesion segmentation could be used to provide relevant information such as lesion load, total lesion count and lesion location, enabling a more accurate and faster diagnosis.[3] However, these algorithms, particularly for MS lesions segmentation, are constrained by the limited annotated training data availability, which decrease segmentation performance.[4] In this context, data augmentation, more specifically, the generation of synthetic MS lesions in brain MRI scans, offers a promising way to improve both performance and generalizability, thereby contributing to more reliable diagnosis.[4]

This thesis aims to enhance deep learning-based segmentation algorithms by generating synthetic MS lesions in pathological MRI scans using a non-deep-learning approach that account for the anatomical structure and location of MS lesions. The method relies on a likelihood map for their locations, a lesion's shape dictionary and a texture generation algorithm. Then, a validation pipeline is described, to first find the best hyperparameters for synthetic lesion's appearance and then to see whether the use of synthetic lesions in the training set can improve the segmentation's accuracy on an in-domain and on an out-of-domain test set.

The work is divided into three parts. First, to better understand the context of this research, an in-depth overview of MS is provided, along with a review of current

techniques for MS lesion segmentation and synthetic lesion generation. Then, the proposed method is described in detail with its validation process. Finally, we report our results and key findings, then outline directions for future improvements.

Chapter 1

State of the art

In this section, several topics are discussed, including multiple sclerosis, automatic segmentation, evaluation metrics for segmentation, and the state-of-the-art in synthetic lesion generation, in order to provide a better understanding of the proposed method and the context in which it was developed.

1.1 Multiple sclerosis

Multiple sclerosis (MS) is the most common chronic neuroinflammatory disorder of the central nervous system (CNS)[1] and the leading cause of non-traumatic disease in the young adult population[2]. It affects women roughly three times more than men [1], with an estimated prevalence of more than 2 million people worldwide [1]. Despite its prevalence, no clear causes of the disease has been yet determined.[2] However, some genetic and environmental risk factors have been found to increase the risk of the disease, such as the contraction of mononucleosis after exposure to the Epstein-Barr virus, tobacco use and obesity.[2][1] Overall, the exact causes of MS, as well as the roles of environmental and genetic factors, are still under investigation.[2]

At its core, multiple sclerosis is an autoimmune demyelinating disease causing focal inflammatory lesion in the brain and spinal cord.[1] This systematic destruction of myelin, which is essential for the propagation of nerve impulses, results in chronic motor and cognitive disabilities that can alter the individual well-being of patients.[5][2]

MS diagnosis

Diagnosis of MS is complex due to the wide variety of symptoms[6], which depend on the affected area of the brain.[6] MRI has proven to be an invaluable help in the diagnosis process[1], as it enables the accurate visualization of lesions in vivo and their evolution over time, revolutionising our understanding of the pathology.[1] Based on MRI images, clinical and laboratory evaluations, McDonald’s criteria provides clinical rules to establish MS diagnosis.[7] It is based on evidence of dissemination in time, reflecting MS progressive course, and dissemination in space, reflecting MS involvement of the whole central nervous system, unlike other diseases that tend to be more localized (i.e. NMOSD, which primarily affects the optic nerve).[7]

Dissemination in space (DIS) refers to the presence of lesions in at least two out of four distinct CNS regions of interest.[7] Four CNS regions are considered as region of interest (ROI) in the context of MS: the periventricular brain region, the cortical or juxtacortical brain region, the infratentorial brain region and the spinal cord.[7] If there are at least two of these ROI touched by the disease, it indicates a multifocal CNS process.[7]

Dissemination in time (DIT) signifies new lesions appearance over time.[7] This is typically assessed by comparing MRI scans, particularly using gadolinium-enhanced images.[7] Lesions enhanced with gadolinium are considered as new active lesions, while non-enhancing lesions are older.[7] This technique provides critical information about the age of lesions, helping to confirm disease activity.[7] Additionally, DIT can also be demonstrated through the presence of IgG oligoclonal bands (OCBs) in the CSF, which reflect past neuroinflammatory activity. Alternatively, it can be shown by the appearance of a new lesion (not necessarily contrast-enhancing) on follow-up MRI scans.[8]

Although these guidelines are highly sensitive to MS, their specificity is lower, which can result in other MS-mimicking diseases being misdiagnosed as MS. Therefore, a neurological and radiological expertise is needed to ensure an accurate diagnosis.[7] Inappropriate use of the McDonald criteria or over-reliance on MRI findings can lead to overburdening healthcare system and potentially harming patients.[7]

MS phenotype

Studying MS phenotypes at diagnosis is essential : it helps predict how the disease will evolve and guides treatment choices.[9] At first presentation of a single inflammatory attacks with no prior history, the condition cannot be directly labeled as a MS form, due to a lack of evidence of DIT. In this prodromal phase, if the patient has a clinical event it is termed clinically isolated syndrome (CIS), if there is only radiological evidence (i.e., lesions on MRI without clinical symptoms), it is

termed radiologically isolated syndrome (RIS).[2] If further attacks happened, the diagnosis will be updated to one of the established MS categories.[2] The most common MS categorie is the Relapsing-remitting phenotype (RRMS) [2]. It is characterized by recurring attacks of MS (which can be manifested as neurological events such as vision loss, motor weaknesses, loss of balance...)[9], followed by period of recovery.[9] Over time, patient's condition with RRMS can worsen and transition into a progressive type of MS, namely the secondary progressive type (SPMS)[2], in which recoveries become less complete and attacks are less frequent, while disability slowly accumulates.[9] In contrast, Primary progressive MS (PPMS), considered as a rare form of MS only seen in 10 to 15% of patients[2], is characterized by symptoms worsening progressively over time since the onset of the pathology, with acute attacks that exacerbate the condition compared to the beginning.[9]

1.1.1 Lesions phenotypes and MRI vizualisation

Lesions pathogenesis

MS lesions are focal area of inflammation, demyelination and glial reaction that can appear throughout the brain white matter.[1] It appears after the leakage of the blood brain barrier (BBB), allowing infiltration of peripheral monocytes/microglia in the CNS.[1] Once inside, they activate immune T and B-cells, which perpetuate inflammation in the CNS.[1] Activated macrophages and microglia strip myelin from axons, leading to subsequent axonal loss.[10] In the earliest phase, all newly formed lesions are considered *active*, characterized by the presence of macrophages and microglia within the lesion and ongoing inflammation[10]. Acute inflammation phase lasts around 4 to 8 weeks before transitioning into a chronic phase.[10] Active lesions, which represent the initial phenotype of MS lesions[10], can be subdivided into two categories : active, demyelinating lesions, in which myelin breakdown is still ongoing and active, post-demyelinating lesions, in which demyelination has already ceased but inflammatory cells remain.

Over time, lesion can evolve into *mixed active/inactive*, with macrophages/microglia migrating to the edge of the lesion, while others become completely *inactive*, lacking any residual macrophages/microglia in the lesion.[10] In a subset of patients, this chronic phase is followed by partial or complete remyelination, resulting in repaired myelin sheaths around axons.[10]

This histological heterogeneity underlies the diverse pathological appearances of MS lesions.[10]

MRI and lesion vizualisation

MRI is a powerful tool for in vivo MS lesion vizualisation due to its ability to differentiate soft tissues, allowing a precise brain tissue observation.[11] Different pulse sequences are used to detect microstructural changes associated to MS lesions. Demyelination prolongs both T1 and T2 relaxation times due to myelin's lipid-rich composition which normally produces very short T1 and T2. MS Lesions appear then hypointense on T1-weighted images and hyperintense on T2-weighted images.[12]

Contrast-enhanced T1 and T2 sequences allows to see acute and active inflammation, but also chronic progression of the disease.[13] Additionally, volumetric assessment on T1-weighted (*T1-w*) scans provides information about the following course of the disease by quantifying brain atrophy.[13]

For diagnosis and lesion monitoring, two MRI sequences are particularly valuable[13] : FLAIR and MPRAGE.

FLAIR (fluid attenuated inversion recovery)[14] is a T2-weighted inversion recovery pulse sequence, using a 180 degree pulse to invert the magnetization of the proton in the tissue.[14] After a delay called inversion time (TI), in which the proton recover some of their longitudinal magnetization, a 90 degree radiofrequency pulse is used to excite the protons again.[14] Then, the produced signal is measured with the MRI. Compared to a T2-w sequence, FLAIR use a longer TI to suppress CSF signal, bright in the T2-w and dark in this sequence.[14]

FLAIR sequences excel at detecting MS lesions because they're highly sensitive to fluid.[15] Edema from demyelination stands out as hyperintense, making lesions more visible.[15] It is especially useful for visualizing periventricular lesions, because CSF-signal suppression makes the boundary between the ventricles and lesions more distinct.

MPRAGE (Magnetization Prepared Rapid Gradient Echo Imaging) is an imaging sequence derived from a T1-w acquisition that provides superior contrast between white matter (WM) and gray matter (GM), enabling more accurate brain volumetry and enhanced detection of neurodegeneration.[16] It begins with a 180 degree inversion pulse to invert the magnetization of protons in the tissue.[16] It is then followed by a TI, in which protons recover a certain longitudinal magnetization.[16] Then, a rapid serie of gradient-echoes is used to record datas. This sequence is less sensitive to CSF.

MPRAGE improve the visualization of lesion volumes by providing precise delin-eation of lesion boundaries.[15]

When evaluating MS lesions, any area of abnormal hyper/hypo - intensity signal must be visible on at least two consecutive slices to qualify as a true lesion.[7] However, with slice thicknesses greater than 3 mm, very small lesions may only

appear on a single slice and risk being overlooked.[7]

Lesions phenotypes

Even though MS lesion can affect any region of the CNS, it tend to affect specific WM regions such as the periventricular and juxtacortical WM, the corpus callosum, infratentorial area and the spinal cord.[7] MS lesions form around veins due to blood–brain barrier (BBB) leakage [1], typically resulting in an ovoid shape. While lesions are generally round or ovoid, their shape can vary depending on location: those adjacent to structures such as the ventricles or GM tend to be less uniformly round than lesions in the deep WM.[7] Lesion’s size ranges from a few millimeters to more than one or two centimeters in diameter.[7]

Assessing lesions in different areas is important to evaluate the DIS of the pathology in a patient.[7] For instance, periventricular lesions are defined as T2-hyperintense areas in direct contact with the ventricles. Lesions that touch a ventricle and located in the corpus callosum are included in this category. They form along deep medullary veins, having their axis perpendicular to the lateral ventricles. They are generally defined as *Dawson’s finger* due to their elongated shape.[17][18] [19] They can be detected via a T2-FLAIR sequences with high sensitivity.[7]

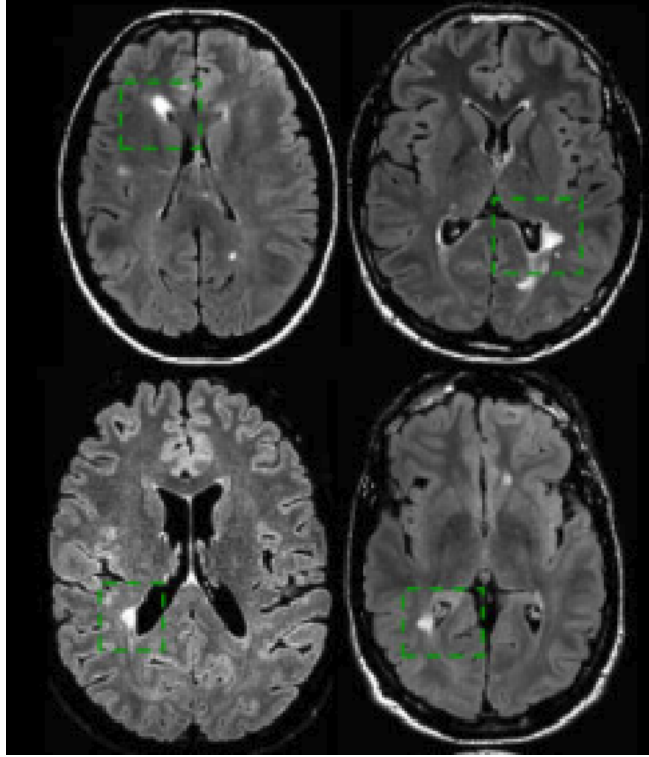


Figure 1.1: Typical periventricular MS lesions, adapted from the article written by Filippi and al. (2019) [7]

Juxtacortical and cortical lesions are T2 hyper-intense regions either in direct contact with the cortex or located within the cortex, without normal WM separating the lesion and the cortex. Juxtacortical lesions involve U-fibers[20] [7], which are short fibers connecting adjacent gyri.[21] These lesions can be located in all brain lobes and in the cerebellum. The detection can be challenging but two mandatory criteria must be checked on double inversion recovery sequences images to ensure the presence of this type of MS lesion : the hyperintensity compared to the surrounding GM and a size of minimum 3 pixels in the images. Juxta-cortical lesion can be tough to find in MPRAGE due to cortex hypointensity compared to the WM in this sequence.[12] Cortical lesions are very specific to MS and allows to recognize CIS patient that have more risks to develop a second clinical attack.[12]

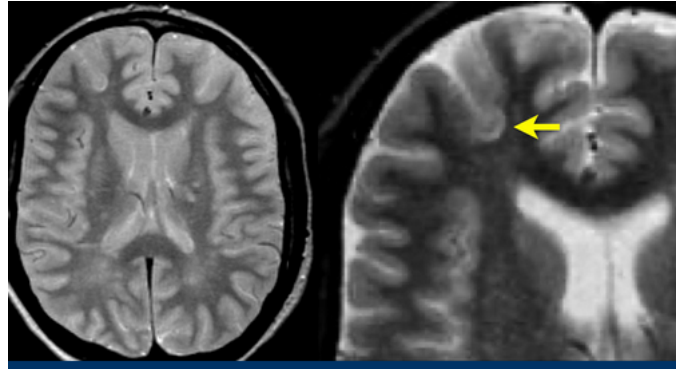


Figure 1.2: "Juxtacortical MS lesion located in the U-fiber" taken from the website [12]

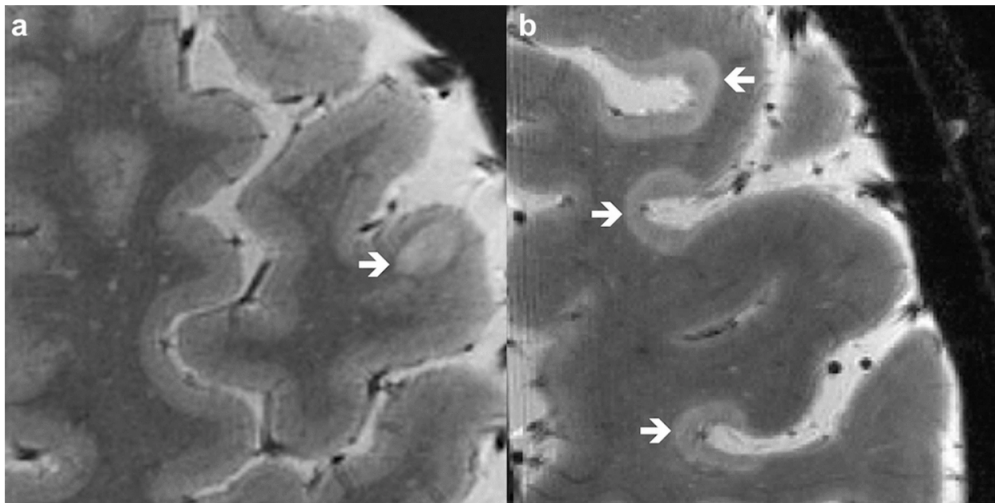


Figure 1.3: "Axial 7T T2*w images showing cortical lesions (arrows) in different patients with MS : a) 41 years old man with SPMS and b) a 31 years old woman with RRMS." Image adapted from the paper by Treaba and al. (2021) [22]

Another type of MS lesion are infratentorial lesions, which are T2 hyperintensities located below the tentorium in the brain stem, the cerebellar peduncles or in the cerebellum.[7] These lesions are often round or ovoid and can range from single well delineated lesions to discrete sub-pial linings along the periphery of the brain stem.[7] Pons are a special case : lesions are often contiguous with the cisterns (small cavity filled with CSF surrounding the brainstem), they involve the floor of the fourth ventricle (CSF filled cavity within the brainstem) and they affect the pontine surface or the trigeminal root entry zone (where the trigeminal nerve

enters the pons).[7]

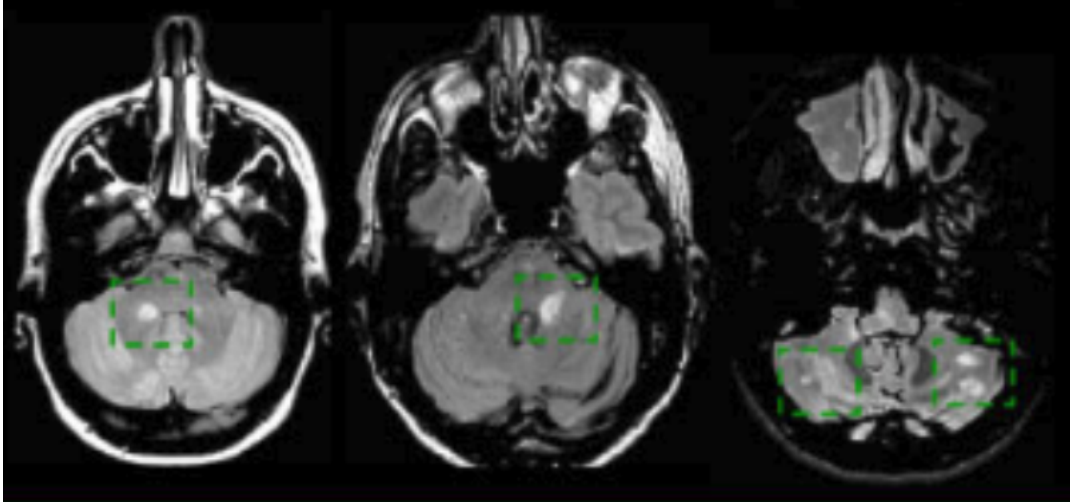


Figure 1.4: Examples of infratentorial MS lesions, adapted from the paper written by Filippi and al. (2019) [7]

Finally, there are spinal cord lesions. Usually multiple and short, involving less than two vertebral segments, these lesions are highly specific to MS and rarely occur in other neuroinflammatory disease.[7] These lesions are hyper-intense in T2-weighted sequences and may occur along the entire spinal cord, even if the cervical portion is more frequently involved. [7] To avoid artifacts, lesions must be visible in at least two sequences or two planes. Generally these lesions are associated with motor disabilities in the disease.[7] [23] [24][25] [26] [27] Techniques of gadolinium enhancement allows a better visualization of the lesions.[7] It is widely used to assess DIT, due to its ability to differentiate between new and old lesions via contrast enhancement, and also to assess DIS by identifying the locations of newly formed lesions.[7]



Figure 1.5: "Proton-density-image of the spinal cord in a patient with MS", adapted from the website[12]

As seen in this subsection, MS lesions exhibit highly heterogeneous shapes and intensities, which can vary depending on their location and the disease phenotype. In addition, specific lesion-related biomarkers, such as paramagnetic rim lesions (PRLs), slowly expanding lesions (SELs), and black holes, reflect different underlying pathological processes, further emphasizing the heterogeneity of MS lesions.

1.2 Medical image segmentation

Image segmentation is a process of partitioning an original image into a set of different regions.[28] When applied in the medical field, image segmentation has multiple applications, including the diagnosis, pathology detection, the study of anatomical structure, treatment planning or computer integrated surgery.[28] Semantic segmentation transforms clinical images into structured data, playing an important role in scientific discoveries and clinical data analysis applications[28] such as radiotherapy planning [29] or treatment response monitoring[30][31] [32]

This subsection begins by reviewing current lesion-segmentation techniques in brain MRI images, along with their key limitations, and then outlines the procedure used to validate these approaches.

1.2.1 MS lesion segmentation

MS lesion segmentation provides information, in particular the lesion load, the total number of lesion and their location, which is important for an accurate diagnosis[3] and disease monitoring[33] [34]. However, manual segmentation can be time-consuming for healthcare professionals[35] and faces several challenges[36], including misclassification of non-MS tissue, confusion with MS-mimicking conditions, and variability in lesion boundaries arising from different MRI sequences and inter-/intra-rater inconsistency.

To overcome all these challenges, several techniques have been proposed to solve this task automatically[36], which speeds up diagnosis and provides a better estimation for lesion load. Early approaches are purely data-driven, relying on simple intensity thresholding of the MR images.[37] Within this class, hierarchical, multi-resolution techniques have also been developed. For example, Pachai et al.[38] introduce a framework built on a pyramidal data structure: at each level of the Laplacian pyramid, the image is locally thresholded to extract hyperintense (lesion) regions, and the binary masks from all levels are then combined, giving the final lesion segmentation. These techniques are easy to implement but can be less robust to noise or partial volume effect.[37] To overcome robustness problem while facing noise, statistical methods can be employed.[37] These approaches explicitly model the intensity distributions of different tissue classes.[37] A well-known example is the Expectation-Maximization (EM) algorithm, which iteratively estimates the probability that each voxel belongs to a given Gaussian component.[37] EM algorithm often achieves strong performance on noisy data, by accounting for Gaussian variance, but can be biased when outliers occurs.[37] To adress the precedent problems,

AI-driven methods have been developed.[37] Deep convolutional networks have then outperformed the state of the art in term of automatic segmentation tasks.[39] [40] In this section, the focus will be put on automatic image segmentation with deep convolutional networks, such as U-Net, the most widely used medical image segmentation algorithm.[41] [42] 1.6

UNet

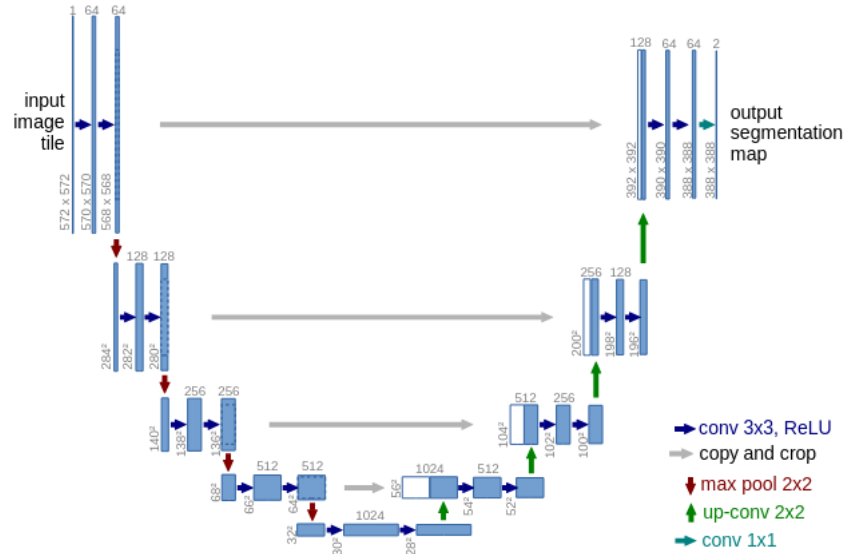


Figure 1.6: U-Net architecture, with the blue boxes representing a multi channel feature map and white boxes the copied feature maps. Arrows represents different operations. Adapted from Ronneberger and al. (2015) [41]

UNet is a CNN-based architecture originally designed for semantic segmentation.[41] It consists of two symmetrical paths: the encoder, a contracting path with convolutional and pooling layers used to capture the image's contextual features, and the decoder, an expanding path composed of transposed convolutions and up-sampling layers to generate the final segmented image.[41] The encoding and decoding paths are linked together by skip connections, which propagates spatial details, enabling the decoder to better interpret the information learned in the encoder part.[41] Finally, the bottleneck, which is the shallowest part between the encoder and the decoder, transforms the abstract information extracted by the encoder into a format that can be interpreted by the decoder.[41]

Although U-Net has demonstrated excellent performance on natural and medical image segmentation tasks[41], fine-tuning and designing a complete deep-learning

pipeline around U-Net can be challenging. It often requires extensive trial-and-error, since choices of parameters (i.e. preprocessing steps, network depth, patch size, loss function, and post-processing methods) can significantly impact the model’s performance.[43] To address this, a framework called nn-UNet and built around the U-Net architecture was developed.[43] Its a deep-learning based algorithm with an U-Net architecture, but the specificity is its self-configuration, including preprocessing, network architecture and post-processing. Its automated configuration covers all the pipeline, allowing for less human impact by reducing the manual decision.[43] nnU-Net has a pipeline based on three steps :

- Fixed parameters which are not adaptable. These parameters have been declared by the original authors as optimal in all their studied applications.
- Rule-based parameters which use the dataset fingerprint to adapt some parameters based on hard-coded heuristic rules.
- Empirical parameters which are chosen by trial-error. For instance the best U-Net configuration.

Other studies have already demonstrated the suitability of MS lesion segmentation using nnU-Net [44] [45], which has outperformed many other segmentation algorithms and exhibits strong generalization capabilities, enabling more accurate delineation.

1.2.2 Performance evaluation of automated segmentation methods

Selection and application of the right evaluation metric is essential for validating the segmentation (i.e. the image-analysis pipeline, particularly in biomedical segmentation) since they drive decisions about a method’s suitability and its clinical applicability.[46] Generic metrics do not always fit biomedical problems, which can compromise the validation of the method’s performance.[46] In this part, a brief overview of several metrics used in the literature to evaluate automatic MS lesion segmentation methods is done. They fall in two subcategories: voxels-based metrics and lesions-based metrics (more specific to MS).

First, **voxel-based metrics** are metrics that assess classification performance at the voxel-level. To do so, these metrics identify True Positives (TP), True Negatives (TN), False Positive (FP) and False Negative (FN). Following up on these identifications, precision and specificity can be computed, referred in MS as *True positive rate (TPR)*, defined by:

$$\text{TPR} = \frac{TP}{TP + FN} \quad (1.1)$$

Measuring the proportion of actual positives correctly identified[47], and *False positive rate (FPR)*, defined by:

$$\text{FPR} = \frac{FP}{TN + FP} \quad (1.2)$$

Measuring the proportion of actual negatives incorrectly labeled as positive. [48] TPR assessed how well the model finds positive voxels, while FPR evaluate the number of negative pixels classified as positive. An effective model seeks a high TPR (few missed positives) and a low FPR (few false alarms). Using both together gives a clearer picture of segmentation performance.[48]

The most popular voxel-based metric for assessing MS lesion segmentation is the *Dice similarity coefficient (DSC)* [49] or dice score. DSC is widely used in medical image segmentation because of the extreme imbalance between background and lesion voxels.[50] Equation 1.3 is the mathematical formulation of DSC:

$$\text{DSC} = \frac{2TP}{FP + FN + 2TP} \quad (1.3)$$

Despite its popularity, the DSC is biased by the prevalence of the positive class in the ground truth.[51] In MS lesion segmentation, this implies that achieving accurate segmentation for a patient with a high lesion load has a greater impact on the overall DSC than achieving the same level of accuracy for a patient with a low lesion load.[51] To address this issue, the *normalized dice similarity coefficient (nDSC)* was introduced. The nDSC corrects for class-imbalance bias by incorporating a normalization factor (κ) that accounts for both the relative sizes of the predicted classes and an expected prevalence.[51] Equation 1.4 is the mathematical formulation of nDSC:

$$\text{nDSC} = \frac{2}{2 + \kappa p + n} \quad (1.4)$$

with $p = \frac{FP}{TP}$, $n = \frac{FN}{TP}$, $\kappa = h(r^{-1} - 1)$. Where h is the ratio between the positive and the negative classes in the ground truth, r is the expected positive-class prevalence.[51]

The other way to assess the performance of MS lesion segmentation methods is through the use of **lesion-level metrics**. These metrics begin by identifying lesion-wise TP, FP and FN. Specifically, a lesion-wise TP is an overlap of at least one voxel between a lesion in the segmented map and in the ground-truth map. In the same way, lesion-wise FN and FP are defined by no overlap between the two maps.[35] Since these metrics follow more a detection strategy, the TN cannot be computed. The two main lesion-specific measures are:

- *Lesion true positive rate (LTPR)* which is the ratio between the lesions that overlap the ground truth in the prediction divided by the number of ground truth lesions. [52] Equation 1.5 is the mathematical formulation of LTPR:

$$\text{LTPR} = \frac{LTP}{LTP + LFN} \quad (1.5)$$

Where LTP is lesion true positive and LFN lesion false negative. [52]

- *Lesion false positive rate (LFPR)* which is characterized by the ratio of lesions not overlapping the ground truth in the prediction, divide by a sum of all predicted lesions. [52] Equation 1.6 is the mathematical formulation of LFPR:

$$\text{LFPR} = \frac{LFP}{LFP + LTP} \quad (1.6)$$

Where LFP is lesion false positive.

As for TPR and FPR, these metrics are complementary, and similarly, to obtain an optimal result LTPR and 1-LFPR must be maximized.

1.3 Synthetic lesion generation

One of the main remaining challenges in deep learning methods is the lack of available labeled training data, especially for medical image segmentation.[4] A way to overcome this issue is by using data augmentation, as it increases the diversity and quantity of training data, enabling a better generalization and higher segmentation performance.[4]

In the context of MS lesion segmentation, clinical datasets are often lacking of representative data due to the wide range of MS phenotypes (some common, others rare cf. section 1.1), the variability in lesion shapes, sizes and anatomical locations (i.e. lesion’s phenotypes subsection 1.1.1) or even the variability in MRI acquisition protocols, making the generalizability of the segmentation more difficult.[53]

One way to perform data augmentation in this setting is by creating synthetic lesions, which proved its efficacy to improve deep learning-based method[54] [55] [56] by allowing more variability in the dataset and, in turn, improving the robustness of segmentation algorithms.

This subchapter describes different methods proposed to generate synthetic lesions and their limitations, starting with deep learning based methods including U-Net and Generative Adversial Networks (GAN), and following with non-deep methods.

1.3.1 Deep learning based methods

The state-of-the-art in term of deep learning-based data augmentation methods can be implemented in two ways, by transforming already existing lesions or by creating completely new ones, and to achieve either strategy, the most commonly used architectures in the literature are U-Net subsection 1.2.1 and Generative Adversarial Network (GAN).

Mostafa Salem and al. (2019) method

In the method proposed by Mostafa Salem and al. [57] (Figure 1.7), the main objective is to add synthetic lesions, coming from a pathological image, to an healthy brain MRI image using an U-Net based architecture. The first step of the algorithm is to compute eight different binary masks created by thresholding the intensity of a pathological image. The model receives as input the T1w and FLAIR healthy image along with the eight binary masks created. Each mask is spatially registered to the healthy image to ensure precise alignment. Lesion patches are extracted from the pathological image and slightly dilated to modify their original shapes. The output is a synthetic pathological image where lesions are pasted at the same place than in the original pathological image.[57]

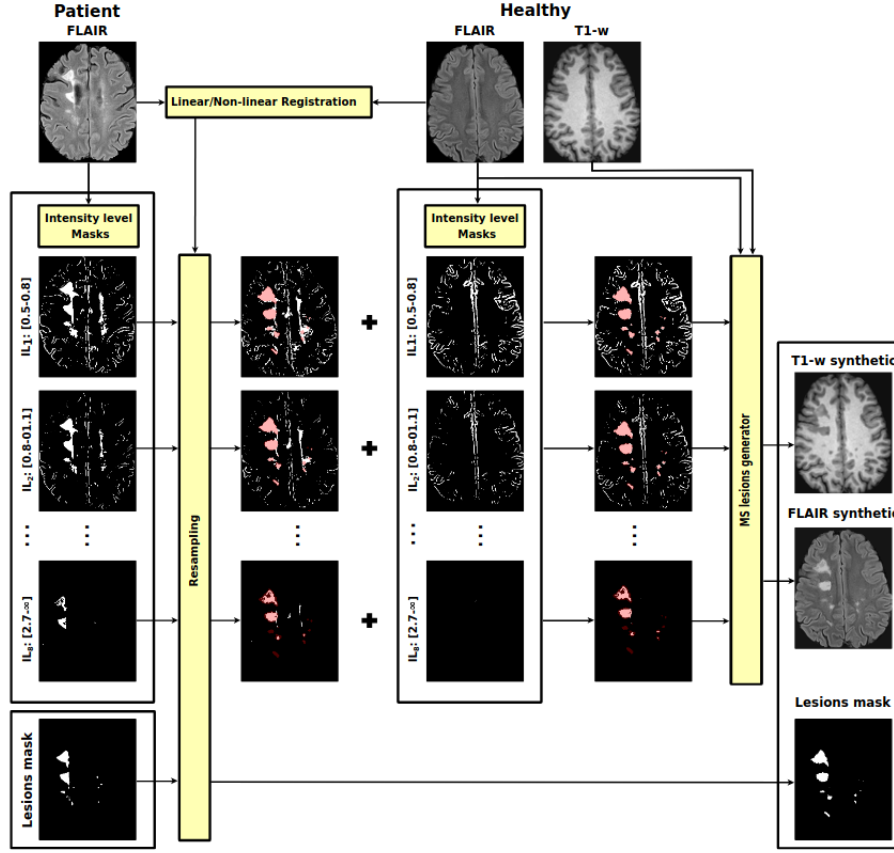


Figure 1.7: Pipeline of the method proposed by Mostafa Salem and al. (2019) [57], an U-Net based method to generate synthetic MS lesions in healthy brain MRI

The method was validated on two independent datasets.[57] The first, an in-house clinical MS cohort, includes 65 patients with CIS or early RRMS and 15 healthy controls. The second is the ISBI2015 challenge set[58], comprising 5 training subjects (4–5 time-points each, for a total of 21 scans) and 14 test subjects (61 scans in total). Segmentation performance was measured with the DSC, the sensitivity and the precision metrics.[57]

For both cohorts, the training set comprised one real pathological scan supplemented by two synthetic scans created by registering the real-lesion mask, using both linear and nonlinear registration, onto two healthy scans drawn from the same dataset as the one used for testing. On the clinical cohort, adding synthetic lesions at a low total lesion load (0.34 mL) improved the DSC by 11 %. Likewise, on the ISBI2015 set, the same low-load augmentation showed a 13 % DSC gain [57]. nDSC and lesion-wise metrics were not computed.

The key advantage of this synthesis approach is that it does not require binary masks as input. New lesions are generated by applying a threshold on the intensity of a pathological MRI image.[57] A limitation of this method is that generated lesions on the healthy image are always the same as the ones from the pathological image. An improvement would be to build a lesion dictionary, allowing for greater diversity and variability in the generated synthetic lesions.

Basaran and al. (2022) method

Another technique for generating synthetic lesions is the use of GANs, which consist of a discriminator and a generator working together. The generator creates new fake images and the discriminator tries to identify the real images from the fake ones. They compete with each other since the goal of the generator is to trick the discriminator into believing that the fake image is real. This process evolves continuously until the probability of the discriminator to be right about the image is fifty percent, which means complete randomness.[59][60] This way, GANs enable the controlled generation of images. [59][60][61]

The method proposed by Basaran et al. (2022) [54] (Figure 1.8) enables the creation of new lesions on healthy MRI scans without relying on pre-existing labeled data. The network can also generate healthy images using a pathological image as input. This architecture uses two generators, G_P and G_H . The first takes a healthy image as input and generates a pathological image, while the second performs the reverse operation. Each generator consists of an encoder and two decoders, which contribute respectively to synthesizing the foreground and background of the output image. Finally, three discriminators, one for the lesion mask (D_F), one for the healthy image (D_H), and one for the pathological image (D_P), attempt to distinguish real images from synthetic ones.

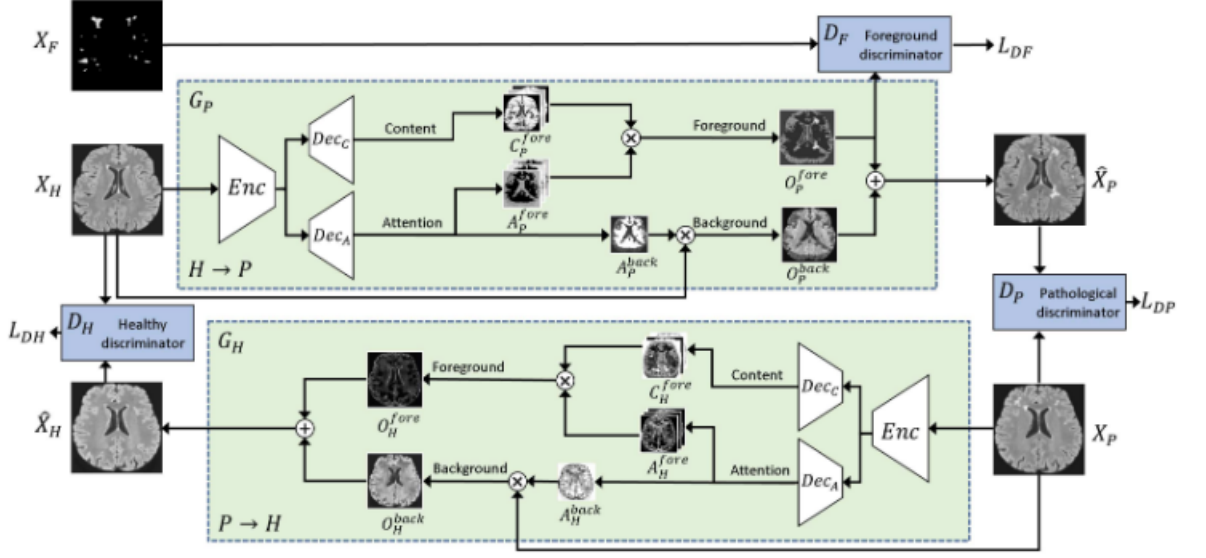


Figure 1.8: Image adapted from Basaran and al. (2022) [54], this figure illustrates the pipeline for generating both healthy and pathological images with a GAN

To evaluate the data augmentation method, several datasets were used. First, the GAN was trained on a private dataset comprising FLAIR MRI scans from 120 subjects, half-healthy and half-pathological. For evaluation, the GAN was trained on the MICCAI 2016 MS lesion database MSSEG2016[62], and tested on both the MICCAI 2008 MS lesion database MSSEG2008[63], and the ISBI2015 challenge set.[58] The proposed augmentation method was then compared to the traditional data augmentations (TDA) provided by nnU-Net[43] and to Carvemix[55] augmentation subsection 1.3.2, using various sizes of the original training set. The evaluation metric was the DSC, nDSC and lesion-wise metrics were not computed. The authors method outperformed all baselines, with particularly large gains on small datasets (e.g., traditional data augmentation achieved a DSC of 37 %, Carvemix + TDA 39 %, whereas the proposed method achieved 56.6 % on the ISBI2015 dataset).

Despite improvements in lesion segmentation DSC, employing a GAN can result in unstable convergence due to the interplay between the generative and discriminative networks.[64] [65] Moreover, GANs are heavy and needs a training period, which is time consuming and not always accessible.

1.3.2 Non deep-learning based methods

In this section, we introduce lesion-generation methods that do not rely on deep learning. These include “mix-and-match” techniques (i.e. where lesions from one patient’s MRI are pasted onto another’s scan), approach that synthesize lesions without using any labeled data, and a method that draws from a dictionary of real lesions to augment healthy MRIs.

Cutmix, Carvemix and Lesionmix

CutMix, CarveMix and LesionMix are data augmentation techniques sequentially proposed between 2019 and 2024, with CarveMix and LesionMix more specific to lesion-wise augmentation.

The first method introduced was CutMix [66], in 2019, which augments a training set by cutting a rectangular patch from one image and pasting it into another. The main advantage is its simplicity and its strong performance.[66] However, it was developed initially for classification tasks, so the method does not adapt well to spatial localization requirement found in medical image segmentation. [55] Moreover, this technique is not really adapted for brain lesion segmentation, being completely unaware of lesions or brain structure. [55]

A second algorithm, this time developed for segmentation tasks, was namely CarveMix [55]. It is a methodology introduced in 2023, which extracts a lesion-containing region from one pathological MRI scan and inserts it into the same anatomical location in a second scan, making the method truly lesion-aware, compared to CutMix. It also allow a better DSC than CutMix, especially on smaller size dataset in which the DSC is 30% higher in CarveMix. This approach preserves realistic lesion shapes and intensities. Its main limitation is the lack of realism in some synthesized images. However, in practice, perfect realism is not always required to boost model performance.[67]

Finally, the most up-to-date algorithm is LesionMix [56], published in 2023. The input of the model is a pathological MRI image and its associated binary mask. The method allows for either lesion population or lesion inpainting following the lesion’s load (which is sampled from a distribution chosen by the user). This adaptability allow the user to have a good class imbalance in the dataset : if there are too many pathological images he can choose to create synthetic healthy image and vice-versa.

The population branch will extract first a lesion from the image. It will proceed to a lesion augmentation level, by flipping, rotating, adding gaussian noise, resizing or even do an elastic transformation to the lesion. A position is drawn from a lesion likelihood map[68], which encodes the probability of lesion locations based

on previously observed lesions, and the lesion is reinserted at that location. The process of lesion population is iteratively repeated until the right volume is reached. In the inpainting branch, a selected lesion is filled with WM to erase it from the current image. To do so, each axial slice of the image is extracted and inpainted individually. Once complete, inpainted slices are reinserted into their original positions. A Gaussian blurring is then applied to produce a more natural result, avoiding visible boundaries between the WM and the inpainted lesion. The pipeline of the procedure is explained on the Figure 1.9.

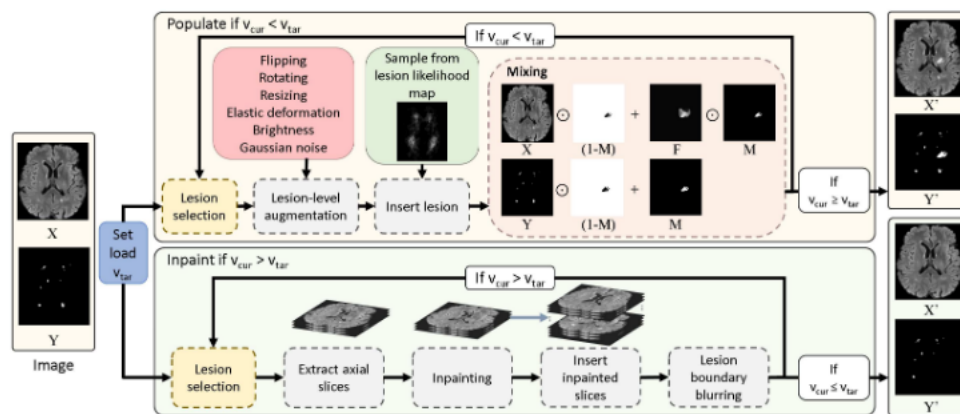


Figure 1.9: Image adapted from LesionMix (Basaran et al. 2023) [56] representing the generating and the inpainting branch present in the method to create synthetic lesions or inpaint real lesions following the lesion’s load targeted.

To validate LesionMix, four independent datasets were used : MICCAI 2008 MS lesion database MSSEG2008 (20 subjects) [63], ISBI2015 challenge set (21 subjects) [58], MICCAI 2016 MS lesion database MSSEG2016 (15 subjects)[62], MICCAI 2017 White matter hyperintensities database MSSEG2017 (60 training, 110 testing) [69]

From the MICCAI 2017 training set, a single image was selected and augmented 100 times; these synthetic examples were added to the original training set to train the model, while the other three datasets were reserved for testing.

The authors evaluated performance with several lesions load added and compared against no augmentation. An uniform volume of lesions shows the best benefice. Then, the data augmentation technique was tested against the TDA included in nnU-Net, the CutMix method and the CarveMix one. LesionMix shows better performance in practically every experiment, with DSC 2 to 5 % higher than the other techniques and greater significance is obtained while using a small size of original dataset. nDSC and lesion-wise metrics were not computed. Unlike its

predecessor, LesionMix leverages a likelihood map to maintain spatial awareness[56] and introduces efficient lesion inpainting, a feature absent in earlier algorithms. Consequently, it offers improved control over lesion load in images.[56]

Tahghighi and al. method

The method proposed by Tahghighi and al. [70] can create completely new synthetic lesions without using labeled data by extracting WM from the brain MRI to create a new lesion.

Figure 1.10 shows a visualization of the pipeline:

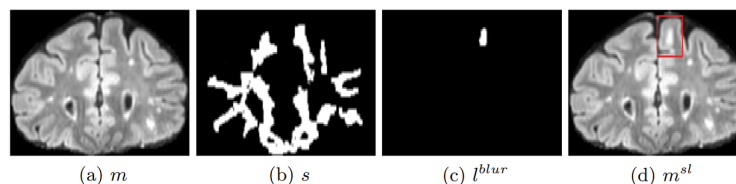


Figure 1.10: Adapted from Tahghighi and al.(2024) [70], (a) brain MRI without synthetic lesion, (b) WM mask, (c) generated lesion, (d) brain MRI scan with the synthetic lesion.

The input is a healthy brain MRI, in which WM voxels are extracted by thresholding the image. A second threshold then removes non-relevant areas, and a Sobel filter isolates only the internal WM by eliminating boundary regions. These three thresholds allow the creation of a likelihood map for lesion locations. To create new lesion shapes, the method employs smoothed cuboids, which are finally inserted back into the image as synthetic lesions.

Two dataset were used for evaluation, the MSSEG-2 challenge set [45], consisting of 20 MRI FLAIR scans pairs (baseline and follow-up) from MS patients, coming from 15 different scanners and the MICCAI-16 (MSSEG)[71] dataset with 53 FLAIR MRI scans coming from 4 different scanners. To evaluate their synthetic lesion generation pipeline, the authors first trained a lesion segmentation model on scans containing both real and synthetic lesions. They then repeated the process using images that contained only synthetic lesions. DSC dropped from 56 % to 52 %, demonstrating that synthetic lesions alone can deliver comparable performance. nDSC and lesion-wise metrics were not computed.

This pipeline allows for a very fast algorithm and a low computational cost. Despite the low similarity between real lesions and the fake ones, a type of data

augmentation such as the one proposed in this pipeline provides better results than no data augmentation at all.[70][67]

MS-MIST

MS-MIST [72] enables the generation of synthetic lesions in healthy brain images based on a specified lesion load. Lesions are sampled from a lesion database, and a new binary lesion mask is generated. This database is constructed by extracting lesions from existing pathological MRI scans registered in the MNI space. When the lesion masks are applied to healthy brain MRIs, they are transformed back into each subject’s native space. A filtering step removes any lesions incorrectly positioned in regions such as the ventricles or cortical areas, ensuring that only lesions within the WM are retained. Lesion texture is simulated using Gaussian convolution, which creates smooth transitions from lesion centers to their borders. Additionally, a random component drawn from a Gaussian distribution is incorporated to mimic the natural variability in lesion texture.

The methodology presented in the paper relies on a database derived from 82 MS patients, 52 patients diagnosed with secondary progressive MS (SPMS) from a local hospital and 30 patients from a MICCAI2016[62] and MICCAI2018 database (<http://wmh.isi.uu.nl/>, and <http://www.ia.unc.edu/MSseg/>, respectively).

To evaluate their technique, the authors asked radiologists to validate if synthetic lesions exhibit realistic transitions with the surrounding white matter, which was the case. For a more quantitative evaluation, the correlation between simulated lesions and their automatic segmentation by the Lesion Segmentation Tool (LST) was evaluated. Correlation was in general high between the segmentation results and the synthetic lesions ($R=0.977$). The best results were obtained for a high lesion’s load, with a DSC of 0.897, a specificity of 0.999, a sensitivity of 0.9953 and an accuracy of 0.9984, voxel-wise.[72] nDSC and lesion-wise metrics were not computed.

The main advantages of the MS-MIST approach include the realistic variability of simulated lesions, compatibility with multiple MRI modalities, and its availability as an open-source resource.[72] However, the methodology presents some limitations [72]: The lesion database currently consists exclusively of lesions from SPMS patients, potentially limiting generalizability, detailed texture modeling is absent, and the synthetic lesion appearance does not vary while changing MRI modalities.[72] Moreover, the method is suited to add lesions only on healthy brains scans, which can present a limitation.

Other techniques also exist for generating synthetic lesions, such as using a denoising diffusion implicit model (DDIM) [73], but we will not provide further explanation here.

1.4 Texture Analysis/Generation

1.4.1 Medical image texture analysis

Texture refers to the appearance of an object and the spatial arrangement of elements within it. In clinical data, analyzing texture can reveal the positioning of features within an image. Most techniques use texture analysis to classify regions of interest, for instance distinguishing healthy tissue from lesions or identifying different anatomical areas. In MRI, texture analysis examines variations in grey-level intensities throughout the image.[74]

A commonly used tool for texture analysis is the Gray-Level Co-occurrence Matrix (GLCM), which enables the extraction of information about the intensity distribution of pixel pairs within an image.

The GLCM is particularly useful in texture analysis because it examines how pairs of grey levels spatially relate and how these intensities are distributed throughout the image. The more dispersed these values are, the more complex or random the texture becomes. [74]

There exist many others tools for texture analysis such as the grey-level histogram, the run-length matrix, auto-regressive models or the wavelet transforms,...

1.4.2 MS lesion texture analysis

Texture analysis of MS lesions plays an essential role in quantifying changes that are not detectable by conventional MRI techniques.[75] It enables the identification of subtle tissue alterations, thereby supporting earlier diagnosis of the disease.[76] By using a GLCM, comparisons can be made between pathological and healthy brain tissue. Structural differences between the lesion white matter (i.e. hyperintensities associated with lesions in the white matter), normal appearing white matter (white matter surrounding lesions that appears normal on MRI but may already exhibit microstructural abnormalities), and normal white matter can be highlight. These distinctions can reveal early signs of disease progression.[75]

1.4.3 Texture generation with an High Order Singular Value Decomposition

One method for generating new dynamical textures is through the use of High Order Singular Value Decomposition (HOSVD)[77]. While traditional Singular Value Decomposition (SVD) is applied to matrices, HOSVD operates on tensors, making it suitable for decomposing video frames or matrices with more than two dimensions. Unlike conventional SVD, HOSVD preserves the correlations across

the multiple dimensions of the data. A key limitation of HOSVD is the absence of optimal decomposition. In contrast, standard SVD achieves the best solution in term of least square error by truncating the first n columns, but this optimality does not hold in higher-dimensional settings. However, in practice, the results produced by HOSVD are often very close to the optimal solution. Similarly to traditional SVD, the main advantage of using an HOSVD is to reduce the noise by truncating the higher dimension.[77]

1.4.4 Texture generation with THGOP

TGHOP is a technique used to generate texture, allowing the creation of entirely new synthetic textures derived from patches of existing ones.[78]

Fine-to-coarse analysis

The process begins with the patches extraction from an original texture image. These patches are represented as vector in an high dimensional space, denoted as S_0 .

Now, it is needed to reduce the space of the vector. To do so, a Subspace Approximation with Adjusted Bias (SAAB) transform, which is a variant of PCA, is employed. It retains components associated with the highest energy, which correspond to the most significant features of the image. This transformation is iteratively applied to reduce the space until a minimal space is reached. The result is a condensation of the most important information associate to the image, conserving both global and local texture patterns.

Core sample generation

When the lower space is reached, the tensors derived from the patches are clustered using the K-means algorithm. Then, a PCA is applied to the clusters to whitened them followed by an ICA to ensure statistical independence among the clusters. A Gaussian distribution is fitted to each clusters in order to model the data distribution. New samples are randomly drawn from these distributions, preserving the statistical characteristics of the original texture.

To ensure the generated samples maintain a similar visual appearance to the original image, histogram matching is applied to align the new samples histograms with those from the original image.

Coarse-to-fine analysis

To reconstruct the data in its original high-dimensional form, inverse SAAB transforms are applied to the newly generated samples.

Then, the generated patches are assembled together to form a complete textured image. To minimize visible artifacts at the patch borders, techniques such as patch blending can be applied.

The entire pipeline of the method is the following :

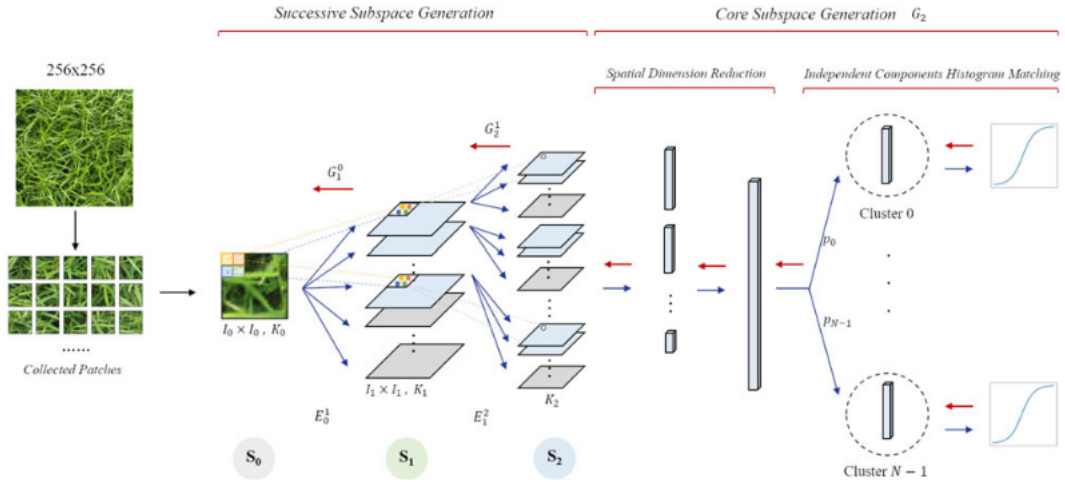


Figure 1.11: Adapted from [78]: pipeline of TGHOP, S_n being the reduced space, E and G respectively the forward and backward transformations.

Chapter 2

Methodology and implementation

In this section, a method designed to enhance the performance of a segmentation's algorithm is presented. As previously said, data augmentation is a powerful tool to improve an algorithm's generalizability, particularly in medical image segmentation, where data are lacking. We begin by providing an intuitive overview of the approach, followed by detailed subsections describing our proposed method for generating synthetic lesions. Finally, we explain how the method is evaluated.

2.1 Proposed method to generate synthetic lesion

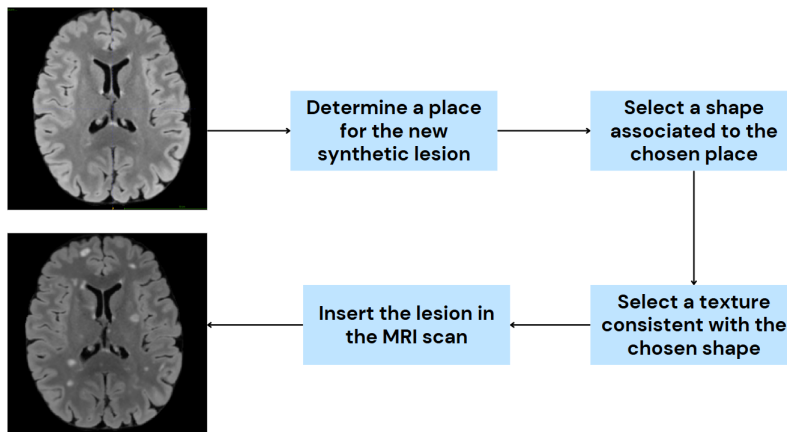


Figure 2.1: Intuitive idea.

As seen in the MS description section 1.1, MS lesions can appear everywhere in the WM and are very heterogeneous in terms of shape and texture. When creating

synthetic lesions on an MRI image, certain questions may then arise :

- *Where should the lesion be placed?*
- *Which shape should the lesion have?*
- *What texture should the lesion have?*

An intuitive understanding of the methodology is obtained by answering these questions.

Where should the lesion be placed? MS lesions can occur anywhere within the WM, but some regions are far more likely than others (cf. subsection 1.1.1). To model this, we first derive a “lesion-occurrence” probability map from an MS population dataset. During synthesis, the lesion centroid is randomly sampled according to the probability map, ensuring that the generated lesion is placed in anatomically relevant locations.

The second question, i.e. **Which shape should the lesion have?**, is tricky due to the MS lesion shape heterogeneity (cf. subsection 1.1.1). Even though MS lesion shape is typically ovoid following its formation around veins, it can be influenced by external anatomical features like cortex or ventricles. To address this, we draw pre-existing lesion shapes coming from the dataset, each associated with its anatomical region, and corrected by external anatomical masks. This approach guarantees that our model reflects the diverse lesion morphologies found across different brain regions (cf. subsection 1.1.1).

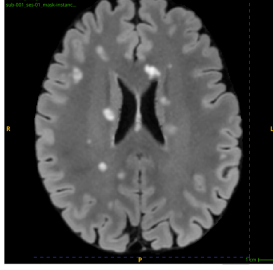
Finally, **What texture should the lesion have?** Again, that question might be difficult to answer due to the heterogeneity of MS lesion texture resulting from different pathological processes. To synthesize realistic new lesion textures, we implemented a texture generation algorithm that takes as input existing MS lesion textures, ensuring consistency with real lesion appearance.

When all these questions are answered, everything can be put back together and new lesions are generated in four simple steps:

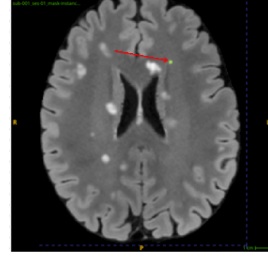
- **Placement** : A centroid is randomly drawn from the lesion occurrence map
- **Shape selection** : Following the placement, a shape is randomly drawn from the library corresponding to that region
- **Texture assignment** : Based on lesion’s volume, a synthetic texture is generated

- **Blurring** : The lesion is inserted in the MRI scan and blurred to ensure a realistic incorporation.

Figure 2.2 illustrates synthetic lesions, by highlighting the only real lesion present in the MRI scan.



(a) MRI scan with augmented lesions.



(b) Highlight of the only real lesion present in the image.

Figure 2.2: Synthetic lesion generation.

2.1.1 Lesion placement

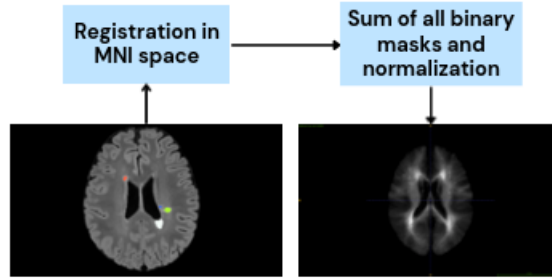
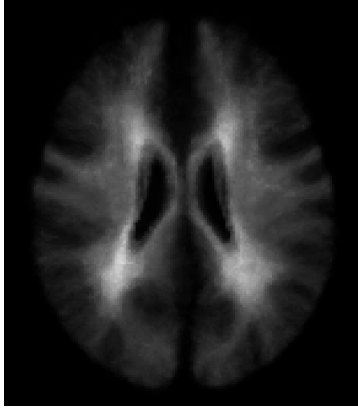


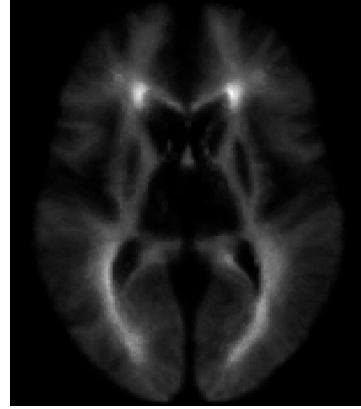
Figure 2.3: Pipeline for the likelihood map creation.

First, to account for the typical locations where lesions occur, a probability map is generated, which associates each anatomical location with a probability value, representing the likelihood of finding a lesion at a specific location[68]. To generate this map, we used binary lesion masks, coming from expert manual annotations, from two distinct sources: our private dataset and the MS-MIST dataset, both described at the end of this section (subsection 2.2.2). These binary masks were registered using ANTs [79] into the MNI152 space (i.e. a normalized space created with an average of 152 healthy young adult brains[80]), enabling a comparison between them. Masks were summed, and a normalization step was applied to scale the value between 0 and 1. This resulted in a probability map in the MNI space

where each voxel value indicates the likelihood of a lesion occurring at that location (cf. Figure 2.3). Since lesion can appear throughout the whole WM and due to the small size of our training set, a small probability (i.e. 20%) to find lesion there was added. Figure 2.4 shows an axial view of the lesion-occurrence map at two different level.



(a) Likelihood map at the level of the top first and second ventricle.



(b) Likelihood map at the level of the thalamus.

Figure 2.4: Axial view of the likelihood map at two different level.

2.1.2 Lesion shape

To define the shape of newly synthesized lesions, a dictionary of lesion shapes was created. The shapes included in this dictionary originate from two separate sources: our private dataset and the MS-MIST dataset, both described at the end of this section (subsection 2.2.2).

Since lesion morphology can vary depending on anatomical location (subsection 1.1.1), each shape in the dictionary is associated with specific regions sampled from the previously described likelihood map. Each selected likelihood point represent a key associated with the shapes of the closest lesions, drawn from the training database and previously registered in the MNI space.

These shapes comes from two separated sources: firstly from the instance masks in our dataset, and secondly from the [50-100mL] and [100-500mL] (i.e. smallest lesions) MS-MIST lesion database[72]. Since this database is using lesion's load binary masks, several shapes were extracted from a mask. For the smaller lesion load database (i.e., [50–100 mL]), an erosion operation is applied to ensure the separation if there are closely adjacent lesions, followed by a dilation to recover

a smoother approximation of the original shape. For the [100–500 mL] database, lesion convexity is also assessed by retaining only the lesions with the highest ratio between the lesion area and the area of its surrounding convex polyhedron, trying to avoid the presence of two merged lesions.

Figure 2.5 is an example of MS lesion mask present in the MS-MIST database and its separation to integrate lesions in our database:

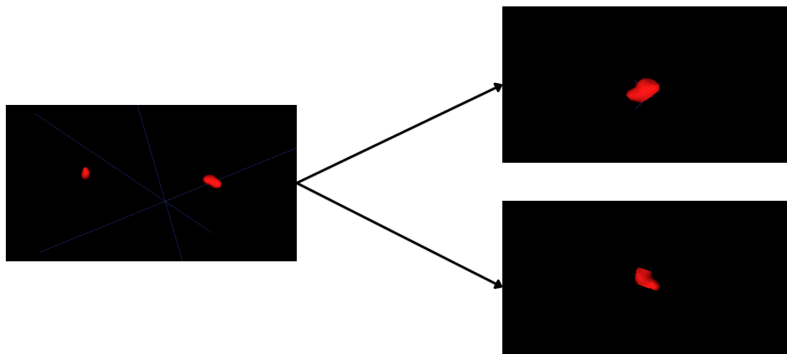


Figure 2.5: Lesion volume separation, with the first mask coming from MS-MIST[72] database.

2.1.3 Lesion texture and intensity

In this part of the method, each lesion coming from our dataset, described at the end of this section (subsection 2.2.2), is processed individually through a two-step pipeline:

First, the lesion intensity is analyzed to characterize the contrast between the lesion and the surrounding WM. This analysis yields a ratio representing the proportional relationship between the lesion and the adjacent WM intensity.

Next, to generate a new texture coming from this lesion, a texture patch is extracted from it. This patch is first denoised using an HOSVD (see subsection 1.4.3). The denoised patch is then passed through the TGHOP algorithm (see subsection 1.4.4) to generate a new texture based on the original one.

The newly generated texture patch is then associated with the previously computed intensity ratio and categorized based on the size of the lesion from which the patch is extracted. If the lesion is small, the patch is assigned to the first group (i.e. small lesion’s texture group); otherwise, it is placed in the second group (i.e. big lesion’s texture group). These groups will later be used during the

reconstruction of synthetic lesions to ensure consistency between the lesion shape’s size and texture/intensity.

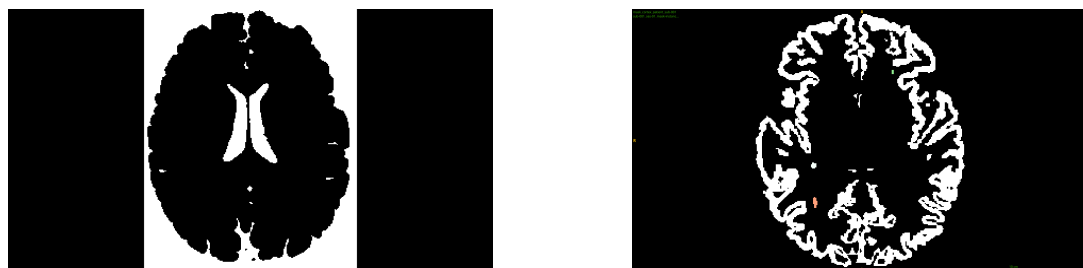
2.1.4 Generation of a synthetic lesion

Finally, once the likelihood map, shape dictionary, and the two lesion texture groups have been established, a new synthetic lesion can be generated.

First, a position is selected within the likelihood map in the MNI space. Based on the area of this position, a shape is chosen from the shape dictionary. The lesion shape is then placed at the chosen position.

To ensure that lesion placement is anatomically plausible, exclusion masks are registered and then applied in MNI space to prevent synthetic lesion generation in three specific regions: the ventricles, the cortex (only juxtacortical lesions were generated for simplicity), and areas with existing lesions. The ventricles and CSF were segmented using an Otsu threshold, leveraging their low intensity on FLAIR images (cf. subsection 1.1.1). Cortical masks were obtained via Freesurfer [81]. Each image in the dataset is associated with both a lesion mask and a cortex mask. Finally, to avoid overlapping lesions (since we are generating synthetic lesions on pathological MRI scans), pre-existing lesions were excluded using the pre-computed lesion binary mask. Although lesions can naturally grow close together, directly superimposing a new synthetic lesion on an existing one is not pathologically relevant.

Figure 2.6 are examples of masks used as avoiding area for each image:



(a) Otsu generated avoiding mask, parts in white are avoided.

(b) Cortex mask generated with Freesurfer [82] in white and in colors a lesion’s mask, used as avoiding mask while generating synthetic lesions.

Figure 2.6: Avoiding masks.

After the correct placement of the new synthetic lesion, the generated binary mask is registered back into the patient’s space.

A texture patch is randomly selected from the appropriate group, depending on

whether the chosen shape is classified as large or small. The patch is then resampled to match the size of the synthetic lesion. Its intensity is adjusted using the lesion-to-WM intensity ratio, ensuring consistent lesion brightness within the image. Finally, the lesion is inserted into the original MRI image and blurred using a Gaussian filter to produce a realistic and smooth result. The amount of blurring, with σ ranging from 0.7 to 1.8, is adapted based on several factors: lesion intensity (i.e., higher intensity leads to stronger blurring), lesion placement (i.e., blurring is reduced when the lesion is close to a brain structure to preserve realism), and lesion size (i.e., smaller lesions receive less blurring).

2.2 Experiments

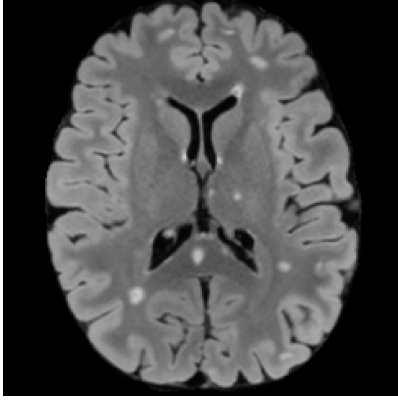
To evaluate our proposed synthetic lesion generation method, we trained a deep learning automatic segmentation method, namely nn-UNet [43], to segment images with and without synthetic lesions, in order to determine whether or not the addition of synthetic lesions improves segmentation performance on validation and test sets.

To do so, we first trained nnU-Net on a dataset without synthetic lesions, comprising 63 subjects: 50 used for training and 13 for testing. During training, we employed 5-fold cross-validation, where each fold consisted of 40 subjects for training and 10 for validation.

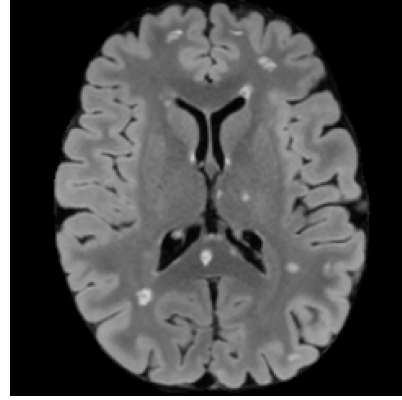
For the experiment involving synthetic lesions, the same dataset and cross-validation split were retained. In this case, the training set was augmented with images containing synthetic lesions, increasing the number of MRI scans to 90 per fold. Within each fold, 10 subjects without synthetic lesions were reserved for validation, while the remaining 40 subjects (each with both their original and synthetic-lesion-augmented MRI scans, which gives 80 images in total) were used for training, resulting in a mix of real and augmented data.

To understand the impact of each hyper parameters of the synthetic lesions on the segmentation’s accuracy, several experiments were run :

- **Blurring:** comparing high versus low levels of Gaussian blur on bigger lesions



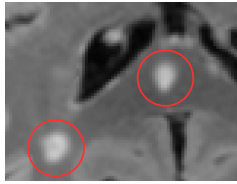
(a) $0.7 \leq \sigma_{\text{small lesion}} \leq 1.1$ and $1.2 \leq \sigma_{\text{big lesion}} \leq 1.8$, following the intensity and the location of the lesion.



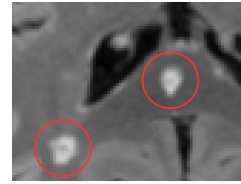
(b) $0.7 \leq \sigma_{\text{small lesion}} \leq 1.1$ and $1 \leq \sigma_{\text{big lesion}} \leq 1.5$, following the intensity and the location of the lesion.

Figure 2.7: Comparison between high blur and low blur lesion on the MRI scan.

The experiment was conducted by varying the blurring on bigger lesions.



(a) Example of high blur lesions.



(b) Example of low blur lesions.

Figure 2.8: Comparison between high blur and low blur lesion.

- **Subject type:** Comparing synthetic lesions added on healthy subjects scans versus applied on pathological scans
- **Lesion's number :** Evaluating segmentation accuracy while adding 50, 100, 150, 300, or 600 synthetic lesions

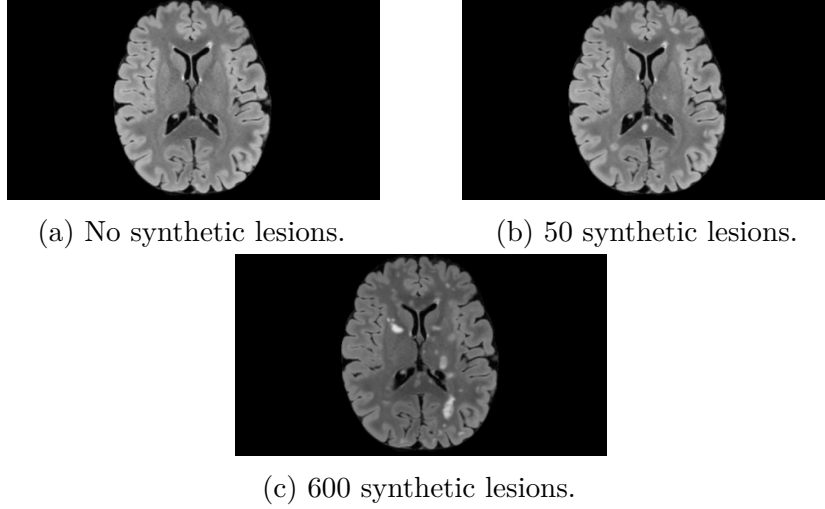


Figure 2.9: Comparison between a low load and an high load of lesion.

When the best configuration has been found following the results on the validation set, results on the test set in-domain were assessed. To understand further the impact of these synthetic lesions, the two model (i.e. trained with and without synthetic lesions) were assessed on external tests sets, MSLesSeg[83] and MSShifts[84], to see their accuracy out-of-domain. We used DSC and nDSC (cf. subsection 1.2.2) to evaluate the segmentation performance, these metrics being the most appropriate for semantic-segmentation evaluation.[46]

2.2.1 Statistical analysis

To assess if the improvement in DSC/nDSC is significant, the Wilcoxon signed rank test [85] is applied. This non-parametric test is suited for paired data and has the advantage of requiring fewer assumptions than the paired t-test, in particular regarding the normality of the data distribution. In the in-domain test set, only 13 MRI scans are available, which is insufficient to assume normality. In this context, the Wilcoxon test is more appropriate than a t-test.

For the out-of-domain test set, which contains more data, a paired t-test is applied, as the assumption of normality is more likely to hold due to the larger sample size. [86]

2.2.2 Material

Dataset

Dataset contains 69 brain MRI examinations acquired at Saint-Luc University Hospital in Brussels between January 2021 and April 2023, but only 63 MS patients (40 women, 23 men) and 4 healthy subjects were retained. It comprises, according to the 2017 McDonald criteria, 45 relapsing-remitting MS, 11 secondary-progressive MS and 7 primary-progressive MS MRI scans. Patient’s mean age was 42.4 years (SD 12.1; median 42.0; range 21.5–67.3), mean EDSS score 2.7 (SD 2.0; median 2.0; range 0–7) and mean disease duration 7.4 years (SD 8.0; median 5.9; range 0–33.3). The 4 healthy subjects scans were used to validated our pipeline.

To complete our lesion dictionary, we incorporated the MS-MIST lesion database (cf. Figure 1.3.2) [72] (<https://github.com/CSIM-Toolkits/Slicer-LesionSimulatorExtension>). We retained only masks in the [50–100] and [100–500] volume folders. The final MS-MIST cohort consisted of 82 MS patients in total: 52 secondary-progressive cases under long-term treatment (≥ 10 years) recruited at Ribeirão Preto Medical School (University of São Paulo, Brazil), plus 30 cases drawn from the publicly available MICCAI 2016 [62] and MICCAI 2018 [87] challenge datasets [72].

In order to evaluate our method on out-of-domain data, two external datasets were used: MSLesSeg [83] and MSShift [84].

MSLesSeg includes 75 subjects, composed of 48 women and 27 men, with a mean age of 37 years (± 10.3 years). Each subject has between one and four longitudinal MRI scans. The distribution of MS phenotypes within this dataset was not specified in the original publication.

MSShift is a composite dataset comprising images from three different sources: the ISBI [58], MSSEG-1 [62], and PubMRI[88] databases. The ISBI dataset contains 21 patients acquired in Best, Netherlands, covering various MS phenotypes such as RRMS, PPMS, and SPMS. The MSSEG-1 dataset includes 52 patients, with 15 scans acquired in Rennes, 8 in Bordeaux, and 29 in Lyon, France, although the MS phenotype distribution was not reported. Finally, the PubMRI dataset consists of 25 patients from Ljubljana, Slovenia, and includes various phenotypes such as RRMS, SPMS, PPMS, and CIS.

Imaging details

Scans coming from our in-domain dataset were acquired on a SIGNA™ Premier 3 T whole-body MRI system (GE Healthcare, Milwaukee, WI) using a 48-channel head coil. The protocol consisted of three sagittal 3D acquisitions: first, a T1-weighted

MPRAGE sequence (TR = 2186 ms, TE = 3 ms, TI = 900 ms; FOV = 256×256 mm²; 156 slices; 1 mm isotropic voxels), then a 3D FLAIR sequence (TR = 500 ms, TE = 105 ms, TI = 1532 ms; FOV = 256×256 mm²; 170 slices; 1 mm isotropic voxels), and finally a high-resolution 3D EPI scan (TR = 80.2 ms, TE = 35 ms, flip angle = 18°; FOV = 256×256 mm²; 355 slices; 0.67 mm isotropic voxels). All data were managed and organized with the BIDS Managing and Analysis Tool (BMAT) [89]. In ANTs [79], the 3D FLAIR images were first rigidly registered to the EPI magnitude images, the MPRAGE images were then aligned to these transformed FLAIRs, and brain masks extracted from the FLAIR scans, using synthstrip [90], were applied to crop every image from the dataset.

The MSLesSeg database is composed of T1-weighted, T2-weighted, and FLAIR MRI scans for each subject at every time point. MRI scans were acquired using various MRI scanners. All images underwent the same preprocessing pipeline, starting with co-registration to the 1 mm³ isotropic MNI152 template using FMRIB’s Linear Image Registration Tool (FLIRT) [91]. Brain extraction was then performed using the Brain Extraction Tool (BET) [92].

In the ISBI dataset, all MRI scans were acquired using a 3T Philips Medical scanner, with a native resolution of $0.82 \times 0.82 \times 2.20$ mm³ for the FLAIR scans. In the MSSEG-1 dataset, FLAIR scans from Rennes were acquired using a 3T Siemens Verio scanner, with a resolution of $0.5 \times 0.5 \times 1.1$ mm³. FLAIR scans from Bordeaux were acquired using two scanners: a 3T GE Discovery scanner ($0.47 \times 0.47 \times 0.9$ mm³) and a 1.5T Siemens Aera scanner ($1.03 \times 1.03 \times 1.25$ mm³). FLAIR scans from Lyon were acquired using a 3T Philips Ingenia scanner, with a resolution of $0.74 \times 0.74 \times 0.70$ mm³. Finally, the PubMRI dataset includes FLAIR scans acquired using a 3T Siemens Magnetom Trio scanner, with a resolution of $0.47 \times 0.47 \times 0.80$ mm³.

All MRI scans in the MSShift database (T1-w and FLAIR) were preprocessed using the same pipeline, which included denoising, skull stripping, bias field correction, and interpolation to 1 mm isotropic voxel space.

Manual segmentation

Lesions from our in-domain dataset were manually segmented on cross-sectional 3D FLAIR images by a trained neurobiologist and subsequently reviewed by an experienced neurologist, with 3D MPRAGE and EPI phase images serving as anatomical references. Infratentorial and cortical lesions, as well as any lesions under 3 mm in diameter, were excluded. All annotations and visualizations were carried out in ITK-SNAP[93]. In total, 1 104 lesions were delineated, corresponding to a mean of 15.4 lesions per subject (median 14.5; range 5–60; standard deviation 8.7). In average, a participant has 5 162.5 mm³ of lesion volume (median 3 714.0

mm³; range 162.0–33 675.0 mm³; standard deviation 5 876.7 mm³), and the average volume for one lesion is 112.6 mm³.

For the MS-MIST dataset, expert radiologists manually delineated lesions, which were then fused using STAPLE to produce a single reference mask. All the masks were registered to MNI152 space (affine + B-spline via BRAINSFit[94]). The resulting lesion database comprised 2,229 individual lesion volumes, categorized into five ranges (50–100, 100–500, 500–1000, 1000–5000, > 5000 mL), but only lesions from [50-100 mL] and [100-500 mL] were kept.

Lesions from the MSLesSeg database were manually segmented by three experts using FLAIR images registered to the 1 mm³ isotropic MNI template, with reference to the T2-w and T1-w modalities. One of the three experts was a junior rater trained by the two others: a senior neuroradiologist and a senior neurologist. The junior expert performed the initial annotation of all images using the JIM9 segmentation tool (<https://xinapse.com/j-im-9-software/>), after which the two senior experts reviewed the segmentations. When a consensus was reached, the resulting mask was defined as the ground truth.

In the MSShift dataset, ground truth segmentations vary depending on the source. For the MSSEG-1 dataset, manual segmentation masks were generated through consensus among seven raters. For the ISBI dataset, the ground truth was based on the agreement of two independent raters. For the PubMRI dataset, three experts discussed and agreed upon the final binary lesion segmentations.

Chapter 3

Results

In this section, results of the proposed method are presented. As a reminder, synthetic lesions produced with our method are inserted in original MRI scans and add to our training dataset, then the segmentation accuracy is evaluated on a validation/test set.

3.1 Assessing the best synthetic lesion’s parameters

In this section, several configurations for synthetic lesions are evaluated by training nnU-Net (cf. subsection 1.2.1) with and without synthetic lesions (cf. section 2.2) and testing on the validation set to determine the optimal configuration. Results are assessed using the DSC and nDSC metrics.

3.1.1 Lesion’s blurring

The first hyperparameter analyzed is the lesion blurring used when inserted in the pathological MRI scan.

	DSC	nDSC
No synthetic lesions added	0.67256 ± 0.0449	0.71243 ± 0.0448
Low blurring for bigger lesion	0.66972 ± 0.056	0.71243 ± 0.065
High blurring for bigger lesion	0.67704 ± 0.042	0.71790 ± 0.044

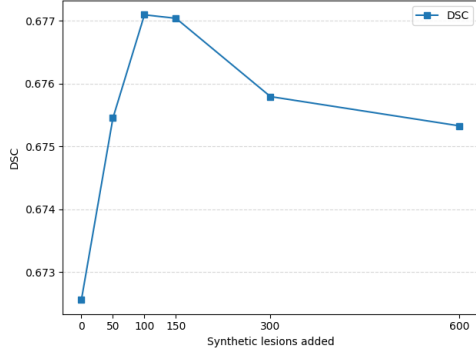
Table 3.1: Mean DSC and nDSC for the 5 validation sets, when trained with synthetic lesions (low blur and high blur) or without synthetic lesions.

In the *low blurring* setting, smaller lesions were blurred with a σ in the range $0.7 \leq \sigma \leq 1.1$ while *larger lesions* used σ values between $1 \leq \sigma \leq 1.5$. The exact value of σ is based on lesion intensity (higher intensities led to higher σ values) and its proximity to anatomical structures (lower σ values were used when the lesion was close to brain structures to preserve realism). Similarly, in the *high blurring* experiment, larger lesion sigma is comprised between $1.2 \leq \sigma \leq 1.8$.

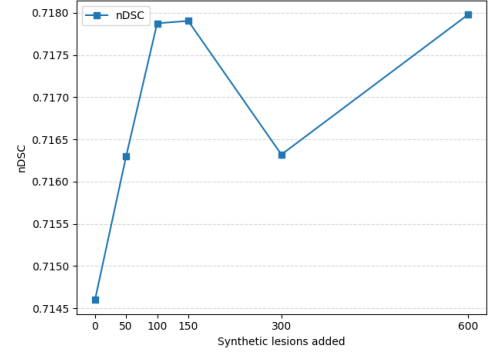
Based on the results obtained in Table 3.1, increasing the blurring in bigger lesions improve the DSC and the nDSC compared to the case without synthetic lesions. Conversely, applying low blurring to larger lesions results in a decrease in both DSC and nDSC relative to the baseline.

3.1.2 Lesion's number

Lesion load and its impact on the DSC and nDSC were then evaluated using the optimal blurring parameter identified in the previous section (i.e $0.7 \leq \sigma_{\text{small lesion}} \leq 1.1$ and $1.2 \leq \sigma_{\text{big lesion}} \leq 1.8$). To assess the performance of nnU-Net, 50, 100, 150, 300, and 600 synthetic lesions were added to pathological MRI scans and compared to the baseline condition (i.e., *0 synthetic lesions added* in Figure 3.1, means 0 synthetic lesions added in the training set).



(a) Mean DSC as a function of lesion's number.



(b) Mean nDSC as a function of lesion's number.

Figure 3.1: DSC and nDSC results as a function of the lesion's number.

Based on these results, an higher lesion load does not lead to better performance. Instead, there appears to be an optimal range, with the best DSC results obtained with approximately 100-150 synthetic lesions added on MRI pathological scans.

3.1.3 Healthy images augmentation

Another hyperparameter explored was the generation of synthetic lesions on healthy images, rather than on pathological ones as done previously. To this end, four healthy control brain MRI scans were augmented by adding 150 synthetic lesions using the best blurring parameter (i.e $0.7 \leq \sigma_{\text{small lesion}} \leq 1.1$ and $1.2 \leq \sigma_{\text{big lesion}} \leq 1.8$). To maintain the same training set size (90 images), multiple lesion placement configurations were generated for each healthy image. For each fold, 10 pathological images without synthetic lesions were retained, while the remaining 80 images consisted of a mix of real pathological images and synthetic pathological images generated by placing synthetic lesions on healthy subjects.

	DSC	nDSC
No synthetic lesions added	0.67256 ± 0.0449	0.71243 ± 0.0448
150 synthetic lesions added on pathological images	0.67704 ± 0.042	0.71790 ± 0.044
150 synthetic lesions added on healthy images	0.67052 ± 0.0524	0.71233 ± 0.053

Table 3.2: Mean DSC and nDSC for the 5 validation sets, when trained with synthetic lesions added on pathological or healthy images or without synthetic lesions.

Table 3.2 shows that adding lesions to pathological images leads to better performance than adding them to healthy control images.

3.2 Results on the test set

In this section, the optimal parameters for synthetic lesions, identified in the previous section using the validation set, are used to evaluate segmentation accuracy on test sets, with synthetic lesions added to the in-domain training set. The best configuration parameters, as mentioned before, are: a Gaussian blurring with σ values of $0.7 \leq \sigma_{\text{small lesion}} \leq 1.1$ for small lesions and $1.2 \leq \sigma_{\text{big lesion}} \leq 1.8$ for large lesions, the addition of 100 synthetic lesions, and the use of pathological MRI scans. The results are assessed using DSC, nDSC and lesion-wise metrics (i.e LTPR and LFPR). Table 3.3 shows all the DSC, nDSC and lesion-wise metrics associated to the segmentation of test set in-domain and out-of domain.

Dataset	DSC		nDSC		LTPR		LFPR	
	Baseline	SynLes	Baseline	SynLes	Baseline	SynLes	Baseline	SynLes
In-domain	0.692 ± 0.12	0.704 ± 0.11	0.713 ± 0.12	0.727 ± 0.12	0.719 ± 0.19	0.722 ± 0.18	0.284 ± 0.15	$0.25 \pm 0.16(*)$
MSLesSeg	0.437 ± 0.18	$0.448 \pm 0.18(**)$	0.442 ± 0.18	$0.452 \pm 0.18(**)$	0.424 ± 0.19	$0.388 \pm 0.18(***)$	0.08 ± 0.12	0.089 ± 0.14
MSShift	0.35 ± 0.15	$0.307 \pm 0.16(***)$	0.354 ± 0.15	$0.311 \pm 0.16(***)$	0.355 ± 0.21	$0.337 \pm 0.19(**)$	0.044 ± 0.09	0.065 ± 0.14

Table 3.3: Results obtained when testing both model (*Baseline* is the model without synthetic lesions and *SynLes* the model trained with synthetic lesions) on the test set in domain, MSLesSeg and MSShift. (*) $p < 0.05$, (**) $p < 0.01$, (***) $p < 0.001$.

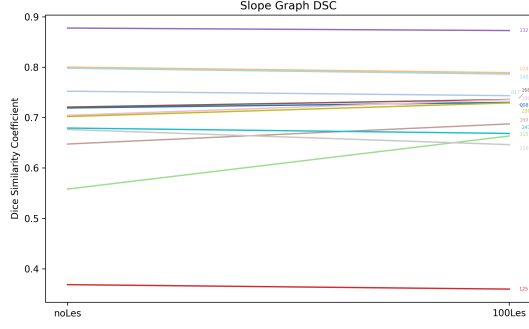
3.2.1 Results on the test set in domain

Here, segmentation performance on the test set in domain when adding synthetic lesions in the training set is assessed.

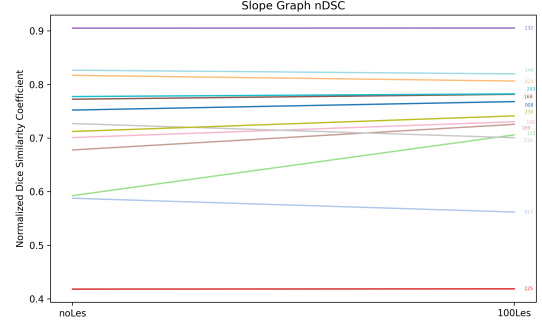
Table 3.4 shows the DSC, nDSC results on the test set.

DSC		nDSC		LTPR		LFPR	
Baseline	SynLes	Baseline	SynLes	Baseline	SynLes	Baseline	SynLes
0.692 ± 0.12	0.704 ± 0.11	0.713 ± 0.12	0.727 ± 0.12	0.719 ± 0.19	0.722 ± 0.18	0.284 ± 0.15	0.25 ± 0.16

Table 3.4: DSC and nDSC associated to the in-domain test set segmentation, with and without synthetic lesions (*Baseline* is the model without synthetic lesions and *SynLes* the model trained with synthetic lesions) added in the training set.



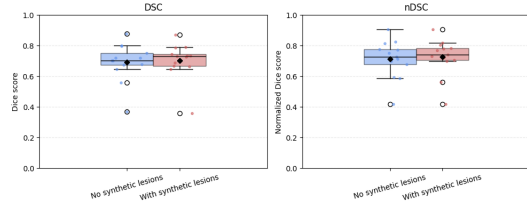
(a) Slope graph comparing the DSC associated to the in-domain test set segmentation with or without synthetic lesion in the training phase.



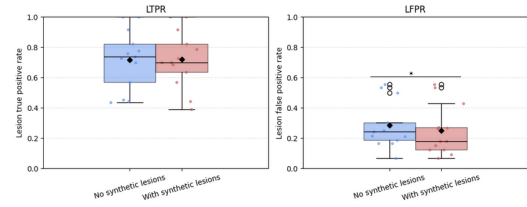
(b) Slope graph comparing the DSC/nDSC associated to the in-domain test set segmentation for the case without synthetic lesions on the left, and with synthetic lesions in the training set on the right.

Figure 3.2: Comparison of DSC (left) and nDSC (right) for in-domain test set segmentation with or without synthetic lesions in the training phase.

Even though the use of a training set containing synthetic lesions increases the DSC and the nDSC, the difference is not statistically significant according to the Wilcoxon test. For the lesion-wise metrics, improvement in LFPR with the use of synthetic lesions in the training set is significant, which means that the segmentation’s precision increase when synthetic lesions are used. Figure 3.3 shows the differences in DSC, nDSC, LTPR, and LFPR on the test set when training nnU-Net with or without synthetic lesions.



(a) Boxplots representing the DSC and nDSC metrics on in-domain the test set while using synthetic lesions (in red) or not in the training set.



(b) Boxplots representing the LTPR and LFPR metrics on the in-domain test set while using synthetic lesions (in red) or not in the training set.

Figure 3.3: Boxplots for the DSC, nDSC and lesion-wise metrics associated to the in-domain test set segmentation.

To further investigate whether adding synthetic lesions to the training set leads better segmentation results on test images with high or low lesion load, several

analyses were conducted. First, Figure 3.4 presents the difference in DSC and nDSC on the test set when training with or without synthetic lesions, as a function of the total lesion load in the segmented MRI scans. Since the p-values associated with the regression lines are all greater than 0.05, there is no statistically significant correlation between segmentation performance and lesion load, regardless of whether synthetic lesions were used during training. However, we can observe a tendency toward decreased segmentation performance on images with higher lesion load.

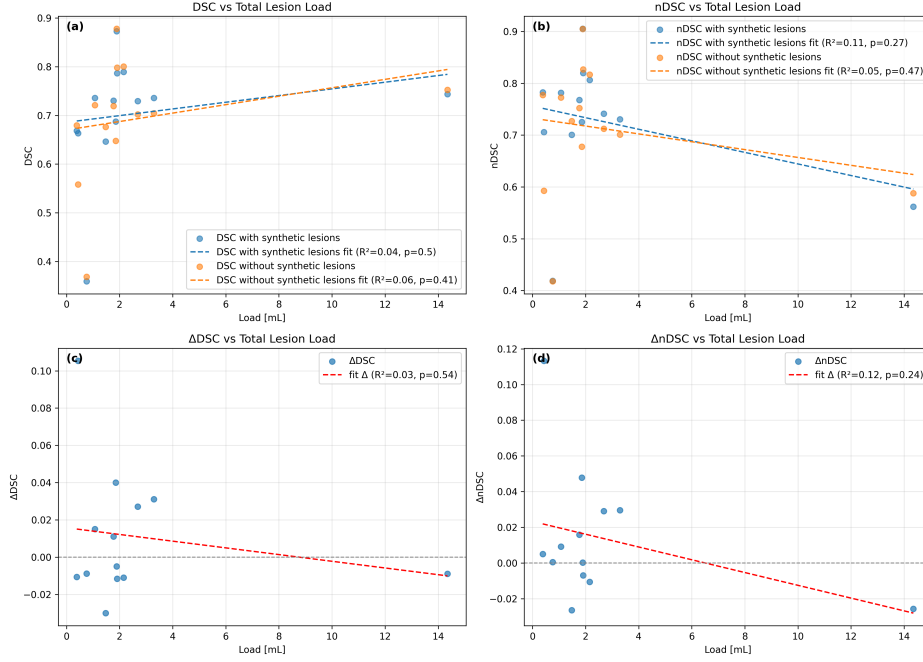


Figure 3.4: (a) DSC obtained for the segmentation of our test set in-domain as a function of lesion load, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of our test set in-domain as a function of lesion load, with and without synthetic lesions added to the training set. (c) Difference in DSC obtained on the test set in-domain between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC obtained on the test set in-domain between the model trained with synthetic lesions and the model trained without.

The same analysis was conducted, this time considering the number of lesions (Figure 3.5). Again, the p-values associated with the regression lines are all greater than 0.05, indicating that there is no significant correlation between a high number of lesions and any change in DSC or nDSC, but we can observe a tendency toward a decreased in performance segmentation on images with higher lesion number.

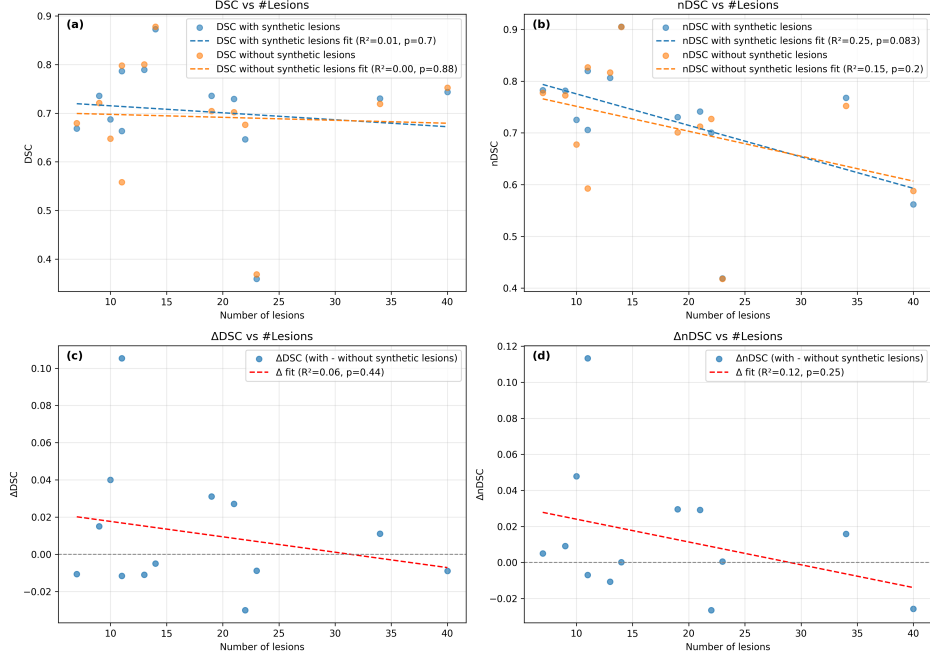


Figure 3.5: (a) DSC obtained for the segmentation of our test set in-domain as a function of the lesion’s number, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of the in-domain test set as a function of lesion’s number, with and without synthetic lesions added to the training set. (c) Difference in DSC obtained on the test set in domain between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC obtained on the test set in domain between the model trained with synthetic lesions and the model trained without.

3.2.2 Results on the test set out-of-domain

In this subsection, segmentation’s performance of nnU-Net on test sets out-of-domain when synthetic lesion’s are added in our training set in-domain are assessed.

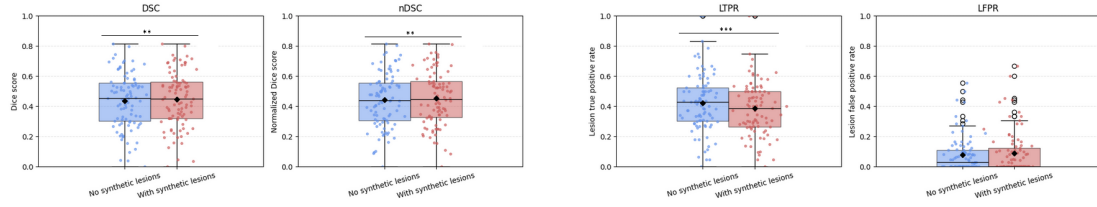
Results with the MSLesSeg database[83]

The first out-of-domain dataset used to evaluate the model is MSLesSeg. Images from both the training and testing sets were combined to increase the size of the test dataset. Table 3.5 presents the DSC, nDSC, LTPR and LFPR metrics associated with the segmentation performance on this dataset.

DSC		nDSC		LTPR		LFPR	
Baseline	SynLes	Baseline	SynLes	Baseline	SynLes	Baseline	SynLes
0.437 ± 0.18	0.448 ± 0.18	0.442 ± 0.18	0.452 ± 0.18	0.424 ± 0.19	0.388 ± 0.18	0.08 ± 0.12	0.089 ± 0.14

Table 3.5: DSC and nDSC associated to MSLesSeg database segmentation, with and without synthetic lesions added in the training set (*Baseline* is the model without synthetic lesions and *SynLes* the model trained with synthetic lesions)

As seen on Figure 3.6, results of the test set segmentation show that using synthetic lesions in the training dataset significantly improves the DSC, nDSC. For the lesion-wise metrics, LTPR is significantly worse when synthetic lesions are used in the training set in-domain. LFPR is increased but it is not significant.



(a) Boxplots representing the DSC and nDSC metrics associated to the MSLesSeg database segmentation while using synthetic lesions (in red) or not in the training set.

(b) Boxplots representing the LTPR and LFPR metrics associated to the MSLesSeg database segmentation while using synthetic lesions (in red) or not in the training set.

Figure 3.6: Boxplots for the DSC, nDSC and lesion-wise metrics associated to the MSLesSeg database segmentation

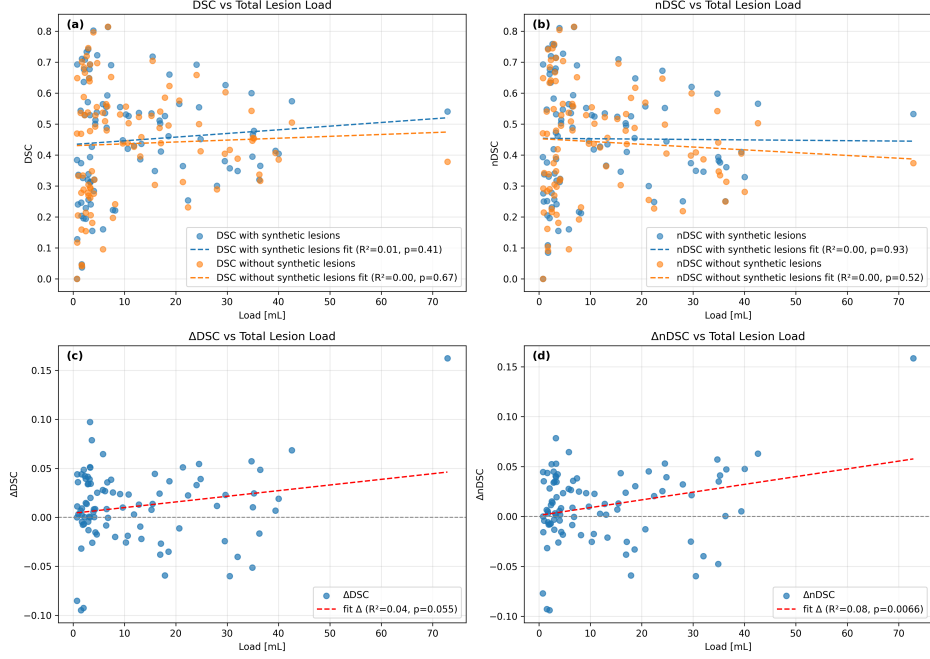


Figure 3.7: (a) DSC obtained for the segmentation of MSLesSeg as a function of lesion load, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of MSLesSeg as a function of lesion load, with and without synthetic lesions added to the training set. (c) Difference in DSC associated to the MSLesSeg database segmentation between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC associated to the MSLesSeg database segmentation between the model trained with synthetic lesions and the model trained without.

Figure 3.7 shows the improvement in segmentation performance on the MS-LesSeg dataset when using synthetic lesions in the training set, compared to the baseline, as a function of lesion load. Lower plots represents the difference in DSC and nDSC between models trained with and without synthetic lesions, plotted against the lesion's load. As seen in the $\Delta nDSC$ graph, the p-value associated to the line is < 0.05 , indicating a significant correlation between the use of synthetic lesions during training and improved segmentation performance on high-lesion-load test images.

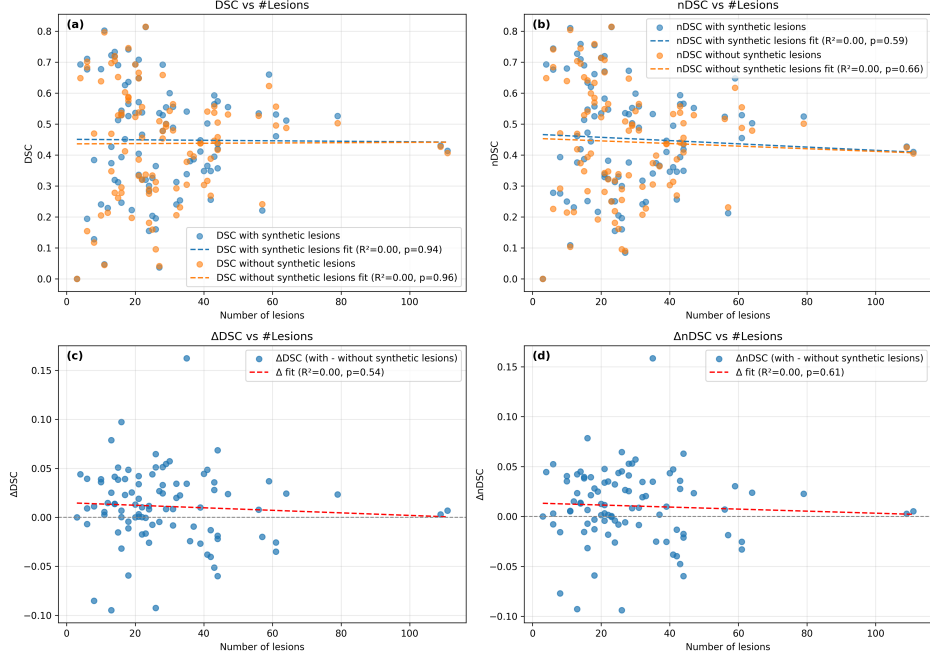


Figure 3.8: (a) DSC obtained for the segmentation of MSLesSeg as a function of the lesion's number, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of MSLesSeg as a function of lesion's number, with and without synthetic lesions added to the training set. (c) Difference in DSC associated to the MSLesSeg database segmentation between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC associated to the MSLesSeg database segmentation between the model trained with synthetic lesions and the model trained without.

Figure 3.8 shows the improvement of MSLesSeg segmentation when using synthetic lesions in the training set, as a function of the number of lesions. Lower graphs displays the difference in DSC and nDSC between models trained with and without synthetic lesions. As shown in the graphs, none of the p-values are significant, indicating that there is no correlation between an higher number of lesions and improved segmentation performance resulting from the use of synthetic lesions in the training set.

Results with the MSShifts database[84]

In this section, datasets from training and testing, coming from MSShifts (which is a composite of three datasets : ISBI [58], MSSEG-1 [62] and PUBMRI [88]) were mixed and used as test set.

DSC		nDSC		LTPR		LFPR	
Baseline	SynLes	Baseline	SynLes	Baseline	SynLes	Baseline	SynLes
0.35 ± 0.15	0.307 ± 0.16	0.354 ± 0.15	0.311 ± 0.16	0.355 ± 0.21	0.337 ± 0.19	0.044 ± 0.09	0.065 ± 0.14

Table 3.6: DSC and nDSC associated to the test set segmentation, with and without synthetic lesions added in the training set (*Baseline* is the model without synthetic lesions and *SynLes* the model trained with synthetic lesions).

As shown in Table 3.6, the DSC, nDSC and LTPR scores for the segmentation of the MSShift dataset are higher when no synthetic lesions are used in the training set, compared to when synthetic lesions are included. LFPR is also lower when using the baseline configuration.

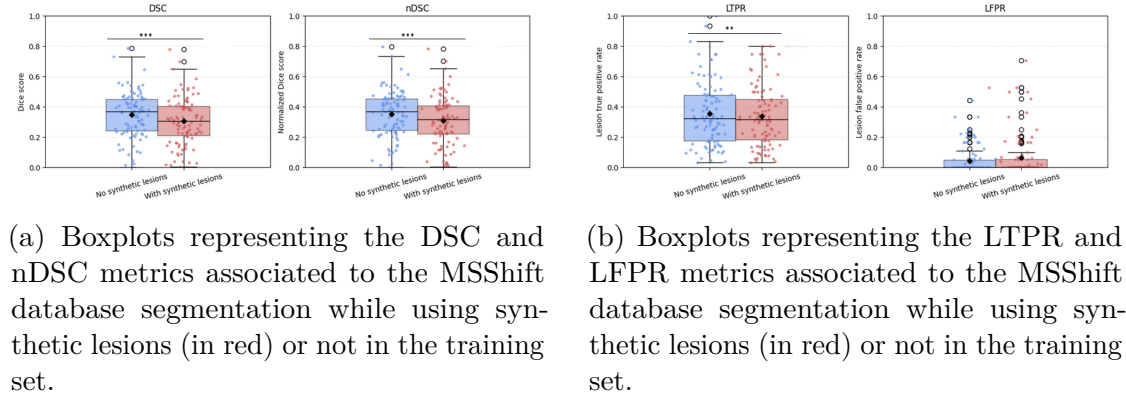


Figure 3.9: Boxplots for the DSC, nDSC and lesion-wise metrics associated to the MSShift database segmentation

As seen in the Figure 3.9, the diminution in DSC, nDSC and LTPR while using synthetic lesions in the training set is significative.

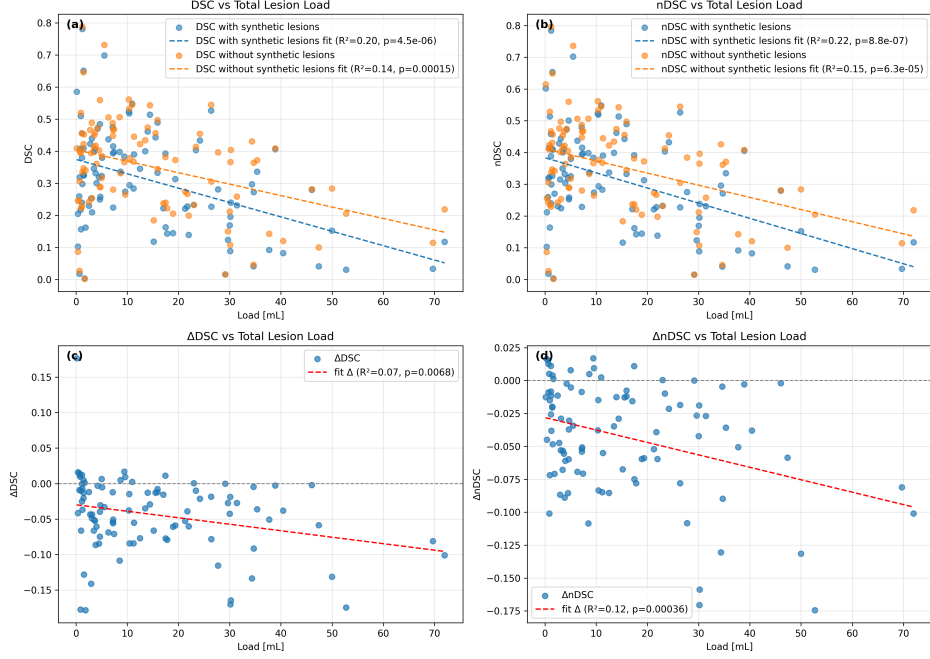


Figure 3.10: (a) DSC obtained for the segmentation of MSShift as a function of lesion load, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of MSShift as a function of lesion load, with and without synthetic lesions added to the training set. (c) Difference in DSC associated to the MSShift database segmentation, between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC associated to the MSShift database segmentation, between the model trained with synthetic lesions and the model trained without.

Figure 3.10 shows the difference in DSC and nDSC of the MSShift segmentation when using or not synthetic lesions in the training set, as a function of the total lesion load in the segmented images. Lower graphs display the difference in DSC and nDSC between the two training configurations. As shown in the figure, the decrease in segmentation accuracy with increasing lesion load is statistically significant for both configurations. Moreover, using synthetic lesions in the training set results in significantly poorer segmentation performance on high-lesion-load images compared to the model trained without synthetic lesions.

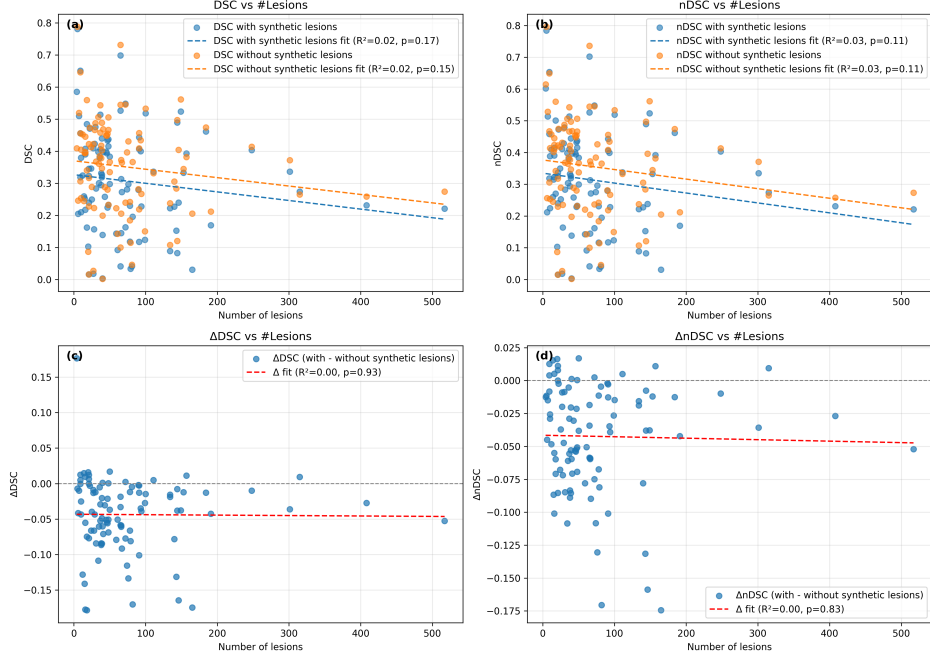


Figure 3.11: (a) DSC obtained for the segmentation of MSShift as a function of the lesion's number, with and without synthetic lesions added to the training set. (b) nDSC obtained for the segmentation of MSShift as a function of lesion's number, with and without synthetic lesions added to the training set. (c) Difference in DSC associated to the MSShift database segmentation, between the model trained with synthetic lesions and the model trained without. (d) Difference in nDSC associated to the MSShift database segmentation, between the model trained with synthetic lesions and the model trained without.

Figure 3.11 shows the difference in DSC and nDSC for MSShift segmentation when using or not synthetic lesion in the training set, as a function of the total number of lesions in the segmented images. A tendency toward decreased segmentation accuracy with a higher number of lesions is observed, but it is not statistically significant. The lower graphs displays the difference in DSC and nDSC between the two configuration, but no significative correlation between the number of lesions and an improvement or deterioration in segmentation performance when using synthetic lesion in the training set is observed.

Chapter 4

Discussion

The main objective of this master’s thesis was to create synthetic lesions on MRI scans to improve the segmentation accuracy of a deep learning algorithm, namely nnU-Net. The best configuration for generating these synthetic lesions combined high blurring, a low-medium lesion load (i.e. 100 synthetic lesions added on the MRI scan), and the use of pathological MRI scans instead of healthy brain MRIs for synthetic lesion generation. The segmentation performance of the model trained with synthetic lesions improved on the in-domain test set compared to the model trained without them, although this improvement was not statistically significant. Extended results on out-of-domain datasets, however, showed mixed directions: segmentation improved on the MSLesSeg dataset [83] but decreased on the MSShift dataset [84] when synthetic lesions were used compared to the baseline model. On both datasets, segmentation accuracy dropped significantly. Finally, there was a noticeable trend of reduced segmentation performance with higher lesion loads, a phenomenon that was not observed when increasing the number of lesions.

The best hyperparameters for synthetic lesion appearance were selected to optimize performance on our in-domain dataset, with adapted blurring values, lesion loads, and the type of MRI scans used for augmentation (i.e., pathological or healthy scans). We observed an improvement in segmentation accuracy when using a low-to-medium lesion load (i.e., 100 synthetic lesions), which may be explained by the fact that the in-domain dataset predominantly consists of patients with a low lesion burden. As a result, models trained with a similarly low number of synthetic lesions on pathological images are better aligned with the underlying lesion load distribution, leading to improved performance. Compared to other methods in the literature ([57], [54], [70], [72]) which generate synthetic pathological images by adding lesions to healthy brain scans, we found that using pathological scans instead

of healthy ones led to improved segmentation accuracy. One possible explanation is that pathological scans better reflect structural alterations present in MS patients, such as brain atrophy and diffuse microstructural changes in normal-appearing white matter, which are absent in healthy individuals.

For in-domain results, we observed a trend toward increased DSC and nDSC when segmenting the test set using models trained with synthetic lesions. Although the increases in DSC and nDSC were not statistically significant, which might be likely due to a too small test set size (i.e. only 13 MRI scans are used for testing), the reduction in LFPR was significant, suggesting that the use of synthetic lesions improved segmentation precision. This might be explained by a better learning of lesion specific characteristics, which may have enhanced the model’s generalizability on our in-domain test set.

We also evaluated our approach on out-of-domain test sets to assess its clinical applicability on heterogeneous datasets.[95] Specifically, we chose to evaluate it on MSLesSeg and MSShift, the latter being a composite dataset composed of three smaller datasets, thus introducing heterogeneity in the MRI scans, as explained in subsection 2.2.2. Interestingly, the results on MSLesSeg and MSShift differed when using synthetic lesions during training. On MSLesSeg, the model with synthetic lesion added in the training set led to improved segmentation accuracy compared to the baseline model, whereas on MSShift, the opposite trend was observed. This discrepancy might be explained by differences in preprocessing. MSLesSeg dataset includes registration to the MNI space, which induces blurring in MRI scans. This transformation appears to benefit our synthetic lesion approach, possibly because lesions are integrated into pathological MRI scans using Gaussian blurring to enhance realism. In contrast, MSShift dataset follows different acquisition protocol and preprocessing steps, resulting in less blurry images with sharper boundaries between white matter and lesions. This may have contributed to the decreased segmentation performance of our model trained with synthetic lesions. Additionally, the low-to-medium lesion load used during synthetic lesion generation was not optimized for MSShift, which includes patients with a higher lesion burden, compared to MSLesSeg dataset. Both the blurring parameters and lesion load were fine-tuned to enhance segmentation performance on our in-domain dataset. This tuning, however, resulted in decreased performance on MSShift dataset, reflecting limited generalization of our model to out-of-domain data. We attribute this lack of generalizability to insufficient diversity in acquisition protocols and preprocessing steps within our training dataset.

We observed a substantial decrease in segmentation performance for both mod-

els out-of-domain, compared to the segmentation performance in-domain. The training set, whether augmented with synthetic lesions or not, did not contain sufficient data variability to improve generalization and adapt to out-of-domain segmentation tasks. Furthermore, manual annotations in the in-domain dataset, on which both models were trained, excluded cortical lesions, infratentorial lesions, lesions smaller than 3 mm in diameter, and Dirty Appearing White Matter (DAWM, i.e., T2 hyperintensities with fuzzy borders). These exclusion have a direct impact while testing on the out-of domain test set, which used more extended annotation protocols, leading to a lot of missed lesions, either in the model without or with synthetic lesions.

Another observation is the tendency of both models (with or without synthetic lesions) to perform worse on test images with higher lesion loads. In contrast, the lesion’s number have a lower influence on the segmentation’s results, which shows a bad segmentation of bigger lesions on out-of-domain dataset. This likely results from the limited lesion load represented in the in-domain dataset, which primarily included patients at early disease stages.

A first limitation of our method, especially our validation pipeline, compared to other synthetic lesion generation methods from the literature, is the lack of direct comparison with existing techniques. This prevents us from drawing clear conclusions regarding the superiority or inferiority of our approach. In future work, benchmarking our method against the most similar technique, namely MS-MIST[72], could offer valuable insights into the performance gains brought by our synthetic lesions. Another limitation of our method lies in a too small and homogeneous training set, which is an issue already highlighted in the literature.[95] The small sample size of our training set, combined with hyperparameter tuning for synthetic lesions based solely on this dataset, led to lower generalization performance on out-of-domain data (with dataset containing higher lesion load and different acquisition protocol). This could be improved by incorporating small samples from multiple datasets instead of relying on a single dataset with consistent acquisition protocols, annotations, and preprocessing steps (i.e., domain generalization [96]). Another way to enhance generalizability is to use techniques such as domain adaptation [97], which adapts training on a source domain to improve performance on a target domain, quantile generalization as a preprocessing step [95], which normalizes MRI intensities to a common template, or the use of test-time training, a domain adaptation technique in which the algorithm continues to learn during the testing phase, thereby reducing the domain shift between the training and testing dataset.[98] The exclusive use of FLAIR MRI is another limitation. In contrast to other works that support multiple MRI sequences, such as MS-MIST[72] or the method proposed

by Mostafa Salem et al. (2019)[57], our approach relies solely on FLAIR. A future direction would be to extend the method to support other acquisition protocols.

The proposed method has nevertheless demonstrated a tendency to improve segmentation accuracy on the in-domain test set, which was our first objective. Furthermore, the method can be easily extend to instance segmentation[99] [100], rather than merely generating binary masks. This is possible thanks to our lesion shape dictionary, which, unlike the lesion load masks used in MS-MIST, enables a straightforward extension of the method toward instance-level lesion segmentation. Instance segmentation is essential for identifying specific lesion subtypes such as those exhibiting the central vein sign (CVS), paramagnetic rims, or cortical involvement. These biomarkers have all been shown to improve the specificity of MS diagnosis, particularly when used in combination[101] [102]. The automated analysis of such biomarkers requires precise localization and delineation of individual lesions[103][104], which is typically achieved through instance-level segmentation masks.

Conclusion

The main goal of this master’s thesis was to improve the semantic segmentation accuracy of MS lesions for a deep learning-based algorithm with the broader aim of enhancing MS diagnosis. To achieve this, a non-deep-learning approach was proposed to generate synthetic lesions, starting with an evaluation of MS lesion characteristics and a review of the state-of-the-art in synthetic lesion generation. Based on these insights, we developed a method to generate synthetic lesions on pathological MRI scans, relying on a lesion likelihood map, an MS lesion shape dictionary, and a synthetic texture dictionary derived from an existing MS population training dataset. Additionally, we proposed a validation pipeline that first identifies the hyperparameter settings producing optimal results via five-fold cross-validation, and then evaluates if integrating synthetic lesions in the training set improves segmentation performance. The model trained with the optimal parameters was then evaluated on both in-domain and out-of-domain test sets and compared to a baseline model trained without synthetic lesions, to provide a broader perspective on the performance of the proposed method.

This thesis introduces a novel approach for generating synthetic lesions with promising results for improving MS lesion segmentation. The proposed method is a non-deep learning approach for generating synthetic lesions on pathological MRI scans. While other methods have previously introduced similar approaches, based on generation on healthy images or based on lesion load databases[72], this is, to our knowledge, the first method that combines a lesion shape dictionary, a novel texture generation mechanism, and the insertion of synthetic lesions into pathological images. Our results suggest that this strategy improves segmentation accuracy in-domain.

However, challenges remain, particularly regarding the generalizability of the approach to out-of-domain datasets. This issue could potentially be mitigated by increasing the variability within the training dataset.

As future work, extending this technique to instance segmentation holds great promise for improving the accuracy of MS diagnosis, especially by enabling the

identification of specific lesion subtypes.

AI tool utilization during this work

In accordance with UCLouvain reglementations, I declare the use of an AI tool (i.e., ChatGPT), which was employed to improve the readability of the manuscript and assist with Python code. All generated content was carefully reviewed and edited when necessary. I take full responsibility for the content of this work.

Références et Bibliographie

- [1] Daniel S. Reich, Claudia F. Lucchinetti, and Peter A. Calabresi. “Multiple Sclerosis”. In: *The New England Journal of Medicine* 378.2 (2018), pp. 169–180. DOI: 10.1056/NEJMr1401483.
- [2] Dejan Jakimovski et al. “Multiple sclerosis”. In: *The Lancet* 403.10396 (2024), pp. 183–202. DOI: 10.1016/S0140-6736(23)01473-3.
- [3] E. Kontopodis et al. “Emerging deep learning techniques using magnetic resonance imaging data applied in multiple sclerosis and clinical isolated syndrome patients (Review)”. In: *Experimental and Therapeutic Medicine* 22 (2021). DOI: 10.3892/etm.2021.10583.
- [4] Z. Eaton-Rosen et al. “Improving Data Augmentation for Medical Image Segmentation”. In: (2018).
- [5] A.H. Crawford, C. Chambers, and R.J.M. Franklin. “Remyelination: The True Regeneration of the Central Nervous System”. In: *Journal of Comparative Pathology* 149.2 (2013), pp. 242–254. ISSN: 0021-9975. DOI: <https://doi.org/10.1016/j.jcpa.2013.05.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0021997513000856>.
- [6] D. Tafti, M. Ehsan, and K. L. Xixis. *Multiple Sclerosis*. [Updated 2024 Mar 20]. 2025. URL: <https://www.ncbi.nlm.nih.gov/books/NBK499849/> (visited on 05/03/2025).
- [7] Massimo Filippi et al. “Assessment of Lesions on Magnetic Resonance Imaging in Multiple Sclerosis: Practical Guidelines”. In: *Brain: A Journal of Neurology* 142.7 (July 2019), pp. 1858–1875. DOI: 10.1093/brain/awz144. URL: <https://doi.org/10.1093/brain/awz144>.
- [8] J. Maroto-García et al. “Biochemical biomarkers for multiple sclerosis.” In: *Clinica chimica acta; international journal of clinical chemistry* (2023), p. 117471. DOI: 10.1016/j.cca.2023.117471.
- [9] National Institute of Neurological Disorders and Stroke. *Multiple Sclerosis*. Consulted on November 2, 2024. URL: <https://www.ninds.nih.gov/health-information/disorders/multiple-sclerosis>.

- [10] T. Kuhlmann et al. “An updated histological classification system for multiple sclerosis lesions”. In: *Acta Neuropathologica* 133.1 (Jan. 2017). Epub 2016 Dec 17. PMID: 27988845, pp. 13–24. DOI: 10.1007/s00401-016-1653-y.
- [11] R. F. Kilcoyne et al. “Magnetic resonance imaging of soft tissue masses”. In: *Clinical Orthopaedics and Related Research* 228 (Mar. 1988). PMID: 3342555, pp. 13–19. URL: <https://pubmed.ncbi.nlm.nih.gov/3342555>.
- [12] Radiology Assistant. *Multiple Sclerosis 2.0: Diagnosis and Differential Diagnosis*. n.d. URL: <https://radiologyassistant.nl/neuroradiology/multiple-sclerosis/diagnosis-and-differential-diagnosis-3> (visited on 05/03/2025).
- [13] Mike P Wattjes et al. “Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis—establishing disease prognosis and monitoring patients”. In: *Nature Reviews Neurology* 11.10 (Oct. 2015), pp. 597–606. DOI: 10.1038/nrneuro.2015.157.
- [14] MRIMASTER. *FLAIR MRI Sequence Physics, Image and Pathology Appearance*. 2023. URL: <https://mrimaster.com/characterise-image-flair/> (visited on 05/03/2025).
- [15] Flavia M. Nelson et al. “3D MPRAGE Improves Classification of Cortical Lesions in Multiple Sclerosis”. In: *Multiple Sclerosis* 14.9 (Nov. 2008), pp. 1214–1219. DOI: 10.1177/1352458508094644.
- [16] MRIMASTER. *MPRANGE MRI / 3D FGRE / 3D FAST SPGR / 3DFAST FE Sequence Physics and Image Appearance*. 2023. URL: <https://mrimaster.com/characterise-image-ultrafat-gradient-3d/> (visited on 05/03/2025).
- [17] Massimo Filippi et al. “MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines”. In: *The Lancet Neurology* 15.3 (Mar. 2016). Epub 2016 Jan 26, pp. 292–303. DOI: 10.1016/S1474-4422(15)00393-2.
- [18] Massimo Filippi et al. “Prediction of a multiple sclerosis diagnosis in patients with clinically isolated syndrome using the 2016 MAGNIMS and 2010 McDonald criteria: a retrospective study”. In: *The Lancet Neurology* 17.2 (Feb. 2018). Epub 2017 Dec 21. PMID: 29275979, pp. 133–142. DOI: 10.1016/S1474-4422(17)30469-6.
- [19] A. J. Thompson et al. “Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria”. In: *The Lancet Neurology* 17.2 (Feb. 2018). Epub 2017 Dec 14. PMID: 29275980, pp. 162–173. DOI: 10.1016/S1474-4422(17)30470-2.

- [20] D. Pareto et al. “Juxtacortical Lesions and Cortical Thinning in Multiple Sclerosis”. In: *American Journal of Neuroradiology* 36.12 (Dec. 2015). Epub 2015 Oct 8. PMID: 26450537; PMCID: PMC7964289, pp. 2270–2276. DOI: 10.3174/ajnr.A4485.
- [21] Wikilivres. *Neurosciences/La substance blanche* — Wikilivres, La bibliothèque libre. [En ligne; accédé le 12-mars-2025]. 2023. URL: https://fr.wikibooks.org/w/index.php?title=Neurosciences/La_substance_blanche&oldid=693063.
- [22] Constantina A. Treaba et al. “The relevance of multiple sclerosis cortical lesions on cortical thinning and their clinical impact as assessed by 7.0-T MRI”. In: *Journal of Neurology* 268 (Feb. 2021), pp. 2473–2481. DOI: 10.1007/s00415-021-10400-4. URL: <https://doi.org/10.1007/s00415-021-10400-4> (visited on 05/03/2025).
- [23] Geert Lycklama et al. “Spinal-cord MRI in multiple sclerosis”. In: *The Lancet Neurology* 2.9 (Sept. 2003), pp. 555–562. DOI: 10.1016/S1474-4422(03)00504-0.
- [24] J. C. J. Bot et al. “Spinal cord abnormalities in recently diagnosed MS patients: added value of spinal MRI examination”. In: *Neurology* 62.2 (Jan. 2004), pp. 226–233. DOI: 10.1212/WNL.62.2.226.
- [25] Katrin Weier et al. “Biplanar MRI for the assessment of the spinal cord in multiple sclerosis”. In: *Multiple Sclerosis Journal* 18.11 (Nov. 2012). Epub 2012 Apr 26. PMID: 22539086, pp. 1560–1569. DOI: 10.1177/1352458512442754.
- [26] Achim Gass et al. “MRI monitoring of pathological changes in the spinal cord in patients with multiple sclerosis”. In: *The Lancet Neurology* 14.4 (Apr. 2015). Epub 2015 Mar 4. PMID: 25748099, pp. 443–454. DOI: 10.1016/S1474-4422(14)70294-7.
- [27] Olga Ciccarelli et al. “Spinal cord involvement in multiple sclerosis and neuromyelitis optica spectrum disorders”. In: *The Lancet Neurology* 18.2 (Feb. 2019). PMID: 30663608, pp. 185–197. DOI: 10.1016/S1474-4422(18)30460-5.
- [28] Reza Azad et al. “Medical Image Segmentation Review: The Success of U-Net”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46.12 (2024), pp. 10076–10095. DOI: 10.1109/TPAMI.2024.3435571.
- [29] Shujun Liang et al. “Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning”. In: *European Radiology* 29.4 (Apr. 2019). Epub 2018 Oct 9, pp. 1961–1967. DOI: 10.1007/s00330-018-5748-9.

- [30] Dawit Assefa et al. “Robust texture features for response monitoring of glioblastoma multiforme on T1-weighted and T2-FLAIR MR images: a preliminary investigation in terms of identification and segmentation”. In: *Medical Physics* 37.4 (Apr. 2010), pp. 1722–1736. DOI: 10.1118/1.3357289. URL: <https://doi.org/10.1118/1.3357289>.
- [31] Michela Antonelli et al. “The Medical Segmentation Decathlon”. In: *Nature Communications* 13.1 (July 2022), p. 4128. DOI: 10.1038/s41467-022-30695-9. URL: <https://doi.org/10.1038/s41467-022-30695-9>.
- [32] Reza Azad et al. “Medical Image Segmentation Review: The Success of U-Net”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46.12 (2024), pp. 10076–10095. DOI: 10.1109/TPAMI.2024.3435571.
- [33] P D Molyneux et al. “Precision and reliability for measurement of change in MRI lesion volume in multiple sclerosis: a comparison of two computer assisted techniques”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 65.1 (1998), pp. 42–47. ISSN: 0022-3050. DOI: 10.1136/jnnp.65.1.42. eprint: <https://jnnp.bmj.com/content/65/1/42.full.pdf>. URL: <https://jnnp.bmj.com/content/65/1/42>.
- [34] Antria Nicolaou et al. “A Systematic Review of Quantitative MRI Brain Analysis Studies in Multiple Sclerosis Disease”. In: *IEEE Access* 12 (2024), pp. 162359–162391. DOI: 10.1109/ACCESS.2024.3489798.
- [35] Alessandro Pasquale De Rosa et al. “Consensus of Algorithms for Lesion Segmentation in Brain MRI Studies of Multiple Sclerosis”. In: *Scientific Reports* 14 (2024), p. 21348. DOI: 10.1038/s41598-024-72649-9.
- [36] Daniel García-Lorenzo et al. “Review of Automatic Segmentation Methods of Multiple Sclerosis White Matter Lesions on Conventional Magnetic Resonance Imaging”. In: *Medical Image Analysis* 17.1 (2013), pp. 1–18. DOI: 10.1016/j.media.2012.09.004. URL: <https://www.sciencedirect.com/science/article/pii/S1361841512001338>.
- [37] D. Mortazavi, A. Z. Kouzani, and H. Soltanian-Zadeh. “Segmentation of multiple sclerosis lesions in MR images: a review”. In: *Neuroradiology* 54.4 (Apr. 2012). Epub 2011 May 17. PMID: 21584674, pp. 299–320. DOI: 10.1007/s00234-011-0886-7.
- [38] C. Pachai et al. “A pyramidal approach for automatic segmentation of multiple sclerosis lesions in brain MRI”. In: *Computerized Medical Imaging and Graphics* 22.5 (1998), pp. 399–408. ISSN: 0895-6111. DOI: [https://doi.org/10.1016/S0895-6111\(98\)00049-4](https://doi.org/10.1016/S0895-6111(98)00049-4). URL: <https://www.sciencedirect.com/science/article/pii/S0895611198000494>.

- [39] O. Ronneberger. “Invited Talk: U-Net Convolutional Networks for Biomedical Image Segmentation”. In: (2017), p. 3. DOI: 10.1007/978-3-662-54345-0_3.
- [40] Md. Eshmam Rayed et al. “Deep learning for medical image segmentation: State-of-the-art advancements and challenges”. In: *Informatics in Medicine Unlocked* 47 (Apr. 2024). Available online 22 April 2024, p. 101504. DOI: 10.1016/j.imu.2024.101504. URL: <https://www.sciencedirect.com/science/article/pii/S2352914824000601>.
- [41] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *arXiv preprint arXiv:1505.04597* (2015). arXiv:1505.04597v1 [cs.CV], 18 May 2015. URL: <https://arxiv.org/abs/1505.04597>.
- [42] Raphael Kassel. *U-NET : le réseau de neurones de Computer Vision*. June 2021. URL: <https://datascientest.com/u-net> (visited on 05/06/2025).
- [43] Fabian Isensee et al. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. In: *Nature Methods* 18.2 (Feb. 2021), pp. 203–211. DOI: 10.1038/s41592-020-01008-z. URL: <https://doi.org/10.1038/s41592-020-01008-z>.
- [44] Eloy Martínez-Heras et al. “Longitudinal Segmentation of Multiple Sclerosis Lesions Using nnU-Net Architecture”. In: Oct. 2023. ISBN: 9781643684482. DOI: 10.3233/FAIA230679.
- [45] Olivier Commowick et al. “Multiple Sclerosis New Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure”. In: *MICCAI 2021 MSSEG-2 Challenge Proceedings*. 2021.
- [46] Lena Maier-Hein et al. “Metrics reloaded: recommendations for image analysis validation”. In: *Nature Methods* 21.2 (Feb. 2024). Epub 2024 Feb 12. PMID: 38347141; PMCID: PMC11182665, pp. 195–212. DOI: 10.1038/s41592-023-02151-z.
- [47] Peter A. Flach. “ROC Analysis”. In: (2017), pp. 1109–1116. DOI: 10.1007/978-1-4899-7687-1_739.
- [48] Kolena Metrics Documentation. *False Positive Rate (FPR)*. 2025. URL: <https://docs.kolena.com/metrics/fpr/> (visited on 05/06/2025).
- [49] F. Spagnolo et al. “How far MS lesion detection and segmentation are integrated into the clinical workflow? A systematic review”. In: *NeuroImage : Clinical* 39 (2023). DOI: 10.1016/j.nicl.2023.103491.

- [50] Michael Yeung et al. “Calibrating the Dice Loss to Handle Neural Network Overconfidence for Biomedical Image Segmentation”. In: *arXiv preprint arXiv:2111.00528* (2021). Revised version (v2) submitted 1 Nov 2022. URL: <https://arxiv.org/abs/2111.00528>.
- [51] Vatsal Raina et al. “Tackling Bias in the Dice Similarity Coefficient: Introducing nDSC for White Matter Lesion Segmentation”. In: *arXiv preprint arXiv:2302.05432* (2023). Submitted 10 Feb 2023; accepted at ISBI 2023. URL: <https://arxiv.org/abs/2302.05432>.
- [52] Nils Gessert et al. “Multiple sclerosis lesion activity segmentation with attention-guided two-path CNNs”. In: *Computerized Medical Imaging and Graphics* 84 (Sept. 2020). Epub 2020 Aug 8. PMID: 32795845, p. 101772. DOI: 10.1016/j.compmedimag.2020.101772.
- [53] R. A. Kamraoui et al. “DeepLesionBrain: Towards a broader deep-learning generalization for multiple sclerosis lesion segmentation”. In: *Medical image analysis* 76 (2020), p. 102312. DOI: 10.1016/j.media.2021.102312.
- [54] Berke Doga Basaran et al. “Subject-Specific Lesion Generation and Pseudo-Healthy Synthesis for Multiple Sclerosis Brain Images”. In: *arXiv* (Aug. 2022). arXiv:2208.02135. URL: <http://arxiv.org/abs/2208.02135>.
- [55] Xinru Zhang et al. “CarveMix: A simple data augmentation method for brain lesion segmentation”. In: *NeuroImage* 271 (2023), p. 120041. ISSN: 1053-8119. DOI: <https://doi.org/10.1016/j.neuroimage.2023.120041>. URL: <https://www.sciencedirect.com/science/article/pii/S1053811923001878>.
- [56] Berke Doga Basaran et al. “LesionMix: A Lesion-Level Data Augmentation Method for Medical Image Segmentation”. In: *arXiv* (Aug. 2023). arXiv:2308.09026. URL: <http://arxiv.org/abs/2308.09026>.
- [57] Mostafa Salem et al. “Multiple Sclerosis Lesion Synthesis in MRI Using an Encoder-Decoder U-NET”. In: *IEEE Access* 7 (2019), pp. 25171–25184. DOI: 10.1109/ACCESS.2019.2900198.
- [58] Aaron Carass et al. “Longitudinal multiple sclerosis lesion segmentation: Resource and challenge”. In: *NeuroImage* 148 (2017), pp. 77–102. DOI: 10.1016/j.neuroimage.2016.12.064. URL: <https://doi.org/10.1016/j.neuroimage.2016.12.064>.
- [59] Hamed Alqahtani, Manolya Kavakli-Thorne, and G. Kumar. “Applications of Generative Adversarial Networks (GANs): An Updated Review”. In: *Archives of Computational Methods in Engineering* 28 (2019), pp. 525–552. DOI: 10.1007/s11831-019-09388-y.

- [60] Antonia Creswell et al. “Generative Adversarial Networks: An Overview”. In: *IEEE Signal Processing Magazine* 35 (2017), pp. 53–65. DOI: 10.1109/MSP.2017.2765202.
- [61] Fabio Garcea et al. “Data Augmentation for Medical Imaging: A Systematic Literature Review”. In: *Computers in Biology and Medicine* 152 (Jan. 2023), p. 106391. DOI: 10.1016/j.combiomed.2022.106391.
- [62] Olivier Commowick, Audrey Istace, Michaël Kain, et al. “Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure”. In: *Scientific Reports* 8.1 (2018), p. 13650. DOI: 10.1038/s41598-018-31911-7. URL: <https://doi.org/10.1038/s41598-018-31911-7>.
- [63] Martin Styner, J. Lee, B. Chin, et al. *3D Segmentation in the Clinic: A Grand Challenge II: MS Lesion Segmentation*. Nov. 2008. DOI: 10.54294/lmkqvm. URL: <https://doi.org/10.54294/lmkqvm>.
- [64] Zeeshan Ahmad et al. “Understanding GANs: fundamentals, variants, training challenges, applications, and open problems”. In: *Multimedia Tools and Applications* (2024). DOI: 10.1007/s11042-024-19361-y.
- [65] PrepBytes. *The disadvantages of generative adversarial networks (GANs)*. <https://www.prepbytes.com/blog/artificial-intelligence/the-disadvantages-of-generative-adversarial-networks-gans/>. [Consulté le 3 mars 2025].
- [66] Sangdoo Yun et al. “CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features”. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (2019), pp. 6022–6031. DOI: 10.1109/ICCV.2019.00612.
- [67] A. Mumuni and F. Mumuni. “Data augmentation: A comprehensive survey of modern approaches”. In: *Array* 16 (2022), p. 100258. DOI: 10.1016/j.array.2022.100258.
- [68] Paul M Thompson and Arthur W Toga. “36 - Warping Strategies for Intersubject Registration”. In: *Handbook of Medical Imaging*. Ed. by ISAAC N. BANKMAN. Biomedical Engineering. San Diego: Academic Press, 2000, pp. 569–601. ISBN: 978-0-12-077790-7. DOI: <https://doi.org/10.1016/B978-012077790-7/50042-4>. URL: <https://www.sciencedirect.com/science/article/pii/B9780120777907500424>.

- [69] Hugo J. Kuijf et al. “Standardized Assessment of Automatic Segmentation of White Matter Hyperintensities and Results of the WMH Segmentation Challenge”. In: *IEEE Transactions on Medical Imaging* 38.11 (2019), pp. 2556–2568. DOI: 10.1109/TMI.2019.2905770. URL: <https://doi.org/10.1109/TMI.2019.2905770>.
- [70] Peyman Tahghighi et al. “Enhancing New Multiple Sclerosis Lesion Segmentation via Self-Supervised Pre-Training and Synthetic Lesion Integration”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. Ed. by Marius George Linguraru et al. Vol. 15008. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2024, pp. 263–272. DOI: 10.1007/978-3-031-72111-3_25. URL: https://doi.org/10.1007/978-3-031-72111-3_25.
- [71] Olivier Commowick et al. “Multiple sclerosis lesions segmentation from multiple experts: the MICCAI 2016 challenge dataset”. In: *NeuroImage* 244 (2021), p. 118589. DOI: 10.1016/j.neuroimage.2021.118589. URL: <https://doi.org/10.1016/j.neuroimage.2021.118589>.
- [72] Antonio Carlos da Silva Senra Filho et al. “Multiple Sclerosis multimodal lesion simulation tool (MS-MIST)”. In: *Biomedical Physics & Engineering Express* 5.3 (Mar. 2019), p. 035003. DOI: 10.1088/2057-1976/ab08fc.
- [73] Jinwei Zhang et al. *Bi-Directional MS Lesion Filling and Synthesis Using Denoising Diffusion Implicit Model-based Lesion Repainting*. 2024. arXiv: 2410.05027 [eess.IV]. URL: <https://arxiv.org/abs/2410.05027>.
- [74] Gabriela Castellano et al. “Texture analysis of medical images”. In: *Clinical Radiology* 59.12 (2004), pp. 1061–1069. DOI: 10.1016/j.crad.2004.07.008.
- [75] Weifang Liu, Zhou Xiaoxia, and Tong Longzheng. “Texture Research on MRI in Patients with Multiple Sclerosis Based on Synthesized Gray Level Co-Occurrence Matrix”. In: *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*. Ed. by Olaf Dössel and Wolfgang C. Schlegel. Berlin, Heidelberg: Springer, 2010, pp. 2058–2061. DOI: 10.1007/978-3-642-03882-2_546. URL: https://doi.org/10.1007/978-3-642-03882-2_546.
- [76] Yunyan Zhang. “MRI Texture Analysis in Multiple Sclerosis”. In: *International Journal of Biomedical Imaging* 2012 (2011). URL: <https://api.semanticscholar.org/CorpusID:12578044>.
- [77] Roberto Costantini, Luciano Sbaiz, and Sabine Süsstrunk. “Higher Order SVD Analysis for Dynamic Texture Synthesis”. In: *IEEE Transactions on Image Processing* 17.1 (Jan. 2008), pp. 42–51. DOI: 10.1109/TIP.2007.910956.

- [78] Xuejing Lei et al. “TGHop: An Explainable, Efficient and Lightweight Method for Texture Generation”. In: *CoRR* abs/2107.04020 (2021). arXiv: 2107.04020. URL: <https://arxiv.org/abs/2107.04020>.
- [79] Nicholas J. Tustison et al. “The ANTsX ecosystem for quantitative biological and medical imaging”. In: *Scientific Reports* 11.1 (Apr. 2021), p. 9068. ISSN: 2045-2322. DOI: 10.1038/s41598-021-87564-6. URL: <https://doi.org/10.1038/s41598-021-87564-6>.
- [80] Lead-DBS. *About the MNI space(s)*. <https://www.lead-dbs.org/about-the-mni-spaces/>. Accessed: 2025-05-10.
- [81] B. Fischl et al. “Whole Brain Segmentation Automated Labeling of Neuroanatomical Structures in the Human Brain”. In: *Neuron* 33 (2002), pp. 341–355. DOI: 10.1016/S0896-6273(02)00569-X.
- [82] B. Fischl. “FreeSurfer”. In: *NeuroImage* 62 (2012), pp. 774–781. DOI: 10.1016/j.neuroimage.2012.01.021.
- [83] Alessia Rondinella et al. “ICPR 2024 Competition on Multiple Sclerosis Lesion Segmentation - Methods and Results”. In: (2024), pp. 1–16. DOI: 10.48550/arXiv.2410.07924.
- [84] A. Malinin et al. “Shifts 2.0: Extending The Dataset of Real Distributional Shifts”. In: *ArXiv* abs/2206.15407 (2022). DOI: 10.48550/arXiv.2206.15407.
- [85] Robert F. Woolson. “Wilcoxon Signed-Rank Test”. In: 2005. URL: <https://api.semanticscholar.org/CorpusID:119763117>.
- [86] Statistics Solutions. *Paired T-Test*. <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/paired-sample-t-test/>. Accessed: 2025-05-19. n.d.
- [87] Section for Biomedical Image Analysis (SBIA), Perelman School of Medicine at the University of Pennsylvania. *MICCAI BraTS 2018: Data*. [Accessed: 2025-05-10]. 2018. URL: <https://www.med.upenn.edu/sbia/brats2018/data.html>.
- [88] Žiga Lesjak et al. “A Novel Public MR Image Dataset of Multiple Sclerosis Patients With Lesion Segmentations Based on Multi-rater Consensus”. In: *Neuroinformatics* 16.1 (2018), pp. 51–63. ISSN: 1559-0089. DOI: 10.1007/s12021-017-9348-7. URL: <https://doi.org/10.1007/s12021-017-9348-7>.

- [89] Colin Vanden Bulcke et al. “BMAT: An open-source BIDS managing and analysis tool”. In: *NeuroImage: Clinical* 36 (2022), p. 103252. ISSN: 2213-1582. DOI: <https://doi.org/10.1016/j.nicl.2022.103252>. URL: <https://www.sciencedirect.com/science/article/pii/S2213158222003175>.
- [90] Andrew Hoopes et al. “SynthStrip: skull-stripping for any brain image”. In: *NeuroImage* 260 (2022), p. 119474. ISSN: 1053-8119. DOI: <https://doi.org/10.1016/j.neuroimage.2022.119474>. URL: <https://www.sciencedirect.com/science/article/pii/S1053811922005900>.
- [91] M. Jenkinson et al. “Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images”. In: *NeuroImage* 17 (2002), pp. 825–841. DOI: 10.1006/nimg.2002.1132.
- [92] Stephen M. Smith. “Fast robust automated brain extraction”. In: *Human Brain Mapping* 17 (2002). DOI: 10.1002/hbm.10062.
- [93] Paul A. Yushkevich et al. “User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability”. In: *Neuroimage* 31.3 (2006), pp. 1116–1128.
- [94] H. Johnson, G. Harris, and K. Williams. “BRAINSFit: Mutual Information Registrations of Whole-Brain 3D Images, Using the Insight Toolkit”. In: (Oct. 2007). DOI: 10.54294/hmb052.
- [95] Liviu Badea and Maria Popa. “Toward Generalizable Multiple Sclerosis Lesion Segmentation Models”. In: *ArXiv* abs/2410.19623 (2024). DOI: 10.48550/arXiv.2410.19623.
- [96] Jindong Wang et al. “Generalizing to Unseen Domains: A Survey on Domain Generalization”. In: *IEEE Transactions on Knowledge and Data Engineering* 35 (2021), pp. 8052–8072. DOI: 10.1109/TKDE.2022.3178128.
- [97] Wouter M. Kouw and M. Loog. “A Review of Domain Adaptation without Target Labels”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43 (2019), pp. 766–785. URL: <https://api.semanticscholar.org/CorpusID:198898096>.
- [98] Benoît Gérin et al. “Exploring viability of Test-Time Training: Application to 3D segmentation in Multiple Sclerosis”. In: *2024 IEEE Conference on Artificial Intelligence (CAI)*. 2024, pp. 557–562. DOI: 10.1109/CAI59869.2024.00110.
- [99] Maxence Wynen et al. “Conflunet: Improving Confluent Lesion Identification In Multiple Sclerosis With Instance Segmentation”. In: *2024 IEEE International Symposium on Biomedical Imaging (ISBI)* (2024), pp. 1–5. DOI: 10.1109/ISBI56570.2024.10635707.

- [100] Maxence Wynen et al. *ConflUNet: Multiple sclerosis lesion instance segmentation in presence of confluent lesions*. 2025. arXiv: 2505.22537 [eess.IV]. URL: <https://arxiv.org/abs/2505.22537>.
- [101] Maxence Wynen et al. *Machine Learning-Based Combination of the Central Vein Sign, Cortical Lesions and Paramagnetic Rim Lesions: A Web-Based Tool for the Diagnosis of Multiple Sclerosis*. <https://ssrn.com/abstract=5192817>. <http://dx.doi.org/10.2139/ssrn.5192817>. 2024.
- [102] Serena Borrelli et al. “Central Vein Sign, Cortical Lesions, and Paramagnetic Rim Lesions for the Diagnostic and Prognostic Workup of Multiple Sclerosis”. In: *Neurology: Neuroimmunology & Neuroinflammation* 11.4 (July 2024), e200253. DOI: 10.1212/NXI.0000000000200253. URL: <https://doi.org/10.1212/NXI.0000000000200253>.
- [103] Francesca La Rosa et al. “Cortical lesions, central vein sign, and paramagnetic rim lesions in multiple sclerosis: Emerging machine learning techniques and future avenues”. In: *NeuroImage: Clinical* 36 (2022), p. 103205. DOI: 10.1016/j.nicl.2022.103205. URL: <https://doi.org/10.1016/j.nicl.2022.103205>.
- [104] Maxence Wynen et al. “Longitudinal Automated Assessment of Paramagnetic Rim Lesions in Multiple Sclerosis Using RimNet”. In: *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM)*. ISMRM. 2021, pp. 0–0.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/epl