

# Deep learning-based analysis of multiple sclerosis lesions with high and ultra-high field MRI

Présentée le 25 mars 2022

Faculté des sciences et techniques de l'ingénieur  
Laboratoire de traitement des signaux 5  
Programme doctoral en génie électrique

pour l'obtention du grade de Docteur ès Sciences

par

**Francesco LA ROSA**

Acceptée sur proposition du jury

Dr G. Boero, président du jury  
Prof. J.-Ph. Thiran, Dr M. Bach Cuadra, directeurs de thèse  
Dr J. Oh, rapporteuse  
Prof. M. Reyes, rapporteur  
Dr M. Liebling, rapporteur





To my parents...



# Acknowledgements

First, I would like to express my sincere gratitude to my PhD supervisors Prof. Jean-Philippe Thiran and Dr. Merixtell Bach Cuadra. Thank you, Jean-Philippe, for believing in me and giving me the opportunity to pursue my PhD in your group. Thank you, Meri, for your daily supervision and insightful feedbacks that pushed me to sharpen my thinking. Our frequent discussions helped me better understand how to target our technical developments to clinical needs, and this is what is driving my research today.

I am very thankful to the jury members of my PhD defense, Dr. Jiwon Oh, Prof. Mauricio Reyes, Dr. Michael Liebling, and the committee president Prof. Giovanni Boero. Thanks to you all for dedicating the time to evaluate my work and for the insightful discussion.

Since the beginning of my PhD, I split my time between the Signal Processing Laboratory 5 (LTS5) and the Medical Image Analysis Laboratory (MIAL). I am grateful to all members of both groups for welcoming me and making this an exceptional journey. Thomas, Juan Luis, Jony, Gab, Elda, Erick, Marco, David, Muhamed, Christophe, Saeed, Saleh, Remy, Samuel, Christian, Tommaso, Hamza, Helene, Priscille, Cristina, Yasser, Emeline, and all the others. From football matches to yogurt breaks, I have fantastic memories with you that made the working days incredibly enjoyable.

A special mention to the students I had the pleasure to co-supervise during these years. Germán, Cristina, Imen, and Maxence, thanks for pushing our research further and allowing me to see our problems from a different perspective.

I was extremely fortunate to be part of TRABIT and I want to thank all the ESRs for the fun moments together. I hope we will be able to organize a reunion and meet up again soon.

I am grateful to all my friends here in Switzerland, back in Italy, and spread around Europe and the US. Gian Franco, Adele, Cipo, Chiara, Filippo, Harshal, Algirdas, Daniele, Luca, Paco, Alex, Iker, Mila, Manuel, Marata, Medo, Barda, Alessandro, Roberto, Fabio, Jonathan, Neng, Maxime, Simone, Ivano, Sanja, Maria Teresa, Kow, Gaston, and all the others. I have unforgettable memories with you from these past years and I am sure that many more are to come.

Thank you, Victoria, for your trust, patience, and for standing by my side.

## Acknowledgements

---

Lastly, I would like to express my deepest gratitude to my family. This achievement would not have been possible without your constant and unconditional support.

*Lausanne, March 2022*

Francesco

# Abstract

Multiple sclerosis (MS) is the most common demyelinating disease of the central nervous system and affects almost 3 million people worldwide. There is currently no cure for MS, and its symptoms, starting with fatigue and weakness, often progress over time into disabilities and cognitive impairments. Hence, early diagnosis of the disease is crucial to start appropriate treatment and delay MS progression. Diagnosis, prognosis, and follow-up of MS patients rely mostly on magnetic resonance imaging (MRI) findings such as white matter lesions. This conventional biomarker, however, lacks specificity in differentiating MS from mimic conditions and does not strongly correlate with disability progression. Recently, advanced imaging biomarkers, including cortical lesions and paramagnetic rim lesions, have emerged as being specific to MS and predicting disease severity. Nevertheless, their assessment requires highly specialized experts and is extremely time-consuming. Thus, there is currently an unmet need for decision support tools to automate the analysis of these biomarkers that can ultimately enhance the understanding of the disease and improve patient care.

The present thesis describes the development and validation of different deep learning approaches for the automated analysis of MS lesions at high and ultra-high field MRI, focusing on cortical and paramagnetic rim lesions. The main contributions are the following. First, an automated method for the segmentation of cortical and white matter lesions with only two 3T MRI sequences is proposed. Based on a patch-wise convolutional neural network, this method is validated on a large multi-site cohort of MS patients, outperforming existing machine learning approaches. Second, we translated our approach to ultra-high field MRI, exploring specialized MRI contrasts. Cortical Lesion AI-based assessment in MS (CLAIMS) is a fully-automated framework providing cortical lesion detection and classification at 7T MRI. Third, the issue of missing MRI modalities was tackled. As the MP2RAGE MRI sequence is highly desirable to detect cortical lesions, yet not often acquired in clinical practice, a generative adversarial network is introduced to synthesize MP2RAGE images from MPRAGE. This approach generates realistic MP2RAGE images which, compared to the MPRAGE, significantly improve the lesion and brain tissue segmentation masks computed by automated

## Abstract

---

tools. Fourth, the first deep learning-based approach for the classification of paramagnetic rim lesions is proposed (RimNet). RimNet has been evaluated with different MRI contrasts at 3T and 7T, and its performance is close to that of an expert.

To summarize, this thesis aims at providing robust deep learning-based approaches for the automated analysis of MS lesions. All methods developed have been designed to meet specific clinical needs. A minimal number of clinically relevant MRI sequences was proven sufficient to outperform existing techniques and approach the experts' manual evaluation. As these methods generalize well to multi-site datasets at both high and ultra-high field MRI, they represent the first step toward an automated and standardized assessment of advanced MS biomarkers.

**Keywords:** multiple sclerosis, magnetic resonance imaging, machine learning, deep learning, cortical lesions, paramagnetic rim lesions, segmentation, detection, ultra-high field MRI.

# Sommario

La sclerosi multipla (SM) è la più comune malattia demielinizzante del sistema nervoso centrale e quasi 3 milioni di persone nel mondo sono affette da SM. Non esiste una cura per la SM e i suoi sintomi, iniziando con stanchezza e debolezza, spesso progrediscono nel tempo in gravi disabilità e deterioramenti cognitivi. Per questo, la diagnosi precoce della malattia è fondamentale per iniziare un trattamento appropriato e rallentare la sua progressione. La diagnosi, la prognosi e il follow-up dei pazienti con SM si basano principalmente sui risultati della risonanza magnetica (RM) come le lesioni della sostanza bianca. Questo biomarker convenzionale, tuttavia, ha una bassa specificità nel differenziare la SM da malattie mimiche e non è fortemente correlato alla progressione della disabilità. Recentemente, biomarkers avanzati, tra cui lesioni corticali e placche croniche attive, sono risultati essere specifici della SM e predire la gravità della malattia. La loro valutazione, tuttavia, richiede esperti altamente specializzati ed un lungo tempo. Di conseguenza, c'è la necessità di strumenti di supporto decisionale per automatizzare l'analisi di questi biomarkers che possono migliorare la comprensione della malattia e la cura del paziente.

La presente tesi descrive lo sviluppo di diversi approcci di deep learning per l'analisi automatizzata delle lesioni della SM, in particolare lesioni corticali e placche croniche attive. I principali contributi sono i seguenti. In primo luogo, viene proposto un metodo automatizzato per la segmentazione delle lesioni corticali e della materia bianca con la RM a 3T. Basato su una rete neurale convoluzionale patch-wise, questo metodo è stato validato su un ampio dataset multi-sito di pazienti con SM, superando metodi di machine learning esistenti. In secondo luogo, questo lavoro è stato adattato alla RM ad altissimo campo, esplorando contrasti avanzati. CLAIMS (Cortical Lesion AI-based assessment in MS) è un algoritmo automatizzato che fornisce il rilevamento e la classificazione delle lesioni corticali dalla RM a 7T. In terzo luogo, è stato affrontato il problema delle sequenze di RM mancanti. Poiché la sequenza di RM MP2RAGE è fortemente desiderata per rilevare le lesioni corticali, ma non è spesso acquisita nella pratica clinica, viene introdotta una rete generativa avversariale per sintetizzare immagini MP2RAGE da MPRAGE. Questo approccio genera immagini MP2RAGE realistiche

che, rispetto ad MPRAGE, migliorano significativamente la segmentazione delle lesioni e del tessuto cerebrale ottenuta con strumenti automatici. In quarto luogo, viene proposto il primo approccio di deep learning per la classificazione delle placche croniche attive (RimNet). RimNet è stato validato con diverse sequenze di RM a 3T e 7T e le sue prestazioni sono risultate vicine a quelle degli esperti.

Per riassumere, questa tesi mira a fornire approcci di deep learning per l'analisi automatizzata delle lesioni della SM. Tutti i metodi sviluppati sono stati pensati per soddisfare specifiche esigenze cliniche. Un numero minimo di sequenze di RM clinicamente rilevanti si è dimostrato sufficiente per superare le tecniche esistenti e avvicinarsi alla valutazione manuale degli esperti. Poiché questi metodi generalizzano bene in datasets multi-sito di RM sia ad alto che ad altissimo campo, essi rappresentano il primo passo verso una valutazione automatizzata e standardizzata dei biomarkers avanzati della SM.

**Parole chiave:** sclerosi multiple, risonanza magnetica, machine learning, deep learning, lesioni corticali, placche croniche attive, segmentazione, rilevamento, RM ad altissimo campo.



# Contents

<b>Acknowledgements</b>	<b>i</b>
<b>Abstract (English/Italiano)</b>	<b>iii</b>
<b>List of figures</b>	<b>xi</b>
<b>List of tables</b>	<b>xiii</b>
<b>List of acronyms</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Thesis rationale . . . . .	2
1.3 Thesis outline . . . . .	2
<b>2 Multiple sclerosis and magnetic resonance imaging</b>	<b>5</b>
2.1 Multiple sclerosis . . . . .	5
2.1.1 Epidemiology and risk factors . . . . .	5
2.1.2 Pathology . . . . .	6
2.1.3 Phenotypes . . . . .	7
2.2 Diagnostic criteria of multiple sclerosis . . . . .	9
2.3 MRI in multiple sclerosis . . . . .	10
2.3.1 Basic MR physics . . . . .	10
2.3.2 Conventional MRI contrasts and their relative MS biomarkers . . . . .	12
2.3.3 Advanced MRI sequences . . . . .	13
2.3.4 Ultra-high field MRI . . . . .	14
2.4 Advanced imaging biomarkers in multiple sclerosis . . . . .	15
2.4.1 Cortical lesions . . . . .	15
2.4.2 The central vein sign . . . . .	17
2.4.3 Paramagnetic rim lesions . . . . .	18
	vii

<b>3</b>	<b>Automated analysis of MS lesions: state-of-the-art</b>	<b>21</b>
3.1	White matter lesion segmentation . . . . .	21
3.1.1	Unsupervised approaches . . . . .	22
3.1.2	Supervised techniques . . . . .	22
3.1.3	Current limitations . . . . .	24
3.2	Automated assessment of advanced biomarkers . . . . .	25
3.2.1	Cortical lesions . . . . .	25
3.2.2	The central vein sign . . . . .	26
3.2.3	Paramagnetic rim lesions . . . . .	27
3.2.4	Challenges for machine learning methods . . . . .	28
<b>4</b>	<b>Cortical and WM lesion segmentation at 3T MRI</b>	<b>31</b>
4.1	Introduction . . . . .	32
4.2	Materials and Methods . . . . .	35
4.2.1	Datasets . . . . .	35
4.2.2	Methodology . . . . .	36
4.2.3	Evaluation . . . . .	39
4.3	Results . . . . .	41
4.3.1	Cross-validation . . . . .	41
4.3.2	Independent test set . . . . .	42
4.4	Discussion . . . . .	44
<b>5</b>	<b>Synthetic MP2RAGE</b>	<b>49</b>
5.1	Introduction . . . . .	50
5.2	Materials and method . . . . .	52
5.2.1	Subjects . . . . .	52
5.2.2	Imaging . . . . .	52
5.2.3	Manual segmentation . . . . .	53
5.2.4	Generative adversarial network . . . . .	53
5.2.5	Pre-processing . . . . .	55
5.2.6	Training details . . . . .	55
5.2.7	Inference . . . . .	55
5.3	Evaluation . . . . .	56
5.3.1	Qualitative evaluation . . . . .	56
5.3.2	Quantitative evaluation . . . . .	56
5.3.3	Automatic segmentation . . . . .	57

## Contents

---

5.3.4	Statistical analysis . . . . .	57
5.4	Results . . . . .	58
5.5	Discussion . . . . .	61
<b>6</b>	<b>Cortical lesion detection and classification at ultra-high-field MRI</b>	<b>67</b>
6.1	Introduction . . . . .	68
6.2	Materials and methods . . . . .	70
6.2.1	Datasets . . . . .	70
6.2.2	Manual segmentation . . . . .	71
6.2.3	Pre-processing . . . . .	71
6.2.4	Convolutional neural network . . . . .	72
6.2.5	Data augmentation . . . . .	73
6.2.6	Experiments . . . . .	74
6.2.7	Evaluation and statistical analysis . . . . .	75
6.3	Results . . . . .	76
6.3.1	Single vs multi-image models . . . . .	76
6.3.2	Evaluating CLAIMS on standard MP2RAGE images . . . . .	77
6.3.3	Independent test set . . . . .	78
6.4	Discussion . . . . .	80
6.4.1	Limitations . . . . .	83
6.4.2	Future work . . . . .	83
6.4.3	Conclusion . . . . .	83
<b>7</b>	<b>Classification of paramagnetic rim lesions</b>	<b>85</b>
7.1	RimNet . . . . .	86
7.1.1	Dataset . . . . .	86
7.1.2	Pre-processing . . . . .	87
7.1.3	CNN architecture . . . . .	88
7.1.4	Training strategies . . . . .	89
7.1.5	Results . . . . .	91
7.1.6	Discussion . . . . .	94
7.2	RimNet evaluation with 3T and 7T MP2RAGE . . . . .	97
7.2.1	Methods . . . . .	98
7.2.2	Results . . . . .	98
7.2.3	Discussion . . . . .	99

<b>8 Conclusion</b>	<b>101</b>
8.1 Discussion . . . . .	101
8.1.1 Main contributions . . . . .	101
8.1.2 Scientific and clinical impact . . . . .	102
8.1.3 Limitations . . . . .	103
8.2 Future research directions . . . . .	104
8.2.1 Extensive validation . . . . .	104
8.2.2 Transfer learning . . . . .	104
8.2.3 Single framework for multiple biomarkers . . . . .	105
8.2.4 Explainable AI . . . . .	105
<b>A Appendix</b>	<b>107</b>
A.1 Other contributions . . . . .	107
A.1.1 CVSnet . . . . .	107
A.1.2 FLAWS for multiple sclerosis . . . . .	109
<b>Bibliography</b>	<b>111</b>
<b>Curriculum Vitae</b>	<b>129</b>
<b>List of publications</b>	<b>133</b>
Articles in peer-reviewed journals . . . . .	133
Peer-reviewed conference papers . . . . .	135
Abstracts in proceedings of international conferences . . . . .	135

# List of Figures

2.1	World map of multiple sclerosis prevalence . . . . .	6
2.2	Demyelination process . . . . .	7
2.3	Multiple sclerosis phenotypes . . . . .	8
2.4	Alignment of the hydrogen atoms with $B_0$ . . . . .	11
2.5	Conventional and advanced MRI contrasts . . . . .	13
2.6	Cortical lesions . . . . .	16
2.7	Central vein sign . . . . .	17
2.8	Paramagnetic rim lesion . . . . .	18
3.1	Inter-rater variability segmenting MS lesions . . . . .	22
4.1	Example of manual lesion segmentation. . . . .	34
4.2	Distribution of the lesion volume and count per patient. . . . .	37
4.3	Scheme of the 3D U-Net <sup>+</sup> architecture proposed. . . . .	38
4.4	Scheme of the two evaluation scenarios. . . . .	39
4.5	Visual illustration of results for all three methods considered. . . . .	42
4.6	Four examples of correctly detected cortical lesions by our method. . . . .	42
4.7	Boxplots of the patient-wise metrics. . . . .	43
4.8	Correlation plots between the manual lesion volume and the one automatically segmented. . . . .	43
4.9	Detection rate for cortical and white matter lesions. . . . .	44
4.10	Boxplots of the patient-wise metrics on the independent test set. . . . .	44
5.1	Scheme of the generator of the GAN proposed. . . . .	54
5.2	Scheme of the generative framework proposed. . . . .	56
5.3	Qualitative results of the synMP2RAGE. . . . .	59
5.4	Motion artefacts observed in the qualitative analysis. . . . .	60
5.5	synMP2RAGE residual images. . . . .	61

5.6	Cumulative histogram. . . . .	62
5.7	Visual results of the tissue and lesion segmentation obtained with MP2RAGE, synMP2RAGE, and MP2RAGE. . . . .	63
5.8	Tissue segmentation boxplots. . . . .	64
5.9	Lesion detection boxplots. . . . .	65
6.1	Cortical lesions at 7T. . . . .	72
6.2	Scheme of the proposed 3D UNet. . . . .	73
6.3	Visual results of cortical lesion detection. . . . .	75
6.4	Correlation between the manual and automated CL count. . . . .	78
6.5	CL detection rate for the three different CL lesion types. . . . .	79
6.6	False positive and detection rate in an external testing dataset. . . . .	80
7.1	Visual examples of PRL. . . . .	87
7.2	Scheme of the process used to annotate rim+ and rim- lesions. . . . .	90
7.3	RimNet architecture. . . . .	91
7.4	ROC and PR curves for all network configurations. . . . .	92
7.5	ROC and PR curves of RimNet in the inter-center study. . . . .	92
7.6	Patient-wise analysis of RimNet results. . . . .	93
7.7	RimNet errors analysis. . . . .	95
7.8	Examples of PRL at 7T. . . . .	97
7.9	ROC and PR curves of RimNet at 7T. . . . .	99
7.10	Lesion-wise confusion matrices. . . . .	99
A.1	Scheme of CVSnet. . . . .	108
A.2	Cortical lesions on MP2RAGE and FLAWS. . . . .	109

## List of Tables

2.1	Requirements of the 2017 McDonald criteria for MS diagnosis . . . . .	9
2.2	MRI sequences used for advanced MS biomarkers . . . . .	15
3.1	List of the main white matter lesion segmentation approaches . . . . .	23
4.1	Metrics obtained for the different methods on the cross-validation evaluation .	41
4.2	Metrics obtained for the different methods on the independent test set . . . . .	43
5.1	Quantitative metrics for MPRAGE and synMP2RAGE compared to the reference MP2RAGE. . . . .	59
5.2	Cortical and white matter lesion detection rates. . . . .	62
6.1	Median metrics for the different models . . . . .	76
6.2	Detection rate on a patient-wise level. . . . .	77
6.3	Correlation between manual and automated CL count. . . . .	79
6.4	Median metrics for the different models . . . . .	80
7.1	RimNet datasets description . . . . .	88
7.2	RimNet MRI protocol . . . . .	89





## List of Acronyms

<b>2D</b>	2-Dimensional
<b>3D</b>	3-Dimensional
<b>AI</b>	Artificial intelligence
<b>ANTs</b>	Advanced normalization tools
<b>AVD</b>	Absolute volume difference
<b><math>B_0</math></b>	External magnetic field
<b>CNS</b>	Central nervous system
<b>CSF</b>	Cerebrospinal fluid
<b>CVS</b>	Central vein sign
<b>CL</b>	Cortical lesion
<b>CLAIMS</b>	Cortical Lesion AI-based assessment in Multiple Sclerosis
<b>CNN</b>	Convolutional neural network
<b>DL</b>	Deep learning
<b>DIR</b>	Double inversion recovery
<b>DSC</b>	Dice coefficient
<b>EDSS</b>	Expanded Disability Status Scale
<b>EPI</b>	Echo planar imaging
<b>FLAIR</b>	Fluid Attenuated Inversion Recovery
<b>GAN</b>	Generative adversarial network
<b>Gd</b>	Gadolinium
<b>GRE</b>	Gradient echo
<b>IQR</b>	Interquartile range
<b>IR-SWIET</b>	Inversion recovery SWI with enhanced T2-weighting
<b>LFPR</b>	Lesion-wise false positive rate
<b>LTPR</b>	Lesion-wise true positive rate

<b>M</b>	Magnetization
<b>ML</b>	Machine learning
<b>MPRAGE</b>	Magnetization-Prepared RApid Gradient Echo
<b>MP2RAGE</b>	Magnetization-Prepared 2 RApid Gradient Echoes
<b>MRI</b>	Magnetic resonance imaging
<b>MS</b>	Multiple sclerosis
<b>NRMSE</b>	Normalized root mean square error
<b>PD</b>	Proton density
<b>PPMS</b>	Primary-progressive multiple sclerosis
<b>PPV</b>	Positive predicted value
<b>PRL</b>	Paramagnetic rim lesion
<b>PSIR</b>	Phase-sensitive inversion-recovery
<b>PSNR</b>	Peak signal-to-noise ratio
<b>RF</b>	Radio frequency
<b>RIS</b>	Radiologically isolated syndrom
<b>RRMS</b>	Relapsing-remitting multiple sclerosis
<b>synMP2RAGE</b>	Synthetic MP2RAGE
<b>SPMS</b>	Secondary-progressive multiple sclerosis
<b>SSIM</b>	Mean structural similarity index
<b>SWI</b>	Susceptibility weighted imaging
<b>T</b>	Tesla
<b>T1</b>	Longitudinal relaxation time
<b>T1 map</b>	T1 relaxation map
<b>T2</b>	Transverse relaxation time
<b>TA</b>	Acquisition time
<b>TE</b>	Echo time
<b>TI</b>	Inversion time
<b>TR</b>	Repetition time
<b>UNI</b>	MP2RAGE $T_1$ -weighted uniform contrast
<b>WM</b>	White matter
<b>WML</b>	White matter lesion

# 1 Introduction

## 1.1 Motivation

Multiple sclerosis (MS) is the most common demyelinating disease of the central nervous system and it currently affects almost 3 million people worldwide, with a rising prevalence over the last decades. With an average age of onset of approximately 32 years old, both younger and older people can experience MS. Its symptoms start with fatigue and weakness and frequently progress into serious disabilities and cognitive impairments. As of today, there is no cure for MS, although several treatments are able to slow down its progression. Thus, an early diagnosis and frequent follow-up assessments are of crucial importance. Currently, MS is diagnosed considering different analyses, including neurological exams, blood tests, lumbar puncture, and imaging. Magnetic resonance imaging (MRI) is the technique of choice to identify brain abnormalities such as the presence of white matter lesions which represent the main conventional biomarker of MS. Present-time diagnostic criteria, however, lack specificity as they were suggested to identify subjects with a high likelihood of MS rather than distinguish MS from other neurological disorders. Recently, studies have shown that advanced MRI biomarkers such as cortical lesions and paramagnetic rim lesions are highly specific to MS and can help differentiate MS from mimic diseases. Moreover, they can potentially support prognosis, as both correlate with the severity of patients' symptoms stronger than conventional biomarkers. However, to assess these advanced biomarkers dedicated MRI sequences at high (3T) or ultra-high (7T) magnetic field as well as highly specialized experts are required.

In the last years, the boom of deep learning has revolutionized automated image analysis tools, including those applied to biomedical images. Several approaches have been developed for the segmentation of white matter MS lesions using conventional MRI sequences. Nevertheless, deep learning techniques have barely been applied to advanced imaging biomarkers or ultra-

high magnetic field imaging, which both present additional challenges. As such, there is an unmet need of developing robust and automated tools that could contribute to standardizing the assessment of these advanced biomarkers and facilitate their analysis in research and clinical environments.

In this thesis, several deep learning-based techniques are proposed for the automated analysis of MS lesions, including advanced biomarkers such as cortical lesions and paramagnetic rim lesions. For this scope, we considered conventional and specialized MRI sequences acquired at both high and ultra-high magnetic field in different clinical centers as well as the manual biomarkers' assessment performed by multiple experts.

## 1.2 Thesis rationale

The main hypotheses of the present thesis are the following:

- A minimal number of conventional and specialized MRI sequences is sufficient for automated approaches to accurately assess advanced MS lesional biomarkers.
- Deep learning-based methods have the potential to support experts in the analysis of advanced biomarkers for the diagnosis and prognosis of MS.
- These methods generalize well with multi-site datasets, different sequences, and high and ultra-high field MRI.

## 1.3 Thesis outline

The content of this thesis is organized as follows:

**Chapter 2:** an introduction to MS and MRI is provided. After a brief summary of MS epidemiology and pathology, the current diagnostic criteria are discussed. MRI is then introduced, including conventional and specialized MRI sequences as well as ultra-high field MRI. This chapter ends with a description of three advanced MS biomarkers: cortical lesions, the central vein sign, and paramagnetic rim lesions.

**Chapter 3:** an overview of the current state-of-the-art methods for the automated assessment of MS lesions is proposed. First, automated approaches performing white matter lesion segmentation, either in a supervised or unsupervised way, are introduced. Second, methods aiming to detect cortical lesions based on either 3T or 7T MRI are presented,

and their strengths and weaknesses are discussed. Finally, the few machine learning-based approaches proposed in the literature for the analysis of the central vein sign and paramagnetic rim lesions are discussed (review paper currently under revision in *Neuroimage: Clinical*).

**Chapter 4:** a novel and fully automated framework for the segmentation of both cortical and white matter lesions with 3T MRI is presented. The framework is based on a convolutional neural network that takes as input a conventional MRI contrast (FLAIR) and a specialized one (MP2RAGE). This work was published as "Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE" in *Neuroimage: Clinical*, 2020.

**Chapter 5:** a generative approach for synthesizing MP2RAGE images from MPRAGE ones is proposed. This method was trained with both MS patients and healthy controls and evaluated testing automated segmentation pipelines. This work was published as "MPRAGE to MP2RAGE UNI translation via generative adversarial network improves the automatic tissue and lesion segmentation in multiple sclerosis patients" in *Computers in Biology and Medicine (CIBM)*, 2021.

**Chapter 6:** an automated approach for the detection and classification of cortical lesions with 7T MRI is presented. This was the outcome of my secondment in Daniel Reich's lab at the National Institutes of Health (NIH). The work was initially presented at the virtual International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) in 2020, and then further improved taking into account clinicians' feedbacks. The relative manuscript is currently under revision as "Multiple sclerosis cortical lesion detection with deep learning at ultra-high-field MRI" in *Nuclear Magnetic Resonance (NMR) in Biomedicine*.

**Chapter 7:** our proposed approach (RimNet) for the semi-automated classification of paramagnetic rim lesions is presented. In this study, we developed a patch-based convolutional neural network that takes as input different MRI contrasts, including FLAIR and phase 3D-EPI, and classify each lesion as having a paramagnetic rim or not. Moreover, our method was evaluated with MP2RAGE uni and T1map images acquired at both 3T and 7T (work presented at ISMRM in 2021).

**Chapter 8:** to conclude, a summary of the main contributions of the present thesis, its limitations, and future research directions are presented.

This thesis includes the work conducted by Francesco La Rosa in four years of his Ph.D. at the

Signal Processing Laboratory 5 (LTS5), École polytechnique fédérale de Lausanne (EPFL) and at the Medical Image Analysis Laboratory (MIAL), Lausanne University Hospital (CHUV) and University of Lausanne (UNIL). Collaborations with the following national and international institutions were crucial for the development of the present work: University Hospital Basel and University of Basel, Basel, Switzerland; University of Bern, Bern, Switzerland; Advanced Clinical Imaging Technology, Siemens Healthcare AG, Lausanne, Switzerland; Université catholique de Louvain (UCL), Louvain-la-Neuve, Belgium; National Institutes of Health (NIH), Bethesda, United States; University of Pennsylvania, Philadelphia, United States; Cedars-Sinai Medical Center, Los Angeles, United States.

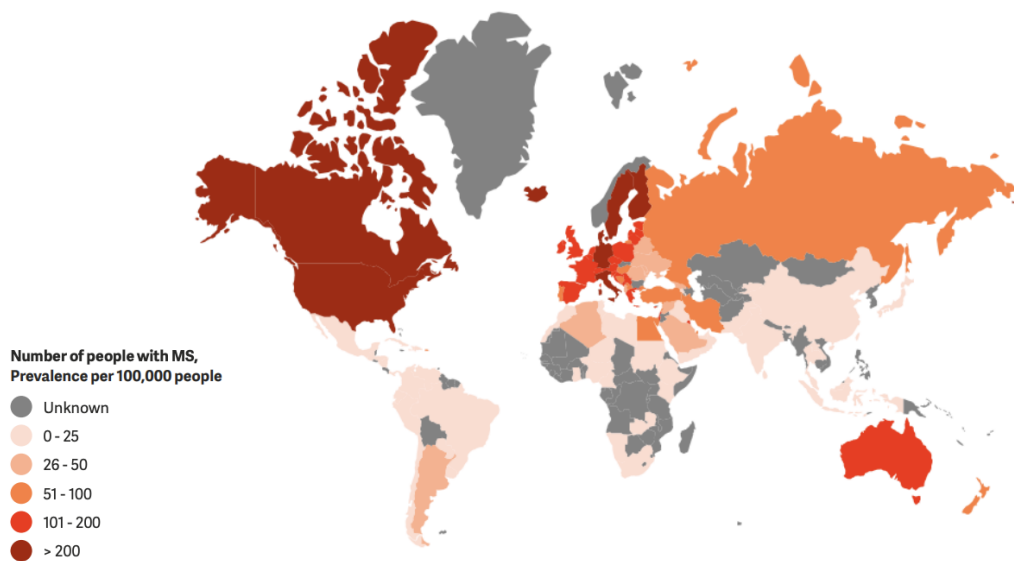
This work was supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie project TRABIT (agreement No 765148) and the Novartis Research Foundation. We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Ecole Polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG).

## **2 Multiple sclerosis and magnetic resonance imaging**

### **2.1 Multiple sclerosis**

#### **2.1.1 Epidemiology and risk factors**

Multiple sclerosis (MS) is the most common demyelinating disease affecting the central nervous system (CNS) (Reich et al., 2018). It consists of an inflammatory disorder of the brain and spinal cord in which focal lymphocytic infiltration causes myelin and axonal injury (Compston & Coles, 2008). This results primarily in the formation of lesions, also called plaques, in the CNS. MS can develop at any age but mostly occurs in young adults, with a mean age of diagnosis of 32 years (Walton et al., 2020). During its long clinical course, MS has a huge impact on patients' quality of life, causing cognitive and physical impairments. As of today, approximately 2.8 million people worldwide are suffering from MS. Epidemiological studies on MS mortality have demonstrated that about 70-88% of patients are still alive after 25 years of the disease onset. Their life quality, however, is seriously degraded and a reduction in life expectancy of approximately 10 years is observed (Oh et al., 2018). Females are 2 to 3 times more likely to be affected by MS than males. Currently, there is no cure for MS, and an early diagnosis of the disease is crucial to start treatment and have the best chances of slowing down the disease progression (Thompson, Banwell, et al., 2018). Misdiagnosis, however, remains a persisting problem, despite efforts over the past decades to improve the diagnostic criteria (Brownlee & Solomon, 2021; Solomon & Corboy, 2017). A single etiologic trigger to MS has not been found yet, but various genetic and environmental factors have been linked to its development (Compston & Coles, 2008; Reich et al., 2018). Studies have examined over 200 gene variants that increase the risk of experiencing the disease, with the predominant HLA DRB1\*1501 haplotype raising the risk approximately 3-fold. Focusing on



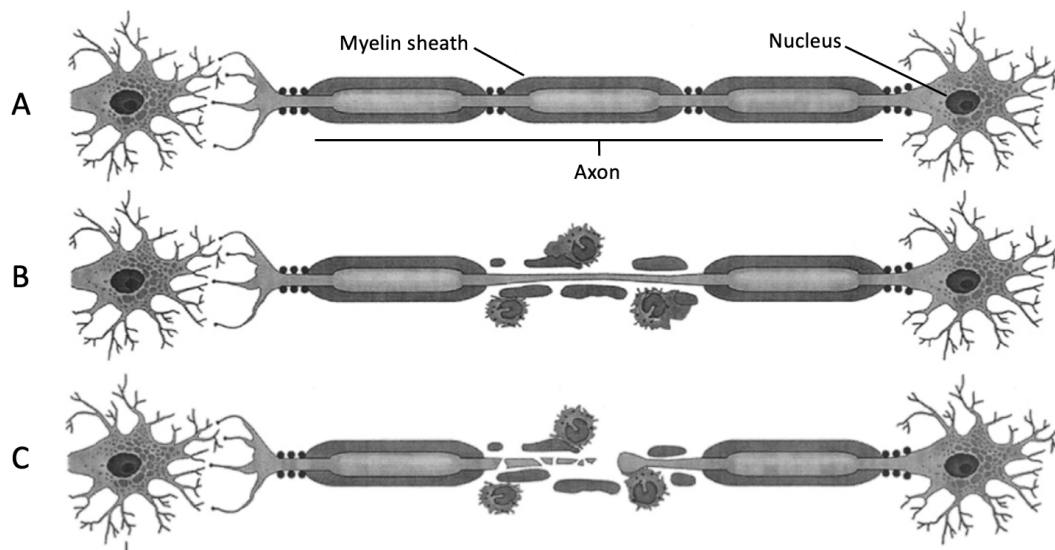
**Figure 2.1** – World map of MS prevalence per 100,000 people updated in 2020. Image retrieved from (Walton et al., 2020).

the environmental factors, it has been observed that the distribution of MS decreases with a decreasing distance from the equator (see Figure 2.1), which might suggest a contribution of sunlight exposure changing the vitamin D levels in the human body. Other enhanced risk factors include Epstein–Barr virus (EBV) infection, mononucleosis, obesity, smoking, and tobacco exposure (Jacobs et al., 2020).

### 2.1.2 Pathology

The biological process of MS is thought to start with T-cells crossing the blood-brain barrier and triggering an autoimmune response in the CNS. As a consequence, there is a progressive inflammatory process that causes demyelination, gliosis, and neuroaxonal degeneration (see Figure 2.2). In some cases, this disrupts as well the neuronal signalling (Dendrou et al., 2015). Multiple areas of scar tissue (referred to as lesions) arise both in the gray (GM) and in the white matter (WM) and these are the hallmark of MS. These pathological abnormalities directly relate to clinical symptoms (Thompson, Baranzini, et al., 2018). For instance, an inflammatory lesion showing a loss of the myelin layers surrounding the optic nerve might result in optic neuritis, whose symptoms are visual impairments such as blurring and blind spots. Moreover, demyelinated axons might become hyperexcitable and randomly start impulses resulting in the perception of flashing light. In certain cases, these clinical symptoms are to a certain extent reversible thanks to the spontaneous processes of remyelination and neuronal plasticity.





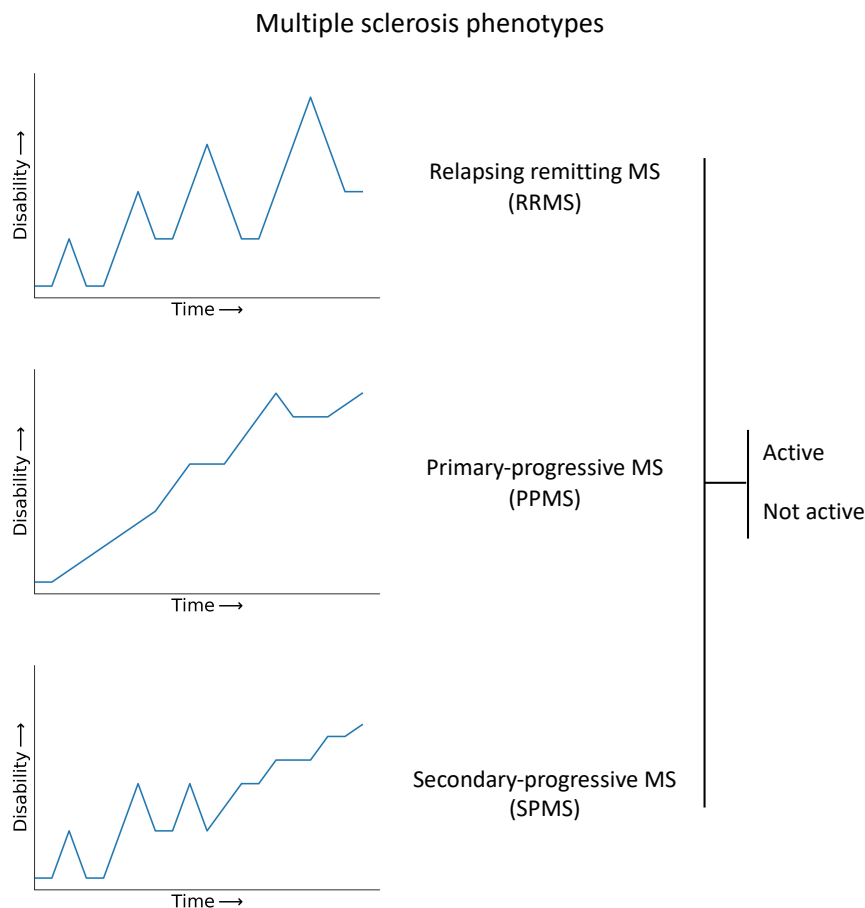
**Figure 2.2** – Demyelination process due to MS. A: normal axon. B: inflammatory demyelination. C: axonal transection. Figure adapted with permission from Trapp et al., 1999.

Remyelination is the mechanism by which new myelin sheaths are naturally generated around axons in the CNS, restoring nerve conduction. It is very efficient in healthy subjects and occasionally occurs also in MS lesions, with a higher frequency for these formed in younger people (Reich et al., 2018). By neuronal plasticity, it is meant the ability of the brain to adapt dynamically to environmental and pathological stimuli, such as, for example, a cortical reorganization to maintain clinical functions. Studies have shown that this happens also in MS patients and, therefore, it is a potential target for future therapies which aim to accelerate this process (Keough & Yong, 2013).

### 2.1.3 Phenotypes

An additional challenge of MS is given by its heterogeneous phenotype, consisting in a broad spectrum of symptoms experienced by the patients. At the early stages, the main ones are represented by fatigue and headache, whereas as the disease progresses they often range from visual problems to ambulatory and cognitive impairments. Overall, MS is the main cause of non-traumatic disability for young adults (Koch-Henriksen & Sørensen, 2010). To quantify the level of disability and monitoring changes over time, the Expanded Disability Status Scale (EDSS) has been established (Kurtzke, 1983). The EDSS ranges from 0 to 10 representing increasing neurological impairments. At 0 the patient experiences no symptoms, whereas 10 means death due to MS. Three main clinical phenotypes of MS have been identified by an international consensus (Lublin et al., 2014): relapsing-remitting (RRMS), primary-progressive

(PPMS), and secondary-progressive (SPMS) (see Figure 2.3). The majority of patients is affected by RRMS, meaning that they experience some sudden symptoms or "attacks" followed by periods of partial or total recovery that can last months or even years. About half of RRMS patients eventually become affected by a steady progression of the symptoms and this represents the SPMS stage. On the other hand, the PPMS phenotype is characterized by a steady accumulation of symptoms without relapses or remissions. All these three disease phenotypes can be further characterized either as active, if an MRI shows new lesions or other MS biomarkers, or not active when the symptoms increase, but no new lesions are visible (Lublin et al., 2014). In addition to these three main MS courses, individuals with MRI findings that suggest MS, such as white matter lesions (WML), but no symptoms related to MS, are referred to as experiencing radiologically isolated syndrome (RIS) (Okuda et al., 2009). These individuals are likely to be affected by subsequent symptoms related to MS in a later stage.



**Figure 2.3** – Scheme of the three main MS phenotypes. Each disease course is considered as active if clinical relapses or MRI findings are observed, otherwise it is referred to as not active (Lublin et al., 2014).

**Table 2.1** – Lesion dissemination in space and time requirements of the 2017 McDonald criteria for MS diagnosis. All lesions must have objective clinical evidence. Table adapted with permission from (Thompson, Banwell, et al., 2018)).

Disease course	Dissemination in time	Dissemination in space
<b>RRMS</b>	≥ 2 clinical attacks	≥ 2 lesions.
	≥ 2 clinical attacks	1 lesion (as well as clear-cut historical evidence of a previous attack involving a lesion in a distinct anatomical location).
	≥ 2 clinical attacks	1 lesion. Dissemination in space demonstrated by an additional clinical attack implicating a different CNS site or by MRI.
	1 clinical attack	≥ 2 lesions. Dissemination in time demonstrated by an additional clinical attack OR by MRI or demonstration of CSF-specific oligoclonal bands.
<b>PMS</b>	1 clinical attack	1 lesion. Dissemination in space demonstrated by an additional clinical attack implicating a different CNS site or by MRI AND dissemination in time demonstrated by an additional clinical attack or by MRI OR demonstration of CSF-specific oligoclonal bands.
	≥ 1 year of disability progression (retrospectively or prospectively determined) independent of clinical relapse.	Two of the following criteria: <ul style="list-style-type: none"> <li>• One or more T2-hyperintense lesions characteristic of MS in one or more of the following brain regions: periventricular, cortical or juxtacortical, or infratentorial.</li> <li>• Two or more T2-hyperintense lesions in the spinal cord.</li> <li>• Presence of CSF-specific oligoclonal bands.</li> </ul>

## 2.2 Diagnostic criteria of multiple sclerosis

Jean Martin Charcot delineated the first MS diagnostic criteria in 1868 (Poser & Brinar, 2004). While delivering a series of lectures, the French neurologist introduced MS as a distinct disease and described common MS symptoms such as tremor and cognitive problems. In the following years, together with his colleague Edme Vulpian, he identified the characteristic

plaques or lesions of MS and observed the typical demyelination process (Pearce, 2005). Since then, MS diagnostic criteria have evolved quickly, primarily thanks to advances in laboratory examinations. In 1981, with the introduction of MRI that allowed to see demyelinating lesions in-vivo, they were revolutionized, giving much importance to MRI findings (Young et al., 1983). In 2001, an international committee has established the McDonald criteria, describing MRI as the imaging technique of choice to demonstrate lesion dissemination in space and time (McDonald et al., 2001). These criteria have been extensively revised in 2005, 2010, and 2017. In the latest revision (Thompson, Banwell, et al., 2018), the international consensus increased the relative importance of cerebrospinal fluid (CSF) analysis and instructed practitioners and researchers to pay additional attention to misdiagnosis and differential diagnosis. Moreover, not only WML but also cortical lesion (CL) count was recommended to be considered for the lesion dissemination in space and time criteria (Thompson, Banwell, et al., 2018). A summary of the 2017 McDonald criteria is proposed in Table 2.1.

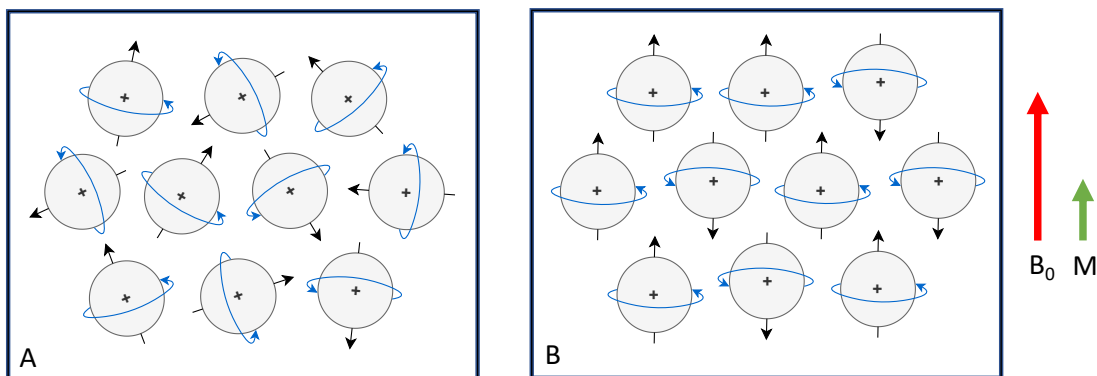
## **2.3 MRI in multiple sclerosis**

In this section, MRI in the context of MS is presented. After a brief introduction to MR physics, the different MRI sequences acquired to diagnose and monitor MS over time are described. Finally, ultra-high field MRI, meaning that the acquisition is performed using a  $\geq 7\text{T}$  scanner, is discussed.

### **2.3.1 Basic MR physics**

MRI is a non-invasive imaging modality that provides highly detailed images of the organs and tissues of the human body, capturing the density and magnetic properties of hydrogen atoms. MRI scanners acquire images creating a strong magnetic field and magnetic field gradients. They are commonly regarded as a safer option compared to computed tomography (CT) whose scanners generate radiation. Typically, clinical MRI scanners have a magnetic field strength between 0.5 and 3T, although 7T scanners have recently been approved for clinical use as well (Cosottini & Roccatagliata, 2021). Once a patient enters the scanner, the hydrogen atoms within its body align with the static magnetic field ( $B_0$ ) created by the machine (see Figure 2.4). According to the classical description of the effect, the protons of these atoms arrange themselves either in line or in the opposite direction of the magnetic field (two energy eigenstates), while spinning around their main axes (precession). The magnetic moments of these two states almost cancel each other out, however, a small unbalanced magnetization  $M$ , aligned with  $B_0$ , remains, and this is the magnetic field measured using MRI (Figure 2.4). At

this point, the protons precess at the "Larmor" frequency, which is given by the gyro-magnetic ratio of the particles multiplied by the magnitude of  $B_0$ . The protons' spins can then be "excited", i.e. their net polarization vector flipped to the orthogonal plane or even reversed, applying radio frequency (RF) pulses at the same Larmor frequency. This also synchronizes the phases of all atomic dipoles. When the RF pulses are turned off, the protons slowly realign themselves with  $B_0$ . The process of recovering their longitudinal net magnetization is referred to as T1 relaxation, whereas the spins gradually losing phase coherence in the transverse plane is termed T2 relaxation. T1 is then represented by the time the longitudinal magnetization needs to reach equilibrium, whereas T2 reflects the time taken by the spins to go out of phase with each other in the transverse plane. The T2 relaxation time is usually much shorter than T1. During the realignment process, some electromagnetic energy is released and measured by the receiver coils of the scanner. The RF pulses are then repeated several times to acquire signals from the entire area of the tissue of interest. Finally, the signal is processed and gray scale images are generated applying the inverse Fourier transforms. The magnitude of the signal, and therefore the image contrast, depends on several components, including the imaging setup and timing, the density of the excited nuclei, and their different relaxation times. T1 and T2 vary considerably for different tissue types and this allows to visualize the anatomy or pathology of interest. Changing the set of acquisition parameters, extremely different contrasts are generated, mainly termed T1-weighted, T2-weighted, or proton density (PD) weighted, highlighting the T1, T2, and proton density of the tissue, respectively. Images can be acquired both in 2D or 3D depending on the way the spatial information of the spins is encoded.



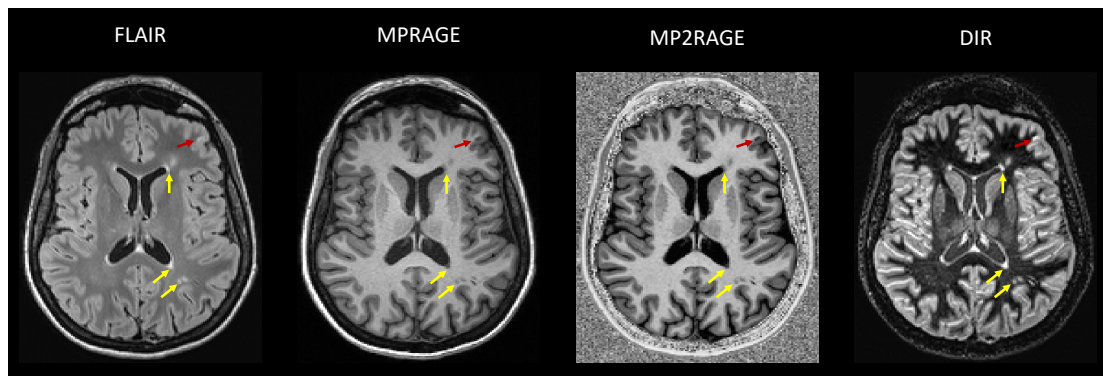
**Figure 2.4** – Alignment of the hydrogen atoms with  $B_0$ . A: no external magnetic field is present and the atoms are randomly oriented. B: when placed in a strong magnetic field ( $B_0$ ), the atoms align themselves either parallelly or anti-parallelly to  $B_0$ . In both cases, they precess around their main axis (blue arrows). In B, their frequency of precession is the Larmor frequency.

### **2.3.2 Conventional MRI contrasts and their relative MS biomarkers**

MRI, first introduced in the hospitals in the 1980s, has revolutionized the diagnosis of MS, allowing to visualize in vivo tissue abnormalities and lesion activity in the brain and spinal cord. As the technology advanced, the visualization of inflammatory lesions became fundamental to diagnose the disease and monitor its progression, making MRI a routine clinical examination for MS. As of today, MRI plays a crucial role in the diagnosis, follow-up, and monitoring of the disease, although its conventional findings are not highly specific of MS (Solomon et al., 2019). The MRI sequences acquired to visualize MS lesions can be divided into two main categories: "conventional" and "specialized" or "advanced" (Hemond & Bakshi, 2018). The latter represents pulse sequences that have been recently developed at high or ultra-high magnetic fields (3T or 7T) to increase the specificity of MS diagnosis. The former includes the protocols incorporated in the consensus recommendations on the use of MRI with MS patients (Wattjes et al., 2021), which are currently acquired in most clinics. Those are T2-weighted, fluid-attenuated inversion recovery (FLAIR), proton density weighted (PD), and T1-weighted pre and post-gadolinium contrast pulse sequences. An example of MS WML observed in different sequences is shown in Figure 2.5. Regarding the spatial distribution in the brain, MS lesions appear with higher frequency in the periventricular WM (Compston & Coles, 2008). Other regions where the lesions are frequently located are the corpus callosum, subcortical region, brain stem, U-fibers, optic nerves, and visual pathway. At the early stages of the disease they are thin and linear, but then over time they potentially change shape, acquiring an ovular appearance (Compston & Coles, 2008). The FLAIR sequence is a conventional T2-weighted acquisition with long TE and TR in order to suppress the signal from the CSF and therefore enhance abnormalities in the WM. It is one of the most common sequences used by physicians for visualizing lesions in the WM, where those appear hyperintense compared to the surrounding tissues. Despite its high lesion sensitivity, FLAIR main drawback consists in the lack of specificity (Hemond & Bakshi, 2018). Lesions caused by several pathological processes, including demyelination, inflammation, and edema, all appear as T2 hyper-intensities. Furthermore, it has been shown that the number and volume of these lesions do not directly correlate with the severity of patients' symptoms and the clinical stage of the disease, as expressed, for instance, by the EDSS score (Barkhof, 1999). T2-weighted pulse sequences, however, remain crucial for the diagnosis, but further acquisitions are needed in order to have a better understanding of the disease.

T1 sequences allow visualizing the contrast between water-predominant structures and those with a high content of fat, such as lipid and myelin. As a consequence of demyelination

processes, which reduce the fat content of axons, MS lesions appear as hypointense spots on T1-weighted sequences. A typical 3D T1 sequence acquired in the clinical routine is the magnetization-prepared rapid gradient-echo (MPRAGE) (Brant-Zawadzki et al., 1992). In addition, routine protocols often include also the injection of a contrast agent, usually gadolinium. T1 post-contrast sequences offer a measure of disease activity. Gadolinium (Gd) enhancement is the result of the blood-brain barrier breakdown and shows the active inflammation of the lesions. New MS lesions present an inflammatory phase which usually lasts 2-6 weeks and in this period their detection can be crucial for an early disease diagnosis (Compston & Coles, 2008). In certain cases, however, some lesions have a T1-hypointensity on pre and post-gadolinium acquisition, while always appearing as hyperintense on T2-weighted sequences. These lesions are commonly referred to as "black holes" and show an area of severe demyelination and axon damage (Tam et al., 2012). Observing black holes is a way of measuring tissue destruction, with the signal becoming darker with greater injury and less hypointense in case remyelination happens or the edema is reduced. These measurements, however, can be highly influenced by artifacts, especially when thick slices of 3 mm or more are acquired. Furthermore, a direct relationship between the black hole volume and the impairments of the patients has not been confirmed yet. Overall, both T1 and T2-weighted sequences are clinically employed to determine the total lesion volume and number of MS lesions. The combination of these two values is usually referred to as lesion burden and provide a way of monitoring disease progression and treatment effects over time.



**Figure 2.5** – Example of an axial view of both conventional and advanced MRI contrasts from an MS patient. Yellow arrows point to white matter lesions and the red arrow to a cortical lesion.

### 2.3.3 Advanced MRI sequences

As briefly mentioned above, conventional MRI contrasts have several limitations. They do not allow to distinguish different pathologies such as areas of inflammation, demyelination, and

edema, they modestly characterize the level of injury in demyelinated regions, and hardly show cortical demyelination (Fox et al., 2011). For these reasons, several specialized MRI sequences and techniques have been developed in the last few years. Restricting our analysis to structural MRI, various contrasts have been proposed, for instance, to increase the sensitivity to cortical lesions at both 3T and 7T. These include phase-sensitive inversion recovery (PSIR), double inversion recovery (DIR), magnetization prepared with 2 rapid gradient echoes (MP2RAGE) (Marques et al., 2010), and inversion recovery susceptibility weighted imaging with enhanced T2-weighting (IR-SWIET) (Beck et al., 2020). Often a combination of multiple sequences was shown to increase the visual cortical lesion detection compared to the analysis of a single contrast (Beck et al., 2020; Beck et al., 2018). Since 2017, when CL count was included in the MS diagnostic criteria (Thompson, Banwell, et al., 2018), these sequences have raised great interest, and some of them are now suggested as optional acquisitions by an international committee (Wattjes et al., 2021). Recently, susceptibility weighted imaging (SWI) has also gained attention. This refers to T2\* contrasts that having a long echo time are sensitive to veins, hemorrhage, and iron storage. Both the T2\* signal loss and the phase accumulation are considered in order to generate two different images, referred to as the T2\* magnitude and phase images, respectively. An example of SWI sequence used in MS research studies is the Segmented 3D Echo-Planar Imaging (3D EPI) (Sati et al., 2014), which has lately been explored for CL detection (Beck et al., 2018), central vein sign classification (Sati et al., 2016), and the identification of paramagnetic rim lesions (Absinta et al., 2019). A summary of the main MRI sequences used to detect these three advanced imaging biomarkers is presented in Table 2.2.

### 2.3.4 Ultra-high field MRI

Over the last years, ultra-high field (7T) MRI has become an increasingly important tool in investigating MS in-vivo. Since 2017, two 7T scanners from different manufacturers have been approved for clinical use (Cosottini & Roccatagliata, 2021). Their main advantages, compared to 3T MRI, are a higher spatial resolution and increased signal-to-noise and contrast-to-noise ratio that are improving our understanding of MS physiopathology (Bruschi et al., 2020). Importantly, 7T MRI allows detecting smaller WML lesions, often at earlier stages of the disease. Moreover, it has considerably improved the in-vivo detection and characterization of cortical lesions (Maranzano et al., 2019), which are associated with disability progression, and helped describe the common and widespread cortical demyelination in MS patients (Bruschi et al., 2020). In particular, studies have shown that 3T MRI allows a similar detection of leukocortical lesions than 7T MRI, however, it is significantly less sensitive to intracortical and subpial lesions (Maranzano et al., 2019). In addition, ultra-high field MRI has improved



the characterization of the lesion formation process and inflammation within the meninges. Finally, 7T T2\*-weighted sequences have been crucial to study advanced imaging biomarkers such as the CVS and PRL that have the potential of reducing MS misdiagnosis rate (Absinta et al., 2013; Tallantyre et al., 2011). Although 7T technology has been advancing quickly, reducing image artifacts and accelerating high-resolution scan protocols (Sati & Reich, 2021), challenges to its widespread use in clinical practice remain. Other than the economical burden for the installation of a 7T scanner, the use of ultra-high-field MRI is currently still limited by technical issues and the need of changing common protocols (Cosottini & Roccatagliata, 2021). In this sense, guideline recommendations to image MS patients at 7T, which are currently missing, would be very extremely helpful.

**Table 2.2** – Summary of the main MRI sequences currently used to manually detect cortical lesions, the central vein sign, and paramagnetic rim lesions at both 3T and 7T (Beck et al., 2020; Magliozzi et al., 2018; Ng Kee Kwong et al., 2021; Sati et al., 2016).

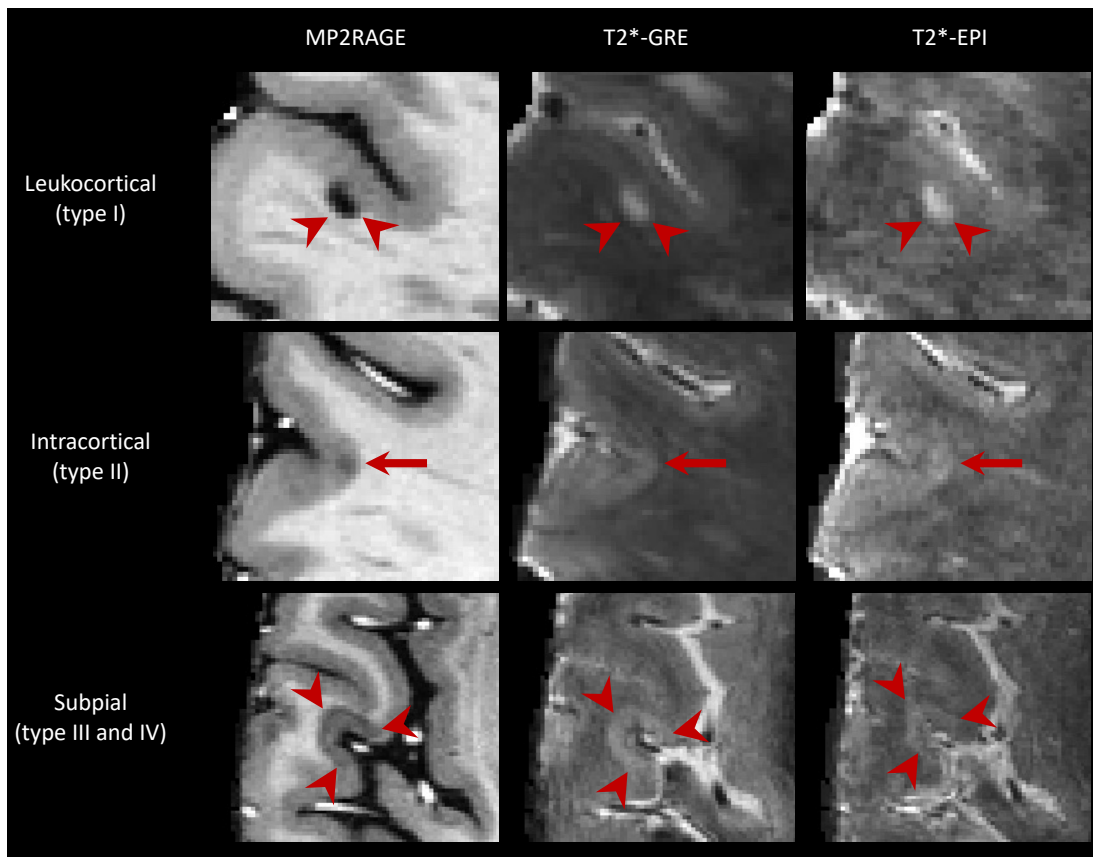
	3T	7T
Cortical lesions	MPRAGE, MP2RAGE, PSIR, DIR, IR-SWIET.	MP2RAGE, T2* GRE, T2* EPI, IR-SWIET.
Central vein sign	FLAIR*, T2* GRE, T2* EPI, SWI.	SWI, T2* GRE, T2* EPI.
Paramagnetic rim lesions	SWI, QSM, phase T2* GRE, phase T2* EPI.	SWI, QSM, phase T2* GRE, phase T2* EPI.

## 2.4 Advanced imaging biomarkers in multiple sclerosis

In this section, three novel biomarkers which have recently gained attention in the MS research and clinical community are discussed. First, cortical lesions, which since 2017 are included in the MS diagnostic criteria. Second, the central vein sign, a biomarker that could potentially help distinguish MS from mimic diseases and reduce MS misdiagnosis rate. Finally, paramagnetic rim lesions, potentially a prognostic biomarker of MS useful for identifying more aggressive disease courses and monitoring the effects of treatment.

### 2.4.1 Cortical lesions

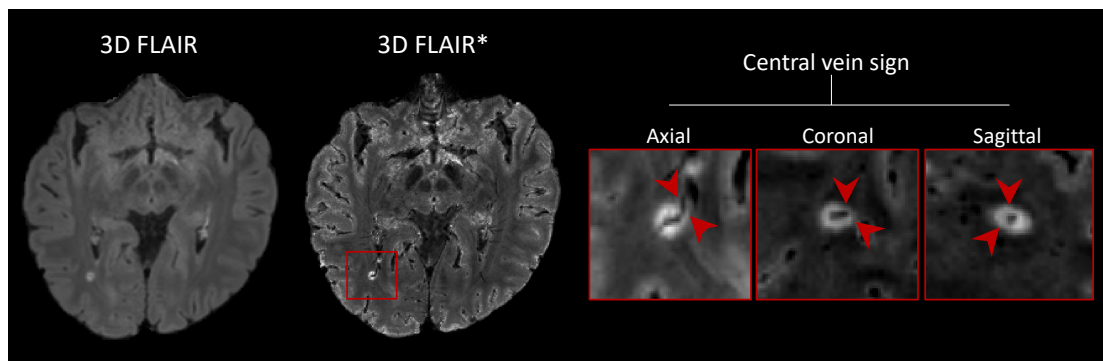
CL are a type of MS lesions that involve, at least partially, the gray matter. They have been classified into three main categories (see Figure 2.6) (Calabrese et al., 2010): leukocortical lesions



**Figure 2.6** – Example of the three main subtypes of cortical lesions imaged at 7T with MP2RAGE, T2\*-GRE and T2\*-EPI. The MP2RAGE shown was obtained as an average of four acquisitions.

are located at the interface between WM and gray matter (GM) (type I), intracortical lesions are purely in the cortex and do not reach the pial surface (type II), and subpial lesions located over one gyrus (type III) or reach the pial surface (type IV). They represent an interesting piece of information for diagnosing MS because of several reasons. First, studies have shown that they can be detected in the early stages of MS (Calabrese et al., 2009; Kidd et al., 1999). Second, the number of CL is linked with physical and cognitive disabilities of MS patients (Harrison et al., 2015), positively correlating with EDSS score and with WM T2 lesion volume, and brain atrophy (Favaretto et al., 2016). Third, it has been observed that CL are associated with disease progression stronger than WM lesions (Calabrese et al., 2009; Favaretto et al., 2016). Fourth, they were found in patients with any of the major MS phenotypes, including 35-40% of patients with CIS (Madsen et al., 2021). All these factors make the understanding of CL crucial in increasing our knowledge of MS and improving its early diagnosis. As mentioned above, a classification approach involving four types of cortical lesions is currently accepted (Calabrese et al., 2010). There is a moderate regional dependence in the distribution of CL. Extensive demyelination is observed in MS patients in the cingulate gyrus and in the temporal

and frontal cortices. The hippocampus is also another preferential area that is believed to show that active brain regions might be more prone to inflammatory or neurodegenerative pathological processes. However, further studies are needed before all factors that cause this specific lesion localization in the brain can be fully understood. Regarding the inflammatory pathological processes, CL independently contribute with several different factors from WML. In particular, CL show low levels of cell infiltration, do not present a local blood-brain barrier dysfunction, and involve a low myelin density. For these reasons, they do not appear usually in routine MRI sequences; even the FLAIR has proved to be insufficient for their detection. A postmortem study showed that 3D FLAIR could detect about 41% of mixed GM-WM lesions and 5% of the purely cortical ones (Geurts, Bo, et al., 2005). This supports the need of specialized MRI techniques for identifying CL (Beck et al., 2020; Kober et al., 2012).

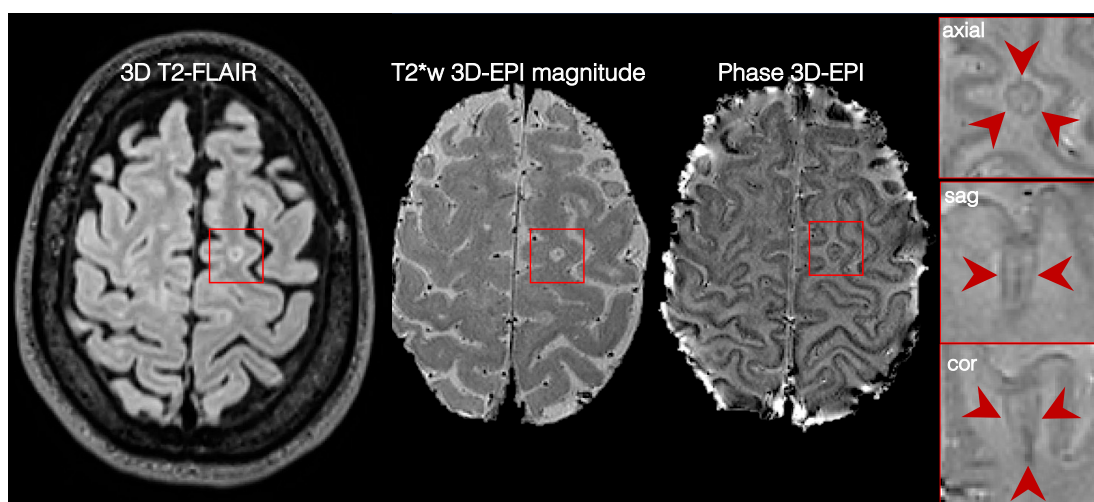


**Figure 2.7** – A central vein running through a lesion visible in the three planes (magnified views) in a 3D FLAIR\* obtained at 3T. Resampling was applied to the magnified images for visualization purposes.

### 2.4.2 The central vein sign

The diagnosis of MS based on imaging findings is not straightforward, as there are mimic diseases that cause similar inflammatory lesions in the WM. Vascular diseases and age-related variations are two other main causes of WM abnormalities. Therefore, if a patient shows symptoms possibly caused by different neurological conditions, several factors have to be carefully considered for the diagnosis, including the lesion count and distribution both in the brain and in the spinal cord. Recent studies at 3T and 7T have shown that an MRI-detectable vein inside WML might be a marker of undergoing pathological processes specific of MS (Maggi et al., 2018; Sati et al., 2016). This feature, commonly referred to as "central vein sign" (CVS) has gained great attention in the last years as it could help differentiate MS from its mimics. Small cerebral veins can be detected with susceptibility-based MRI sequences and in particular T2\*-weighted images (Sati et al., 2014), taking advantage of the magnetic properties

of blood. As the signal from the blood is mostly suppressed, veins appear hypointense on these images and can be easily observed in vivo. T2\* images, however, do not offer a suppression of the CSF as in FLAIR, and as a consequence MS lesions are more challenging to be visualized. Among the solutions proposed, a successful approach is the one known as FLAIR\* (Sati et al., 2012), which combines a FLAIR and a T2\* image, responsible respectively for lesions and vein detection. Several studies have shown that the detection of the CVS in FLAIR\* images can reliably differentiate MS from other diseases, including migraine (Maggi et al., 2018; Solomon et al., 2016). In particular, on a patient-wise level, different criteria have been proposed, some based on the percentage of lesions presenting the CVS (from 35% to 60%) and others simply on the CVS lesion count (3-lesion or 6-lesion rule) (Maggi et al., 2018; Sati et al., 2016; Solomon et al., 2018). Undergoing clinical trials aim at validating these results in larger cohorts of patients as the CVS is a promising candidate to be included in the MS diagnosis criteria in the near future.



**Figure 2.8** – Example of an axial view of 3D T2-FLAIR, T2\*-w EPI magnitude and phase with a PRL. The magnified views show the PRL in the three planes.

### 2.4.3 Paramagnetic rim lesions

Studies have demonstrated that about 30% of chronic lesions are pathologically characterized by perilesional accumulation of iron-laden microglia and macrophages, showing evidence of active demyelination and axonal loss around an inactive core (Absinta et al., 2019; Frischer et al., 2015; Luchetti et al., 2018). This type of MS lesions has been defined as "chronic active lesions" or "paramagnetic rim lesions" (PRL). The peripheral paramagnetic rim can be depicted on in-vivo susceptibility-based MRI (T2\*-weighted magnitude and phase images) at

both 3T and 7T (see Figure 2.8). PRL volume is more unlikely to change over time compared to other MS lesions, and an increased PRL burden has been linked to disability accumulation (Absinta et al., 2019). In particular, it has been shown that PRL are associated with a more aggressive disease course and patients with 4 or more PRL reached motor and cognitive impairment at an earlier age (Absinta et al., 2016). This seems to be caused by the fact that PRL are destructive and do not remyelinate (Absinta et al., 2016; Kolb et al., 2021). Moreover, PRL seem to be a specific feature of MS as, using MRI, they have been rarely detected in patients with other neurological conditions (Maggi, Sati, et al., 2020). Therefore, PRL have the premise of becoming a clinically relevant biomarker to improve MS diagnosis, characterize pathological processes causing disability progression, and monitor treatment effects over time. Currently, however, there is not yet a standardized MRI protocol for PRL analysis and this poses additional challenges for its wide clinical use.



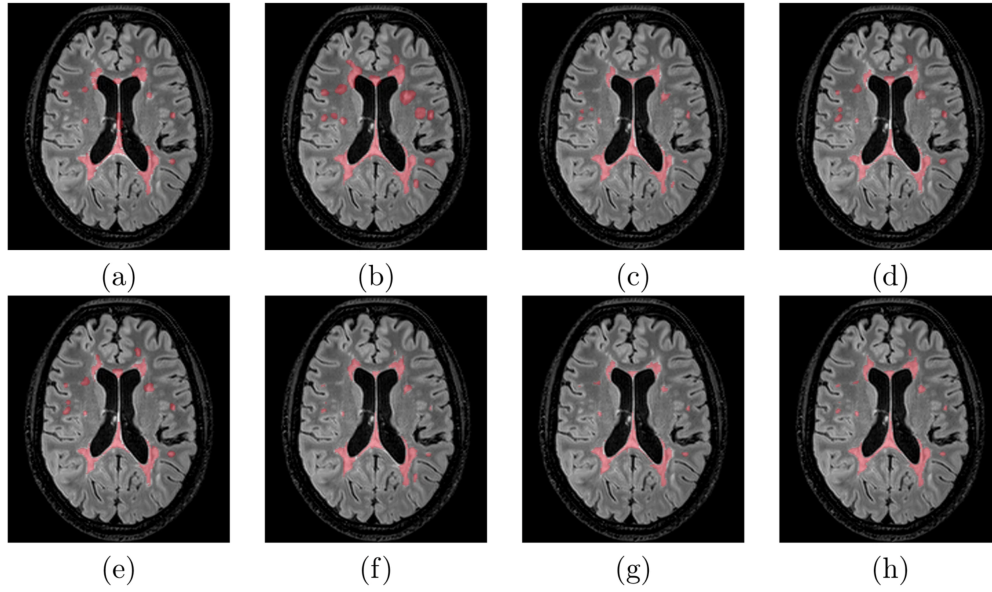
## 3 Automated analysis of MS lesions: state-of-the-art

*The content of section 3.2 of this chapter is currently under revision in Neuroimage: Clinical as "Cortical lesions, central vein sign, and paramagnetic rim lesions in multiple sclerosis: emerging machine learning techniques and future avenues" Francesco La Rosa, Maxence Wynen, Omar Al-Louzi, Erin S Beck, Till Huelnhagen, Pietro Maggi, Jean-Philippe Thiran, Tobias Kober, Russell Takeshi Shinohara, Pascal Sati, Daniel S. Reich, Cristina Granziera, Martina Absinta, Meritxell Bach Cuadra. The first author (author of this thesis) contributed to the conceptualization, methodology, and writing of the study.*

Manual segmentation or classification of MS lesions is time-consuming and characterized by a modest inter-rater reliability (see Figure 3.1) (Commowick et al., 2018; Danelakis et al., 2018). Automated methods are therefore needed to facilitate the analysis of MRI and support clinical decisions regarding the diagnosis, prognosis, and follow-up of MS patients. In this chapter, a summary of automated methods for the assessment of MS imaging biomarkers is presented. First, both model-driven and data-driven approaches for WML segmentation are described. Then, methods for advanced MS biomarkers, such as CL, CVS, and PRL, are introduced. This includes a few works to which I contributed during my doctoral work.

### 3.1 White matter lesion segmentation

In recent years, several methods have been proposed to automatize the WML segmentation process (Danelakis et al., 2018; Zeng et al., 2020). These can be broadly classified into two main categories: unsupervised and supervised approaches.



**Figure 3.1** – Example of the inter-rater variability delineating MS lesions. (a-g) lesion segmentation of seven experts overlaid on a 3D FLAIR image. (h) consensus segmentation. Major differences are observed for small lesions. Image retrieved from (Commowick et al., 2018).

### 3.1.1 Unsupervised approaches

Unsupervised approaches generally aim at modelling the distribution of the image intensities and detecting the lesions as outliers compared to the healthy tissue. They do not require labelled data and typically generalize fairly well to unseen data acquired in a different site or with a different protocol. This category includes methods based on intensity thresholding (Souplet et al., 2010), probabilistic atlases of healthy subjects (Shiee et al., 2008), clustering algorithms (Tomas-Fernandez & Warfield, 2015), Markov random field (Van Leemput et al., 2001), and Gaussian mixture models (Cerri et al., 2021). The reader is referred to the following reviews available in the literature for an extensive description of these methods (Danelakis et al., 2018; Lladó et al., 2012). Unsupervised approaches have achieved decent performance, in some cases comparable with supervised techniques (Cerri et al., 2021). Over the last years, however, supervised approaches have won all public grand challenges (Carass et al., 2017; Commowick et al., 2021; Commowick et al., 2018), and established themselves as the state-of-the-art techniques for WML segmentation.

### 3.1.2 Supervised techniques

Supervised techniques exploit a labeled dataset and are based on a classifier responsible for taking a decision. Depending on the type of classifier, they can be divided into two sub-classes:



**Table 3.1** – Overview of the two main categories of automated methods for WML segmentation. Some of them, such as the clustering algorithms, can be applied both in a supervised and unsupervised manner. Only the main approaches are cited, for a more comprehensive review refer to (Lladó et al., 2012; Zeng et al., 2020)).

White matter lesion segmentation approaches	
Unsupervised methods	Supervised methods
<ul style="list-style-type: none"> <li>• Intensity thresholding</li> <li>• Probabilistic atlases</li> <li>• Clustering algorithms</li> <li>• Markov random Field</li> <li>• Gaussian mixture models</li> </ul>	<ul style="list-style-type: none"> <li>• Support Vector Machine</li> <li>• Random Forest</li> <li>• k-Nearest Neighbour</li> <li>• Convolutional neural networks</li> </ul>

methods requiring hand-crafted features and deep learning (DL) approaches. The firsts are represented, for instance, by methods considering as classifier Support Vector Machines, Random Forest or k-Nearest Neighbour (k-NN) (Fartaria et al., 2016). These approaches estimate a set of features based on intensity or texture analysis and feed these to the classifier. In recent years, however, the second sub-class, represented by deep learning-based methods has been predominant (Zeng et al., 2020). A major advantage of deep learning methods, particularly of convolutional neural networks (CNN), is that the most relevant imaging features are automatically extracted during the training phase from multiple MRI contrasts, without any prior modelling required, thus avoiding the need for feature engineering. The main downside is that these methods require a large amount of high quality labeled data, which is often very difficult to obtain, particularly in medical imaging. In order to compare the performance of different methods and with the aim of increasing the available training data, publicly available datasets have been released to the medical imaging community in recent years. In particular, the Medical Image Computed and Computer-Assisted Intervention (MICCAI) society and the IEEE International Symposium on Biomedical Imaging (ISBI) conference have proposed four MS WML segmentation challenges: 2008 MICCAI MS Lesion challenge (Styner et al., 2009), 2015 ISBI Longitudinal MS Lesion Segmentation Challenge (Carass et al., 2017), 2016 MICCAI MS segmentation challenge (MSSEG) (Commowick et al., 2018), and 2021 MICCAI-endorsed new MS lesions segmentation challenge (MSSEG-2) (Commowick et al., 2021). The top-performing methods of all these challenges have been supervised approaches and, recently,

there has been a proliferation of deep learning methods. In 2016, Brosch et al. first proposed a CNN for MS lesion segmentation (Brosch et al., 2016), achieving results comparable to state-of-the-art methods even on small datasets (e.g. the 2008 MICCAI challenge). Their network was simply composed of three layers (input, convolutional and deconvolutional) and processed the entire volume of a scan. Since then, several other approaches based on CNNs, with progressively deeper and more complex architectures, have been proposed (Kaur et al., 2021). Valverde et al. proposed a cascade of two patch-wise convolutional neural networks that aimed at reducing the number of false positives (Valverde et al., 2017). This method is able to learn from small datasets and achieved the best result on the MICCAI 2008 challenge, while resulting between the top-performing methods in the ISBI 2015 and MSSEG 2016 challenges. We have also evaluated a framework based on a similar cascade of two CNNs (La Rosa et al., 2019) and verified that it outperformed, in terms of WML detection, a previously proposed approach based on a k-NN (Fartaria et al., 2016). Roy et al. introduced a CNN, made of two separate convolutional pathways, that accepts as input 2D patches from multi-channel images (Roy et al., 2018). This approach achieved state-of-the-art results in both a private dataset and on the 2015 ISBI challenge. Other recent works (Aslani et al., 2019; McKinley et al., 2021) also showed competitive results only considering whole-brain 2D slices, and avoiding the more computational expensive 3D convolutions. Furthermore, Hashemi et al. introduced a novel asymmetric loss function in order to overcome the problem of data unbalance, which is common in lesion segmentation tasks (Hashemi et al., 2019). In particular, they proposed a 3D fully convolutional neural network with large overlapping patches, trained with an asymmetric loss function based on the Tversky index (Salehi et al., 2017). Their framework achieved at the time the best score on the 2015 ISBI challenge with the lowest false positive rate. Recently, the U-Net architecture (Ronneberger et al., 2015) has emerged as the most successful backbone for segmentation CNN, achieving the first place in several medical imaging segmentation challenges (Isensee et al., 2021). It has been widely applied also for MS WML segmentation, often outperforming previous models (Gabr et al., 2020; hu et al., 2020; La Rosa, Abdulkadir, et al., 2020)

### **3.1.3 Current limitations**

Despite these efforts, and the fact that most of the above mentioned methods provide competitive WML segmentations, not a single approach is yet widely employed in clinical practice (Zeng et al., 2020). One first prominent challenge is given by the diversity of MRI data: the performance of these methods often drops for images acquired with a different scanner not seen during the training phase (Carass et al., 2017). An additional problem is given by the

fact that the segmentation metrics considerably decrease when analyzing patients with a low lesion load (Commowick et al., 2021). Moreover, the complex regulatory environment for medical imaging algorithms remains a critical barrier (Benjamens et al., 2020). Finally, very few of these methods consider also CL, which are on the contrary of great interest to clinicians.

## **3.2 Automated assessment of advanced biomarkers**

### **3.2.1 Cortical lesions**

ML-based methods able to automatically segment CL have been explored with both 3T and 7T MRI. The first work present in the literature considered a postmortem MS brain imaged at 3T with high resolution (0.35mm isotropic) (Tardif et al., 2010). Tardif et al. proposed to first identify the cortical and white matter surfaces, then extract laminar profiles between the two tissues, and finally apply a k-means classifier to the profile intensity and shape features to parcellate the cortex and detect lesions. Although showing promising results on one postmortem MS brain, this method was never validated with larger cohorts nor in-vivo data. A few years later, Fartaria et al. (Fartaria et al., 2016) proposed the first automated method for the segmentation of both WM and cortical lesions. In their study, they analyzed a cohort of 39 early-stage MS patients and considered both conventional (FLAIR, MPRAGE) and advanced (MP2RAGE, DIR) MRI sequences at 3T. In a nutshell, their method consisted of coregistering the different MRI contrasts, leveraging prior tissue probability maps from existing brain atlases of healthy subjects, and finally classifying each voxel either as being a lesion or healthy tissue with a k-nearest neighbor (k-NN) algorithm. Additionally, as post-processing, lesions smaller than 3.6 $\mu$ L were discarded, and a region-growing algorithm was applied to improve the lesion delineation. Results were promising, showing a CL detection rate of 62% when advanced imaging (FLAIR, MP2RAGE, and DIR) was included. An extension of this segmentation framework with a Bayesian partial volume (PV) estimation method was presented by the same authors (Fartaria et al., 2017). They argued that CL, being generally small and located at the interface between WM and GM, suffer from strong PV effects. The addition of this PV model indeed improved the delineation of CL, while the detection rate remained similar (58%) (Fartaria et al., 2017). The same research group also proposed a different segmentation method for WML and CL using only 7T MP2RAGE images (called MSLAST: Multiple Sclerosis Lesion Analysis at Seven Tesla) (Fartaria et al., 2019). MSLAST computes tissue concentration maps with a PV algorithm and unifies them based on topological constraints. A connected-components analysis is then performed on gray matter and cerebrospinal fluid maps, and small components are classified as MS lesions. This method was evaluated with 25 MS patients'

scans from two research centers and reached a 58% patient-wise CL detection rate (when  $6\mu\text{L}$  was considered as minimum lesion volume) with a false positive rate of 40%. Moreover, it showed promising scan-rescan repeatability within the same session, with a mean total lesion volume difference (WML and CL combined) of 0.29 mL (mean total lesion volume 5.52 mL), vs 0.13 mL for the manual segmentations. More recently, we have proposed two DL-based approaches as well. The first study aims to automatically segment WML and CL at 3T using FLAIR and MP2RAGE (see Chapter 4), whereas in the second we propose a framework for the segmentation and classification of CL with 7T MRI (refer to Chapter 6).

### 3.2.2 The central vein sign

The automated and standardized detection of the central vein sign could potentially be of clinical utility in improving the accuracy of MS diagnosis. As of today, two automated methods for the classification of MS lesions as CVS+ (MS lesions showing the presence of the CVS) or CVS- (MS lesions without the CVS) have been proposed in the literature (Dworkin et al., 2018; Maggi, Fartaria, et al., 2020). Dworkin et al. (Dworkin et al., 2018) proposed a probabilistic method based on the Frangi vesselness filter (Frangi et al., 1998). They first perform an automated WML segmentation using T1 and FLAIR 3D MRI volumes acquired at 3T81,82 and obtain a map of the veins by applying the vesselness filter to a  $T2^*w$  3D EPI image. Confluent lesions are then separated and lesion centers detected by textural analysis (Dworkin et al., 2019). Periventricular lesions are discarded as suggested by consensus guidelines (Sati et al., 2016), and a permutation algorithm is applied to verify whether veins occur at the lesions' centers more often than would be expected due to random chance. Finally, to account for scan motion, the single lesion CVS+ probabilities are weighted by the noise in their  $T2^*w$  3D-EPI intensities, and averaging across the total number of lesions, a patient-wise CVS value is obtained. This method was evaluated on a cohort of 31 adults, of whom 16 had MS. When considering a 40% cutoff rule, the method yielded a sensitivity of 0.94 and a specificity of 0.67 on a patient-wise classification level. The performance of the method on a lesion-wise level was not assessed. Although still far from experts' performance, this was a first attempt to automatize the CVS assessment and encouraged further improvements. Maggi, Fartaria et al. (Maggi, Fartaria, et al., 2020) introduced an optimized CNN for the automated CVS assessment, called CVSnet. CVSnet is inspired by the VGGnet (Simonyan & Zisserman, 2015) but composed of only three convolutional layers followed by rectified linear unit (ReLU) activations. Dropout was applied in each layer, and then two fully-connected layers of size 32 and 2, respectively, were added to provide the output. The authors selected 3D patches of size  $21 \times 21 \times 21$  voxels as input for the network, where each patch was centered on an MS

lesion and FLAIR\* was the only MRI contrast used. Moreover, an ensemble of 10 networks with the same architecture was trained and the probability outputs were averaged to provide the final prediction. This study considered a cohort of 80 patients imaged at three different sites, of whom 42 had MS, 35 an MS-mimic, and 5 an unknown diagnosis. On the test set, CVSnet significantly outperformed the vesselness filter (Frangi et al., 1998) ( $p < 0.001$ ), reaching a sensitivity, specificity, and accuracy of 0.83, 0.75, and 0.79, respectively, approaching expert performance. However, as argued by the authors, these results are not directly comparable with those of Dworkin et al. (Dworkin et al., 2018), as the CVSnet considered different exclusion criteria to pre/select the lesions, and the initial lesion segmentation was performed manually. Refer to the Appendix (section A.1.1) for more details about this method.

### 3.2.3 Paramagnetic rim lesions

To our knowledge, only two methods have been proposed so far for the detection of rim-like features and classification of PRL (Barquero et al., 2020; Lou et al., 2021). Barquero et al. introduced a DL-based approach (called RimNet) for the semi-automated classification of PRL, which considered 3D FLAIR and T2\*w 3D-EPI and phase 3D-EPI images (refer to Chapter 7.) Lou et al. (Lou et al., 2021), on the other hand, proposed a fully automated method for PRL assessment. They considered a cohort of 20 subjects with MS imaged with 3D FLAIR, 3D MPRAGE, and T2\*-w 3D-EPI and phase 3D-EPI images. One neurologist inspected the T2\* magnitude and unwrapped phase images and annotated PRL (113 PRL over the entire cohort). The automated pipeline, after some pre-processing steps that included lesion segmentation (Valcarcel, Linn, Khalid, et al., 2018; Valcarcel, Linn, Vandekar, et al., 2018), lesion center detection (Dworkin et al., 2019), and lesion labeling, consisted of extracting 44 different lesion-wise radiomic features. A random forest classifier was then fitted on these features, and its ability to classify PRL was evaluated on a test set of 4 patients. Sensitivity and specificity of 0.75 and 0.81, respectively, were achieved. Although fully automated, this study has three limitations. First, the extremely small testing dataset (4 patients only with 47 PRL), annotated by a single expert, does not guarantee the generalization of the proposed method. Second, all patients analyzed had at least one PRL, and this might add a bias to the trained model. Finally, as acknowledged by the authors, about 65% of misclassified lesions were confluent, highlighting the need for a better solution to address these lesions.

### 3.2.4 Challenges for machine learning methods

From a ML perspective, the automated segmentation or classification of advanced MS MRI biomarkers faces additional challenges compared to their WML counterparts.

**Imaging guidelines** - The first obstacle is represented by the lack of consensus guidelines for imaging protocols. Although efforts have been made to standardize the use of MRI in clinical practice for conventional biomarkers (Wattjes et al., 2021), guidelines are still in a preliminary stage for advanced biomarkers. CL were included in the MS diagnostic criteria in 2017 (Thompson, Banwell, et al., 2018), but, currently, there is no single gold standard sequence at 3T for their detection in a clinical setting. PSIR, DIR, and MP2RAGE are all recommended by an international consensus (Filippi et al., 2019). However, these contrasts remain primarily acquired in research settings and are not yet widely used in clinical routine. Regarding the CVS, in 2016 a consensus statement by the North American Imaging in MS Cooperative (NAIMS) has proposed a standard radiological definition and suggested specific MRI acquisitions (Sati et al., 2016). Following these recommendations, recent studies have shown that high-resolution T2\*w 3D-EPI or FLAIR\* improve the detection of the CVS compared to the clinical standard (Castellaro et al., 2020; Suh et al., 2019). Nevertheless, a standardized clinical protocol for CVS detection is still missing. Among all advanced biomarkers, PRL is probably the one at the earliest stages. Although recent studies support the feasibility of its assessment on clinical scans and its utility in improving the diagnosis and prognosis of MS (Maggi, Sati, et al., 2020), there are not yet international guidelines for its definition nor a standardized MRI protocol for its analysis. Several different imaging modalities have been used for the PRL assessment, including phase 3D-EPI, susceptibility weighted imaging (SWI), QSM, and multi-echo T2\* GRE at both 3T and 7T (Absinta et al., 2018; Absinta et al., 2016). However, there is a paucity of studies that have systematically compared the sensitivity of these acquisition techniques for PRL detection, especially when implemented at different field strengths. These evolving or unclear criteria for CL, the CVS, and PRL, wide variety of imaging settings, and lack of clear guidelines for standardized protocols clearly jeopardize the development and wide use of these biomarkers and targeted ML techniques.

**Limited datasets** - An additional limitation, particularly for supervised DL-based approaches, is the scarcity and limited size of datasets in which these biomarkers were manually annotated. CL, CVS, and PRL all require for their assessment advanced MRI sequences at high or ultra-high magnetic field and experienced raters, and this makes it difficult to have large multi-site datasets. Although national MS registries exist in most countries, the data sharing of MRI in MS is still limited and often includes only conventional sequences (Vrenken et al., 2021).

Moreover, the CVS or the rim-shape in PRL are visible only on a few slices per lesion, reducing even more the data available to train a supervised approach. The presence of motion or other MRI artifacts can also result in the exclusion of poor quality scans reducing the data available for training/testing even further.

**Expert annotations** - Even for experts, the task of segmenting CL, detecting the CVS, or classifying PRL is intrinsically more challenging than segmenting WML. CL are generally smaller in size and more affected by partial volume (PV) effects, compared to WML. The cortex is convoluted, so lesion shape is not as regular as in WM, and traditional methods of radiological evaluation (scrolling through an image stack) are less effective in this context. The detection of the CVS requires susceptibility-based MRI and its exclusion criteria need to be carefully considered when performing its assessment (Sati et al., 2016). T2\*w phase images used to detect PRL present variability in the susceptibility signal and several artifacts, therefore experienced raters are needed. Moreover, as these three biomarkers have been so far mainly studied in research settings, clinicians do not commonly see them in clinical practice and might need specific training and dedicated time to perform a proper assessment.

**Inter-rater variability** - The lack of standardization for the definition and imaging of these biomarkers combined with the difficulty of the task, contributes to a modest inter-rater variability. Barquero et al. (Barquero et al., 2020) showed that, in a cohort of 124 MS patients, approximately 38% of PRL needed a consensus review from two raters classifying PRL independently (Kappa score of 0.73). Absinta et al. observed similar inter-rater agreement between three experts at 3T (Fleiss coefficient of 0.71), with somewhat higher intra-rater reliability (Cohen k of 0.77) (Absinta et al., 2018). Similar values were reported at 7T for the same set of patients, whereas the agreement between 3T and 7T annotations was substantial (Cohen k of 0.78). In a similar way, the inter-rater agreement was shown to be moderate for the segmentation of CL (Harrison et al., 2015; Nielsen et al., 2012) and moderate to high for the CVS (Kau et al., 2013; Maggi, Fartaria, et al., 2020). Imaging quality and motion artifacts are other factors to consider as they can result in inconspicuity of all three biomarkers and, therefore, contribute to poor inter-rater agreement. Overall, the inter-rater variability represents an additional challenge for the development of automated approaches, as there might be large inconsistencies in the annotations of the training or testing set due to different raters performing the manual assessment. Thus, to reduce this variability, a higher number of annotators would be needed, but this is often not possible due to a limited number of experts or lack of resources (time/money).





## 4 Cortical and WM lesion segmentation at 3T MRI

*The content of this chapter is adapted from the postprint version of the following article:*

*Francesco La Rosa, Ahmed Abdulkadir, Mário João Fartaria, Reza Rahmanzadeh Po-Jui Lu, Riccardo Galbusera, Muhamed Barakovic, Jean-Philippe Thiran, Cristina Granziera, Merixtell Bach Cuadra: "Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE." Neuroimage: Clinical 27 (2020): 102335; doi: [10.1016/j.nicl.2020.102335](https://doi.org/10.1016/j.nicl.2020.102335). The first author (author of this thesis) contributed to the conceptualization, methodology, software, validation, and writing of the study.*

### Abstract

The presence of cortical lesions in multiple sclerosis patients has emerged as an important biomarker of the disease. They appear in the earliest stages of the illness and have been shown to correlate with the severity of clinical symptoms. However, cortical lesions are hardly visible in conventional magnetic resonance imaging (MRI) at 3T, and thus their automated detection has been so far little explored. In this study, we propose a fully-convolutional deep learning approach, based on the 3D U-Net, for the automated segmentation of cortical (mostly type I) and white matter lesions at 3T. For this purpose, we consider a clinically plausible MRI setting consisting of only two MRI contrasts: one conventional T2-weighted sequence (FLAIR), and one specialized T1-weighted sequence (MP2RAGE). We include 90 patients from two different centers with a total of 728 and 3856 gray and white matter lesions, respectively. We show that two reference methods developed for white matter lesion segmentation are inadequate to detect small cortical lesions, whereas our proposed framework is able to achieve a detection rate of 76% for both cortical and white matter lesions with a false positive rate of 29% in comparison to manual segmentation. Further results suggest that our framework generalizes

well for both types of lesion in subjects acquired in two hospitals with different scanners.

## **4.1 Introduction**

Multiple Sclerosis (MS) is a chronic demyelinating disease involving the central nervous system (CNS). An estimated 2 million people are currently having the disease worldwide (Reich et al., 2018). MS is characterized by sharply delimited lesional areas with primary demyelination, axonal loss, and reactive gliosis, both in the white and in the grey matter. However, the pathological process is not confined to these macroscopically visible focal areas but is generalized in the entire central nervous system (Compston & Coles, 2008; Kuhlmann et al., 2017).

Magnetic resonance imaging (MRI) is the imaging tool of choice to detect such lesions in both the WM and GM of the CNS. The current diagnostic criteria (McDonald criteria (Thompson, Banwell, et al., 2018)) are also based on the count and location of lesions in MRI. Common MRI protocols currently include T1-weighted (T1w), T2-weighted (T2w), and fluid-attenuated inversion recovery T2 (FLAIR) sequences. For many years, the main focus in research and clinical practice has been set on white matter lesions (WML), clearly visible in the above-mentioned conventional MRI sequences. In the last decade, however, cortical damage has emerged as an important aspect of this disease. Recent studies have shown that the amount and location of cortical lesions (CL), visible mostly on advanced MRI sequences at high and ultra-high magnetic fields (3 or 7T), correlate better with the severity of the cognitive and physical disabilities than those of WML (Calabrese et al., 2009). Since 2017, CL are also included in the above-mentioned MS diagnostic criteria (Thompson, Banwell, et al., 2018). Consequently, specialized MR sequences with greater sensitivity to detect CL, such as magnetization-prepared 2 rapid acquisition with gradient echo (MP2RAGE) (Kober et al., 2012; Marques et al., 2010) and double inversion recovery (DIR) (Wattjes et al., 2007), are now more and more used in the clinical setting (Filippi et al., 2019).

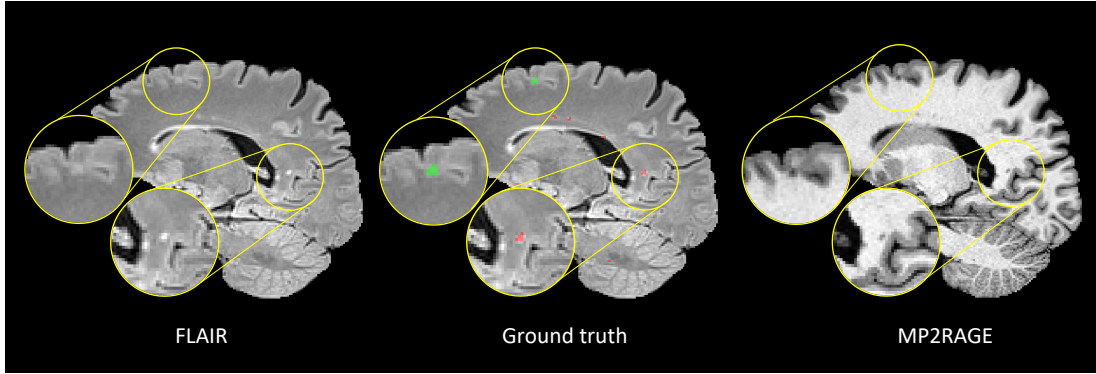
Currently, manual segmentation on clinical MRI is considered the gold standard for MS lesion identification and quantification. However, given how time-consuming this process is, several methods for automated MS lesion segmentation have been proposed in the literature (Kaur et al., 2020). These can be broadly classified into supervised and unsupervised approaches. The former ones rely on a manually labeled training set and aim at learning a function that maps the input to the desired output. The latter do not require manual annotations as they are based on generative models that rely on modeling the MRI intensities values of different

brain tissues and lesions (Lladó et al., 2012).

Deep learning algorithms are particularly suited for image segmentation tasks and dominate leader-boards of biomedical imaging processing challenges, including the segmentation of MS WML (Carass et al., 2017). Specifically, several convolutional neural network (CNN) architectures have been tailored for the segmentation of MS WML (Kaur et al., 2020). Some of them employ 2D convolutional layers (Aslani et al., 2019; Roy et al., 2018) whereas others employ 3D convolutional layers to incorporate information from all three spatial directions simultaneously (Hashemi et al., 2019; Valverde et al., 2017; Valverde et al., 2019). The clear edge these methods have over classical approaches is the capability of automatically extracting the relevant features for the task. Their application and generalization in clinical datasets, however, remains to be proved. They have often considered only 2D MRI sequences and segmentations were performed with a large minimum lesion volume threshold; for instance, Valverde et al. (Valverde et al., 2017) set this value to 20 voxels for the clinical MS datasets. Moreover, apart from (Valverde et al., 2019), all these deep learning methods are currently not publicly available. Finally, with the exception of our previous work (La Rosa et al., 2019), they have not been evaluated on CL.

Compared to WML, imaging of CL faces additional challenges due to their pathological features (Filippi et al., 2019). While WML can be automatically detected with high accuracy (Carass et al., 2017) from conventional MRI sequences, such as MP2RAGE and FLAIR, CL, which affect mostly the more superficial and less myelinated layers of the cortex (Filippi et al., 2019), have a low contrast to surrounding tissue with clinical MS MRI protocols. As mentioned above, the detection of CL, at least at 3T, requires specialized imaging such as MP2RAGE and 3D DIR (Kober et al., 2012; Wattjes et al., 2007), and the number of lesions visible is still low in comparison to histopathology (about 20%) (Calabrese et al., 2010). These advanced imaging requirements limit the access to large training datasets. Furthermore, the automated detection of CL is challenging as the number, volume, and location of CL varies substantially across subjects (see Figure 4.2).

To the best of our knowledge, only three publications explored the simultaneous segmentation of CL and WML at 3T (Fartaria et al., 2016; Fartaria et al., 2017; La Rosa et al., 2019). Our first approach was based on a k-nearest neighbors (k-NN) classifier (Fartaria et al., 2016), which was later on improved by including a partial volume tissue segmentation framework (Fartaria et al., 2017) to better delineate the lesions. In Fartaria et al. (Fartaria et al., 2016), we considered a multi-modal MRI framework including 3D MP2RAGE, 3D FLAIR, and 3D DIR sequences. Compared to manual segmentation, an overall WML and CL detection rate of



**Figure 4.1** – From left to right a sagittal slice of the FLAIR, manual lesion segmentation mask overlaid on the FLAIR, and MP2RAGE contrast. Colorcode of overlay: WML in red and CL in green. The zoomed-in WML is clearly visible in FLAIR, whereas the MP2RAGE shows a higher contrast for the CL.

respectively 77% and 62% was achieved. Recently, we have also explored the ability of deep learning architectures (La Rosa et al., 2019). We proposed an original 3D patch-wise CNN that improved lesion-wise results with respect to Fartaria et al (Fartaria et al., 2016). The main limitation of both approaches, however, was that training and testing them without the DIR sequence caused a significant drop in performance (CL detection rate from 75% to 58% in (La Rosa et al., 2019)). Unfortunately, the DIR sequence is not widely acquired in clinics due to its long acquisition time (about 13 minutes) and frequent artefacts (Filippi et al., 2019), thus being for now mostly used for research purposes.

This study complements the literature with an evaluation of a deep learning method to segment CL and WML based on two MRI sequences only (3D FLAIR and MP2RAGE, see Figure 4.1) acquired at 3T. This choice reflects a clinically plausible set of input data that is not disruptive of established processes. Our aim is to provide a segmentation framework for different types of MS lesions, large and small, with a minimum lesion size of 3 voxels as recommended in the guidelines for MS CL (Geurts, Pouwels, et al., 2005). We propose a fully-convolutional architecture inspired by the 3D U-Net (Çiçek et al., 2016). Compared to our previous studies, we significantly extend our cohort of patients to 90 subjects from two different clinical centers. We evaluate the method firstly with a 6-folds stratified cross validation over the entire cohort and secondly with a train-test split.

## 4.2 Materials and Methods

### 4.2.1 Datasets

In this study, we consider two datasets for a total of 90 subjects overall. Dataset I includes patients at different stages of the disease, whereas Dataset II only at the early stages. Refer to Figure 4.2 for an analysis of the differences between the datasets in terms of lesion volume and lesion count per patient.

**Dataset I** Images for Dataset I were acquired at Basel University Hospital from 54 patients (35 female / 19 male, mean age  $44 \pm 14$  years, age range [22-73] years). According to the revised McDonald criteria (Thompson, Banwell, et al., 2018), 39 of these were diagnosed with relapsing remitting MS, 8 with primary progressive MS, and 7 with secondary progressive MS. Expanded Disability Status Scale (EDSS) scores ranged from 1 to 7 (mean  $3.2 \pm 1.9$ ). Imaging was performed on a 3T MRI scanner (MAGNETOM Prisma, Siemens Healthcare, Erlangen, Germany) with a 64-channel head and neck coil. The following 3D sequences were acquired with a 1 mm<sup>3</sup> isotropic spatial resolution: 3D-FLAIR (TR/TE/TI= 5000,386,1800 ms, acquisition time= 6 min), and a prototype MP2RAGE (TR/TE/TI1/TI2= 5000,2.98,700,2500, acquisition time= 8 min). The study was approved by the Ethics Committee of our institution, and all patients gave written informed consent prior to participation.

**Dataset II** Images for Dataset II were acquired at Lausanne University Hospital from 36 patients (20 female / 16 male, mean age  $34 \pm 10$  years, age range [20-60] years) diagnosed with relapsing remitting MS. Expanded Disability Status Scale (EDSS) scores ranged from 1 to 2 (mean  $1.5 \pm 0.3$ ). Imaging was performed on a 3T MRI scanner (MAGNETOM Trio, Siemens Healthcare, Erlangen, Germany) with a 32-channel head coil. The following 3D sequences were acquired with a resolution of 1 x 1 x 1.2 mm<sup>3</sup>: 3D-FLAIR (TR/TE/TI= 5000,394,1800 ms, acquisition time= 6 min), and a prototype MP2RAGE (TR/TE/TI1/TI2= 5000,2.89,700,2500, acquisition time= 8 min). The study was approved by the Ethics Committee of our institution, and all patients gave written informed consent prior to participation.

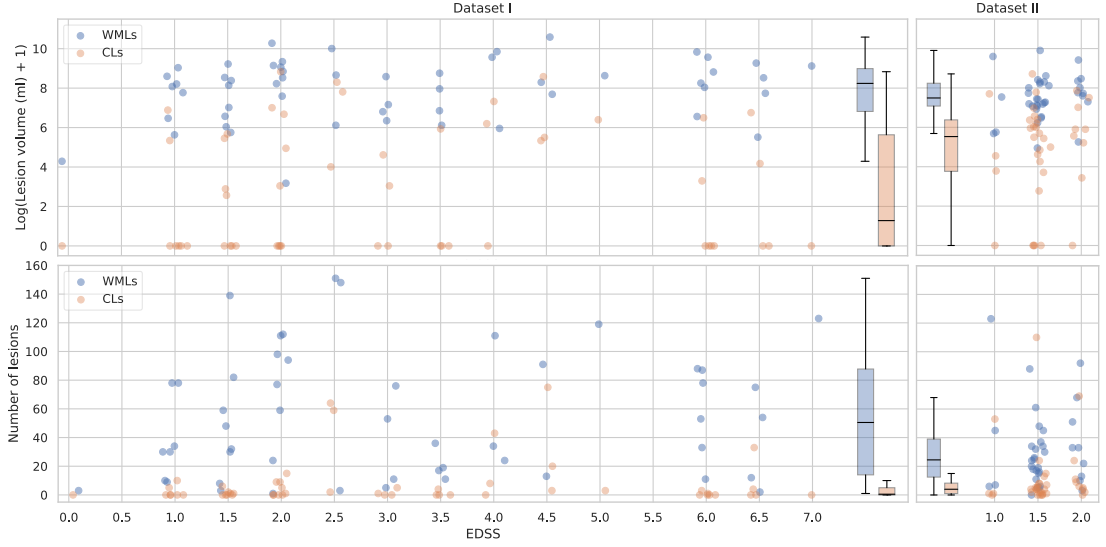
**Manual segmentation** WML appear as hyperintense areas in T2w images and as hypointense in T1w images and are usually well visible in conventional sequences at 3T. On the contrary, the majority of CL cannot be clearly seen in FLAIR at 3T and specialized sequences as the MP2RAGE improve their detection (see Figure 4.1 and 4.6). In Dataset I, all lesions were detected and classified by consensus by a neurologist and a medical doctor with 11 and 5 years of experience in MS research, respectively. The medical doctor then manually segmented all

lesions. In Dataset II both WML and CL were manually detected and classified by consensus by the same neurologist and one radiologist with 7 years of experience, using multiple imaging modalities. Their agreement rate prior to consensus was of 97.3%. The lesion borders were then delineated in each image looking at multiple planes by a trained technician. In total in our two datasets, 3856 WML and 728 CL were manually labeled. Furthermore, CL were classified by a single expert in different subtypes according to (Calabrese et al., 2010). Cortical MS lesions can extend across the WM and GM (leukocortical, type I), can be contained entirely in the gray matter without extending to the surface of the brain or to the subcortical WM (pure intracortical, type II), or can be widespread from the outer to the inner layers of the cortex without perivenous distribution and often over multiple gyri (subpial, type III). Within our cohort, the majority (89%) of the cortical lesions identified belonged to type I, 11% to type II, and only 0.01% to type III. Given the high imbalance between different subtypes, in the automated segmentation analysis we pool all CL together.

Lesions smaller than 3 voxels were automatically re-classified as background in the ground truth and in the predictions. This is equal to a volume of 3  $\mu$ L for the lesions in Dataset I and 3.6  $\mu$ L for the lesions in Dataset II. It should be noted that 3  $\mu$ L corresponds to the consensus recommended minimum CL size in the 3D DIR sequence (Geurts, Pouwels, et al., 2005), and we chose this as the minimum lesion size in our study as currently there is no guideline for the MP2RAGE. The distribution of the subject-wise total lesion volume and lesion count in our cohort can be seen in Figure 4.2.

#### **4.2.2 Methodology**

**U-Net** Our network architecture is based on the 3D U-Net (Çiçek et al., 2016; Ronneberger et al., 2015). The U-Net architecture integrates an analysis path, where the number of feature maps are increased while the image resolution is being reduced, and a synthesis path where the resolution is increased and number of features decrease, yielding a semantic segmentation output. Several variants of it have been proposed, for example changing the resolution levels, varying the number of convolution layers or introducing residual blocks. The U-Net has been tested on different biomedical imaging segmentation applications, and methods based on it or on its 3D implementation (Çiçek et al., 2016) have won several segmentation challenges (Isensee & Maier-Hein, 2019; Myronenko, 2018). Two variants of the U-Net were also specifically proposed for MS WML segmentation (Feng et al., 2018; Kumar et al., 2019). Kumar et al. (Kumar et al., 2019) have proposed a dense 2D U-Net and showed promising results on a challenge dataset (Carass et al., 2017), even though the method was not compared

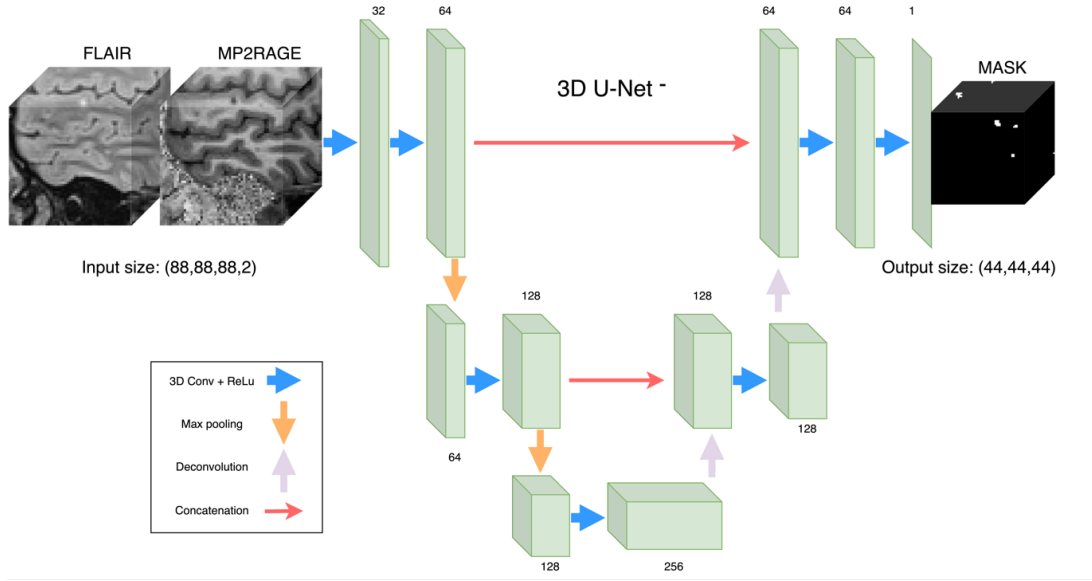


**Figure 4.2** – Distribution of the lesion volume and count per patient in our two datasets, considering WML and CL separately. In the first row each patient has a blue dot corresponding to its WML volume and an orange dot for the CL one. In the second row this is repeated with the lesion count. Boxplots are added to summarize the distribution of each dataset. For visualization purposes jitter is added to the true EDSS value.

to other deep learning approaches. Feng et al. (Feng et al., 2018) have presented a standard 3D U-Net with advanced data augmentation, but again this work lacked an evaluation on a clinical dataset or a proper comparison with other state-of-the-art methods.

**Proposed architecture** Compared to the original implementation of the U-Net, we reduce the number of resolution levels from four to three (therefore 3D U-Net<sup>-</sup>) because of the limited number of available CL. By doing so, we drastically reduce the number of trainable parameters from 18.3M to 3.8M and thus decrease the risk of overfitting. Furthermore, our choice of a 3D architecture is motivated by the fact that the input modalities FLAIR and MP2RAGE are both 3D acquisitions, and therefore we want to fully exploit the volumetric anatomical information. The 3D U-Net<sup>-</sup> architecture integrates a spatial context of 41 voxels. In our experiments, we use an input shape of (88, 88, 88) and because no padding is applied through the network the output size was (48, 48, 48) (see Figure 4.3). In the analysis path, the 3D convolutional filters (each followed by a ReLU activation function) has respectively 32, 64, 64, 128, and 128 filters. The decoder has the following number of feature maps: 256, 128, 128, 64, 64, 1. Skipped connections are present as in the original implementation (see Figure 4.3).

**Pre-processing** The only pre-processing step performed is a rigid registration based on mutual information of the FLAIR image of each subject to the corresponding MP2RAGE (UNI contrast) image with ELASTIX (Klein et al., 2010). Each multi-step pre-processing pipeline,



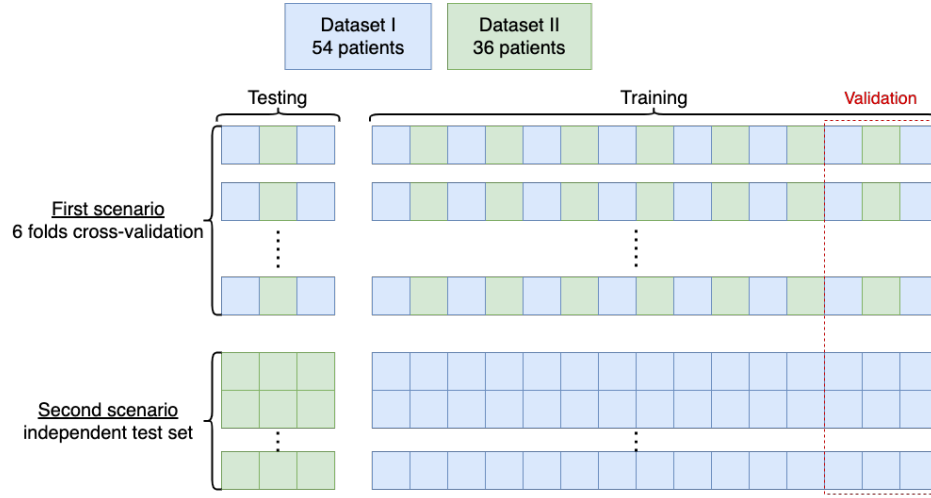
**Figure 4.3** – 3D U-Net<sup>-</sup> architecture proposed. On the left, examples of input patches in the two contrasts used, and on the right the relative lesion mask obtained in output.

for instance, skull stripping and/or bias field correction has its own data dependent effect on the performance. While we cannot exclude the possibility that a certain combination of pre-processing steps would lead to better performance for a given data set, we reduce the model-related sources of variance in performance to the network architecture and its training, which leads to an integrated solution that facilitates evaluation and incremental improvements.

**Training** Prior to training all input volumes are normalized with zero mean and unit variance. We implement a sampling strategy by which each connected component in the ground truth has the same possibility of being sampled, regardless of its size. Thus, we encourage the network to focus also on the smaller structures such as CL. L2 regularization was used with a regularization factor of  $1e-5$ . The network is trained with a pixel-wise weighted cross-entropy loss function with the following weights: background 1, WML 1, CL 5. This is motivated by the fact that there are about 5 times less CL than WML in each fold of our dataset. The learning rate is initially set to  $1e-8$  and gradually increased in the first 2000 iterations to  $1e-4$  in a warm-up phase. Afterwards it is reduced by half every 10000 iterations. The batch size is set to 2, and Adam is used as optimizer. The validation loss is monitored with early stopping to determine when to stop the training.

**Data augmentation** Data augmentation is a well-known technique for deep neural networks to increase the performance in the testing set. In our work, extensive data augmentation is





**Figure 4.4** – Scheme of the two evaluation scenarios.

performed on-the-fly to prevent overfitting. The transformations are carefully chosen as, since we are dealing with large 3D patches, excessive augmentation would slow down the training. More specifically, we apply random rotation of the input volumes around the z axis only up to 90°, random spatial scaling up to 5% of the volume size, and random flipping along all three axes.

**Implementation** The code has been implemented in the Python language in the Niftynet framework (Gibson et al., 2018) based on TensorFlow (Abadi et al., 2016). It is available along with the trained models on our research website. The software requires Python version 3.6. Training also requires CUDA/cuDNN libraries by NVIDIA and compatible hardware.

### 4.2.3 Evaluation

**Comparison with other related methods.** For comparison, we evaluated two state-of-the-art MS WML segmentation methods publicly available:

- LST-LGA is an unsupervised lesion growth algorithm (Schmidt et al., 2012) implemented in the LST toolbox version 3.0.0 (“LST – Lesion segmentation for SPM | Paul Schmidt – freelance statistician”, 2018) for Statistical Parametric Mapping (SPM). LST-LGA has been widely evaluated in the context of MS WML segmentation and used as comparison with more recent approaches (Aslani et al., 2019; Roy et al., 2018; Valverde et al., 2017). In a nutshell, the algorithm performs an initial bias field correction and affine registration of the T1 image (in our case the MP2RAGE) to the FLAIR, and then proceed with the lesion segmentation. Lesions are identified based on a voxel-wise binary regression

with spatially varying parameters. We applied LST-LGA with the default initialization parameters ( $\kappa=0.3$ ,  $\text{MRF}=1$ , maximum iterations=50). The final threshold to obtain a binary segmentation mask was optimized in the validation set by maximizing the dice coefficient.

- **nicMSlesions** is a state-of-the-art deep learning WML segmentation method (Valverde et al., 2017; Valverde et al., 2019). Having reached an excellent performance in a MS lesion segmentation challenge (Carass et al., 2017), it is now a common method to compare with (Aslani et al., 2019; Roy et al., 2018; Weeda et al., 2019). This method selects lesion candidates' voxels based on the FLAIR contrast and extracts  $11 \times 11 \times 11$  patches around them. A double CNN is then trained to first find lesion candidates and then reduce the false positive rate. The pipeline includes as pre-processing steps a registration of the different modalities to the same space, skull-stripping, and denoising. This approach was run with all the default parameters, including a maximum of 400 epochs, early stopping, and the patient value set to 50 epochs.

**Evaluation strategies and metrics.** We evaluated the three methods in two different scenarios (see Figure 4.4). First, we performed a stratified 6-folds cross validation pooling data from Dataset I and II. Second, we trained the supervised methods on the subjects of Dataset I and evaluated using data from Dataset II.

Results were evaluated quantitatively with the manually delineated masks. We computed the following, widely used (Carass et al., 2017; Kaur et al., 2020; Lladó et al., 2012), evaluation metrics: dice coefficient (DSC), absolute volume difference (AVD), voxel-wise positive predicted value (PPV), lesion-wise true positive rate (LTPR), lesion-wise false positive rate (LFPR), WML detection rate (LTPR\_WM), CL detection rate (LTPR\_CL) as defined here (Carass et al., 2017).

Statistical analysis was also performed at the patient-wise level using the SciPy Python library (SciPy 1.0 Contributors et al., 2020). As the distributions violate the normality assumptions, Wilcoxon signed rank test was used to statistically test differences in WM\_LTPR, CL\_LTPR, and LFPR. All tests were adjusted for multiple comparison using a Bonferroni correction. We computed the Pearson's linear correlations between manual and estimated masks to analyze the volume differences.

**Table 4.1** – Median values (IQR) of the metrics obtained for the different methods on the cross-validation evaluation (90 subjects). The minimum lesion size is 3 voxels. The best result per each metric is shown in bold.

Method	Dice	AVD	PPV	LTPR	LTPR_WM	LTPR_CL	LFPR
LST-LGA	0.36 (0.19)	0.60 (0.31)	<b>0.71</b> (0.34)	0.36 (0.24)	0.38 (0.25)	0.10 (0.33)	0.36 (0.35)
nicMSLesions	0.53 (0.28)	0.32 (0.43)	0.52 (0.43)	0.65 (0.26)	0.67 (0.26)	0.53 (0.39)	0.45 (0.40)
3D U-Net-	<b>0.62</b> (0.16)	<b>0.27</b> (0.30)	0.61 (0.23)	<b>0.76</b> (0.20)	<b>0.77</b> (0.22)	<b>0.75</b> (0.50)	<b>0.29</b> (0.25)

## 4.3 Results

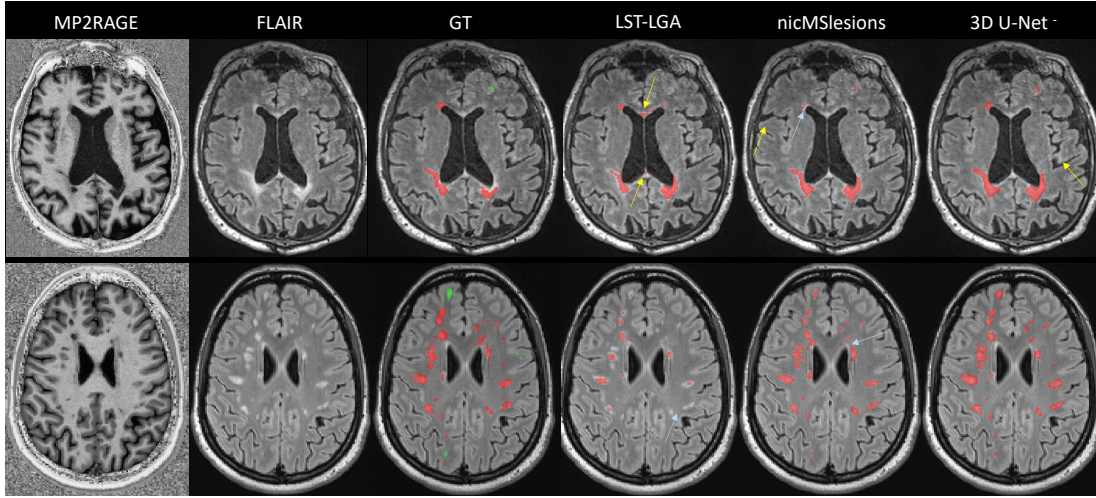
### 4.3.1 Cross-validation

There were 75 training cases from which 20% (15 patients) were used as validation set to determine early stopping and optimize the threshold and 15 testing cases in each fold (see Figure 4.4). The threshold was chosen as the value (with intervals of 0.05) that gave the highest dice coefficient in the validation set. Qualitative results are reported in Figure 4.5, where the first row shows a slice of a subject from Dataset I, and the second row one from a subject of Dataset II. Moreover, in Figure 4.6 we show examples of CL correctly detected by our proposed 3D U-Net-.

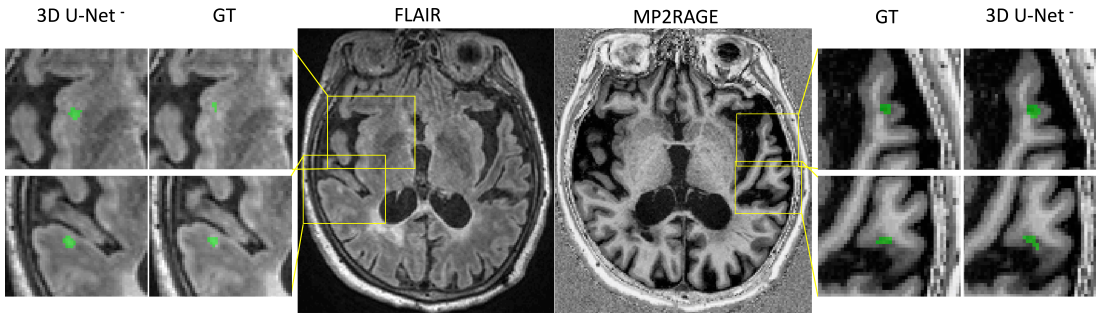
The median of all metrics obtained in the test folds of the cross-validation are shown in Table 4.1. 3D U-Net- achieves a 75% CL detection rate and the best performance in all metrics except for the PPV, for which LST-LGA has the best value (71%). Overall, the results of 3D U-Net- and nicMSLesions are in line with other recent works, both for WML (Carass et al., 2017) and for CL (Fartaria et al., 2016; La Rosa et al., 2019).

We further analyzed the patient-wise LTPR and LFPR by the boxplots in Figure 4.7. 3D U-Net-significantly outperforms the other two methods in the three detection accuracy metrics (refer to Figure 4.7 for the p-values). Moreover, there is only a slight difference between the 3D U-Net- CL and WML detection rate. Figure 4.8 shows the correlation between the manually segmented lesion volumes and the ones automatically estimated. On the identity lines the predicted and estimated volumes are equal an the closer the results are to the line, the more accurate. The solid lines show the linear regression model fitted with these points, and the Pearson's linear correlation coefficient is reported for each method in the legend.

We also analyzed the results per lesion size range. As mentioned above, we decided to stay below the MS WML minimum size recommended in the diagnostic criteria, as our datasets



**Figure 4.5** – Visual illustration of results for all three methods. In the first row is presented a subject from Dataset I, and in the second row a subject from Dataset II. Colorcode of overlay of the ground truth (GT): WML in red and CL in green. The yellow arrows point at false positives of the automatic methods, whereas the light blue arrows at false negatives.



**Figure 4.6** – Four examples of correctly detected cortical lesions by our proposed method. All lesions are better visible in the MP2RAGE sequence compared to the FLAIR.

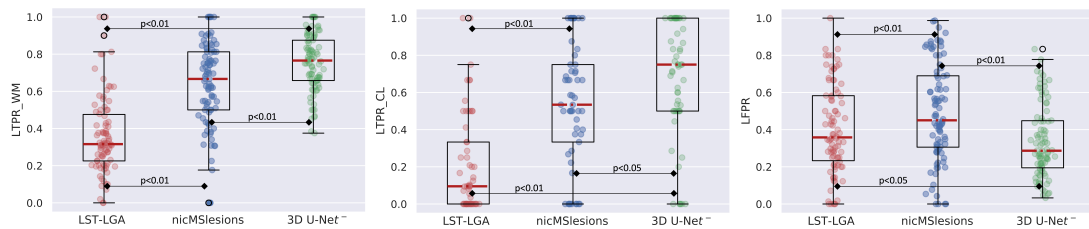
include also CL. In Figure 4.9 is presented the detection rate for WML and CL separately for the three methods.

#### 4.3.2 Independent test set

In the second scenario we aim at exploring the generalization of our segmentation framework by training on one single center while testing on the other one (see Figure 4.4). We trained a model on Dataset I, keeping 20% of the cases for validation ( $n=11$ ), and evaluated the performance on the independent data of Dataset II. This setting simulates a more realistic scenario where our method is evaluated with imaging data acquired outside the training site. Moreover, let us note that patients in Dataset II are at very early stages of MS with EDSS scale

**Table 4.2** – Median values (IQR) of the metrics obtained for the different methods on the independent test set (36 subjects). The minimum lesion size is 3 voxels. The best result per each metric is shown in bold.

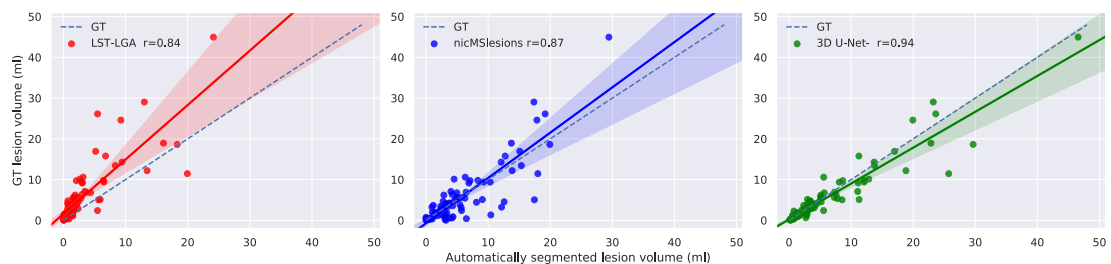
Method	Dice	AVD	PPV	LTPR	LTPR_WM	LTPR_CL	LFPR
LST-LGA	0.32 (0.27)	0.63 (0.25)	<b>0.75</b> (0.29)	0.28 (0.21)	0.30 (0.19)	0.19 (0.20)	0.35 (0.34)
nicMSlesions	0.58 (0.18)	0.27 (0.21)	0.62 (0.26)	0.67 (0.16)	<b>0.69</b> (0.20)	0.59 (0.27)	0.31 (0.37)
3D U-Net <sup>-</sup>	<b>0.60</b> (0.19)	<b>0.13</b> (0.27)	0.64 (0.24)	<b>0.69</b> (0.10)	<b>0.69</b> (0.12)	<b>0.71</b> (0.48)	<b>0.27</b> (0.33)



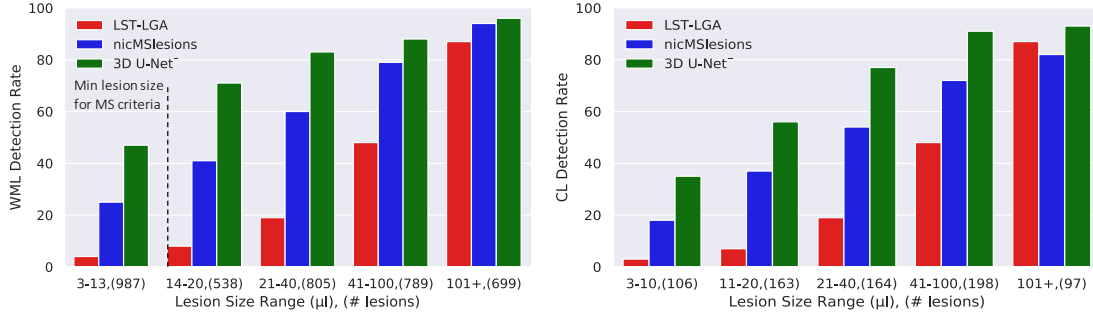
**Figure 4.7** – Boxplots of the patient-wise metrics obtained for the three methods. The  $p$ -values are computed with the Wilcoxon signed rank test.

ranging between 1 and 2 only (see Figure 4.2).

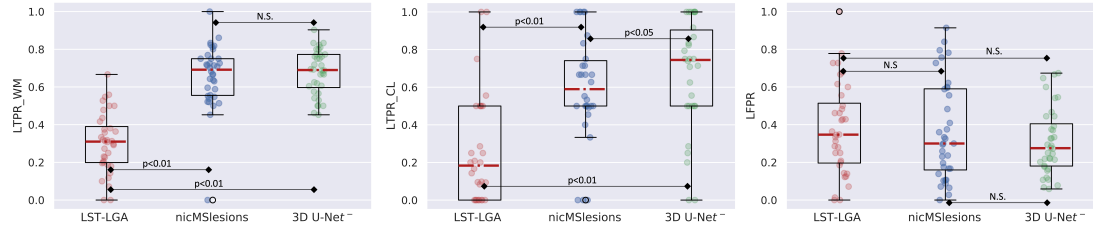
In Table 4.2 we report the quantitative criteria evaluated on an independent test set. Our proposed method achieves a 71% CL detection rate and outperformed the others in all metrics except for the PPV and the WML LTPR, for which it achieves the best results (69%) together with nicMSlesions. Finally, in Figure 4.10 we show the boxplots for the WM and CL LTPR, and the LFPR, with their relative significant differences performed with the Wilcoxon signed rank test.



**Figure 4.8** – Correlation between the manual lesion volume and the automatically segmented one. The solid lines show the linear regression model between the two measures along with confidence interval at 95%. The dashed lines indicate the expected lesion volume estimates. The Pearson's linear correlation coefficient between manual and automatic lesion volume is reported in the legend ( $r=0.84$  for LST,  $r=0.87$  for nicMSlesions,  $r=0.94$  for 3D U-Net<sup>-</sup>).



**Figure 4.9** – Detection rate for WML and CL for the three methods evaluated considering different lesion size ranges. In parenthesis are reported the total number of lesions of each range. The dashed line shows the minimum lesion size established for WML.



**Figure 4.10** – Boxplots of the patient-wise metrics obtained for the three methods in the separate test set. The p-values are computed with the Wilcoxon signed rank test. N.S.: not significant.

## 4.4 Discussion

With a simple deep learning architecture that was trained end-to-end and did not include explicit feature engineering or advanced pre-processing, we segmented CL and WML using one conventional and one specialized MRI contrast with a detection rate of 76% (median) and lesion false positive rate of 29% (median). The proposed prototype 3D U-Net<sup>-</sup> outperformed the baseline methods and proved to generalize well for cases acquired with a different scanner. Consequently, we foresee our proposed method as a useful support tool in a research setting where MS CL and WML need to be segmented in a high number of MRI scans. The automatic segmentation obtained could, for instance, represent a first lesion labelling to be then refined by the experts, allowing them to further standardize and speed up the overall process.

As reflected by being recently introduced in the MS diagnostic criteria (Thompson, Banwell, et al., 2018), CL are of great clinical interest, yet their automated segmentation has been receiving little attention. The fully-convolutional 3D U-Net<sup>-</sup> has a relatively low number of training parameters (3.8 M) and it is fast to run at inference time (about 20 seconds to infer a new subject, not counting initial intra-subject registration). The method was evaluated on two datasets of 54 and 36 subjects, respectively. In order to emulate a realistic clinical setting, we

considered a minimum lesion size of 3 voxels, which is the recommended minimum size of CL for 3D sequences with at least 1 mm voxel spacing (Geurts, Bo, et al., 2005). This is smaller than what previous automatic studies have reported for WM lesions (20 voxels in (Valverde et al., 2017), for instance), and also much smaller than the clinical definition of minimum WML diameter of 3 mm (Grahl et al., 2019; Thompson, Banwell, et al., 2018) corresponding in a spherical approximation to a volume of about  $14 \text{ mm}^3$ .

We compared our proposed approach with two reference methods: an unsupervised lesion growth approach (LST-LGA) and a supervised deep learning technique (nicMSlesions), both of which were originally proposed for MS WML segmentation only. Our first evaluation consisted in a 6-folds per site stratified cross validation including all 90 subjects. In this way, we evaluated the performance over a large dataset, with significant variability of lesion count and volume across subjects. We considered the main metrics evaluated in MS lesion segmentation challenges (Carass et al., 2017). In particular, we focused our study on the CL and WML detection rate, as well as the absolute volume difference, because lesion count and volume are included in the MS diagnostic criteria (Thompson, Banwell, et al., 2018). Among the metrics reported, the PPV is a pixel-wise metric that by definition ignores the false negatives, which in our case represent the missed lesions, and are therefore quite important. Moreover, it should be noted that the widely used dice coefficient is not reflecting well the overall performance in the case of very small structures as in our study. For how it is computed, the dice is naturally biased towards the lesion, penalizing cases with only small structures more severely.

Table 4.1 shows that 3D U-Net<sup>+</sup> outperformed the other methods in all metrics except for the PPV. The high detection rate for both WML and CL proves the capability of our proposed method to detect both types of MS lesions with similar accuracy, even if the latter have a low contrast in the FLAIR images (see Figure 4.1 and 4.5). Performing Wilcoxon signed rank test using a Bonferroni correction for multiple comparisons, 3D U-Net<sup>+</sup> is significantly better than the other approaches in CL and WML detection rate and in false positive rate. Our claims are also supported by an analysis of the correlation between the manual total lesion volume and the one automatically segmented (see Figure 4.7. Moreover, we explored the lesion detection rate per lesion size (Figure 4.8). By increasing the minimum lesion size all methods perform better, and their relative difference in detection rate decreases.

In the independent test scenario, we trained the two supervised methods with the patients of Dataset I and tested the models with the subjects of Dataset II. In this way, we evaluated the methods on cases acquired in another site and with a different scanner. We acknowledge,

however, that the acquisition parameters of both scanners were very similar, thus limiting the generalization to this particular setting. It can be observed in Table 4.2 that also in this independent test setting, the supervised deep learning approaches outperform LST-LGA in terms of lesion detection, dice coefficient, and also volume difference. NicMSLesions significantly improves with respect to the previous scenario its false positive rate (p-value  $< 0.01$ ) and reaches the best WM LTPR together with 3D U-Net<sup>-</sup> (69%). Thus, these results support previous claims of generalizing very well in WML segmentation to cases from different datasets (Aslani et al., 2019; Weeda et al., 2019). Moreover, let us note that in this study we assessed nicMSLesions considering a much smaller minimum lesion size than in any other previous study where it was tested (Valverde et al., 2017; Valverde et al., 2019; Weeda et al., 2019). Compared to a 69% WML detection rate, nicMSLesions CL detection rate was only 59%. We hypothesize this is due to the dependency of its lesion candidate selection on the FLAIR intensity value. As the sensibility of FLAIR to CL is limited, several CL might be discarded for this reason. Interestingly, 3D U-Net<sup>-</sup> performs even better than in the first scenario in terms of volume difference (p-value  $< 0.05$ ), but understandably performs slightly worse in cortical and white matter LTPR (p-value N.S.). In particular, it reaches the same result as nicMSLesions (69%) in terms of WML, but it performs significantly better for the CL detection rate (71% vs 59%, p-value  $< 0.05$ , Wilcoxon signed rank test). Furthermore, is worth noting that in this scenario the ground truth of the training and testing datasets was delineated by different experts. This might contribute to intrinsic differences in the labeled masks, thus posing additional challenges to the automatic methods.

A major advantage of our CL and WML segmentation framework is that it is based on two 3T MRI sequences only. As successfully done in previous automatic segmentation studies (Fartaria et al., 2016; La Rosa et al., 2019), we explore the use of the specialized MP2RAGE sequence instead of the conventional MPAGE. Nonetheless, we acknowledge that this sequence is still not widely acquired in clinical routine for MS. However, in order to visually detect CL, specialized sequences are needed (Calabrese et al., 2009; Filippi et al., 2019), and the MP2RAGE is a promising substitute of the conventional MPAGE. MP2RAGE has increased lesion and tissue contrast compared to the common MPAGE (Kober et al., 2012). Moreover, while currently the MP2RAGE sequence requires about 8 minutes of scanning time protocol, recent developments have shown that its acquisition time can be reduced to less than 4 minutes without compromising the image quality (Mussard et al., 2020). Thus, it could easily be included (additionally to or instead of the MPAGE) in a 3T MRI MS clinical protocol in order to support the CL analysis.



In contrast to other studies, we report the detection accuracy of very small lesions that when evaluated using an overlap measure, such as Dice coefficient, and in presence of large lesions would not contribute strongly to the performance evaluation but may be clinically relevant. However, our method detected very small lesions (between 3-10 voxels, 3-10  $\mu\text{L}$ ) poorly. We believe this is due to partial volume effect and artefacts affecting them and could be improved if more small lesions would be included in the training dataset. In our study, experts agreed by consensus on the lesion detection, but the ground truth masks of each dataset were manually delineated only once and by different experts. This limits our analysis, not allowing us to compare the automatic methods' performance to the inter and intra rater reliability. Moreover, given the 3T MRI settings of our work, the vast majority of CL detected by the experts (89%) are leukocortical/juxtacortical. Thus, also the accurate CL detection of our proposed method is limited to this particular CL type.

In conclusion, we achieve an accurate CL and WML segmentation with a simple 3D fully-convolutional CNN, which operates on data that is not treated with advanced pre-processing, is fast to run at inference time and generalizes well across two different scanners. The considered MRI sequences are close to the ones of a clinical scenario, meaning that the proposed approach could support experts in the lesion segmentation process. Future work will aim at improving the lesion delineation. This might include, for instance, exploring the T1 map acquired together with the MP2RAGE sequence. Moreover, we will tackle the challenging task of providing an output segmentation that classifies the lesions in WML and CL types, as this could have an added clinical value.

Currently, this framework is exploited by our collaborators at Basel University Hospital who are running it for research purposes to obtain a first MS lesion segmentation before manually correcting it and performing additional analysis. To date, our proposed method has been also used in several publications in peer-reviewed journals (Barquero et al., 2020; Bosticardo et al., 2021; Lu, Barakovic, et al., 2021; Lu, Yoo, et al., 2021; Maggi et al., 2021; Rahmanzadeh et al., 2021; Yu et al., 2021).



## 5 Synthetic MP2RAGE

*The work presented in this chapter is adapted from the postprint version of the following article: Francesco La Rosa, Thomas Yu, Germán Barquero, Jean-Philippe Thiran, Cristina Granziera, Merixtell Bach Cuadra: "MPRAGE to MP2RAGE UNI translation via generative adversarial network improves the automatic tissue and lesion segmentation in multiple sclerosis patients." in Computers in Biology and Medicine, Volume 132 (2021); doi: [10.1016/j.combiomed.2021.104297](https://doi.org/10.1016/j.combiomed.2021.104297). The first author (author of this thesis) contributed to the conceptualization, methodology, software, validation, and writing of the study.*

### Abstract

**Background and Objective:** compared to the conventional magnetization-prepared rapid gradient-echo imaging (MPRAGE) MRI sequence, the specialized magnetization prepared 2 rapid acquisition gradient echoes (MP2RAGE) shows a higher brain tissue and lesion contrast in multiple sclerosis (MS) patients. The goal of this work is to retrospectively generate realistic-looking MP2RAGE uniform images (UNI) from already acquired MPRAGE images in order to improve the automatic lesion and tissue segmentation.

**Methods:** for this task we propose a generative adversarial network (GAN). Multi-contrast MRI data of 12 healthy controls and 44 patients diagnosed with MS was retrospectively analysed. Imaging was acquired at 3T using a SIEMENS scanner with MPRAGE, MP2RAGE, FLAIR, and DIR sequences. We train the GAN with both healthy controls and MS patients to generate synthetic MP2RAGE UNI images. These images were then compared to the real MP2RAGE UNI (considered as ground truth) analyzing the output of automatic brain tissue and lesion segmentation tools. Reference-based metrics as well as the lesion-wise true and false positives, Dice coefficient, and volume difference were considered for the evaluation. Statistical

differences were assessed with the Wilcoxon signed-rank test.

**Results:** the synthetic MP2RAGE UNI significantly improves the lesion and tissue segmentation masks in terms of the Dice coefficient and volume difference (p-values < 0.001) compared to the MPRAGE. For the segmentation metrics analyzed, no statistically significant differences are found between the synthetic and acquired MP2RAGE UNI.

**Conclusion:** synthesized MP2RAGE UNI images are visually realistic and improve the output of automatic segmentation tools.

## 5.1 Introduction

Magnetic resonance imaging (MRI) plays a crucial role in multiple sclerosis (MS) as it is the imaging technique of choice for diagnosing the disease and monitoring its progression (Filippi et al., 2016). MRI for MS includes both T1-weighted and T2-weighted sequences. T1-weighted images are commonly preferred for visualizing the anatomy of the brain and quantifying normal appearing gray matter (GM) and white matter (WM), but are also helpful for detecting lesional tissue in MS patients (Thompson, Banwell, et al., 2018). Currently, one of the most used T1-weighted sequences at high magnetic field (3T) is the three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MPRAGE) (Mugler & Brookeman, 1990). This sequence offers accurate anatomical images of the brain in a reasonable acquisition time (about 5 minutes) and is routinely included in the standard clinical protocols. An extension of the MPRAGE is the so-called magnetization prepared 2 rapid acquisition gradient echoes (MP2RAGE) (Marques et al., 2010). This specialized sequence combines two images acquired at different inversion times, creating T1-weighted uniform images (UNI) with excellent tissue contrast and self-correction for B1 - bias field. Moreover, in addition to the UNI, a T1 relaxation map (T1 map) is also concurrently obtained as output of the acquisition. Although the T1 map provides important quantitative information, the UNI image is the one primarily used for both visual and automatic inspection. To the best of our knowledge, the T1 maps are not exploited for tissue or lesion segmentation by any automatic methods. For this reason, we focus on the UNI images, and throughout this work, by MP2RAGE we refer to its UNI image. The MP2RAGE has been shown to yield an improved tissue segmentation compared to the MPRAGE with classical segmentation tools optimized for conventional T1 contrasts (Cuadra et al., 2013; Okubo et al., 2016). Additionally, further studies (Fartaria et al., 2016; Kober et al., 2012) have described its valuable application to MS patients, obtaining an improved lesion visualization and detection, in particular regarding cortical lesions (CL). Currently, despite its promising added value, MP2RAGE remains mainly in research settings as changing the

MRI clinical protocols is a lengthy process. It would thus be highly beneficial if MP2RAGE-like images could be estimated from current MPAGE acquisitions in order to benefit from the enhanced tissue contrast of MP2RAGE without waiting for its clinical adoption.

In the last decade, a class of deep learning algorithms called generative adversarial networks (GANs) (Goodfellow et al., 2014) have emerged as the state-of-the-art technique for generating new synthetic data. Their two main components are a generator, responsible for synthesizing new realistic data from a given input, and a discriminator, whose goal is to distinguish real and synthetic data. The generator and the discriminator are trained simultaneously in competition with each other, and this results in realistic-looking data being produced. GANs are especially effective with images, exploiting convolutional neural networks (CNNs) for their generator and discriminator. Starting from the computer vision field, they are finding several broad applications and have been recently explored also for medical imaging (Kazeminia et al., 2019). Considering that data scarcity and class imbalance often represent an obstacle for training CNNs, GANs have arguably several potential applications also in the medical field. Their main drawback, however, is that in some cases they introduce artifacts or unrealistic details, which cannot be tolerated in the clinical context either for diagnosis or for follow up purposes.

GANs have been explored for MRI for image reconstruction (Shende et al., 2019), to increase the image resolution (Chen et al., 2020), augmenting and increasing dataset size (Shin et al., 2018), as well as generating new contrasts or converting MRI images to CT (and vice-versa) (Lei et al., 2019). Specifically, for MS patients, they have been shown to be an effective method for data augmentation (Salem et al., 2019) and for generating realistic-looking MRI contrasts. Recently, Finck et al. (Finck et al., 2020) have proposed a GANs for the generation of synthetic double inversion recovery (DIR) images starting from standard MRI acquisitions as T1, T2, and fluid-attenuated inversion recovery (FLAIR) of MS patients. Two independent readers then evaluated the images, showing that the synthetic DIR was able to depict significantly more MS lesions compared to the conventional FLAIR. This could have an important application for MS diagnosis. However, as argued by Hagiwara et al. in a recent editorial article (Hagiwara et al., 2020), adding a synthetic contrast implicitly requires additional time by the experts to analyze it. They conclude mentioning that it would be interesting to explore how an automatic lesion segmentation method would perform with it. Notably, CNNs are inspired by the human brain's learning process, but their way of extracting and combining features does not necessarily reflect what experts do.

In this work, we propose a GAN that, through a pixel-by-pixel image translation process, synthesizes the MP2RAGE images corresponding to the input MPAGE images. The GAN is

trained on a dataset consisting of 12 healthy controls and 8 MS patients with axial, sagittal, and coronal 2D slices. The trained model is then tested on 36 MS patients in the early stages of the disease. The synthesised images are evaluated in two ways. First, we compare the MPRAGE images to the synthesized MP2RAGE images using quality metrics. Second, we compare differences in the segmentation of these images using two different automated methods: a supervised MS lesion segmentation CNN (La Rosa, Abdulkadir, et al., 2020) and an unsupervised tissue segmentation approach (B. B. Avants et al., 2011).

## 5.2 Materials and method

### 5.2.1 Subjects

In this study, we consider a cohort of 56 subjects (36 female / 20 male, mean age  $34 \pm 9$ , age range [20,61] years). Of these, 12 are healthy controls, and the remaining 44 MS patients diagnosed with relapsing-remitting MS according to the McDonald criteria (Thompson, Banwell, et al., 2018). These patients were at early stages of the disease: the mean disease duration was  $(1.9 \pm 1.5)$  years and the Expanded Disability Status Scale (EDSS) scores ranged from 1 to 2 (mean  $1.5 \pm 0.3$ ).

### 5.2.2 Imaging

Imaging was acquired on a 3T MRI whole-brain scanner (MAGNETOM Trio, Siemens Healthcare, Erlangen, Germany) with the following 3D sequences (acquired in the sagittal plane) with a resolution of  $1 \times 1 \times 1.2 \text{ mm}^3$ : 3D FLAIR (field-of-view=  $256 \times 240 \times 176$ , TR/TE/TI= 5000,394,1800 ms, flip angle=  $120^\circ$ , acquisition time= 6 min), 3D MPRAGE (field-of-view=  $256 \times 240 \times 160$ , TR/TE/TI= 2300,2.98,900 ms, flip angle=  $7^\circ$ , acquisition time= 5 min), DIR (field-of-view=  $256 \times 240 \times 160$ , TR/TE/TI1/TI2= 10000,218,450/3650 ms, flip angle=  $90^\circ$ , acquisition time= 13 min), and MP2RAGE (field-of-view=  $256 \times 240 \times 176$ , TR/TE/TI1/TI2= 5000,2.89,700,2500, flip angles=  $4^\circ/5^\circ$ , acquisition time= 8 min). In the following sections we will refer to FLAIR and MPRAGE as "conventional sequences", as these are part of the clinical protocol, and to DIR and MP2RAGE as "specialized sequences", as these are mainly used in a research setting. The study was approved by the Ethics Committee of our institution, and all subjects gave written informed consent prior to participation.

### 5.2.3 Manual segmentation

A radiologist and a neurologist, with 7 and 11 years of experience in MS research, respectively, detected white matter lesions (WML) and cortical lesions (CL) by consensus on the MS patients' scans using all four imaging modalities and the three orthogonal planes. A trained technician then manually delineated all lesions, again considering both the conventional and advanced sequences using ITK-SNAP (Yushkevich et al., 2006). Importantly, pathological validation remains the ultimate gold standard for lesion detection and this work does not aim at evaluating the MRI sensitivity compared to pathology.

### 5.2.4 Generative adversarial network

The goal of image-to-image translation networks is to learn the mapping between an input and an output image. GANs have been successfully proposed as a solution for this task, where the generator learns this mapping, while the competing discriminator pushes it to further improve. One of the first and top-performing image-to-image translation networks proposed is the pix2pix architecture (Isola et al., 2018). This is a general-purpose GAN which includes a U-Net-like CNN as generator and a PatchGAN as discriminator. It has been applied to a wide range of tasks, such as translating day photographs to night ones or sketches to photographs (Isola et al., 2018). Pix2pix combines a common pixel-wise L1 loss with the adversarial loss of the discriminator, showing how the latter consistently helps improve the results.

**Proposed framework.** Inspired by the pix2pix architecture (Isola et al., 2018), we propose a pixel-wise translation network that receives as input MP2RAGE images and outputs realistic-looking MP2RAGE ones (synMP2RAGE). The original implementation is adapted with additional residual blocks in the generator increasing the overall number of parameters. Moreover, as proposed by Kupyn et al. (Kupyn et al., 2018), we include a global skip connection between the input and the output of the last layer of the generator. In this way, the CNN learns a residual correction to the input MP2RAGE image. This empirically decreased the training time and improved the generator robustness. The architecture of the generator is illustrated in Figure 5.1. Contrary to pix2pix, our discriminator classifies at each iteration the entire image as either real or fake, and it is composed of five convolutional layers, each one followed by a Leaky ReLU activation function. The downsampling is learned through the convolutions and no max-pooling or fully-connected layers are present, as recommended in a previous work (Radford et al., 2016). The complete fully-convolutional discriminator's architecture has about 50k trainable parameters.

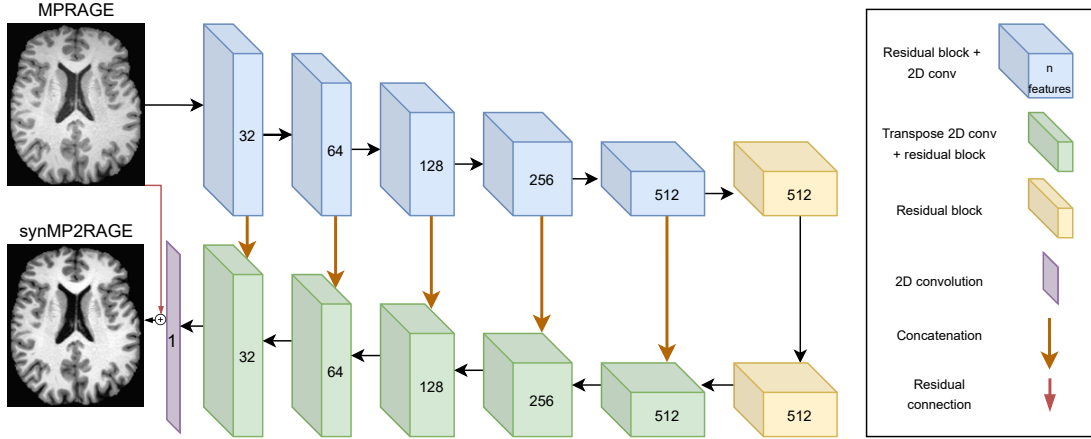


Figure 5.1 – Scheme of the generator of the GAN proposed.

**Loss functions.** The first loss considered is the pixel-wise mean absolute error ( $L_1$  loss) between the images produced by the generator and the target ones, as used in most works in the literature (Chen et al., 2020; Isola et al., 2018; Kupyn et al., 2018). The second loss is given by the ability of the generator of fooling the discriminator. This is computed as the sigmoid cross-entropy (adversarial loss) between the discriminator output and an array of 0s and 1s, where the 0s represent fake images and the 1s real ones. MRI acquisitions, however, are considerably affected by noise, and the  $L_1$  and the adversarial losses are not sufficient to produce realistic-looking images. In order to account also for the visual quality of the produced MP2RAGE considering overall spatial features/textures, we added a perceptual loss. This is computed as the mean absolute error between the feature maps of the fourth convolutional layer of the pre-trained VGG-16 model produced by the real and the synthesized MP2RAGE. The fourth layer was chosen as it gave the best results experimentally and it is as well in accordance with a previous work on MRI (Yang et al., 2018). Deeper layers extract more abstract features and did not seem to be beneficial for our scope. The perceptual loss, together with the adversarial one, is responsible for making the images look more realistic. Empirically, we observed that they prevented the network from over-smoothing the synthesized MP2RAGE, which is also observed to occur in the use of perceptual losses in computer vision (Johnson et al., 2016; Sajjadi et al., 2017). Summing up, the total loss is given by:

$$L_{Tot} = (\alpha * L_1) + (\beta * L_{perceptual}) + L_{adversarial} \quad (5.1)$$

Where  $\alpha$  and  $\beta$  were set to 150 and 5, respectively. Generator and discriminator are trained and updated simultaneously at each iteration with the relative losses, refer to Figure 5.2 for a



scheme of the framework.

### 5.2.5 Pre-processing

First, each MP2RAGE acquisition in our dataset is rigidly registered to the corresponding MP2RAGE UNI using ELASTIX (Klein et al., 2010). Second, the MP2RAGE image is skull-stripped with ANTs (B. B. Avants et al., 2011) and the brain mask is applied to the MP2RAGE image. This step is necessary in order to remove the noise present outside the brain in the MP2RAGE. Finally, both volumes for each subject are normalized to zero mean and standard deviation of one.

### 5.2.6 Training details

For each input 3D MP2RAGE volume, we extracted 150x150 pixel slices across the three orthogonal views of the brain (axial, coronal, and sagittal planes). The 9000 2D slices obtained in this way were then concatenated together. This particular size was chosen as it includes the entire brain for all subjects. Data augmentation is then performed in order to prevent overfitting. In particular, we randomly crop 128x128 or 64x64 images at each iteration. The latter ones are then resized to the 128x128 input size using bilinear interpolation. Moreover, random flipping along both axes is applied. The images obtained are then fed as input to the generator with a mini-batch size of 1. The initial learning rate set is 1e-5 with Adam (Kingma & Ba, 2017) as optimizer for both the generator and the discriminator. The generative framework has been developed in Tensorflow 2.1.0 (Abadi et al., 2016) using one NVIDIA Tesla P100 GPU. The code is deployed as a Jupyter Google Colaboratory<sup>I</sup> notebook which simply runs on the internet browser of any computer taking advantage of the free GPU usage available in Google Colaboratory. The code is available on our research website<sup>II</sup>.

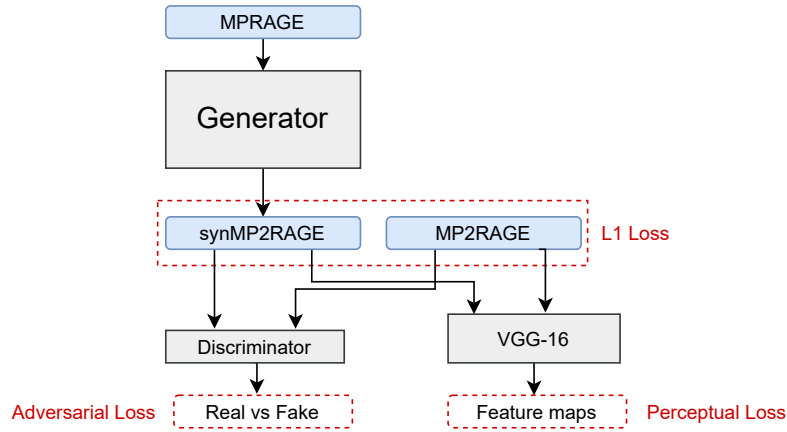
### 5.2.7 Inference

At inference time the MP2RAGE is skull-stripped, normalized to zero mean and unit standard deviation, and zero-padded, obtaining 256x256 pixels slices in each of the three different planes. As the generator’s architecture is fully-convolutional, these inference input dimensions are arbitrary and can be chosen depending on the testing images. The generator is then tested separately with the transversal, coronal and sagittal images. The three output volumes are

---

<sup>I</sup><https://colab.research.google.com>

<sup>II</sup><https://github.com/FrancescoLR/synMP2RAGE>



**Figure 5.2** – Scheme of the generative framework proposed. The three loss functions used are highlighted in red.

finally averaged to obtain the final synthetic MP2RAGE image. The inference process takes about 50 seconds per subject on a UNIX machine equipped with a GPU Tesla P-100 from NVIDIA.

## 5.3 Evaluation

### 5.3.1 Qualitative evaluation

In order to identify possible artificial artefacts introduced by the GAN, an image analysis expert, with more than 20 years of experience in brain MRI, carefully examined, slice-by-slice and in all three orthogonal planes, the 36 synMP2RAGE images of the testing dataset. In a second assessment, the artefacts found were then analyzed in comparison with the corresponding MPRAGE images.

### 5.3.2 Quantitative evaluation

In order to quantitatively compare the synMP2RAGE and the MPRAGE with the original MP2RAGE, we computed three widely used reference-based similarity metrics (Chen et al., 2020; Finck et al., 2020): peak signal-to-noise ratio (PSNR), normalized root mean square error (NRMSE) and mean structural similarity index (SSIM). The metrics were computed per subject, considering the skull-stripped images, and averaged across the entire testing dataset.

### 5.3.3 Automatic segmentation

We propose to test two different automated segmentation approaches in order to objectively evaluate the benefits of analyzing synthetic MP2RAGE images compared to the commonly acquired MP2RAGE. We considered two different methods:

- Atropos (B. B. Avants et al., 2011), an unsupervised approach for brain tissue segmentation distributed with ANTs. This is a Bayesian method that aims at solving the expectation-maximization algorithm modelling the class intensities with either parametric or non-parametric finite mixtures. We initialized the algorithm with k-means and selected 3 classes to be segmented: WM, GM and cerebrospinal fluid (CSF). The likelihood model chosen was a Gaussian and ran for 5 iterations. All other parameters were the default ones. Prior to running Atropos, all scans were skull-stripped and normalized for bias field inhomogeneities with ANTs.
- A CNN recently proposed for MS lesion segmentation (La Rosa, Abdulkadir, et al., 2020). This architecture is based on the 3D U-Net and was specifically adapted to detect both cortical and white matter lesions with high accuracy. We evaluated the CNN, setting all default parameters, with a 3 folds cross-validation over the 36 testing cases. Four different combinations of MRI contrasts as input for the segmentation were considered: FLAIR-MPRAGE, FLAIR-MP2RAGE, FLAIR-DIR, and FLAIR-synMP2RAGE. Prior to training, the second contrast was rigidly registered to the FLAIR space with ELASTIX (Klein et al., 2010).

Automated segmentations were assessed computing the Dice coefficient and volume difference for the brain tissues and the detection rate, Dice coefficient, and volume difference for the cortical and white matter lesions.

### 5.3.4 Statistical analysis

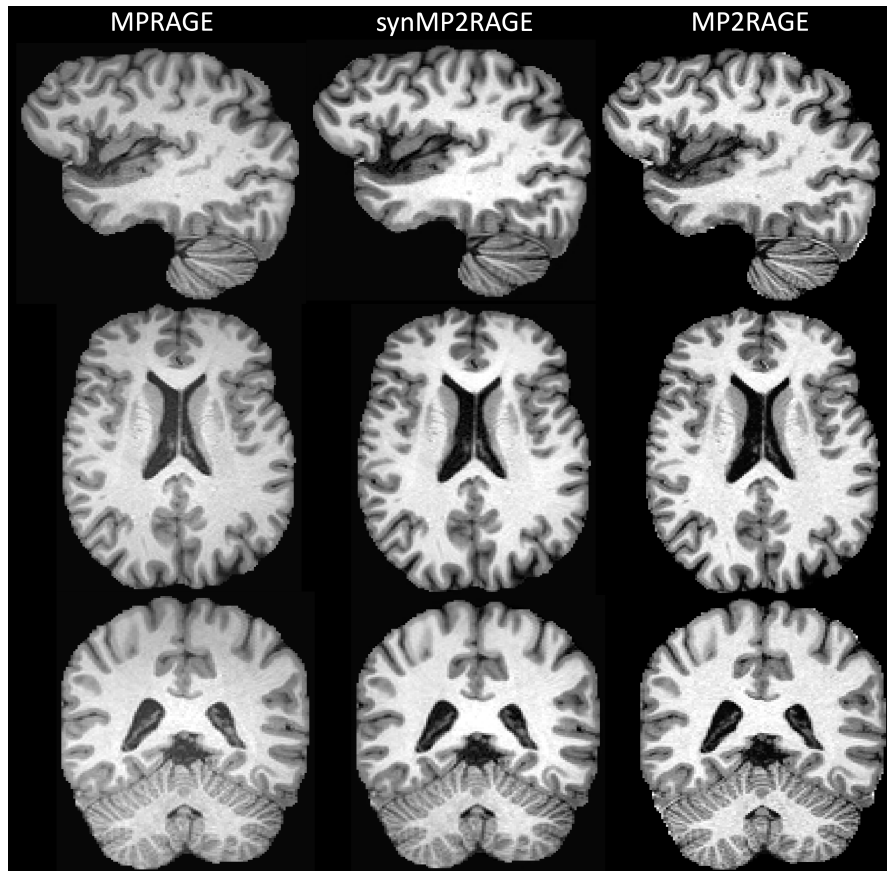
For all metrics considered, the Wilcoxon signed-rank test was computed at the subject-wise level using the SciPy Python library (SciPy 1.0 Contributors et al., 2020). Statistical differences were considered for p-values < 0.05.

## 5.4 Results

The synthetic images inferred from the test set were evaluated both qualitatively and quantitatively. Qualitative results, comparing a slice of each of the three orthogonal planes of MPRAGE, MP2RAGE, and synMP2RAGE, are shown in Figure 5.3. As can be observed, our generative approach synthesizes images that are consistent in the three planes and exhibit a visually evident increase in tissue contrast, with only a slight over-smoothing compared to the real MP2RAGE. In the qualitative evaluation, our image analysis expert judged the synMP2RAGE images of high quality, with good tissue contrast and low noise. The visual assessment revealed the presence of common MRI artefacts such as blurred areas, aliasing artefacts, and checkboard patterns in few slices of 24 out of 36 subjects. However, analyzing the corresponding MPRAGE images, it was verified that all artefacts were already present in the original acquisitions. In Figure 5.4 we illustrate the different types of artefacts found. Interestingly, for one subject the MP2RAGE was affected by strong motion, whereas the MPRAGE and generated synMP2RAGE seem fine. Overall, our expert concluded that the GAN primarily learns an intensity mapping, and no new artefacts are introduced or removed, except for the bias field which is much reduced or even removed.

Residual images are presented in Figure 5.5. Compared to the MPRAGE, we can notice that synMP2RAGE improves both the lesion and tissue contrast, showing high residual values. Looking at the residual with MP2RAGE, however, we see that, as expected, the method is not perfect and between the prevailing random noise, some border voxels along the CSF have high residuals. Moreover, in Figure 5.6 we report the cumulative histogram over the 36 testing cases for MPRAGE, MPRAGE with N4 bias field correction, MP2RAGE, and synMP2RAGE. It can be seen that the bias field correction contributes to increasing the prominence of the two peaks of GM and WM. By contrast, in the synMP2RAGE histogram, the two peaks, besides having greater prominence, are also farther apart from each other compared to the MPRAGE. As a consequence, the synMP2RAGE histogram almost matches that of the real MP2RAGE, confirming the added value of the synthesized images.

Results of the quantitative evaluation considering reference-based similarity metrics are reported in Table 5.1. We compare the metrics between synMP2RAGE and the ground truth MP2RAGE and those between MPRAGE and the ground truth MP2RAGE. For all three metrics, synMP2RAGE outperforms the initial MPRAGE, achieving for the PSNR, NRMSE, and SSIM 31.39, 0.13, and 0.98, respectively. This shows that our method is transforming the MPRAGE input closer to the MP2RAGE.



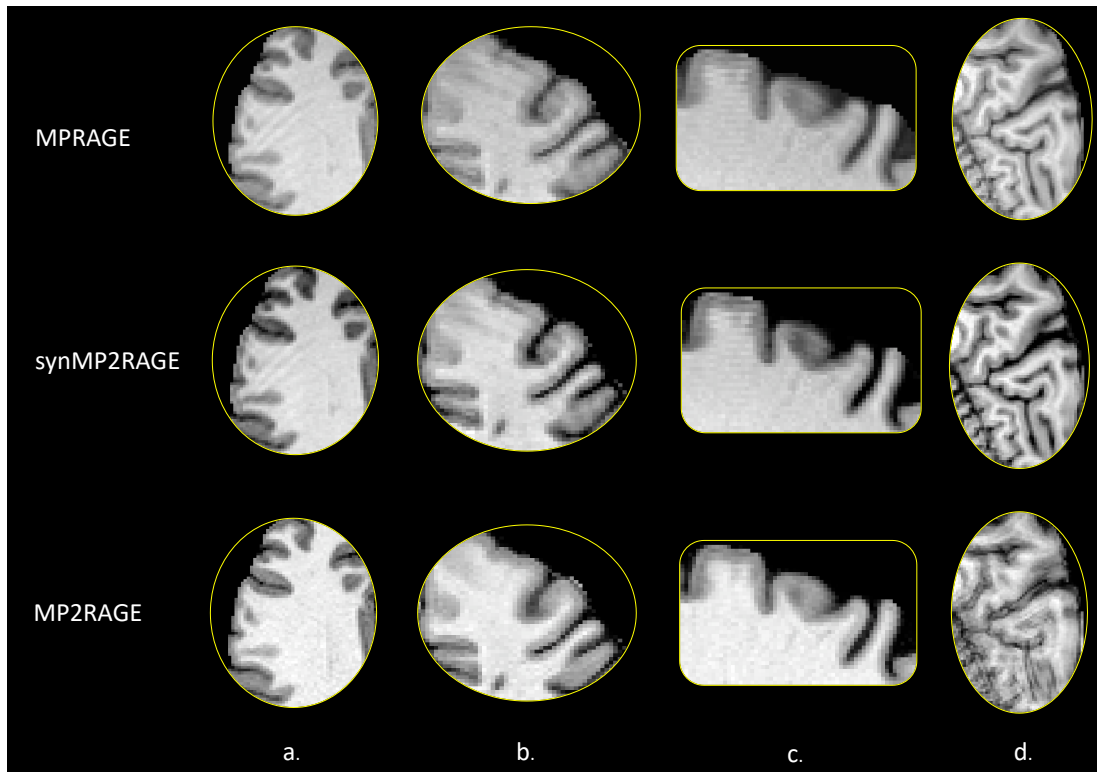
**Figure 5.3** – Qualitative results showing the synMP2RAGE compared to the acquired one and the MPRAGE.

Turning now to the automatic segmentation evaluation, the synMP2RAGE images were assessed in terms of both the lesion and tissue masks obtained. Visual examples of lesion and tissue segmentations for the different contrasts are shown in Figure 5.7. Upon closer inspection of this figure, the improvements of the synMP2RAGE segmentations over those of the MPRAGE can be easily appreciated.

Regarding the automatic brain tissue segmentation results, Figure 5.8 depicts the boxplots of the Dice coefficient for the three main brain tissue types (WM, GM, and CSF). For each tissue

**Table 5.1** – Quantitative metrics and their standard deviation obtained for MPRAGE and synMP2RAGE compared to the reference real MP2RAGE. For all metrics comparison  $p$ -values  $< 0.001$ .  $\uparrow$ : the higher the better,  $\downarrow$  the lower the better.

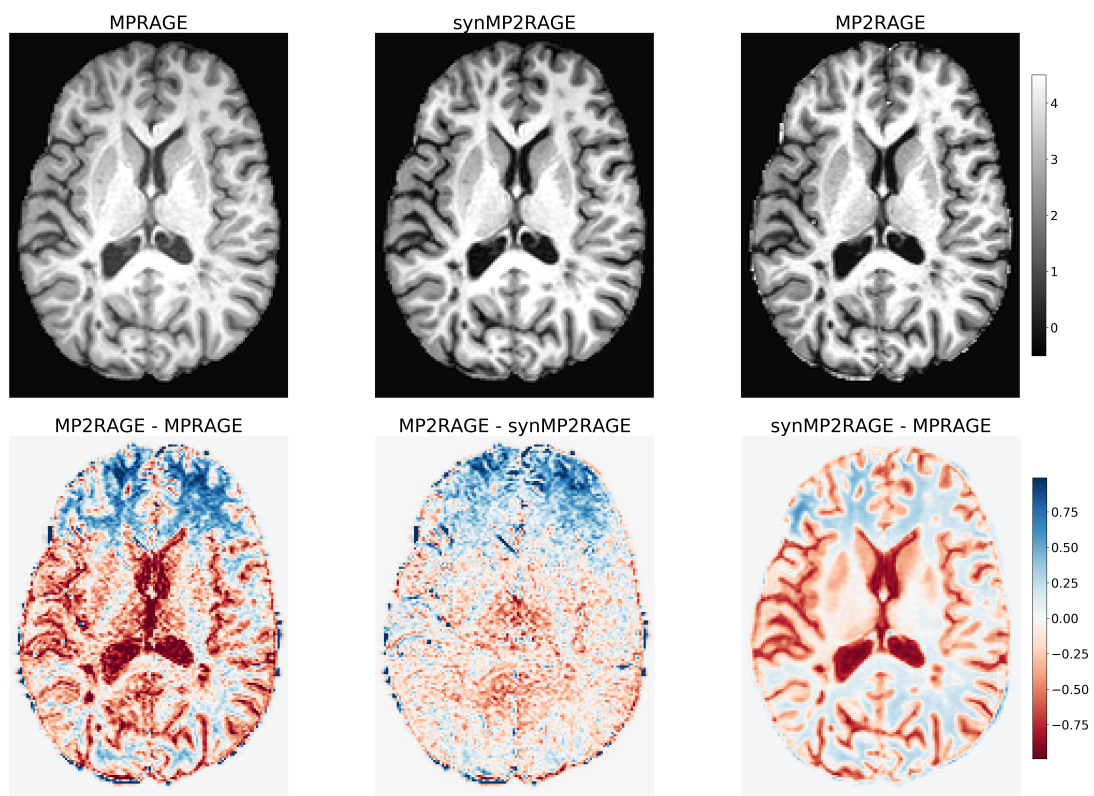
Contrast	PSNR $\uparrow$	NRMSE $\downarrow$	SSIM $\uparrow$
MPRAGE	$29.49 \pm 0.73$	$0.17 \pm 0.01$	$0.97 \pm 0.01$
synMP2RAGE	$31.39 \pm 0.96$	$0.13 \pm 0.01$	$0.98 \pm 0.01$



**Figure 5.4** – Motion artefacts observed in the qualitative analysis. In *a.,b.,c.* different artefacts are present both in the MPRAGE and synMP2RAGE, but not in the MP2RAGE. In *d.* the MP2RAGE is affected by strong motion artefacts which are not observed in the MPRAGE and synMP2RAGE.

type, synMP2RAGE significantly outperformed the MPRAGE (all p-values < 0.0001), reaching in particular for the WM a median value of over 0.94.

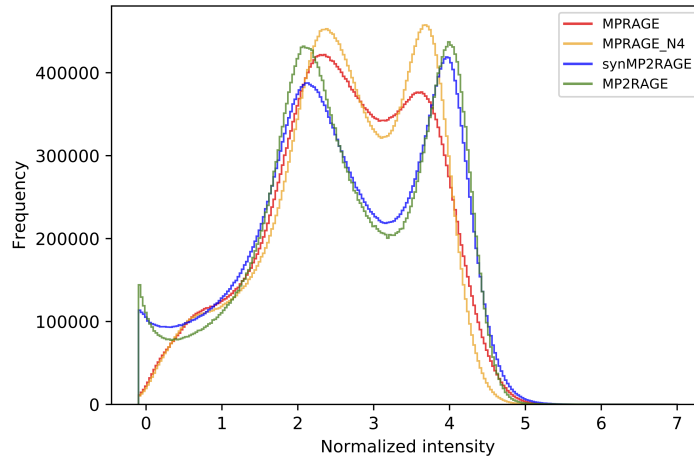
Moving on now to consider the automatic MS lesion segmentation, we considered lesion-wise and voxel-wise metrics for WML and CL. The top row of Figure 5.9 shows the boxplots of lesion-wise true and false positives computed for each patient. No significant differences between any MRI contrast combinations were found. The boxplots of the Dice coefficient and volume difference are presented in the bottom row of the same figure. For both metrics, the segmentation exploiting synMP2RAGE significantly improves compared to both the MPRAGE and the DIR (p-values < 0.001), proving the added value of the synthesized images. Differences were not significant between synMP2RAGE and real MP2RAGE (p-values > 0.05). Finally, the cortical and WM lesion detection rates were computed lesion-wise over the 36 testing cases for the different contrasts are reported in Table 5.2. We can observe that once again in terms of detection rate there are very minimal differences.



**Figure 5.5** – First row: example of a 2D axial slice of MPRAGE, synMP2RAGE, and MP2RAGE. Second row: the three residuals between the images above. The intensity values are normalized with zero mean and unit variance.

## 5.5 Discussion

In this work, we present a tailored Generative Adversarial Network for translating MPRAGE images to realistic-looking MP2RAGE ones. The MP2RAGE is an extension of the MPRAGE MRI sequence which shows a higher brain tissue contrast and is helpful for depicting MS lesions. Regardless of these advantages and its limited acquisition time (about 8 minutes or even less if accelerated (Mussard et al., 2020)), however, it is still not widely acquired in clinical routine MRI acquisitions. Our work aims at supporting MRI studies whenever MP2RAGE was not originally acquired. We evaluated the method on a test set of 36 MS patients. Qualitative results show that the generated synMP2RAGE is visually similar to the acquired MP2RAGE with only a slight over-smoothing (Figure 5.3) and the GAN limits its task mainly to increase the tissue contrast (see the histogram in Figure 5.6). Importantly, our generator did not introduce any artificial artefact compared to the acquired MPRAGE images, supporting our choice of using a combination of L1, adversarial and perceptual loss during training. A quantitative evaluation was then performed considering reference-based



**Figure 5.6** – Cumulative histogram over the 36 cases of the testing set comparing the real contrasts and the generated one. The intensity peak of the background just below zero is omitted for visualization purposes.

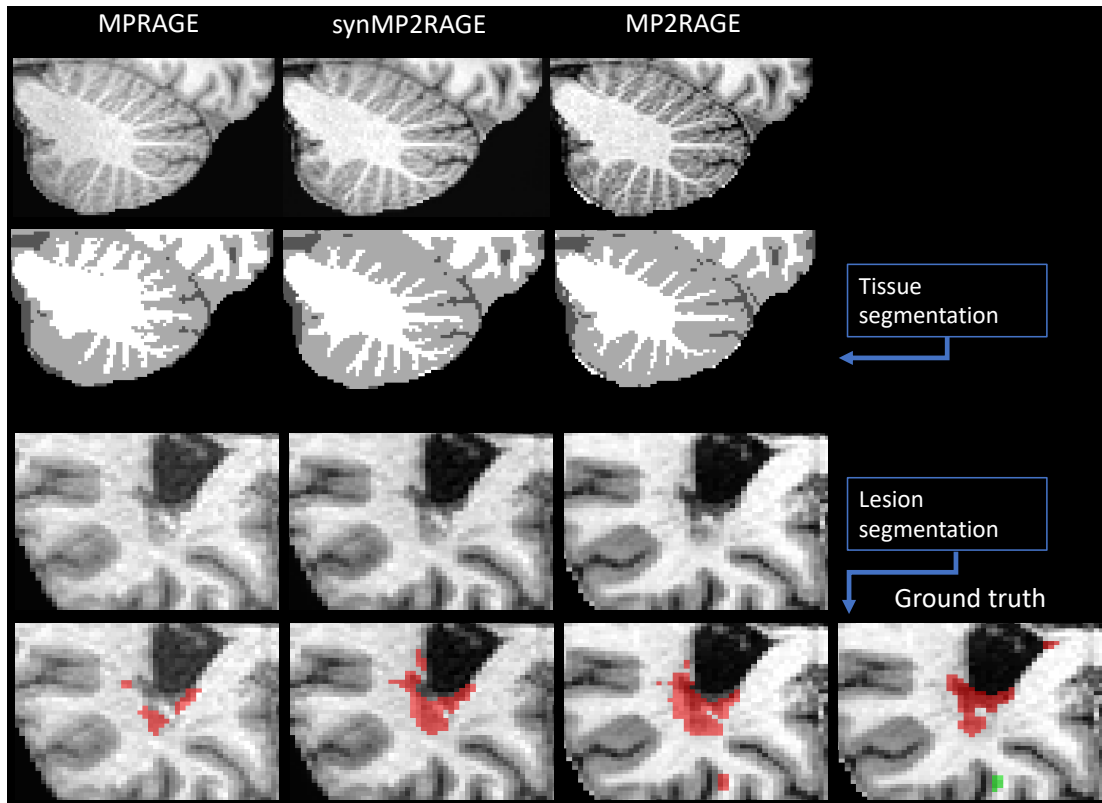
**Table 5.2** – Cortical and white matter lesion detection rates computed lesion-wise over the entire testing cohort for the different contrasts, all considered together with FLAIR. The false positive rate is fixed to 30% in order to easily compare the detection rates.

	WML detection rate (n = 1641)	CL detection rate (n = 373)
MPRAGE	73%	71%
synMP2RAGE	73%	71%
MP2RAGE	75%	71%
DIR	71%	75%

metrics, showing that synMP2RAGE outperformed MPRAGE on all metrics. Two different methods proposed in the literature were explored to evaluate the added value/applicability of the synthetic MP2RAGE images: automatic tissue and lesion segmentation in MS patients. In both cases, using the synMP2RAGE images offered significant improvements in terms of Dice overlap and volume difference (p-values < 0.05 in the patient-wise analysis) over using MPRAGE images. Importantly, in terms of lesion segmentation, synMP2RAGE does not significantly differ from using the ground truth MP2RAGE (p-values > 0.05).

There are several possible explanations supporting the fact that a GAN might be able to learn a mapping that translates MPRAGE image to MP2RAGE ones. For instance, a study has shown that while generating exogenous MRI contrast obtained from a contrast agent might not be possible, synthesizing a missing contrast (such as T1 or T2) from others, is (Lee et al., 2020). In our work, we aim at obtaining a variation of the MPRAGE contrast given as input to the generator. Both MPRAGE and MP2RAGE UNI are T1-weighted sequences with similar acquisition parameters (see Imaging subsection) and limited visual differences. Including a

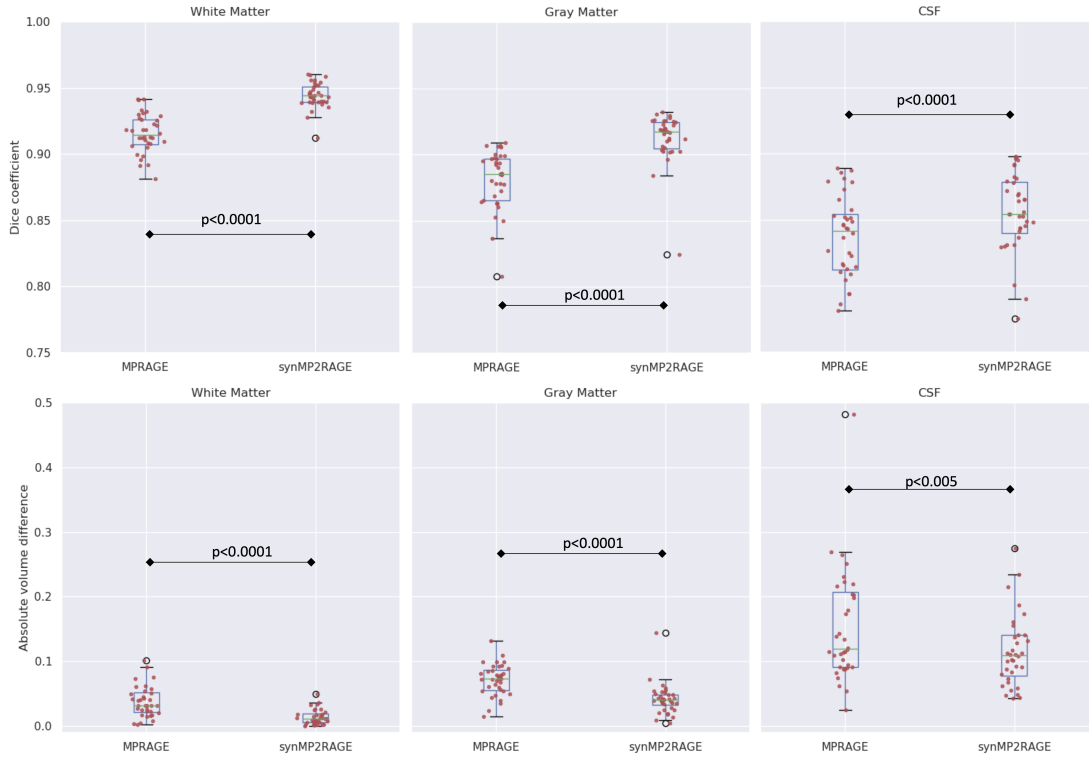




**Figure 5.7** – Comparison of the tissue and lesion segmentation obtained with MPRAGE, synMP2RAGE, and MP2RAGE. In the first row: zoom-in slices of the three contrasts. In the second row the relative tissue segmentation: in white the WM, light gray the GM, and dark gray the CSF. In the last row, the lesion segmentation is compared to the ground truth: red for WML and green for CL.

global skip connection in the GAN ensures that the correction produced is added to the input image, and therefore that the output does not greatly differ from the input. In this work, we do not consider synthesizing the MP2RAGE T1-map images, as these are slightly different visually and are currently not used by automated segmentation methods. However, if clinical studies will promote their usage, future work could focus on the generation of these quantitative images.

We believe that our method of translating MPRAGE acquisitions to synthetic MP2RAGE images has a high practical value for several reasons. Firstly, our evaluation answers the question in Hagiwara et al. (Hagiwara et al., 2020) about whether a synthetic MRI contrast could also improve the performance of an automated analysis tool: independently of the method considered, the synthetic MP2RAGE presented improved the output of both a CNN-based and a Bayesian segmentation approach. Secondly, as the MP2RAGE is now gradually being adopted for clinical use, generating its synthetic version for MRI studies where MP2RAGE

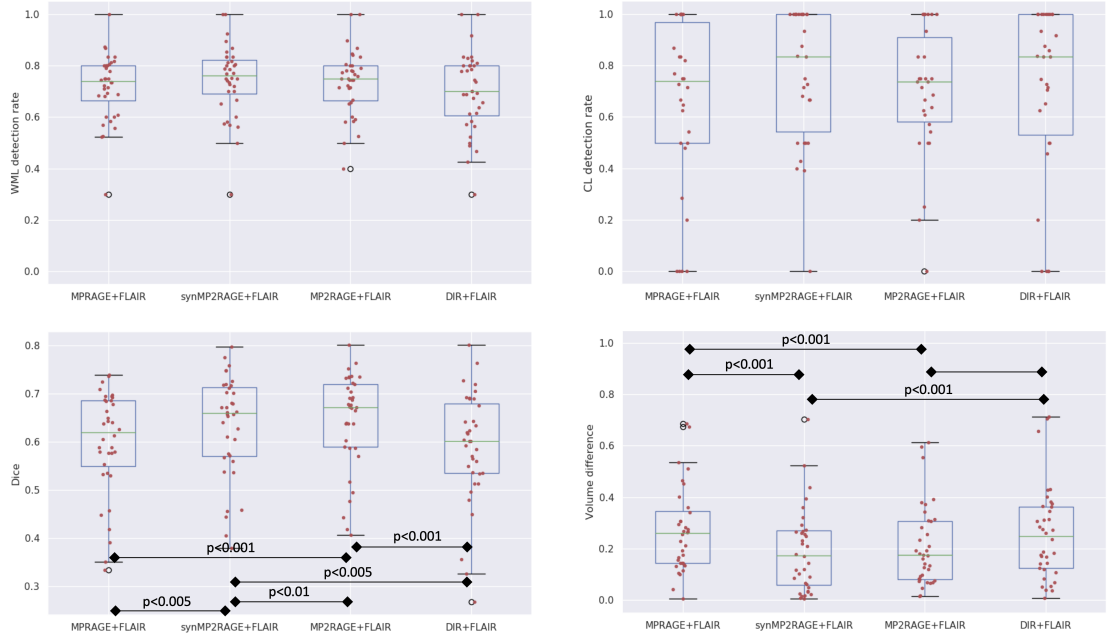


**Figure 5.8** – Metrics of the automated brain tissue segmentation obtained using MPRAGE and synMP2RAGE and considering the MP2RAGE as ground truth. First row: Dice coefficient for WM, GM, and CSF. Second row: the absolute volume difference for the same tissue types.

was not originally acquired would allow homogenizing the datasets for retrospective analysis. Thirdly, experts might visually benefit from its increased tissue and lesion contrast as well.

Furthermore, to the best of our knowledge, we present the first comparison of conventional and specialized MRI sequences for the automatic segmentation of MS lesions using a CNN. Interestingly, no significant differences ( $p$ -values  $> 0.05$ ) in the patient-wise nor in the lesion-wise analysis were found in terms of cortical and white matter lesion detection rate between MPRAGE, MP2RAGE, synMP2RAGE, and DIR. This is in contrast with previous works showing that specialized MRI contrasts improve the detection of CL both visually and automatically (Fartaria et al., 2016; Kober et al., 2012). We hypothesize that this is because of the intrinsic learning process of a CNN, particularly different from classical machine learning approaches previously applied to this task (Goodfellow et al., 2014; Lei et al., 2019). The evaluation set, however, is limited to 36 patients and it does not allow us to draw strong conclusions.

The generalizability of our results is subject to certain limitations. First, for all training and testing cases, the MPRAGE images were acquired with the same scanner and imaging protocol, including the same resolution. Therefore, the question whether the proposed generative



**Figure 5.9** – First row: WML and CL detection rate for the combination of input MRI contrasts considered. Second row: Dice coefficient and absolute volume difference for the same MRI contrasts. All comparisons not shown are not significant.

framework is able to learn a general mapping from different MP2RAGE images to a common MP2RAGE remains open. Second, in this work we only examine MS patients in the early stages of the disease. The performance of our proposed method on patients with a high lesion load remains to be explored. Third, while our qualitative evaluation tends to indicate that no artifacts are introduced in the synMP2RAGE and that its usage supports automated segmentation tools, a radiological analysis of the generated image by experts would be needed in the context of evaluating the clinical value of synMP2RAGE.

In conclusion, we propose a GAN which successfully translates MP2RAGE images to realistic-looking MP2RAGE ones of MS patients. An extensive evaluation of the synthesized images with automatic segmentation tools proved that the generated MP2RAGE significantly improves both the tissue and lesion segmentation. The framework is fast to run at inference time and publicly available, and can, therefore, be useful to MRI and clinical researchers for multiple tasks, even beyond neuroimaging studies of MS patients.



## 6 Cortical lesion detection and classification at ultra-high-field MRI

*The content of the following chapter is currently under review in Nuclear Magnetic Resonance (NMR) in Biomedicine as: Francesco La Rosa, Erin S Beck, Josefina Maranzano, Ramona-Alexandra Todea, Peter van Gelderen, Jacco A. de Zwart, Nicholas J. Luciano, Jeff H. Duyn, Jean-Philippe Thiran, Cristina Granziera, Daniel S. Reich, Pascal Sati, Merixtell Bach Cuadra: "Multiple sclerosis cortical lesion detection with deep learning at ultra-high-field MRI.". The first author (author of this thesis) contributed to the conceptualization, methodology, software, validation, and writing of the study.*

### Abstract

Manually segmenting multiple sclerosis (MS) cortical lesions (CL) is extremely time-consuming, and past studies have shown only moderate inter-rater reliability. To accelerate this task, we developed a deep learning-based framework (CLAIMS: Cortical Lesion Artificial Intelligence-based assessment in Multiple Sclerosis) for the automated detection and classification of MS CL with 7T MRI. Two 7T datasets, acquired at different sites, were considered. The first consisted of 60 scans that include 0.5mm isotropic MP2RAGE acquired 4 times (MP2RAGEx4), 0.7mm MP2RAGE, 0.5mm T2\*-weighted GRE, and 0.5mm T2\*-weighted EPI. The second dataset consisted of 20 scans including only 0.75x0.75x0.9 mm MP2RAGE. CLAIMS was first evaluated using 6-fold cross-validation with single and multi-contrast 0.5mm MRI input. Second, performance of the model was tested on 0.7mm MP2RAGE images after training with either 0.5mm MP2RAGEx4, 0.7mm MP2RAGE, or alternating the two. Third, its generalizability was evaluated on the second external dataset and compared with a state-of-the-art technique based on partial volume estimation and topological constraints (MSLAST). CLAIMS trained only with MP2RAGEx4 achieved comparable results to the multi-contrast model, reaching a

CL true positive rate of 74% with a false positive rate of 30%. Detection rate was excellent for leukocortical and subpial lesions (83%, and 70%, respectively), whereas it reached 53% for intracortical lesions. The correlation between disability measures and CL count was similar for manual and CLAIMS lesion counts. Applying a domain-scanner adaptation approach and testing CLAIMS on the second dataset, the performance was superior to MSLAST when considering a minimum lesion volume of  $6\mu\text{L}$  (lesion-wise detection rate of 71% vs 48%). The proposed framework outperforms previous state-of-the-art methods for automated CL detection across scanners and protocols. In the future, CLAIMS may be useful to support clinical decisions at 7T MRI, especially in the field of diagnosis and differential diagnosis of multiple sclerosis patients.

## **6.1 Introduction**

Multiple sclerosis (MS) is an inflammatory demyelinating disease affecting the central nervous system. It is characterized by focal areas of white matter (WM) demyelination (Calabrese et al., 2010). In recent decades, however, histopathological studies have shown that lesions in the cortex are also common (Calabrese et al., 2010; Harrison et al., 2015; Pirko et al., 2007). Moreover, the increasing use of ultra-high-field MRI has led to the observation that CL are extremely frequent in MS patients, persist over time, correlate with disability and progressive disease (Treaba et al., 2021; Treaba et al., 2019) and may help to differentiate MS from its clinical mimics. CL have been classified into three major types with potentially different etiologies (Calabrese et al., 2010): leukocortical (type 1, located at the interface between WM and GM), intracortical (type 2, involving purely the cortex and not reaching the pial surface), and subpial (including type 3 lesions located entirely in the GM and touching the pial surface, and type 4 lesions, extending from the pial surface, through the cortex, into the white matter). In order to maximize MS diagnostic and prognostic accuracy, it is, therefore, crucial to analyze the clinical implications of cortical lesions and their response to current and novel MS treatments. Compared to lower magnetic fields, ultra-high field MRI allows higher signal-to-noise ratio (SNR) and enhanced magnetic susceptibility contrast, both of which can provide important insights in MS pathophysiology (Ineichen et al., 2021; Springer et al., 2016; Tallantyre et al., 2009). Contrary to white matter lesions (WML), CL are difficult to visualize and require advanced sequences such as magnetization-prepared 2 rapid acquisition gradient echoes (MP2RAGE) (Marques et al., 2010). MP2RAGE has shown promising sensitivity to CL at 3T (Kober et al., 2012), and this significantly improves at 7T (Madsen et al., 2021), especially when combined with a  $T_2^*$ -weighted sequence (Beck et al.,

2018). Overall, 3T methods for visualizing cortical and, in particular, subpial lesions are relatively insensitive compared to 7T methods. For example, compared to combined use of 7T MP2RAGE and T2\*w images, 3T double inversion recovery was 6% sensitive for subpial lesions and 3T MP2RAGE was 5% sensitive (Beck et al., 2018). Thus, 7T imaging is essential for expanding our understanding of cortical demyelination (Tallantyre et al., 2009). With the recent FDA approval of 7T scanners for clinical use, we can expect an increasing number of scans with these devices in the coming years. However, even with the most sensitive imaging methods, identification and segmentation of cortical lesions is very time-consuming and requires significant experience. Despite the promise of 7T for the visualization of CL, studies using 7T MRI have shown only modest inter-rater reliability, both when considering CL detection and CL count, assessed with the Cohen's kappa coefficient ( $k$ ) and the Lin's concordance correlation coefficient (CCC), respectively (Harrison et al., 2015; Nielsen et al., 2012). In Nielsen et al., for instance, two experts analyzing 7T FLASH-T2\* images and a total of 103 CL had a moderate inter-rater agreement ( $k=0.69$ ) (Nielsen et al., 2013). In a similar study, Harrison et al. considered 7T magnetization-prepared rapid acquisition gradient echo (MPRAGE) and assessed the intra-rater and inter-rater reliability in terms of lesion count by two experts (Harrison et al., 2015). Strong intra-rater agreement was found (CCC = 0.96), whereas the inter-rater correlation was weak (CCC = 0.54). An automatic segmentation method for CL will therefore be essential to support large-scale studies and consistent evaluation of cortical lesions in multi-site clinical trials. Deep learning methods have lately shown outstanding performance in MRI segmentation, classification, and synthesis (Mazurowski et al., 2019). Applied to MS, these techniques have achieved state-of-the-art performance for WML segmentation (Kaur et al., 2020; La Rosa et al., 2019; McKinley et al., 2021) as well as for the automated assessment of other novel imaging biomarkers, such as paramagnetic rim lesions and the central vein sign (Barquero et al., 2020; Maggi, Fartaria, et al., 2020). However, while some different approaches have been presented to automatically segment CL based on 3T MRI (Fartaria et al., 2016; Fartaria et al., 2017; La Rosa, Abdulkadir, et al., 2020), to the best of our knowledge, only two methods have been proposed for detection at 7T (Fartaria et al., 2019; La Rosa, Beck, et al., 2020). First, Fartaria et al. have proposed MSLAST (Multiple Sclerosis Lesion Analysis at Seven Tesla), an automated method based on partial volume estimation and topological constraints that segments both WML and CL with a single MP2RAGE scan as input. MSLAST was evaluated on a cohort of 25 individuals with MS imaged in two research centers using 0.7mm isotropic MP2RAGE images and achieved a detection rate of 74% and 58% for WML and CL, respectively, with a false positive rate of 40% when considering a minimum lesion size of 6  $\mu$ L (Fartaria et al., 2019). Second, we previously proposed a deep learning-based method to detect and classify CL in 7T MRI considering three different contrasts: MP2RAGE,

T2\*w GRE, and T2\*w EPI (voxel size of 0.5x0.5x0.5 mm) (La Rosa, Beck, et al., 2020). On a larger cohort of 60 MS patients, the performance of our method showed promise, achieving a CL detection rate of 67% and a false positive rate of 42%. Moreover, about 200 CL detected by our proposed network and initially classified as false positives, were retrospectively judged as lesions by an expert. The contribution of the current work is three-fold. First, building upon our previous method (La Rosa, Beck, et al., 2020), we propose “CLAIMS” (Cortical Lesion AI-based assessment in Multiple Sclerosis), an improved pipeline for the automated detection and classification of CL with either single or multi-contrast MRI at 7T. Second, we assess the relative value of MP2RAGE and T2\*-weighted images, as well as the impact of MP2RAGE image quality, for this automated task. Third, an additional dataset from another institution is considered in order to test CLAIMS’s generalizability and robustness. For this purpose, a domain adaptation approach was performed and evaluated as well.

## **6.2 Materials and methods**

### **6.2.1 Datasets**

Two datasets, acquired at Institution A and B, were analyzed in this study. At Institution A, MRI acquisitions were performed on 60 individuals with MS (43 relapsing remitting, 17 progressive, 63% female,  $49 \pm 11$  (mean  $\pm$  standard deviation) years old, age range 29–77 years) with Expanded Disability Status Scale (EDSS) scores ranging from 0 to 7.5 (median 2.0) and disease duration of  $14 \pm 11$  (range 0–42) years. In addition to EDSS, the clinical assessment included the following disability measures: 9-Hole Peg Test (9-HPT), 25-foot timed walk (25TW), and Symbol Digit Modalities Test (SDMT). Imaging was done on a 7T whole-body research system (Siemens Healthcare, Erlangen, Germany) using a 32-channel head coil. The MRI protocol included: (i) 3D MP2RAGE (TR/TI1/TI2/TE = 6000/800/2700/5 ms, voxel size = 0.5x0.5x0.5 mm) repeated 4 times and averaged to improve image quality (total acquisition time ~40 minutes, denoted as MP2RAGEx4 throughout this paper), (ii) 3D-segmented T2\*w EPI (Sati et al., 2014) (TR/TE = 52/23 ms, voxel size = 0.5x0.5x0.5 mm) acquired in two partially overlapping volumes for whole brain coverage (total acquisition time ~7 minutes), (iii) T2\*w multi-echo GRE (TR/TE1/TE2/TE3/TE4/TE5 = 4095/11/23/34/45/56 ms, voxel size = 0.5x0.5x0.5 mm) acquired in three volumes for nearly full supratentorial coverage and averaged across the echo times (total acquisition time ~35 min). Moreover, for 55 out of the 60 patients, an additional single 3D MP2RAGE acquisition was performed with the following parameters: TR/TI1/TI2/TE = 5000/700/2500/2.9 ms, voxel size = 0.7x0.7x0.7 mm (acquisition time ~10 minutes). At Institution B, 20 patients with early relapsing-remitting MS (RRMS) (75% female,



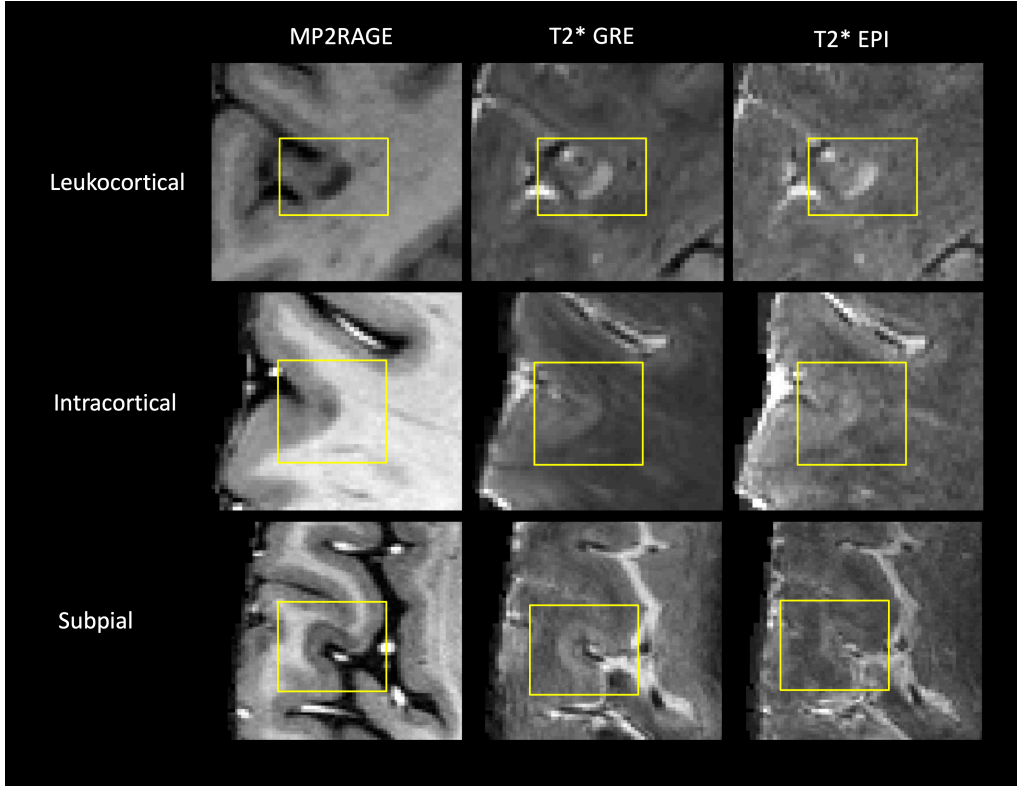
35±7 (mean ± standard deviation) years old, age range 21–46 years) with EDSS scores ranging from 0 to 4 (median 1.5), and disease duration < 5 years were imaged. A 7T research scanner (Siemens Healthcare, Erlangen, Germany) was used, and the protocol included a 3D MP2RAGE (TR/TI1/TI2/TE = 6000/750/2350/2.92 ms, voxel size = 0.75×0.75×0.9 mm). Throughout the manuscript and for both datasets, by MP2RAGE we always refer to its uniform denoised image (Marques et al., 2010). The study was approved by the Institutional Review Board of both institutions, and all patients gave written informed consent prior to participation.

### 6.2.2 Manual segmentation

In the 60 cases from institution A, CL were visually identified and delineated using MP2RAGEx4, T2\*w GRE, and T2\*w EPI images independently by one neurologist (E.B.) and one neuroradiologist (J.M.) (both with several years of experience identifying CL), who subsequently reached consensus in a joint session. The experts classified CL as leukocortical, intracortical, or subpial according to previously described criteria (Calabrese et al., 2009). In total, 2247 CL (21.0 median lesions/case, IQR=54) were segmented, of which 36% were leukocortical, 7% were intracortical, and 57% were subpial. This also includes 192 CL (37/8/147 leukocortical, intracortical and subpial, respectively) that were added after a retrospective analysis by an expert of the “false positives” generated by the CNN in our previous study. Figure 6.1 shows an example of each lesion type. As expected, intracortical lesions are mostly smaller compared to the other lesion types. A WML segmentation was obtained with a semi-automated method (Elliot & Arbel, 2016). In the cases from institution B, the manual segmentations were performed by consensus between 1 radiologist (A.R., 6 years of experience) and 1 neurologist with expertise in MS and neuroimaging (C.G., 13 years of experience). 188 CL (1.0 median lesions/case, IQR=10) were identified and subsequently classified into three types, leukocortical (69%), intracortical (26%), and subpial (5%). For the analysis of this dataset, intracortical and subpial lesions were grouped together.

### 6.2.3 Pre-processing

The images of each subject were linearly registered to the same space (MP2RAGEx4) using ANTs (B. Avants et al., 2008) and subsequently skull-stripped using FSL-Brain Extraction Tool (Jenkinson et al., 2012). All images were then resampled to 0.5×0.5×0.5 mm using a bilinear interpolation for the MRI contrasts and a nearest neighbor interpolation for the lesion masks. Finally, all non-zero voxels were normalized with mean 0 and standard deviation 1. Examining the lesion masks, about 80 CL appeared to be outside of the T2\* contrasts’ FOV and were

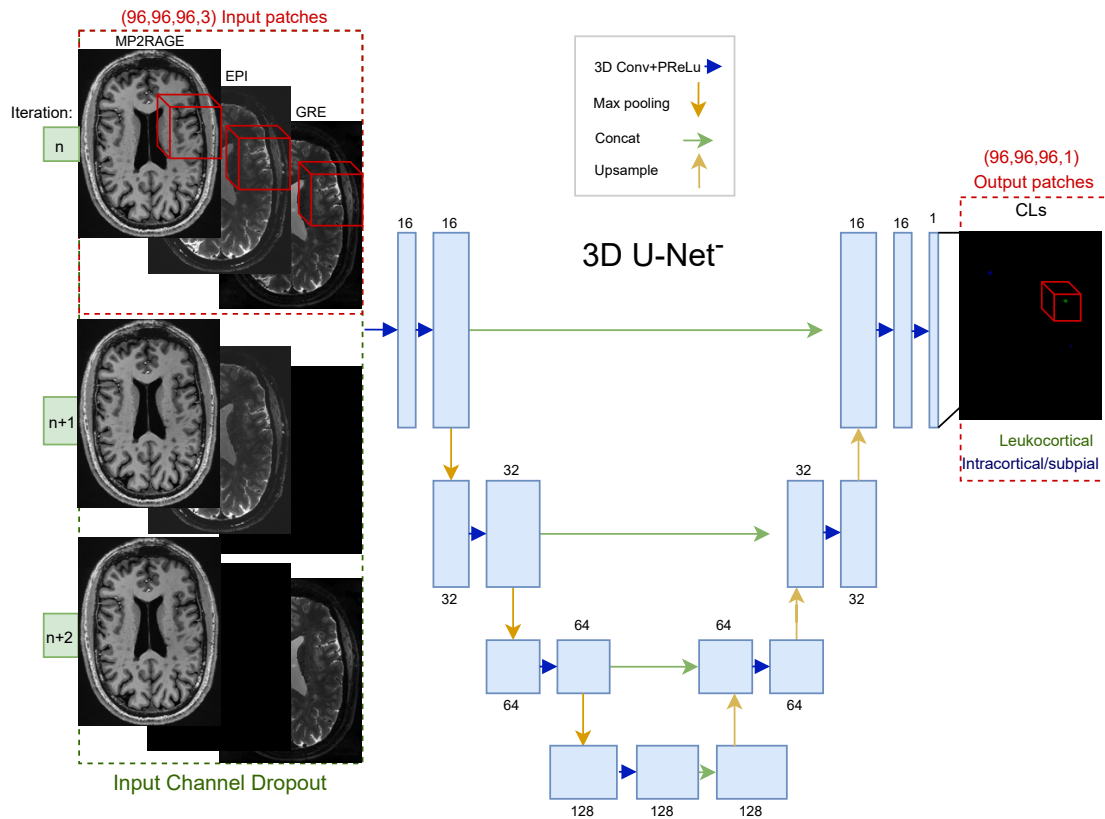


**Figure 6.1** – Examples of the three CL types identified in dataset A shown in the three MRI images considered.

therefore excluded in the analysis that included these contrasts.

#### 6.2.4 Convolutional neural network

CLAIMS relies on our previously proposed CNN architecture (La Rosa, Beck, et al., 2020) with several targeted modifications that boosted its performance. In particular, it is inspired by the 3D U-Net (Çiçek et al., 2016), but we used four resolution levels instead of three, each one with an increasing number of features: 16, 32, 64, and 128. Compared to our previous work, the tissue segmentation output branch was removed, and bigger patches of size 96x96x96 voxels were provided as input to the network (see a scheme in Figure 6.2). As output, two separate labels are provided, one representing leukocortical lesions and another one representing intracortical and subpial lesions. Instead of cross-entropy loss, used in our previous work, we trained the network with focal loss ( $\gamma=2$ ), as it has recently been shown to outperform cross-entropy for a large variety of tasks (Mukhoti et al., 2020). All WML voxels from the semi-automatically obtained WML masks were considered with a weight of 0 during training. Adam optimizer was used, and each model was trained for 400 epochs with an initial learning



**Figure 6.2** – Scheme of the proposed multi-contrast CNN architecture inspired by the 3D U-Net.

rate of  $0.5e-3$ .

### 6.2.5 Data augmentation

All data augmentation techniques used in our previous study (La Rosa, Beck, et al., 2020), including the input channel dropout, were also applied in this work. Additionally, random intensity shift (with an offset of 0.1) and affine transformations (random rotations of up to 15 degrees and random scaling of up to 10% of the image size) were performed.

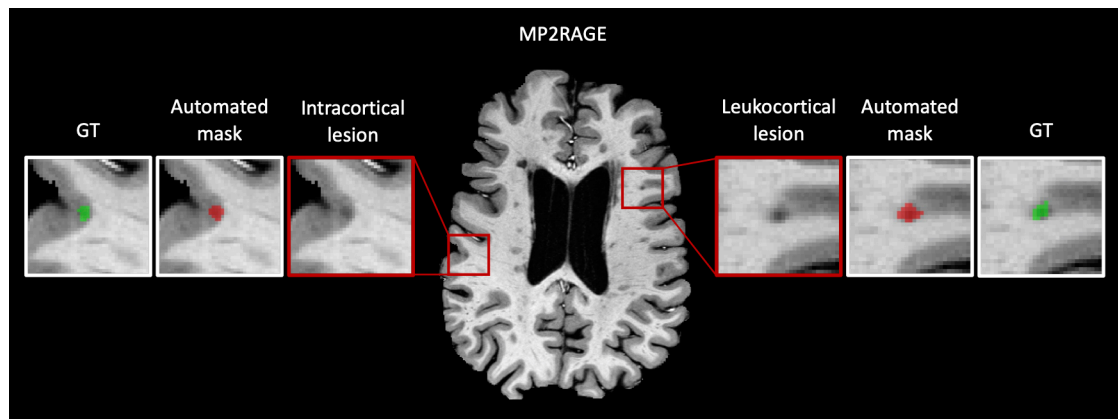
The training was done on an NVIDIA RTX3090 for 400 epochs and took approximately 44 hours. Testing takes about 2 minutes for each single case using the same machine. The code has been implemented in MONAI (Consortium, 2020) running on top of Pytorch and is publicly available on our research website<sup>I</sup>.

<sup>I</sup><https://github.com/Medical-Image-Analysis-Laboratory>

### **6.2.6 Experiments**

Several experiments and ablation studies were performed in order to compare the different MRI contrasts and performance of CLAIMS on separate datasets:

- First, over the 60 cases of Institution A, a 6-fold cross-validation was done to compare the single and multi-image inputs (50 cases for training and 10 for testing in each fold). The following image combinations were considered as input: MP2RAGEx4 + T2\* GRE + T2\* EPI, MP2RAGEx4 alone, T2\* GRE alone, and T2\* EPI alone. The lesion count of the best performing model (MP2RAGEx4) was compared with ground truth's lesion count. Moreover, we assessed the Spearman correlation between four disability measures and both automated and manual lesion counts.
- Second, an experiment was conducted to evaluate the performance of CLAIMS on MP2RAGE images obtained with a single acquisition and slightly bigger voxel size (acquisition time of approximately 10 minutes), which would be closer to a clinical scenario compared to the 0.5mm MP2RAGEx4 (with an acquisition time of about 40 min and therefore intended for research purposes). For this purpose, a model was trained either with the 0.5 mm MP2RAGEx4, alternatively with the MP2RAGE 0.7 mm isotropic images, and in a third approach by randomly providing at each iteration either a 0.5mm MP2RAGEx4 or a 0.7mm single acquisition MP2RAGE to the network. The resulting three models were then all tested on the 0.7mm single acquisition MP2RAGE images. Importantly, averaging four acquisitions increases the signal-to-noise ratio (SNR) by a factor of 2, whereas going from a voxel size of 0.5 to 0.7 increases SNR by a factor of 2.732. Therefore, 0.7mm single acquisition MP2RAGE has a slightly higher SNR than the 0.5mm MP2RAGEx4.
- Third, a training was performed with all cases of institution A (using both MP2RAGE 0.7mm single acquisition and 0.5mm MP2RAGEx4, but no T2\* images) and then tested 14 cases of institution B. Furthermore, a domain adaptation of this model with 6 different cases of institution B was also done, consisting of re-training all CNN's layers starting with the previous weights. In this case, the fine-tuning lasted 50 epochs and had an initial learning rate of  $0.5e-4$ . The results were compared with the previous state-of-the-art method using 14 subjects from the same dataset (Fartaria et al., 2019).



**Figure 6.3** – Visual results showing CL detection with the single input 0.5mm MP2RAGEx4 model. GT=ground truth.

### 6.2.7 Evaluation and statistical analysis

Common detection metrics such as the lesion-wise true positive rate (LTPR) and lesion-wise false positive rate (LFPR) were considered for the evaluation on a lesion-wise level. Due to the extremely small size of CL (starting from  $1\mu\text{L}$  in our datasets), each lesion labeled by the expert is considered as detected if it overlaps by at least one voxel with the automatically generated mask. The median detection rate and false positive rate are considered on a patient-wise level as well. The Dice coefficient (DSC) and volume difference (VD) were computed to quantify the accuracy of lesion delineation and volumetric segmentation for the subjects with at least one CL. The classification accuracy between the two types of CL considered (leukocortical vs intracortical/subpial) was assessed as percentage of correctly classified CL. For all cases from institution A, a minimum lesion size of  $1\mu\text{L}$  was considered for both the automated and manual mask when computing the metrics, whereas for the subjects of institution B we evaluated different minimum lesion volumes (1, 3, 6, 9, and  $15\mu\text{L}$ ). For both datasets, the segmented connected components that overlap with WML are not considered as false positives, as the difference between leukocortical and juxtacortical lesions is subtle. The Wilcoxon signed-rank test was used to compare the metrics on a patient-wise level. The Bonferroni correction was used to adjust significance for multiple comparisons. Differences are considered significant at  $p\text{-value} < 0.05$ . In order to verify the correlation between CL number and disability measures, the Spearman correlation coefficient ( $\rho$ ) was computed. Cohen's kappa coefficient ( $k$ ) ("Interrater reliability: the kappa statistic", 2012) was evaluated to verify the agreement between the CL delineated by the experts and the CL detected by CLAIMS. Lin's concordance coefficient (CCC) (Lin, 1989) was computed to assess the level of correlation between the ground truth and CLAIMS' lesion count.

## 6.3 Results

### 6.3.1 Single vs multi-image models

To assess the contribution of each MRI contrast, we performed the following ablation study. Three networks were trained, each with a single MRI contrast as input (MP2RAGEx4, T2\*GRE, and T2\*EPI), and then compared to the model that received all three images (Table 6.1). Qualitative results are presented in Figure 6.3. In terms of lesion detection, the three images model and the MP2RAGEx4 model achieved similar results, with a LTPR of 82% and 83% for leukocortical lesions, 49% and 53% for intracortical lesions, and 70% and 69% for subpial lesions. Both of them had a LFPR of 30%. They considerably outperformed the single input models trained with T2\*GRE and T2\*EPI (LTPR of 34% and 30%, and LFPR of 45% and 49%, respectively). Similar performance was observed in terms of lesion delineation, with the 0.5mm MP2RAGEx4 achieving the best metrics (0.49/0.80 Dice coefficient and volume difference vs 0.18/4.51 for the T2\*GRE and 0.16/4.57 for the T2\*EPI). Finally, a moderately high CL subtype classification accuracy of over 80% was observed for all four models.

**Table 6.1** – Median metrics for the different input contrasts obtained with a 6-folds cross-validation over the 60 cases of institution A. LTPR, LFPR and classification accuracy are computed on a lesion-level, whereas DSC and VD are considered on a patient-wise level. The best result for each metric is in bold.

Input	Lesion-wise			Patient-wise	
	LTPR	LFPR	Classification accuracy	DSC	VD
MP2RAGEx4, T2* GRE, T2* EPI	<b>0.74</b>	<b>0.30</b>	<b>0.88</b>	<b>0.47</b>	<b>0.87</b>
MP2RAGEx4,	<b>0.74</b>	<b>0.30</b>	<b>0.88</b>	0.49	0.80
T2* GRE	0.36	0.45	0.82	0.18	4.51
T2* EPI	0.32	0.55	0.81	0.16	4.57

On a patient-wise level, the single input 0.5mm MP2RAGEx4 model and the three images model achieved comparable results ( $p>0.05$ ), and both significantly outperformed the single input models trained with T2\*GRE and T2\*EPI ( $p<0.001$ ). The single input 0.5mm MP2RAGEx4 mode reached a Cohen's kappa coefficient of 0.49 when compared with the manually annotated CL. Analyzing as well the correlation between the manual and automated lesion counts, the Lin's concordance coefficient reached 0.91 (see Figure 6.4), indicating substantial agreement. Moreover, we computed the Spearman coefficient to compare the correlation between four disability measures (EDSS, 9-HPT, 25TW, SMDT) and both the ground truth and CLAIMS lesion counts. The manual and CLAIMS lesion counts correlated similarly with each of the disability measures, presented in Table 6.3. Table 6.2 and Figure 6.5 show the detection rate for the three different CL types. As observed in our previous work (La Rosa, Beck, et al.,

## Cortical lesion detection and classification at ultra-high-field MRI

**Table 6.2** – Comparison of lesion detection rate on a patient-wise level for the different models. NS: non-significant; \*:  $p < 0.05$ ; \*\*:  $p < 0.01$ ; \*\*\*:  $p < 0.001$ .

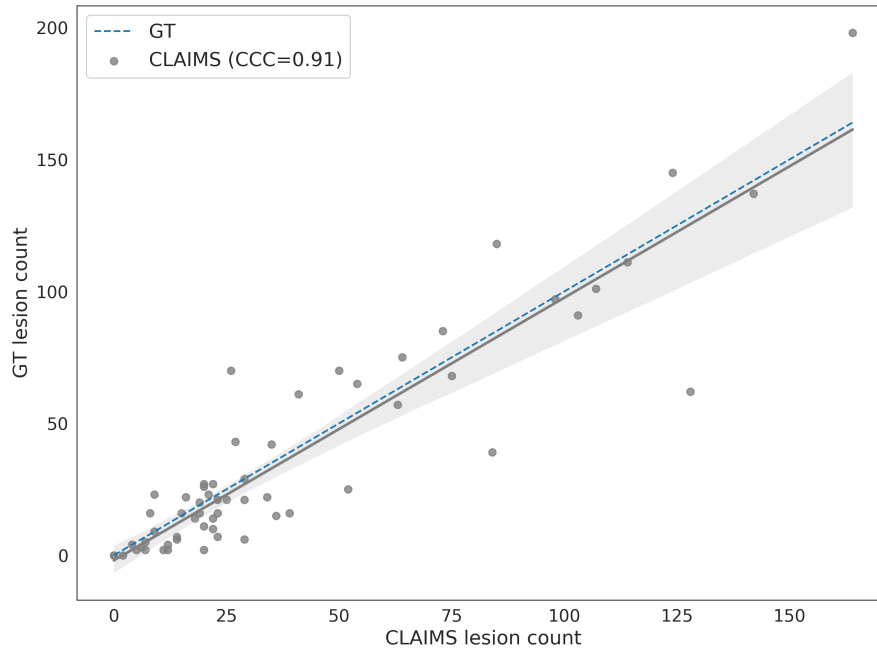
	MP2RAGEx4		T2*GRE		T2*EPI		3 contrasts	
		Per patient		Per patient		Per Patient		Per Patient
	Detected (%)	Median (range, IQR)	Detected (%)	Median (range, IQR)	Detected (%)	Median (range, IQR)	Detected (%)	Median (range, IQR)
All	1649 (74%)	14 (0-163, 36)	748 (36%)	6 (0-141, 10)	692 (32%)	5 (0-128, 11)	1648 (74%)	14 (0-151, 31)
Leukocortical	672 (83%)	5 (0-54, 13)	268 (63%)	4 (0-39, 12)	254 (59%)	4 (0-36, 11)	656 (82%)	6 (0-48, 13)
Intracortical	83 (53%)	1 (0-17, 2)	25 (15%)	0 (0-5, 1)	19 (12%)	0 (0-2, 0)	76 (49%)	1 (0-17, 2)
Subpial	894 (69%)	6 (0-96, 8)	455 (48%)	4 (0-92, 6)	421 (41%)	5 (0-81, 4)	916 (70%)	7 (0-98, 9)

	Significance					
	MP2RAGEx4 vs T2*GRE	MP2RAGEx4 vs T2*EPI	MP2RAGEx4 vs 3 contrasts	3 contrasts vs T2*EPI	3 contrasts vs T2*GRE	T2*GRE vs T2*EPI
All	***	***	N.S.	***	***	N.S.
Leukocortical	***	***	N.S.	***	***	N.S.
Intracortical	***	***	N.S.	***	***	**
Subpial	***	***	N.S.	***	***	*

2020), intracortical lesions remain the most challenging, with a detection rate of 53% in the best scenario (three image model). On the other hand, leukocortical and subpial lesions have a high detection rate of over 80 and 70%, respectively, for both the MP2RAGEx4 and the three images model. Similar behavior is observed between the three CL types for all models with a significant drop, however, in the detection of intracortical lesions for the T2\*GRE and T2\*EPI models (13% and 7%, respectively).

### 6.3.2 Evaluating CLAIMS on standard MP2RAGE images

In order to evaluate CLAIMS on more commonly acquired single acquisition 0.7mm MP2RAGE images, three models were trained with the following inputs: (1) MP2RAGE 0.7mm, (2) MP2RAGEx4 0.5mm, (3) alternating MP2RAGEx4 0.5mm and 0.7mm at each training iter-



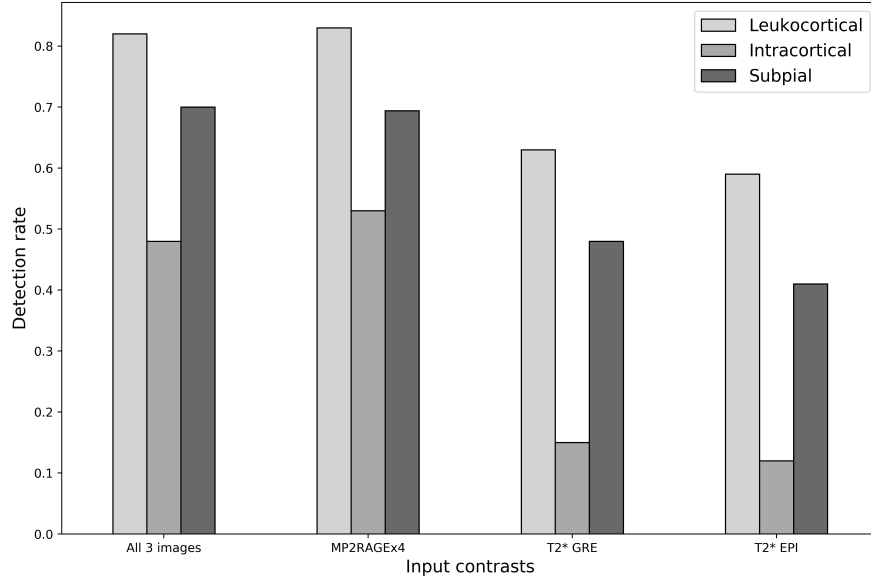
**Figure 6.4** – Correlation between the manual CL count and the one provided automatically by CLAIMS (best model trained with MP2RAGEx4). The solid lines show the linear regression model between the two measures along with a confidence interval at 95%. The dashed lines indicate the expected lesion count estimates. The Lin's concordance correlation coefficient between manual and automatic lesion count is reported in the legend ( $CCC = 0.91$ ) and shows a substantial agreement.

ation. All models were then tested on the single acquisition 0.7mm MP2RAGE. For any given subject, care was taken to ensure that all images were either in the training or testing set, and not split across them. Table 6.4 reports the metrics obtained when evaluating these models on the MP2RAGE 0.7mm images with a 6-fold cross-validation. All model results were compared to the ground truth determined using MP2RAGEx4, T2\*w GRE and T2\*w EPI images. First, compared to the previous results for the 0.5mm MP2RAGEx4 single input model (see Table 6.1), we observed that training and testing on single acquisition 0.7mm MP2RAGE causes a drop in the detection rate from 74 to 53%. Second, between the three models, the one using both 0.5 and 0.7mm images during training achieves the best metrics with a LTPR of 53% and a LFPR 33%. Finally, the worst-performing model is the one trained only with MP2RAGEx4, having a LTPR of 35% and a LFPR of 41%.

### 6.3.3 Independent test set

In our last study, an MP2RAGE-only model was trained with all cases from Institution A (mixing 0.5mm MP2RAGEx4 and 0.7mm MP2RAGE, as this was the top performing input in





**Figure 6.5** – Lesion-wise CL detection rate for the three different CL lesion types considered over the 60 subjects of institution A.

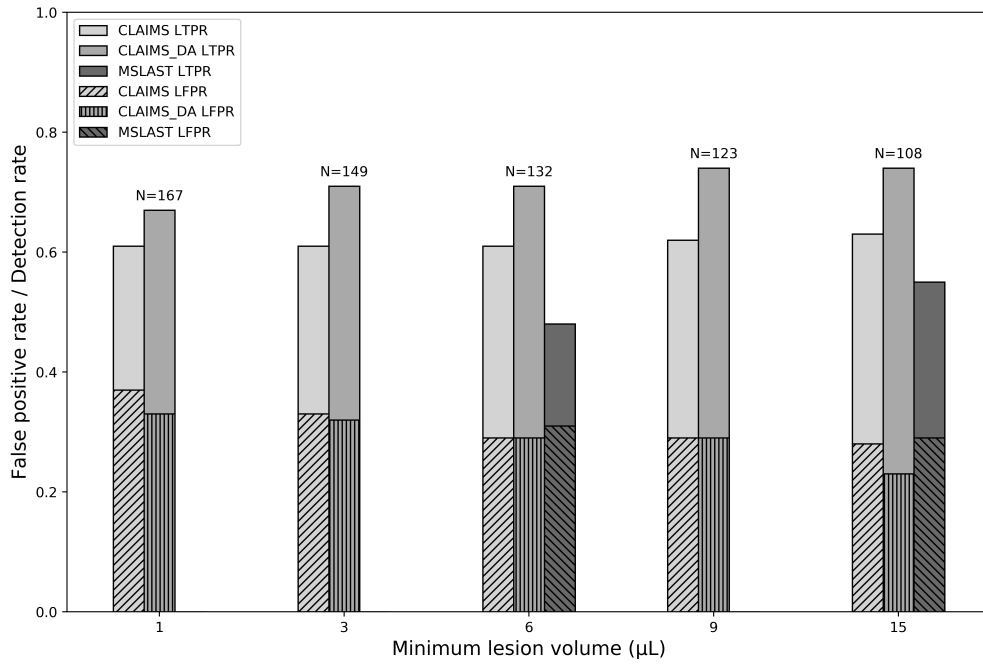
**Table 6.3** – Spearman correlation coefficient  $\rho$  (and its relative  $p$ -value) computed between four disability measures and the manual and automated CL count. Both counts show a moderate correlation for all four measures.

	CLAIMS CL number		Ground truth CL number	
	$\rho$	p-value	$\rho$	p-value
GT CL number	0.855	<0.001	-	-
EDSS	0.454	0.0003	0.426	0.0008
25TW	0.424	0.0008	0.425	0.0008
9HPT	0.497	<0.0001	0.404	0.0014
SDMT	-0.578	<0.0001	-0.531	<0.0001

the previous experiment) and tested on the 14 subjects of Institution B, which include only 0.75x0.75x0.9mm MP2RAGE images and were used also to report results with MSLAST in a previous work (Fartaria et al., 2019). Furthermore, we also propose a domain adapted version of CLAIMS (CLAIMS\_DA), fine-tuning the same model (training all layers) on 6 additional cases from Institution B. Figure 6.6 presents the lesion-wise true and false positive rates for CLAIMS, CLAIMS\_DA, and MSLAST, considering different minimum lesion volumes, even much smaller than the ones proposed in the MSLAST paper (1, 3, and 9 $\mu$ L vs 6 and 15 $\mu$ L). CLAIMS outperforms MSLAST for all volume thresholds and CLAIMS\_DA achieves even better results on a lesion-wise level. Considering a minimum lesion volume of 6 $\mu$ L, it reaches a detection rate of 71% with a false positive rate of 29%, whereas MSLAST has a detection rate of 48% and 31% of false positives. The CL classification accuracy for CLAIMS and CLAIMS\_DA was 81% and 84%, respectively. On a patient-wise level, no statistically significant differences

**Table 6.4** – Metrics obtained for models trained with different inputs (listed in the first column) and tested on the MP2RAGE 0.7mm images. The best result for each metric is in bold.

Input	Lesion-wise			Patient-wise	
	LTPR	LFPR	Classification accuracy	DSC	VD
MP2RAGE 0.7mm	0.52	0.39	0.85	0.24	1.10
MP2RAGEx4 0.5 mm	0.35	0.41	0.81	0.16	1.29
MP2RAGEx4 0.5 mm and 0.7mm	<b>0.53</b>	<b>0.33</b>	<b>0.85</b>	<b>0.29</b>	<b>1.06</b>



**Figure 6.6** – False positive and detection rate in a pure testing scenario on the Institution B dataset for CLAIMS, CLAIMS domain-adapted (CLAIMS\_DA), and MSLAST. Different minimum lesion volumes are considered. N refers to the number of CL in the ground truth for each minimum lesion volume.

were observed between the three models for both LTPR and LFPR ( $p > 0.05$ ).

## 6.4 Discussion

In this work, we explored the value of different 7T MRI contrasts for the automated detection of CL. For this purpose, we trained and tested a novel U-net-based deep learning method we call CLAIMS. We analyzed three MRI contrasts as input (MP2RAGE, T2\*-w GRE, and T2\*-w EPI), different image resolutions, and CLAIMS's generalizability on an external testing dataset. CLAIMS's automated CL count was then compared to the experts' CL number, and its correlation with four disability measures was examined. Furthermore, we compared our

results with the only previous automated CL segmentation method at 7T in the literature (Fartaria et al., 2019). Lastly, we reported the results obtained considering different minimum lesion volumes and analyzing the three CL subtypes.

The results of our ablation study with different input contrasts fed to the model show that the MP2RAGE contrast alone is sufficient to achieve the best performance. Contrary to a past study regarding the manual segmentation of CL (Beck et al., 2018), the addition of two T2\*-weighted images did not contribute significantly to improving the automated CL detection. Moreover, the single input models trained with the T2\*-weighted contrasts performed poorly compared to the MP2RAGE one. This is in line with a previous study where it was shown that the MP2RAGE increases the visual detection of all CL types compared to T2\*-weighted imaging at 7T (Madsen et al., 2021). However, it is important to note that our ground truth was created considering all three MRI images, and some lesions might not be visible on the T2\* contrasts alone. Moreover, even a lower resolution MP2RAGE outperformed the T2\*-weighted models, arguably posing the MP2RAGE as the preferred contrast by CLAIMS. Focusing on the different CL types, we observed that their detection rate varies considerably. As in our previous study (La Rosa, Beck, et al., 2020), intracortical lesions were once again the most challenging type, with a detection rate of only 53% for the best model. On the other hand, both leukocortical and subpial lesions were detected with a high sensitivity of over 80 and 70%, respectively. These values approach the inter-rater agreement of 85% in a previous study considering two raters (Nielsen et al., 2013). In the prior study, Cohen's kappa coefficient was 0.69, showing substantial agreement, whereas ours, computed between CLAIMS output and the manual masks, reached 0.49. However, we considered a dataset with twenty times more CL (2247 vs 103), and this could explain the higher variability. Similarly, the inter-rater reliability can also be estimated by computing Lin's concordance correlation coefficient, which considers the correlation between lesion counts. Analyzing lesion counts from 10 sample scans and by two raters, Harrison et al. showed a Lin's coefficient of 0.54 for all CL (Harrison et al., 2015), meaning that the agreement was weak. On the contrary, our best model achieved a value 0.91, showing a substantial correlation between the manual and automated CL count. This suggests that CLAIMS could be useful to support and speed up experts' CL assessments. To further support this claim, we analyzed the correlation between four disability measures and the manual and automated CL number. Both CL counts correlated with the four measures (Spearman's coefficient between 0.40 and 0.57) in a very similar way. To the best of our knowledge, this is the first time that an automated CL count has proved to correlate with disability scores.

Relying only on the MP2RAGE, we then evaluated CLAIMS on MP2RAGE images that are more feasible to acquire, obtained with a single acquisition of about 10 minutes (vs 40min for MP2RAGEx4) and with a slightly bigger voxel size. Specifically, we compared different models trained with either 0.5mm, 0.7mm, or both 0.5mm and 0.7mm. The 0.5mm isotropic images were obtained as an average of 4 acquisitions and had, therefore, a twice higher SNR compared to a single acquisition at the same resolution, but a lower SNR compared to the 0.7mm MP2RAGE. Importantly, the model trained with 0.5mm images and tested on 0.7mm images (simply interpolated to 0.5mm) performed very poorly, showing the value of the voxel size for the detection of very small structures. Of note, the ground truth was labelled on the 0.5mm images, and this might also cause a partial drop in performance. However, when mixing both 0.5mm and 0.7mm images during training, the metrics improve considerably, even outperforming the model trained with 0.7mm images only. This indicates that CLAIMS can extract useful information during training from the smaller voxel size images and successfully use these at inference time. Overall, however, we notice that going from 0.5mm MP2RAGEx4 to single acquisition 0.7mm MP2RAGE causes a detection rate drop of about 20%, proving the importance of image quality and resolution even for automated methods.

Finally, we tested the performance of CLAIMS trained mixing both 0.5mm MP2RAGEx4 and 0.7mm MP2RAGE on a different dataset where MP2RAGE scans were acquired with a voxel size of 0.75x0.75x0.9mm. This is the same dataset on which the previous state-of-the-art method (MSLAST) was evaluated (Fartaria et al., 2019), and therefore we could carry out a precise comparison including the same subjects. It is important to note that the dataset from Institution B includes subjects in the very early stages of the disease (disease duration < 5 years) who have a much lower lesion burden compared to the subjects of institute A (median lesion count per case of 1 vs 21). Moreover, the CL subtype distribution is different as well, with a majority of leukocortical lesions in dataset A, whereas intracortical lesions are prevalent in dataset B. This could be partially explained by the lower voxel size and the presence of T2\* images in dataset A, which allow higher visual detection of subpial lesions (Beck et al., 2020). Nevertheless, CLAIMS proves to be robust and performs well in this multi-center scenario, with a lesion-wise detection rate of about 61% when different minimum lesion volumes are considered (1, 3, 6, 9, and 15 $\mu$ L). It outperforms MSLAST (both for minimum lesion volume of 6 $\mu$ L and 15 $\mu$ L), and it also classifies CL into two types (leukocortical and intracortical) with an accuracy of about 80%. When decreasing the minimum lesion volume considered, the detection rate remains stable, with only a marginal increase of false positives. For instance, if a minimum lesion volume of 3 $\mu$ L is considered, CLAIMS achieves a detection rate of 61% with a false positive rate of 33%. Moreover, when fine-tuning CLAIMS with six additional

subjects belonging to the same dataset, the lesion-wise performance increased considerably, reaching a detection rate of 71% with a false positive rate of 32%. This confirms the efficacy of domain adaptation for deep learning-based models applied to a different dataset and poses our proposed method as the state-of-art technique for CL detection with 7T MRI.

### 6.4.1 Limitations

Our study also presents some limitations. First, the two datasets considered were acquired with scanners from the same manufacturer and with similar acquisition protocols. Additional differences in the images could arise under more general conditions, potentially causing a drop in performance. In this case, a fine-tuning of the pre-trained model relying on few annotated subjects could help overcome this issue and help regain performance. Second, the detection rate of intracortical lesions remains quite low, even though it improved by more than 10% compared to our previous study (La Rosa, Beck, et al., 2020). These are infrequently visible and extremely small CL, and perhaps a larger training dataset with an increased number of intracortical lesions could mitigate this issue. Third, in this work, we analyzed a single CNN architecture that proved effective for CL detection in our previous study. It was out of the scope of this study to compare several different models and tweak their parameters to optimize the performance. We rather selected the state-of-the-art architecture and analyzed in detail its potential to tackle a clinically relevant problem such as the detection of CL.

### 6.4.2 Future work

Future work could include exploring more advanced deep learning architectures in order to better include the information coming from each single MRI contrast, particularly the T2\*w images. Moreover, the use of larger and additional datasets could further prove the generalizability of the proposed method. Finally, as any MP2RAGE acquisition always comes with a quantitative T1 map (Marques et al., 2010), use of this image type could also be explored for the automated detection of CL, although in our experience most CL are similarly seen on both MP2RAGE uniform denoised and T1 map images.

### 6.4.3 Conclusion

To summarize, we present CLAIMS, a DL framework for the detection and classification of CL with 7T MRI. When CLAIMS is trained only with MP2RAGE, it achieves state-of-the-art performance for all three CL types, and its CL count correlates with disability measures

similarly to experts' visual assessment. If fine-tuned, it adapts extremely well to a different dataset acquired in a different site. As 7T scanners from several manufactures are now being approved for clinical use, CLAIMS could eventually be useful to support clinical decisions, particularly in the field of diagnosis and differential diagnosis of MS patients.

## 7 Classification of paramagnetic rim lesions

*Parts of the content of section 7.1 of this chapter are adapted from the postprint version of the following article: Germán Barquero, Francesco La Rosa, Hamza Kebiri, Po-Jui Lu, Reza Rahmanzadeh, Matthias Weigel, Mário João Fartaria, Tobias Kober, Marie Théaudin, Renaud Du Pasquier, Pascal Sati, Daniel S. Reich, Martina Absinta, Cristina Granziera, Pietro, Maggi, Merixtell Bach Cuadra: "RimNet: A deep 3D multimodal MRI architecture for paramagnetic rim lesion assessment in multiple sclerosis." Neuroimage: Clinical 28 (2020): 102335; doi: [10.1016/j.nicl.2020.102412](https://doi.org/10.1016/j.nicl.2020.102412). The second author (author of this thesis) contributed to the conceptualization, methodology, software, validation, and writing of the study.*

*The work presented in Section 7.2 of this Chapter belongs to the postprint version of the following abstract: Francesco La Rosa, Germán Barquero, Omar Al-Louzi, Bénédicte Maréchal, Tobias Kober, Jean-Philippe Thiran, Pascal Sati, Daniel S. Reich, Pietro, Maggi, Martina Absinta, Merixtell Bach Cuadra, Cristina Granziera: "Automated assessment of paramagnetic rim lesions in multiple sclerosis patients with 3T and 7T MP2RAGE." ISMRM (2021). The first author (author of this thesis) contributed to the conceptualization, methodology, software, validation, and writing of the study.*

As extensively discussed in Chapter 3, an automated method for the assessment of PRL would be highly desirable. We have proposed the first semi-automated approach (termed RimNet) to facilitate this process and evaluated it in a large multi-site dataset acquired at 3T. Moreover, we have tested it with 3T and 7T MP2RAGE images to analyze its performance at ultra-high magnetic field. In this chapter, RimNet and its evaluation at 7T are described and discussed.

## 7.1 RimNet

The objective of this study was to develop and evaluate a CNN architecture (RimNet) for the semi-automated detection of paramagnetic rim lesions in MS employing different MRI contrasts acquired at 3T.

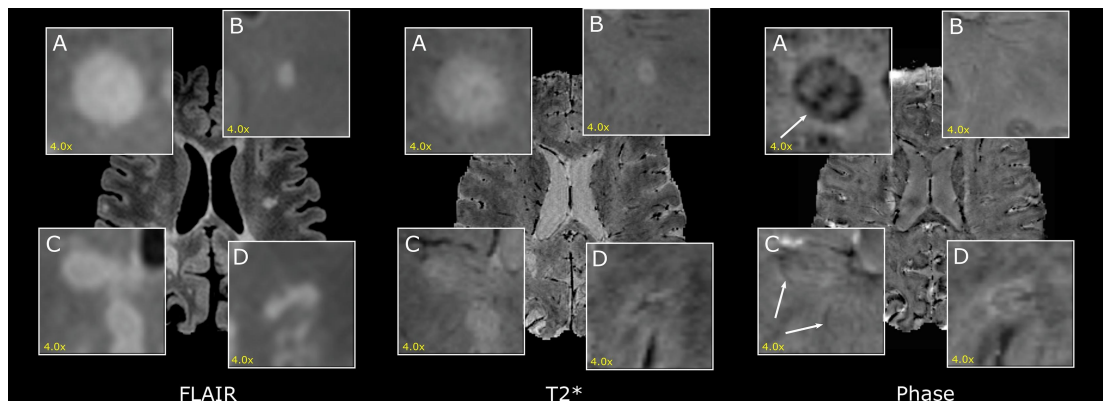
### 7.1.1 Dataset

We retrospectively analyzed MRI scans from MS patients diagnosed according to the revised 2010 McDonald MS criteria (Polman et al., 2011) who were recruited between December 2017 and September 2019 in two academic research hospitals, the Centre Hospitalier Universitaire Vaudois (Lausanne, Switzerland) and the Universitätsspital Basel (Basel, Switzerland). Of the 141 eligible MS patients, 124 were included (11 patients were excluded because of motion artefacts and 6 because of coexisting brain pathologies): 55 patients (33 female, 25–76 years old) in Lausanne and 69 patients (44 female, 22–73 years old) in Basel. Patients' demographic and clinical characteristics are summarized in Table 7.1. The study received approval by the local ethics committee, and all patients gave written informed consent for the retrospective use of their data.

All patients underwent a single brain MRI acquisition at 3T (either MAGNETOM Skyra or MAGNETOM Prisma, Siemens Healthcare, Erlangen, Germany) in Lausanne; MRI in Basel were also acquired at 3T (MAGNETOM Prisma, Siemens Healthcare, Erlangen, Germany). In both centers, three-dimensional segmented echo-planar imaging (3D-EPI) (Sati et al., 2014), giving high-resolution T2\*-w magnitude and phase images, and 3D T2-FLAIR images were acquired (Table 7.2). A 3D T1-weighted MPRAGE sequence was acquired in Lausanne and a 3D MP2RAGE in Basel (Marques et al., 2010). 3D-EPI images were obtained with a resolution of  $0.65 \times 0.65 \times 0.65 \text{ mm}^3$  and  $0.67 \times 0.67 \times 0.67 \text{ mm}^3$  in Lausanne and Basel hospitals, respectively. 3D FLAIR, MPRAGE, and MP2RAGE images were acquired with an isotropic resolution of  $1 \text{ mm}^3$  in both centers.

For training and evaluation of our supervised DL-based method, a ground truth sample of rim+ (lesions presenting a paramagnetic rim) and rim- (lesions not presenting a paramagnetic rim) was obtained. Rim+ lesions were manually annotated by two raters (PM and MA, with imaging research experience of 10 and 14 years, respectively). For exploring the presence of rims, the phase images were primarily used, which helped to identify the phase shift mainly produced by iron-laden macrophages and relative myelin content at the lesion edge (Absinta





**Figure 7.1** – Example of MS lesions on 3D FLAIR (left), 3D-EPI magnitude (center) and phase (right) images. A) and B) are clear examples of lesions with presence (rim+) and absence (rim-) of a paramagnetic rim, respectively. In C) there are two more subtle rim+ lesions. D) is an example of a rim- lesion that has a rim+ like intensity artefact.

et al., 2016; Dal-Bianco et al., 2017). Moreover, 3D FLAIR was used in the annotation process to visually assess whether paramagnetic rims identified on the phase images matched an MS lesion on FLAIR, which allowed discarding potential false positives due to rim-shaped artefacts. After a first screening conducted individually by each expert, 38.3% of the lesions needed consensus review, which was done in a second joint screening by the two experts. Following consensus, 462 rim+ lesions were identified. Rim- lesions were annotated as follows. Each connected component in the segmentation output (corresponding to the connected components by considering a 6-connected-voxels neighborhood) was considered a lesion candidate. All lesions that did not overlap with the rim + map were labeled as rim-. In order to have one lesion candidate paired with only one experts' rim+ annotation, an experienced technician manually separated the rim+ lesions inside confluent ones. A volume analysis performed on the lesions' automatic segmentations revealed that all rim+ lesions included in our dataset were bigger than  $12.3 \text{ mm}^3$ . As a result, we decided to exclude lesions smaller than  $12.3 \text{ mm}^3$  (1671 rim- lesions) from our study, as such small lesions could systematically be classified as rim- lesions.

Overall, our dataset of 124 patients contains 4857 rim- and 462 rim+ annotated lesions (10.5:1 ratio).

### 7.1.2 Pre-processing

Several pre-processing steps were performed. First, MPRAGE or MP2RAGE images were rigidly registered to the FLAIR space. Second, a WML segmentation was obtained using the automated method described in chapter 4 (La Rosa, Abdulkadir, et al., 2020). Third, post-processing and

**Table 7.1** – Number of patients per hospital, mean and standard deviation of patients' age, median and interquartile range of patients' EDSS and number of patients per MS type. Abbreviations: EDSS, expanded disability status scale; RRMS, relapsing-remitting MS; PPMS, primary-progressive MS; SPMS, secondary- progressive MS.

Hospital	Scanner	Patients	Age (years)	EDSS	RRMS	PPMS	SPMS
Lausanne	Skyra/Prisma	55	47.8 ± 11.0	2.0 (1.5–5.0)	36	9	10
	Skyra	28	48.5 ± 10.9	2.8 (1.5–4.9)	18	5	5
	Prisma	27	47.0 ± 11.3	2.0 (1.5–5.0)	18	4	5
Basel	Prisma	69	42.0 ± 14.1	2.5 (1.5–4.0)	51	7	11
Total		124	45.0 ± 13.1	2.0 (1.5–4.5)	87	16	21

registration of unwrapped phase images were performed as previously described (Absinta et al., 2013) and FLAIR images, along with the lesion segmentations, were affinely registered to the T2\* 3D-EPI space. Moreover, anatomical segmentations were generated with FreeSurfer (Fischl, 2012) from MPRAGE or MP2RAGE images and affinely registered to the T2\* 3D-EPI space. All registrations were performed with the SimpleElastix tool (Marstal et al., 2016) by using the adaptive stochastic gradient descent optimizer together with the advanced Mattes mutual information metric (Mattes et al., 2001). Fourth, lesion holes were filled with the lesion-filling function included in the FSL package (Battaglini et al., 2012). Fifth, 3D patches of size 28x28x28 were extracted centered on the center of mass of automatically detected lesions and linearly normalized between  $-1$  and  $1$  (Maggi, Fartaria, et al., 2020). The T2\*-w 3D-EPI magnitude, the unwrapped 3D-EPI phase and the 3D FLAIR images will be hereafter referred as T2\*, phase, and FLAIR, respectively.

### 7.1.3 CNN architecture

Our multimodal framework for the classification of PRL is inspired by the expert's imaging setting, which consists in the visual inspection of phase and FLAIR images. In this way, the prototype RimNet, see Fig. 7.3, is built upon two parallel CNNs based on the Visual Geometry Group Net (VGGNet) (Simonyan & Zisserman, 2015), which has proven to outperform other state-of-the-art architectures in similar multimodal approaches. Each single CNN, or branch, receives a patch and feeds it to a succession of three blocks of two convolutional layers followed by a max-pooling layer. The main branch extracts phase features, which are merged with those extracted from FLAIR after the very first block in order to exploit multimodal low-level features. Finally, the two four-dimensional output tensors of each CNN are concatenated before being fed to a final succession of fully connected layers, which profits from the high-level multimodal

## Classification of paramagnetic rim lesions

**Table 7.2** – Parameters of the MRI acquisition protocol followed in each center.

	Lausanne	Basel
Magnet strength	3T	3T
Manufacturer	Siemens	Siemens
Model	Prisma/Skyra	Skyra
Imaging Plane	Sagittal	Sagittal

	3D T2*-EPI	3D T2-FLAIR	3D T2*-EPI	3D T2-FLAIR
Resolution (mm, isotropic)	0.65	1	0.67	1
N° of slices	288	176	256	176
Repetition time (TR, ms)	64	5000	64	5000
Echo time (TE, ms)	35	391	35	386
Inversion time (TI, ms)	–	1800	–	–
Flip angle (deg)	10	Variable	10	Variable
Averages	1	1	1	1
Acquisition time	6'20"	4'47"	6'19"	5'40"

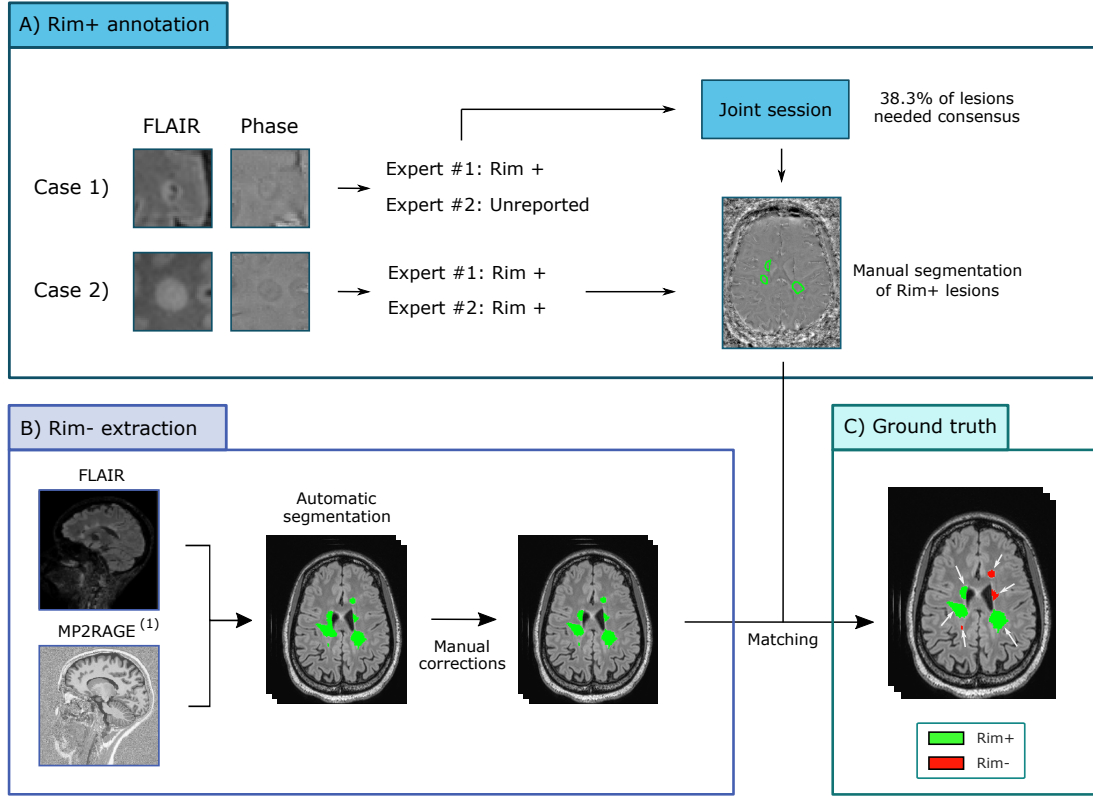
features.

Alternative RimNet configurations were also tested and compared: 1) replacing FLAIR by T2\* (phase + T2\*) and 2) using T2\* instead of phase (T2\* + FLAIR). Along with the proposed multi-modal scenarios, we also evaluated the prediction capability of each contrast separately and used these as baseline models. For this unimodal exploration, we used a network consisting of only one CNN branch of the RimNet directly connected to the cascade of fully connected layers.

### 7.1.4 Training strategies

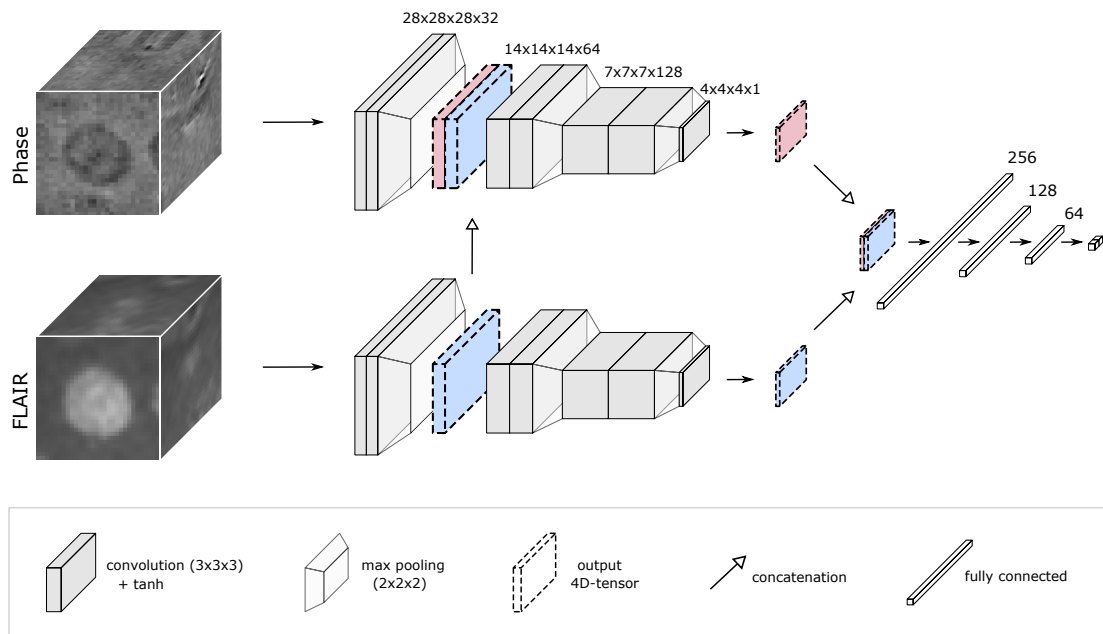
Our data augmentation consisted in rotating each PRL by 90°, 180°, and 270° in the three axes, which led to a tenfold increase of PRL and a 1.03:1 class ratio in the training set. Moreover, three elastically deformed versions of each lesion were generated, effectively quadrupling the training data. We also performed online data augmentation during training, first by flipping the patch along an axis (X, Y, Z, or none) and then by translating it 2 voxels (-2, 0 or +2) towards each of the three possible axes (X, Y, and Z). Both processes were designed to avoid generating repeated patches and hence increase the generalization of our model.

In order to avoid overfitting and to better reflect the performance in a clinical scenario, models were trained following a per-site stratified four-fold nested cross-validation procedure. The stratification process took into consideration the number of samples per class (rim+/-) and per center included in each fold. To perform a patient-wise rim analysis simulating a real



**Figure 7.2** – Description of the protocol used to label and generate our dataset. A) For each patient, two experts visually inspected the 3D-EPI phase and 3D FLAIR images and only reported paramagnetic rim lesions (rim+ lesions). The rim+ lesions detected by one expert and undetected or considered rim- by the other (unreported) went through a joint session where experts provided a final decision. B) Lesion candidates were extracted from the automatic segmentation (corresponding to the connected components by considering a 6-connected-voxels neighborhood) and matched with the rim+ annotations. In order to guarantee that one lesion candidate matched only one rim+ lesion annotation, a technician manually separated the rim+ lesions inside confluent ones.

case scenario, we imposed that all lesions of the same patient had to belong to the same split. This yielded folds each with  $36.3 \pm 1.9$  and  $78.0 \pm 0.0$  rim+ lesions ( $492.8 \pm 4.6$  and  $679.0 \pm 4.9$  rim- lesions) from Lausanne and Basel, respectively. Regarding the number of patients, each fold contained  $13.8 \pm 0.8$  from Lausanne and  $17.2 \pm 3.0$  from Basel. All experiments were trained with this fold configuration. For each fold's distribution, a three-fold inner cross-validation determined the number of epochs trained with each learning rate ( $1.0 \times 10^{-4}$ ,  $5.0 \times 10^{-5}$ ,  $2.5 \times 10^{-5}$ ,  $1.0 \times 10^{-5}$ ) before its decay, which was triggered after three consecutive epochs without a decrease in the validation loss. Early stopping was applied when the last learning rate change was triggered. For all network configurations, training was done with a batch size of 32 and SoftMax cross-entropy as the loss function.



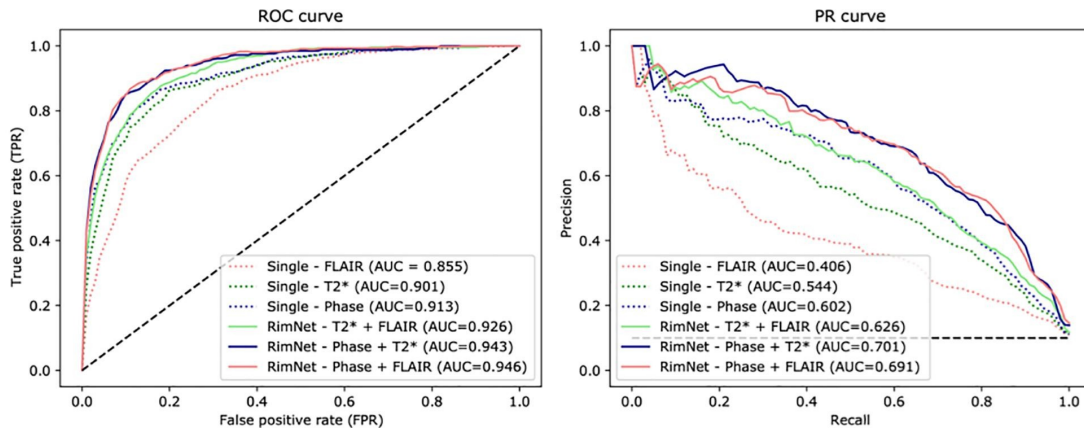
**Figure 7.3** – The architecture of RimNet. Built upon two parallel CNNs inspired by VGGNet, the proposed RimNet favors the multimodal low-level feature extraction by merging the output of the first convolutional block (two convolutions followed by a max pooling) of the second image to that of the main image (3D FLAIR and 3D-EPI phase in the figure, respectively). Finally, high-level multimodal feature maps are exploited through the final cascade of fully connected layers. Abbreviations: *tanh*, hyperbolic tangent function.

In order to evaluate the generalization of RimNet across different clinical centers, we additionally performed an inter-scanner/hospital study where the network was trained with only the Basel patients and then tested on the Lausanne dataset. The cross-validation process yielded four models trained with Basel data, which were used as an ensemble of classifiers to infer the labels for Lausanne patients' lesions.

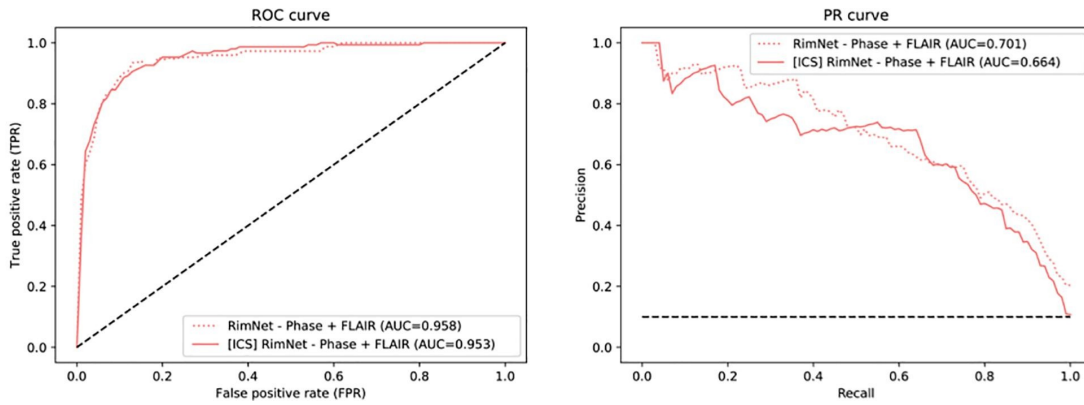
### 7.1.5 Results

Results of the lesion-wise analysis for all single and multimodal tested architectures are shown in Fig. 7.4. The exploration of the prediction capabilities of each individual modality shows that all modalities can, with different contributions, classify PRL substantially better than chance. As expected, both susceptibility-based modalities performed better than FLAIR (AUC = 0.855) at classifying PRL ( $P$ 's < 0.0001). Thus, phase (AUC = 0.913) and T2\* (AUC = 0.901) position themselves as the best sequences for this task ( $P$  = 0.47, DeLong test). Bimodal combinations of phase and either T2\* or FLAIR showed significantly higher prediction capabilities than the T2\* and FLAIR combination ( $P$  values of 0.003 and 0.023, respectively). For the rest of the section, we will call the architecture trained with FLAIR and phase as RimNet. RimNet

trained only with Basel samples showed a performance (AUC = 0.953) indistinguishable from the same network configuration trained with samples from both centers (AUC = 0.958,  $P = 0.47$ ). Regarding the lesion-wise metrics, For the cross-validation configuration, the single modality model performed with a sensitivity of 62.1%, a specificity of 95.0%, a PPV of 53.9%, and an NPV of 96.3%, while RimNet performed with a sensitivity of 70.6%, a specificity of 94.9%, a PPV of 56.9%, and an NPV of 97.1%. In the inter-center study, RimNet achieved a sensitivity of 75.8%, a specificity of 95.1%, a PPV of 52.8%, and an NPV of 98.3% (Fig. 7.5). In both scenarios, lesion-wise performance for both raters was superior to the one of the evaluated models (McNemar's test;  $P$ 's < 0.0001).



**Figure 7.4** – ROC and PR curves for all network configurations. Abbreviations: ROC, receiver operating characteristic; PR, precision-recall; ICS, inter-center study; Single, unimodal network; RimNet, the proposed multimodal network.

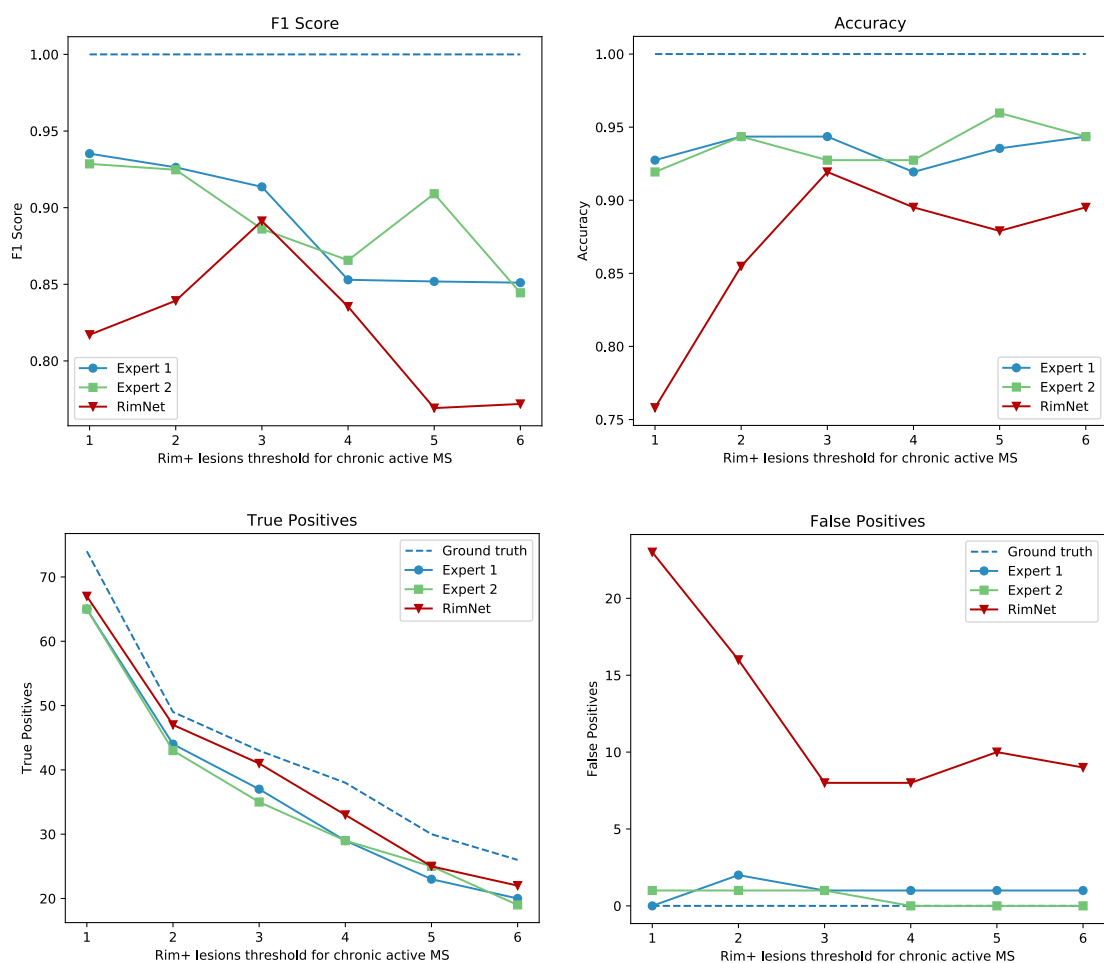


**Figure 7.5** – ROC and PR curves of RimNet in the inter-center study. Results are compared to those of RimNet trained with data from both centers and evaluated in Lausanne ( $P = 0.47$ ). Abbreviations: ROC, receiver operating characteristic; PR, precision-recall; ICS, inter-center study; RimNet, the proposed multimodal network.

We evaluated the RimNet patient classification performance based on the number of PRL per

## Classification of paramagnetic rim lesions

patient. To do so, we kept the operating point at a lesion-wise accepted false positive rate (FPR) of 0.05 and we categorized patients as “chronic active” and “non-chronic active” based on the total number of PRL per patient by using thresholds ranging from 1 to 6. Each patient’s rim analysis took an average of 300 ms on a desktop Intel(R) Core i74790 CPU machine at 3.60 GHz and a GeForce GTX 1080Ti GPU. The results are shown in Fig. 7.6. Significant differences between the model and the raters were proved only for the 1-rim-lesion threshold when compared to expert 1 (for the 1 to 6 thresholds, p-values of 0.0034, 0.0604, 0.2284, 0.1770, 0.1524, 0.1839) and for the 1- and 2-rim-lesions thresholds when compared to expert 2 (p-values of 0.0056, 0.0344, 0.14207, 0.1480, 0.2116, 0.0961).



**Figure 7.6** – Patient-wise analysis depending on the number of paramagnetic rim lesions set to consider a patient as “chronic active.” RimNet (with phase and FLAIR as inputs) is evaluated choosing an FPR of 0.05.

After the first individual screening of the annotation process, 188 lesions needed consensus review. During the joint session, experts agreed to classify 170 as rim+ (90.4%) and 18 as rim-

(9.6%). Fig. 7.7 shows the ratios of RimNet’s mistakes split by whether the lesions required consensus in the annotation process or not. Results show how RimNet (phase + FLAIR) misclassified significantly ( $p < 0.0001$ ) more PRL that required consensus (41.8%) than PRL that did not (22.3%). The same behavior was observed with rim- lesions, for which RimNet misclassified 38.9% of rim- lesions requiring consensus, compared to a miss rate of 4.9% for those with an early agreement ( $p = 0.01$ ).

### **7.1.6 Discussion**

In this work, we propose “RimNet”, a deep learning prototype for semi-automatic assessment of PRL in MS. By exploiting different MRI contrasts in a multi-center and multi-scanner setting, we showed that RimNet performance is at the level of expert readers. Importantly, our prototype achieves remarkably good PRL detection results even when tested in a multi-center scenario, thus supporting its potential for generalization across different clinical centers and datasets.

Among the different MRI techniques so far proposed to detect paramagnetic rim lesions in MS (Absinta et al., 2018; Clarke et al., 2020; Haacke et al., 2009; Walsh et al., 2013; Yao et al., 2012), 3D-EPI derived unwrapped phase images adopted in this study have shown promising performance in depicting the phase shifts produced by iron-laden macrophages and relative myelin content at the lesion edge and have been implemented for visual PRL analysis in clinical MRI studies (Absinta et al., 2018; Absinta et al., 2019). Our automated PRL evaluation is in line with these findings, as illustrated in our experiments on the single modality network, which showed the best performance when using phase as input modality (AUC = 0.913). Although containing susceptibility and morphological information of MS lesions, T2\*-magnitude did not prove as reliable as phase in the manual PRL classification (Absinta et al., 2018). However, the T2\* network (AUC = 0.901) showed a surprisingly good performance, closer than expected to the phase one ( $p = 0.470$ ). Nowadays, the role of FLAIR images during visual rim assessment is mainly restricted to the detection of MS lesions, thus allowing the experts to discard potential false positives due to rim-shaped artefacts. Although one could expect poor performance using only FLAIR as input, our experiments show, on the contrary, that FLAIR’s prediction capabilities are far from insignificant (AUC = 0.855). This relatively good performance of unimodal FLAIR architecture, along with the notably good performance of T2\*, suggests that, beyond the presence or absence of a lesion, morphometric features such as size, shape, and signal intensity, as depicted by these modalities, could play an important role in PRL classification.





**Figure 7.7** – RimNet errors analysis. The proportion of RimNet correct rim+/- predictions based on whether a consensus was required after the individual experts' annotations (consensus needed, bottom row) or not (agreed, top row) is shown. The columns split lesions based on their ground truth label (rim+ and rim- for the presence or absence of a paramagnetic rim, respectively). The total number of lesions of each type is shown inside the pie charts.

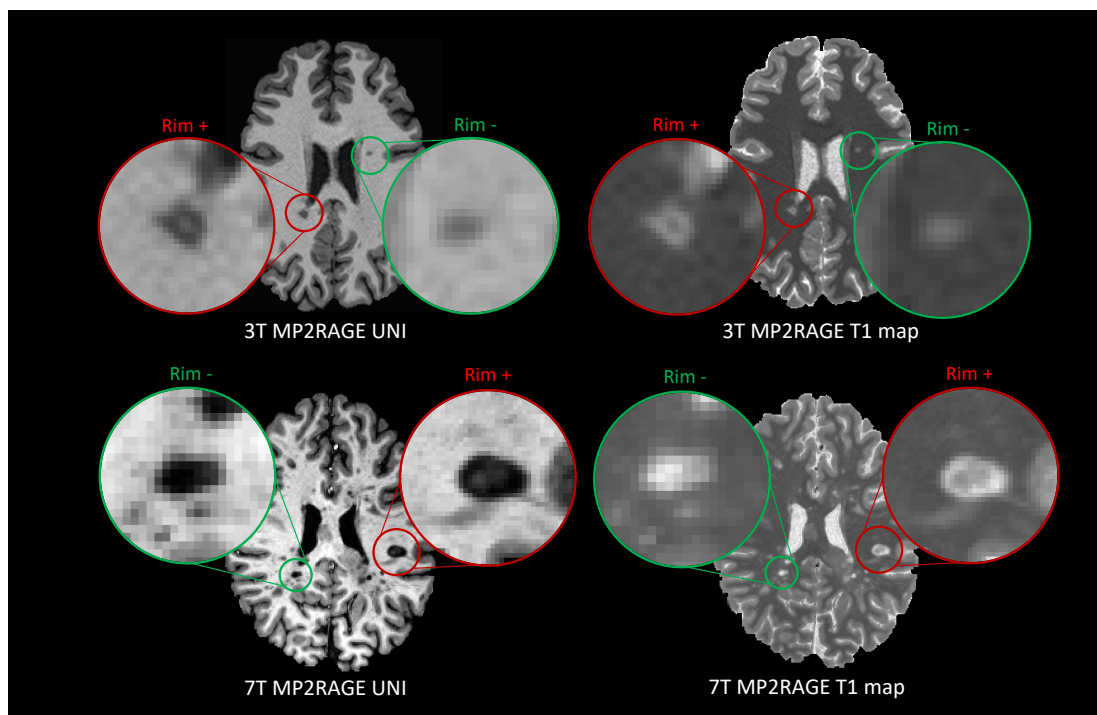
As an improvement on the simple unimodal approaches, we present the prototype RimNet, which relies on 3D multimodal MRI input patches. The early fusion of low-level features extracted from FLAIR and phase allows RimNet to extract low-level multimodal and lesion-aware features from the latter. Simultaneously and thanks to the straightforward parallel flow of the data of both modalities, the network benefits from the prediction capacities of both contrasts. High-level multimodal capabilities are exploited through the last fully connected layers. Results showed the superior performance of RimNet over all unimodal architectures (AUC = 0.946). When replacing phase with T2\*, performance dropped significantly (AUC = 0.926), thus validating the hypothesis that T2\* is not as reliable as phase regarding the extraction of PRL features. Moreover, the almost negligible loss of performance when replacing FLAIR with T2\* (AUC = 0.943) supports the hypothesis that both FLAIR and T2\* can be equally used to extract morphometric features that enhance the overall classification performance. This conclusion suggests the feasibility of performing PRL analysis with only one single MRI acquisition, although we chose to use the conventional FLAIR images in most of the analyses in the current work.

In the lesion-wise analysis, RimNet showed excellent performance with regard to sensitivity

(70.6%) and negative predictive value (NPV, 96.3%), which are values close to those of the experts (77.7% and 97.9%, averaged values for experts' sensitivity and NPV values, respectively). The downside of RimNet is its low positive predictive value (56.9%, compared to 96.5% and 97% of the experts), which is due to a relatively high number of false positives. However, this does not prevent it from becoming an excellent assessment tool to assist the visual rim analysis, as in this scenario lesions classified as PRL would always be checked by an expert. The biggest obstacle that MRI deep learning techniques face in their way toward being included in clinical workflows is the need for robustness against changes in the acquisition hardware and software. In other words, they must prove a good capability of generalizing with data acquired in different settings across healthcare institutions. The outstanding performance of RimNet in the inter-center study (AUC = 0.953) places it as a promising decision support tool for physicians in the rim analysis of MS patients, by providing an accurate PRL assessment in less than a second. This could be integrated with the only requirement of an expert clicking on the lesion in order to automatically extract and analyze the patch, instantly obtaining the decision of an extra rater.

This conception of RimNet is also supported by its results in the patient-wise analysis. Considering the recently proposed clinically meaningful threshold of  $\geq 4$  PRL per patient (Absinta et al., 2019; Maggi, Sati, et al., 2020), RimNet achieves a higher sensitivity (83.5%) and negative predictive value (94.0%) than experts (76.3% and 90.5%, respectively, averaged across experts), as well as similar accuracy (89.5%) and F1 score (83.5%) (92.3% and 85.8%, respectively, averaged across experts). As already depicted from the lesion-wise analysis, the main weaknesses of RimNet are its sensitivity and positive predictive value. Nonetheless, the absence of a drop in performance in the patient-wise inter-center study further supports the potential of RimNet as decision-support tool.

The largest limitation of our method resides in the very nature of any patch-based approach: lesions need to fit in patches of a fixed size. As a result, big lesions and confluent lesions entail big challenges. In the presented approach, the former were fed to the network untouched and the latter were manually split into unique lesions. This represents an obstacle to full automation of the rim analysis, which is highly needed for the inclusion of RimNet in clinical practice. Future work should improve our pipeline, so it becomes a fully automated approach. Another important limitation of our work resides in the lack of notable differences among the acquisition protocols of the scanners included in the inter-center study. Thus, although the inter-center study can be considered a preliminary analysis of the RimNet generalization power, future studies should test the performance of RimNet using data acquired with different



**Figure 7.8** – Examples of PRL and rim- lesions in diferent contrasts. First row: from left to right, 3T T2\*-w segEPI phase image, 3T MP2RAGE UNI, and T1 map. Second row: 7T T2\*-w dual-echo GRE phase image, 7T MP2RAGE UNI, and T1 map.

gradient-echo MRI sequences. Finally, we only focused on FLAIR and T2\*-EPI sequences. Future work should explore other MRI modalities such as T1-weighted or quantitative susceptibility maps (QSM) (Kaunzner et al., 2019; Wisnieff et al., 2015), which could provide RimNet with new information on the tissue properties of PRL.

In conclusion, RimNet is the first deep learning-based framework to semi-automatically classify PRL. Its excellent performance holds great promise for the standardization of the PRL biomarker and its translation into everyday clinical practice.

## 7.2 RimNet evaluation with 3T and 7T MP2RAGE

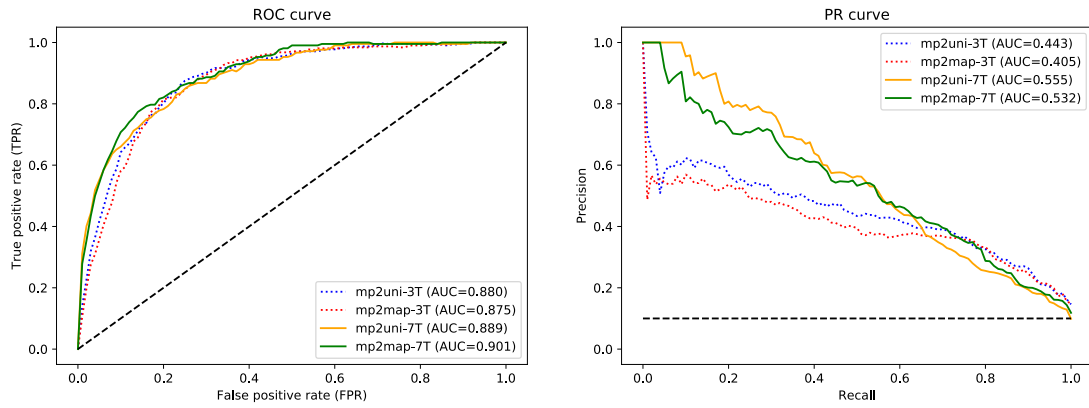
In this work, we evaluated a unimodal RimNet architecture based on either the MP2RAGE uniform contrast or the concurrently obtained T1 map at both 3T and 7T. Its performance was tested in a cross-validation procedure on 76 cases imaged at 3T and, in a separate testing set, on 20 cases imaged at 7T. Results show that predictions improve considerably at 7T, suggesting that 7T MP2RAGE might be helpful for automatically identifying PRL.

### 7.2.1 Methods

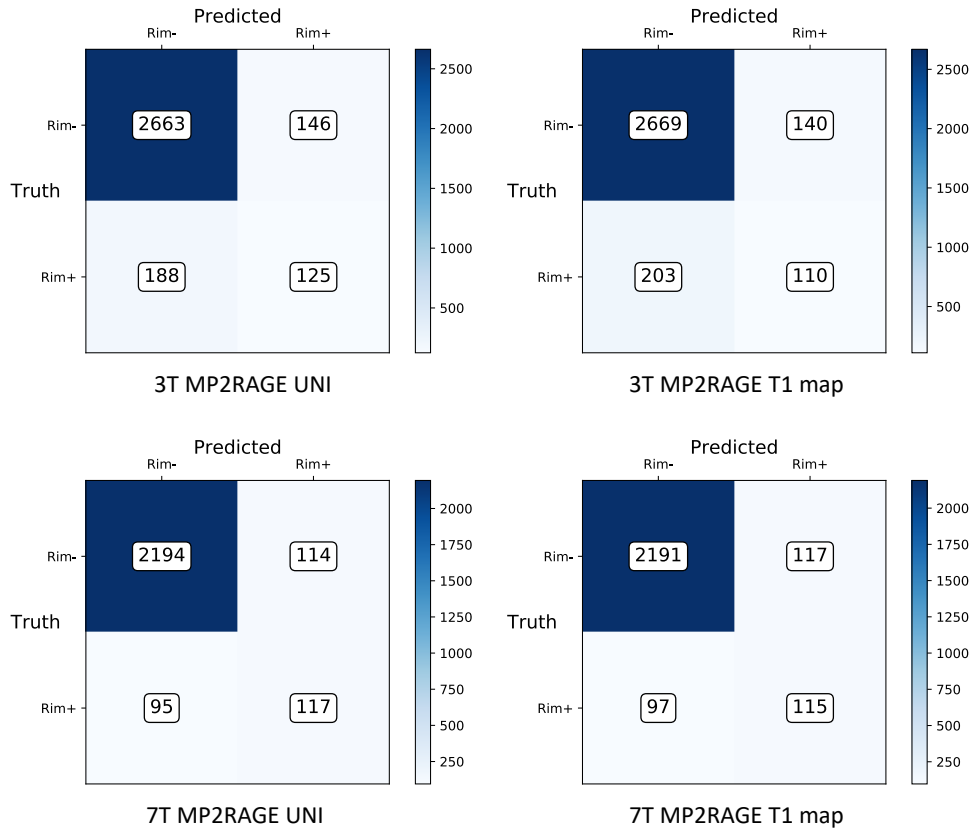
We retrospectively analyzed MRI data from 96 MS cases imaged at two university hospitals. Dataset-1 consisted of 76 cases (55 relapsing-remitting (RRMS) and 21 progressive MS (PMS), 49 female/27 male, mean age  $41 \pm 14$  years, age range [25-76] years, median EDSS 2.5, EDSS range [1.5-4.0]) imaged at 3T MRI (MAGNETOM Prisma, Siemens Healthcare, Erlangen, Germany) using a 64-channel head coil. The protocol included MP2RAGE and T2\*-weighted 3D-EPI. Dataset-2 included 20 cases (13 RRMS/7 PMS, 15 female/5 male, mean age  $44 \pm 11$  years, age range [25-66] years, median EDSS 2.5, EDSS range [0-7.5]) scanned on a prototype 7T system (Siemens Healthcare, Erlangen, Germany) using a 32-channel head coil. The protocol included MP2RAGE and T2\*-weighted dual-echo gradient echo (GRE). In Dataset-1, MS lesions were automatically segmented (La Rosa, Abdulkadir, et al., 2020) and PRL manually annotated by two experts, who reached consensus in a joint session. An automated method optimized for 7T MP2RAGE (Fartaria et al., 2019) was employed for segmenting MS lesions in Dataset-2, where a single expert visually identified PRL. Overall, Dataset-1 had a total of 2809 rim-lesions and 313 PRL, whereas Dataset-2 had 2308 rim-lesions and 212 PRL. MP2RAGE images were affinely registered to the 3D-EPI space using SimpleElastix (Marstal et al., 2016). Histogram matching was performed between 3T and 7T images for both UNI and T1 map, considering only the voxels inside the skull, in order to mitigate the intensity value differences. A unimodal RimNet architecture, implemented as described in the original work (see section 7.1, was trained first with the MP2RAGE uniform image (UNI) and then with the T1-map images from Dataset-1 in a four-fold nested cross-validation procedure. Inference was performed on both Dataset-1 and Dataset-2; for the former, an ensemble of the four trained classifiers was used. The AUC under the ROC and PR curve, as well as lesion-wise confusion matrices, were computed for performance evaluation.

### 7.2.2 Results

The ROC and PR curves, together with their relative AUC, are reported in Figure 7.9 for all four contrasts tested. The two models applied to 7T MP2RAGE UNI and T1-maps achieved the best results, with a PR AUC of 0.56 and 0.53, respectively. They clearly outperformed the 3T MP2RAGE UNI and T1 map models, which reached AUC of 0.44 and 0.41. In Figure 7.10, we present the lesion-wise confusion matrices of the four models. By setting a false positive rate of 0.05, as in the original RimNet paper (Barquero et al., 2020), the top-performing model based on 7T UNI images achieved a sensitivity of 55%, specificity of 95%, positive predictive value of 50%, and negative predictive value of 96%.



**Figure 7.9** – ROC and PR curves obtained for the testing datasets. Abbreviations: ROC, receiver operating characteristic; PR, precision-recall; mp2uni, MP2RAGE UNI; mp2map, MP2RAGE T1 map. 3T refers to the images of dataset-1 and 7T to those of dataset-2.



**Figure 7.10** – Lesion-wise confusion matrices obtained for 3T and 7T UNI and T1-map images.

### 7.2.3 Discussion

In this work, we explored the ability of a unimodal RimNet architecture based on either MP2RAGE uniform images or T1 maps to automatically assess PRL at both 3T and 7T. Results

show that, at both magnetic field strengths, MP2RAGE has some PRL detection capability, and that the UNI images slightly outperform the T1 maps. Moreover, even though RimNet was trained exclusively on 3T images, all evaluation metrics considerably improved for both images when tested at 7T. The best model, represented by the 7T MP2RAGE UNI, reached a PR AUC of 0.56, which is comparable to the AUC of 0.54 reported in the original RimNet paper for the unimodal T2\*-w magnitude image at 3T, but still worse than the AUC for the phase image. This confirms, as we previously suggested (Barquero et al., 2020), that morphometric features such as size, shape, and signal intensity, all well depicted by MP2RAGE, are relevant in the classification of PRLs.

Our contribution is two-fold. First, we show that the UNI and T1-map images of the MP2RAGE sequence acquired at 3T have some PRL prediction capability. Second, we observe that they considerably improve RimNet’s performance at 7T, suggesting that the MP2RAGE might be a suitable sequence for PRL assessment at 7T. Future work will explore integrating the MP2RAGE with the T2\*-weighted magnitude and phase images, as well as the quantitative susceptibility mapping.

## 8 Conclusion

In this thesis, novel DL-based methods were developed for the analysis of MS lesions, including advanced biomarkers such as cortical lesions (CL) and paramagnetic rim lesions (PRL). For this purpose, conventional and specialized MRI contrasts acquired at both 3T and 7T in multiple sites have been considered. This chapter summarizes the main contributions and presents their scientific and clinical impact. Following a discussion of their current limitations, this chapter ends with future research directions for ML-based methods in MS imaging.

### 8.1 Discussion

#### 8.1.1 Main contributions

The main contributions of this thesis are summarized as follows:

- **Cortical and white matter lesion segmentation at 3T.** To the best of our knowledge, we developed the first DL-based method for the simultaneous segmentation of MS cortical and white matter lesions with 3T MRI (La Rosa, Abdulkadir, et al., 2020). Our method has been evaluated on a cohort of 90 patients imaged in two sites and outperformed current state-of-the-art approaches for WML segmentation (see Chapter 4).
- **Synthetic MP2RAGE.** We proposed a generative adversarial network that successfully synthesizes realistic-looking MP2RAGE UNI images from an MPRAGE acquisition (La Rosa et al., 2021). The synthetic MP2RAGE has significantly improved the lesion and brain tissue segmentation masks computed by automated methods compared to the MPRAGE, whereas no statistically significant differences have been found in terms of lesion detection (refer to Chapter 5).

- **Cortical lesion detection with 7T MRI.** We designed an automated framework (Cortical Lesion AI-based assessment in MS, CLAIMS) for the detection and classification of cortical lesions at 7T (La Rosa, Beck, et al., 2020). Our approach has been evaluated on multiple MRI contrasts acquired at 7T and has achieved competitive performance on an external testing dataset (see Chapter 6).
- **Paramagnetic rim lesion classification.** We proposed the first semi-automated approach (RimNet) for the classification of PRL in MS patients at 3T (Barquero et al., 2020). RimNet has been evaluated on a large cohort of 124 MS patients and its performance is close to that of an expert. When tested on 7T MP2RAGE images, RimNet has yielded competitive results as well (refer to Chapter 7).
- **Open science.** All articles were published in open access peer-reviewed journals and methods developed by myself as first author were released publicly<sup>I</sup> in the hope that this could foster further improvements from different research groups.

### 8.1.2 Scientific and clinical impact

The work presented in this thesis has an important scientific impact which might influence clinically relevant advances in the near future. Our major contributions are the following. First, we confirmed our initial hypothesis developing a DL-based framework for the simultaneous segmentation of both cortical and white matter lesions with only two 3T clinically relevant MRI sequences (La Rosa, Abdulkadir, et al., 2020). Our method performs an accurate lesion segmentation, is fast at inference (about one minute per subject), and generalizes well to multiple sites. Our collaborators at Basel University Hospital are extensively using our software for research purposes to obtain a first MS lesion segmentation before manually correcting it and performing additional analyses. This confirms our second hypothesis regarding the utility of deep learning-based tools for supporting manual assessment of MS biomarkers. To date, our proposed method has been used in several clinical research studies and peer-reviewed publications in neurology and neuroimaging journals (Barquero et al., 2020; Bosticardo et al., 2021; Lu, Barakovic, et al., 2021; Lu, Yoo, et al., 2021; Maggi et al., 2021; Rahmanzadeh et al., 2021; Yu et al., 2021). Although further developments are needed to start the process of its deployment to the clinic, this contribution represents an important step for the application of DL methods to MS imaging, particularly for CL. Second, focusing exclusively on CL, we adapted our proposed approach to ultra-high field MRI, exploiting a multi-contrast dataset acquired at 7T (La Rosa, Beck, et al., 2020). To the best of our knowledge, we introduced

<sup>I</sup><https://github.com/orgs/Medical-Image-Analysis-Laboratory>



## Conclusion

---

the first automated framework for the simultaneous detection and classification of CL. As 7T scanners from different vendors have recently been approved for clinical use (Cosottini & Roccatagliata, 2021), our method could be helpful for researchers and clinicians in the field to facilitate CL assessment. This study validated our third hypothesis concerning the translation of our developed methods to ultra-high field MRI and their generalization to multi-site datasets. Third, a generative framework was designed to synthesize MP2RAGE images from MPRAGE. The synthetic MP2RAGE images significantly improved the automated brain tissue and lesion segmentation of automated tools, proving the potential impact of our framework on research studies running these software. Fourth, we proposed RimNet, a semi-automated approach for the classification of PRL (Barquero et al., 2020). Evaluated on a large, multi-site cohort of MS patients, RimNet approached the performance of experts in a fraction of time, confirming again our second hypothesis. Although further work is required to fully automatize the framework, RimNet could already help standardize the assessment of PRL and contribute to their translation in everyday clinical practice.

To summarize, all methods developed in this doctoral work were designed to meet specific clinical needs and represent a first step toward supporting the analysis of advanced MS biomarkers on high and ultra-high field MRI.

### 8.1.3 Limitations

Despite the major methodological contributions of this PhD thesis toward an automated assessment of CL and PRL, there are still some limitations that hinder the translation of the proposed methods to the clinic. These supervised approaches were trained and evaluated on multi-site datasets often acquired with the same scanner in a controlled setting. Furthermore, the MRI protocols used were similar and not representative of the current large diversity of images acquired in the clinics, including scans affected by noise and artifacts or protocols missing certain modalities. Therefore, the automated methods' robustness on larger datasets and on images acquired with different scanners, especially from multiple manufacturers, remains to be proven. This limitation is emphasized by the current lack of standardized acquisition protocols which increases the diversity of the MRI sequences considered for the same biomarkers. As a consequence, this represents also a major hurdle for potential regulatory approval of such methods. In addition, the achieved performance levels of the automated methods are still inferior compared to human experts. Considering the high inter-rater variability and the limited amount of data available, there is also a considerable risk of having methods that perform well on data annotated by a single expert and not as well with annotations from other raters. To mitigate this issue, we have often considered the consensus

segmentation of multiple experts to train the supervised approaches. Moreover, a crucial issue is given by the presence of confluent lesions as they prevent patch-wise approaches to classify each MS lesion singularly. These lesions currently require a manual splitting performed by experts, but this process needs to be automatized to facilitate the wide deployment of the proposed tools. Finally, several hurdles limit the translation of the proposed methods to existing radiological workflow. First, these tools lack graphical interfaces which would be essential for their widespread clinical use (Allen et al., 2019). Second, an integration of these approaches into medical imaging systems currently in use, such as PACS, is required. Third, the approval from regulatory agencies should certify the safety and efficacy of these software providing decision-making support to medical practitioners.

## **8.2 Future research directions**

### **8.2.1 Extensive validation**

The first step needed to facilitate the deployment of the methods described is their validation on large datasets acquired with different scanners, possibly from multiple vendors. This would be a desirable step to prove the robustness of these DL-based methods to the diverse clinical images. For this purpose, a validation of the biomarkers' specific criteria and standardization of the relative MRI protocols is absolutely required. As recently done for WML (Thompson, Banwell, et al., 2018), consensus guidelines should be given to promote standardized imaging protocols at each magnetic field strength. Moreover, having a consensus ground truth of more than two experts would help reduce the inter-rater variability. It would be extremely helpful if research groups organize grand challenges for CL or PRL as already successfully done for WML (Carass et al., 2017; Commowick et al., 2021; Commowick et al., 2018). This would boost the development of several different approaches and concurrently establish benchmarks for their fair assessment.

### **8.2.2 Transfer learning**

Considering the scarcity of large, annotated datasets, an additional strategy that should be explored consists of transfer learning (Zhang et al., 2019). Sharing of neural network weights between research groups could foster interdisciplinary applicability of CNN trained on relatively large datasets towards different purposes, such as advanced MS biomarkers, by fine-tuning the trained models on smaller datasets. Potential advantages would include a shorter training time and robust feature extraction across different MRI device manufacturers

## Conclusion

---

or different pulse sequence acquisition parameters. Similar improvements could be given by federated learning, consisting in training networks on multiple local datasets, while explicitly sharing only the networks' parameters (Li et al., 2020).

### 8.2.3 Single framework for multiple biomarkers

Each automated method presented in this thesis tackled the assessment of a single MS biomarker. Future work may aim at developing a unified framework that, taking as input multiple MRI contrasts, performs an automated analysis of multiple biomarkers simultaneously. This would be extremely useful for research purposes as well as in clinics. A combined assessment of CL and PRL, for instance, could directly provide important prognostic information, whereas adding to these two biomarkers the central vein sign would be helpful for differential diagnosis. Along this line of research, the automated assessment of multiple biomarkers could also be exploited to develop regression models. These approaches could potentially predict clinical outcomes or treatment effects, as shown by a recent study in which, however, biomarkers were assessed manually (Treaba et al., 2021).

### 8.2.4 Explainable AI

The complexity of DL approaches frequently prevents their interpretation, which has led to consider these algorithms as “black boxes”. In the near future, reliable interpretability and explainability techniques will be needed to trust and verify the automated methods' prediction (Reyes et al., 2020). On one hand, interpretable methods are those mapping abstract concepts into a domain that can be understood by the user (Montavon et al., 2018). On the other hand, by explainability, we refer to a set of domain features, such as pixels of an image or human-understandable high-level attributes, that contribute to the output decision of the model and its internal working. Interpretable approaches would increase the confidence of clinicians in these tools, whereas explainable DL would provide new insights into model decisions and help identify any bias. To our knowledge, only two groups have investigated explainable DL in MS. Eitel et al. have explored voxel-wise heatmaps to visualize the locations considered by a CNN for distinguishing between a normal brain and one affected by MS (Eitel et al., 2019). They found that diagnostic success relied on the appearance of both lesions and non-lesional tissue. Nair et al. studied uncertainty measures computed with Monte-Carlo dropout of DL-based lesion segmentation masks (Nair et al., 2020). Interestingly, their results showed that discarding lesions with high estimated uncertainty from the output segmentation would improve the performance of the model. Specifically regarding CL and

PRL, these techniques could improve the respective automated methods performance and suggest additional characteristics such as image texture or morphometric features playing an important role in the biomarkers assessment. Overall, providing an uncertainty measure along with the usual output would increase the transparency of the decisions and, importantly, increase the trust of the end-user. In particular, for MS biomarkers, increasing the clinicians' confidence in automated methods will be crucial to promote their wide deployment.

# A Appendix

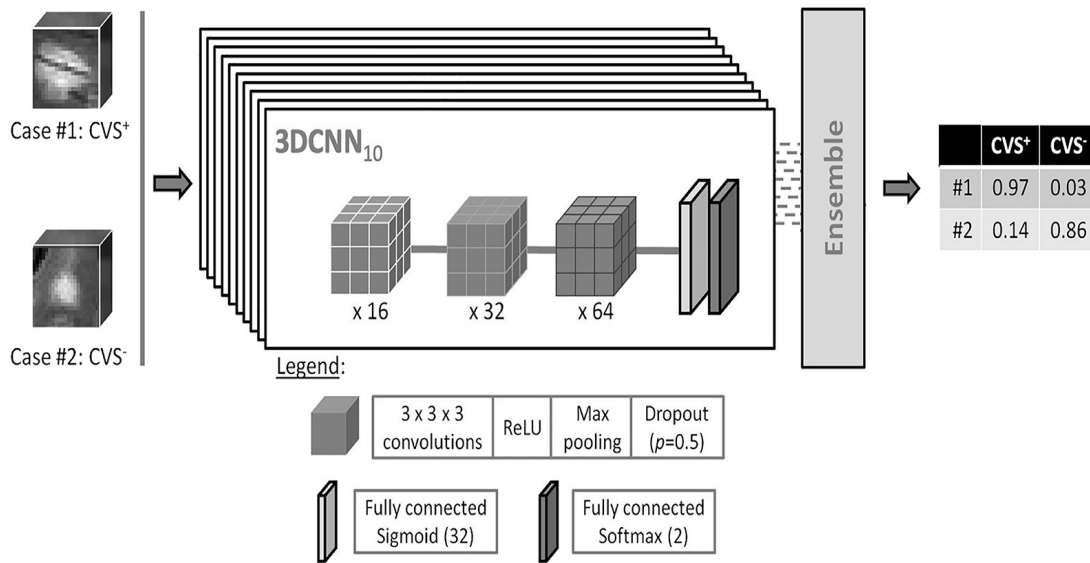
## A.1 Other contributions

During my doctoral thesis, I contributed to two additional works led by our collaborators. The first regards the development of an automated method for the CVS assessment, whereas the second concerns the application of the fluid and white matter suppression (FLAWS) sequence to MS patients, in particular for CL detection. Both works are briefly presented in the following sections.

### A.1.1 CVSnet

*The work presented in this section is adapted with permission from the postprint version of the following article: Pietro Maggi, Mário João Fartaria, João Jorge, Francesco La Rosa, Martina Absinta, Pascal Sati, Reto Meuli, Renaud Du Pasquier, Daniel S Reich, Meritxell Bach Cuadra, Cristina Granziera, Jonas Richiardi, Tobias Kober: "CVSnet: A machine learning approach for automated central vein sign assessment in multiple sclerosis." in NMR in Biomedicine, 2020; doi: [110.1002/nbm.4283](https://doi.org/10.1002/nbm.4283). The author of this thesis contributed to the methodology, software, validation, and writing of the study.*

CVSnet is a 3D convolutional neural network designed to automatically classify single MS lesion as either CVS+ (presence of the CVS) or CVS- (absence of the CVS). Similarly to Rim-Net (see chapter 7.1), it is based on 3D patches centered around the lesion of interest. Its architecture is composed of three convolutional layers followed by two fully connected layers (see Figure A.1 for more details) and it provides as output the probability of the lesion to have the CVS. In this study, an ensemble of 10 CNNs with the same architecture was trained and

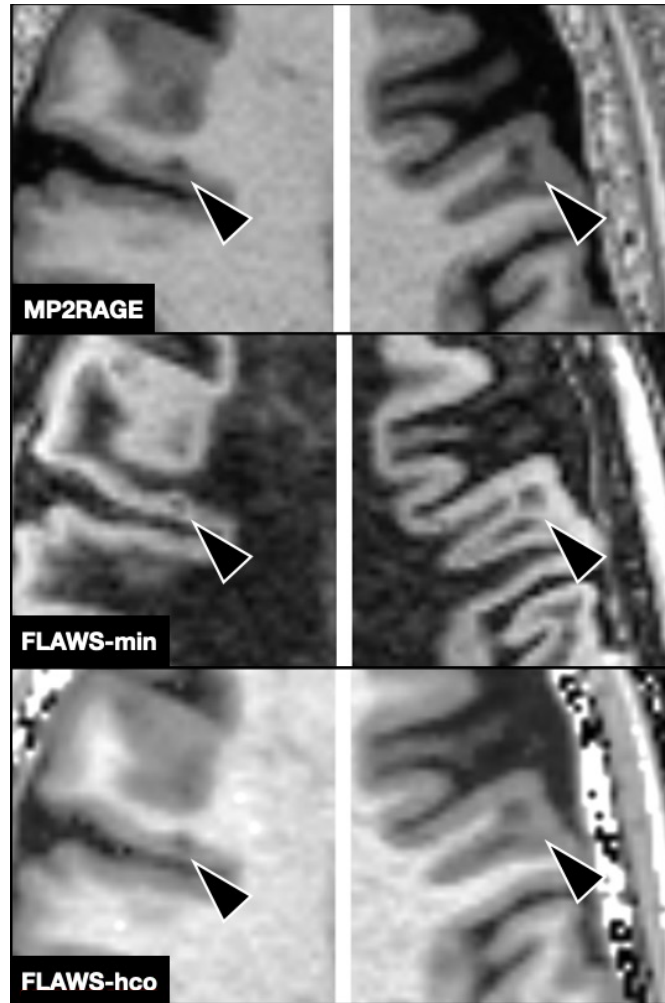


**Figure A.1** – CVSnet, the proposed 3D convolutional neural network, consists of a simple, repetitive, and relatively shallow architecture. On the test set, we use these 10 different versions of the same algorithm to yield 10 different predictions for each incoming test patch and average the output (“ensembling”) to yield a single, more stable, prediction.

at inference time their output was averaged to lower prediction variance for the test set. 80 subjects (42 MS patients, 33 with mimic disease, 5 with unknown diagnosis) were imaged in three different clinical centers. Two experts manually segmented MS lesions and classified CVS+/- lesions by consensus in the FLAIR\* images (Sati et al., 2012). Patients were divided into MS vs. MS mimics based on three previously proposed criteria: i) the “50% rule”, ii) the “6-lesion rule”, and iii) the “3-lesion rule” (Mistry et al., 2016; Solomon et al., 2018; Tallantyre et al., 2011). CSVnet results were evaluated lesion-wise and subject-wise and compared with a state-of-the-art vesselness filtering approach (Frangi et al., 1998). In the test set, CVSnet approached human performance, with lesion-wise median balanced accuracy of 81%, and subject-wise balanced accuracy of 91%. The CVS assessment, in previously manually segmented lesions, was ~600-fold faster using the proposed CVSnet compared with human visual assessment (test set: 4 seconds vs. 40 minutes). The main drawback of the proposed framework is given by the manually obtained MS lesion segmentation, which prevents the tool from being fully-automated. Future work should explore automated lesion segmentation techniques as well as investigate different approaches to consider the lesion exclusion criteria (Sati et al., 2016).

### A.1.2 FLAWS for multiple sclerosis

*The work presented in this section is adapted from the preprint version of the following article: Jannis Müller, Francesco La Rosa, Jeremy Beaumont, Charidimos Tsagkas, Reza Rahmanzadeh, Matthias Weigel, Meritxell Bach Cuadra, Giulio Gambarota, Cristina Granziera; "FLAWS: New Sensitive 3T MRI Contrasts for Cortical Lesion Detection in Multiple Sclerosis." currently under revision in Investigative Radiology. The second author (author of this thesis) contributed to the conceptualization, methodology, software, validation, and writing of the study.*



**Figure A.2** – Example of two cortical lesions (arrow heads) on MP2RAGE (top horizontal row), FLAWS-min (mid horizontal row) and FLAWS-hco (bottom horizontal row).

FLuid And White matter Suppression (FLAWS) provides multiple reconstructed contrasts acquired during a single acquisition (Tanner et al., 2012). These include FLAWS minimum image (FLAWS-min), which provides a high sensitivity to the grey matter signal and therefore

may ease cortical lesion identification, and high contrast FLAWS (FLAWS-hco), which gives a contrast that is similar to MP2RAGE (Beaumont et al., 2021). In this study, we compared the manual detection rate of cortical lesions using MP2RAGE, FLAWS-min and FLAWS-hco in MS patients. We included 30 relapsing-remitting MS patients who underwent MP2RAGE and FLAWS MRI with isotropic spatial resolution of 1mm at 3 Tesla. Cortical lesions were segmented manually by consensus of three trained raters and classified as intracortical or leukocortical lesions (Calabrese et al., 2009) on (i) MP2RAGE uniform images, (ii) FLAWS-min (iii) FLAWS-hco. Additionally, the segmented lesions on FLAWS-min and FLAWS-hco were merged to produce a union lesion map (FLAWS-min+hco). Results showed that the median number and volume of single cortical, intracortical and leukocortical lesions were comparable between MP2RAGE, FLAWS-min and FLAWS-hco (all  $p > 0.05$ ). In patients with cortical lesions (22/30), median cumulative lesion volume was larger on FLAWS-min (587ml, IQR 1405ml) than on MP2RAGE (490ml, IQR 990ml;  $p = 0.04$ ), whereas there was no difference between FLAWS-min and FLAWS-hco, or FLAWS-hco and MP2RAGE. FLAWS-min+hco showed significantly more cortical (median 4.5, IQR 15) and leukocortical (median 3.5, IQR 12) lesions than MP2RAGE (median 3, IQR 10; median 2.5 IQR 7; both  $p < 0.001$ ). Considering Fleiss  $\kappa$ , interrater agreement was moderate on MP2RAGE ( $\kappa = 0.582$ ) and FLAWS-hco ( $\kappa = 0.584$ ), but substantial on FLAWS-min ( $\kappa = 0.614$ ). To summarize, both FLAWS-min and FLAWS-hco provide a good performance in terms of cortical lesion identification with a detection rate that is comparable to one of the current gold standard sequences, MP2RAGE. The combination of FLAWS-min and FLAWS-hco yielded a higher number of cortical lesions than MP2RAGE alone. A major limitation of this work was that white matter lesions were not considered and might have influenced the different cortical lesion counts. Future research should aim at exploring FLAWS contrasts with automated cortical lesion approaches.



# Bibliography

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., ... Zheng, X. (2016). TensorFlow: A system for large-scale machine learning, 21. <https://doi.org/doi.org/abs/1605.08695>
- Absinta, M., Sati, P., Fechner, A., Schindler, M., Nair, G., & Reich, D. (2018). Identification of Chronic Active Multiple Sclerosis Lesions on 3T MRI. *American Journal of Neuroradiology*, 39(7), 1233–1238. <https://doi.org/10.3174/ajnr.A5660>
- Absinta, M., Sati, P., Gaitán, M. I., Maggi, P., Cortese, I. C. M., Filippi, M., & Reich, D. S. (2013). Seven-tesla phase imaging of acute multiple sclerosis lesions: A new window into the inflammatory process: MS Lesions at 7T Phase Imaging. *Annals of Neurology*, 74(5), 669–678. <https://doi.org/10.1002/ana.23959>
- Absinta, M., Sati, P., Masuzzo, F., Nair, G., Sethi, V., Kolb, H., Ohayon, J., Wu, T., Cortese, I. C. M., & Reich, D. S. (2019). Association of Chronic Active Multiple Sclerosis Lesions With Disability In Vivo. *JAMA Neurology*, 76(12), 1474. <https://doi.org/10.1001/jamaneurol.2019.2399>
- Absinta, M., Sati, P., Schindler, M., Leibovitch, E. C., Ohayon, J., Wu, T., Meani, A., Filippi, M., Jacobson, S., Cortese, I. C., & Reich, D. S. (2016). Persistent 7-tesla phase rim predicts poor outcome in new multiple sclerosis patient lesions. *Journal of Clinical Investigation*, 126(7), 2597–2609. <https://doi.org/10.1172/JCI86198>
- Allen, B., Seltzer, S. E., Langlotz, C. P., Dreyer, K. P., Summers, R. M., Petrick, N., Marinac-Dabic, D., Cruz, M., Alkasab, T. K., Hanisch, R. J., Nilsen, W. J., Burleson, J., Lyman, K., & Kandarpa, K. (2019). A Road Map for Translational Research on Artificial Intelligence in Medical Imaging: From the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop. *Journal of the American College of Radiology*, 16(9), 1179–1189. <https://doi.org/10.1016/j.jacr.2019.04.014>
- Aslani, S., Dayan, M., Storelli, L., Filippi, M., Murino, V., Rocca, M. A., & Sona, D. (2019). Multi-branch convolutional neural network for multiple sclerosis lesion segmentation. *NeuroImage*, 196, 1–15. <https://doi.org/10.1016/j.neuroimage.2019.03.068>
- Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26–41. <https://doi.org/10.1016/j.media.2007.06.004>

- Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., & Gee, J. C. (2011). An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics*, 9(4), 381–400. <https://doi.org/10.1007/s12021-011-9109-y>
- Barkhof, F. (1999). MRI in multiple sclerosis: correlation with expanded disability status scale (EDSS). *Multiple Sclerosis Journal*, 5(4), 283–286. <https://doi.org/10.1177/135245859900500415>
- Barquero, G., La Rosa, F., Kebiri, H., Lu, P.-J., Rahmanzadeh, R., Weigel, M., Fartaria, M. J., Kober, T., Théaudin, M., Du Pasquier, R., Sati, P., Reich, D. S., Absinta, M., Granziera, C., Maggi, P., & Bach Cuadra, M. (2020). RimNet: A deep 3D multimodal MRI architecture for paramagnetic rim lesion assessment in multiple sclerosis. *NeuroImage: Clinical*, 28, 102412. <https://doi.org/10.1016/j.nicl.2020.102412>
- Battaglini, M., Jenkinson, M., & De Stefano, N. (2012). Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Human Brain Mapping*, 33(9), 2062–2071. <https://doi.org/10.1002/hbm.21344>
- Beaumont, J., Gambarota, G., Saint-Jalmes, H., Acosta, O., Ferré, J.-C., Raniga, P., & Fripp, J. (2021). High-resolution multi-T<sub>2</sub>-weighted contrast and T1 mapping with low B1+ sensitivity using the fluid and white matter suppression (FLAWS) sequence at 7T. *Magnetic Resonance in Medicine*, 85(3), 1364–1378. <https://doi.org/10.1002/mrm.28517>
- Beck, E. S., Gai, N., Filippini, S., Maranzano, J., Nair, G., & Reich, D. S. (2020). Inversion Recovery Susceptibility Weighted Imaging With Enhanced T2 Weighting at 3 T Improves Visualization of Subpial Cortical Multiple Sclerosis Lesions. *Investigative Radiology*, 55(11), 727–735. <https://doi.org/10.1097/RLI.0000000000000698>
- Beck, E. S., Sati, P., Sethi, V., Kober, T., Dewey, B., Bhargava, P., Nair, G., Cortese, I., & Reich, D. (2018). Improved Visualization of Cortical Lesions in Multiple Sclerosis Using 7T MP2RAGE. *American Journal of Neuroradiology*, 39(3), 459–466. <https://doi.org/10.3174/ajnr.A5534>
- Benjamens, S., Dhunoo, P., & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digital Medicine*, 3(1), 118. <https://doi.org/10.1038/s41746-020-00324-0>
- Bosticardo, S., Schiavi, S., Schaedelin, S., Lu, P.-J., Barakovic, M., Weigel, M., Kappos, L., Kuhle, J., Daducci, A., & Granziera, C. (2021). Microstructure-Weighted Connectomics in Multiple Sclerosis. *Brain Connectivity*, brain.2021.0047. <https://doi.org/10.1089/brain.2021.0047>
- Brant-Zawadzki, M., Gillan, G. D., & Nitz, W. R. (1992). MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence—initial experience in the brain. *Radiology*, 182(3), 769–775. <https://doi.org/10.1148/radiology.182.3.1535892>
- Brosch, T., Tang, L. Y. W., Yoo, Y., Li, D. K. B., Traboulsee, A., & Tam, R. (2016). Deep 3D Convolutional Encoder Networks With Shortcuts for Multiscale Feature Integration Applied to Multiple Sclerosis Lesion Segmentation. *IEEE Transactions on Medical Imaging*, 35(5), 1229–1239. <https://doi.org/10.1109/TMI.2016.2528821>
- Brownlee, W. J., & Solomon, A. J. (2021). Misdiagnosis of multiple sclerosis: Time for action. *Multiple Sclerosis Journal*, 27(6), 805–806. <https://doi.org/10.1177/13524585211005367>

- Bruschi, N., Boffa, G., & Inglese, M. (2020). Ultra-high-field 7-T MRI in multiple sclerosis and other demyelinating diseases: from pathology to clinical practice. *European Radiology Experimental*, 4(1), 59. <https://doi.org/10.1186/s41747-020-00186-x>
- Calabrese, M., Agosta, F., Rinaldi, F., Mattisi, I., Grossi, P., Favaretto, A., Atzori, M., Bernardi, V., Barachino, L., Rinaldi, L., Perini, P., Gallo, P., & Filippi, M. (2009). Cortical Lesions and Atrophy Associated With Cognitive Impairment in Relapsing-Remitting Multiple Sclerosis. *Archives of Neurology*, 66(9). <https://doi.org/10.1001/archneurol.2009.174>
- Calabrese, M., Filippi, M., & Gallo, P. (2010). Cortical lesions in multiple sclerosis. *Nature Reviews Neurology*, 6(8), 438–444. <https://doi.org/10.1038/nrneurol.2010.93>
- Carass, A., Roy, S., Jog, A., Cuzzocreo, J. L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Prados, F., Sudre, C. H., Jorge Cardoso, M., Cawley, N., Ciccarelli, O., Wheeler-Kingshott, C. A., Ourselin, S., Catanese, L., Deshpande, H., Maurel, P., Commowick, O., ... Pham, D. L. (2017). Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. *NeuroImage*, 148, 77–102. <https://doi.org/10.1016/j.neuroimage.2016.12.064>
- Castellaro, M., Tamanti, A., Pisani, A. I., Pizzini, F. B., Crescenzo, F., & Calabrese, M. (2020). The Use of the Central Vein Sign in the Diagnosis of Multiple Sclerosis: A Systematic Review and Meta-analysis. *Diagnostics*, 10(12), 1025. <https://doi.org/10.3390/diagnostics10121025>
- Cerri, S., Puonti, O., Meier, D. S., Wuerfel, J., Mühlau, M., Siebner, H. R., & Van Leemput, K. (2021). A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis. *NeuroImage*, 225, 117471. <https://doi.org/10.1016/j.neuroimage.2020.117471>
- Chen, Y., Christodoulou, A. G., Zhou, Z., Shi, F., Xie, Y., & Li, D. (2020). MRI Super-Resolution with GAN and 3D Multi-Level DenseNet: Smaller, Faster, and Better [arXiv: 2003.01217]. *arXiv:2003.01217 [cs, eess]*. Retrieved May 5, 2020, from <http://arxiv.org/abs/2003.01217>
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation [Series Title: Lecture Notes in Computer Science]. In S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, & W. Wells (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (pp. 424–432). Springer International Publishing. [https://doi.org/10.1007/978-3-319-46723-8\\_49](https://doi.org/10.1007/978-3-319-46723-8_49)
- Clarke, M. A., Samaraweera, A. P., Falah, Y., Pitiot, A., Allen, C. M., Dineen, R. A., Tench, C. R., Morgan, P. S., & Evangelou, N. (2020). Single Test to ARrive at Multiple Sclerosis (STAR-MS) diagnosis: A prospective pilot study assessing the accuracy of the central vein sign in predicting multiple sclerosis in cases of diagnostic uncertainty. *Multiple Sclerosis Journal*, 26(4), 433–441. <https://doi.org/10.1177/1352458519882282>
- Commowick, O., Cervenansky, F., Cotton, F., & Dojat, M. (2021). MSSEG-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure, 1. Retrieved October 28, 2021, from <https://hal.inria.fr/hal-03358968>
- Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., Pop, S. C., Girard, P., Amélie, R., Ferré, J.-C., Kerbrat, A., Tourdias, T., Cervenansky, F., Glatard, T., Beaumont, J., Doyle, S., Forbes, F., Knight, J., Khademi, A., ... Barillot, C. (2018). Objective Evaluation of Multiple Sclerosis

- Lesion Segmentation using a Data Management and Processing Infrastructure. *Scientific Reports*, 8(1), 13650. <https://doi.org/10.1038/s41598-018-31911-7>
- Compston, A., & Coles, A. (2008). Multiple sclerosis. *The Lancet*, 372(9648), 1502–1517. [https://doi.org/10.1016/S0140-6736\(08\)61620-7](https://doi.org/10.1016/S0140-6736(08)61620-7)
- Consortium, T. M. (2020). Project MONAI. <https://doi.org/10.5281/ZENODO.4323059>
- Cosottini, M., & Roccatagliata, L. (2021). Neuroimaging at 7 T: are we ready for clinical transition? - PubMed. *European Radiology Experimental*. Retrieved November 3, 2021, from <https://pubmed.ncbi.nlm.nih.gov/34435257/>
- Cuadra, M. B., Gelin, S., & Roche, A. (2013). Classical Segmentation Methods on Novel MR Imaging: A Study of Brain Tissue Segmentation of MP2RAGE Vs MPRAGE. <http://archive.ismrm.org/2013/0948.html>
- Dal-Bianco, A., Grabner, G., Kronnerwetter, C., Weber, M., Höftberger, R., Berger, T., Auff, E., Leutmezer, F., Trattinig, S., Lassmann, H., Bagnato, F., & Hametner, S. (2017). Slow expansion of multiple sclerosis iron rim lesions: pathology and 7 T magnetic resonance imaging. *Acta Neuropathologica*, 133(1), 25–42. <https://doi.org/10.1007/s00401-016-1636-z>
- Danelakis, A., Theoharis, T., & Verganelakis, D. A. (2018). Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. *Computerized Medical Imaging and Graphics*, 70, 83–100. <https://doi.org/10.1016/j.compmedimag.2018.10.002>
- Dendrou, C. A., Fugger, L., & Friese, M. A. (2015). Immunopathology of multiple sclerosis. *Nature Reviews Immunology*, 15(9), 545–558. <https://doi.org/10.1038/nri3871>
- Dworkin, J., Linn, K., Oguz, I., Fleishman, G., Bakshi, R., Nair, G., Calabresi, P., Henry, R., Oh, J., Papinutto, N., Pelletier, D., Rooney, W., Stern, W., Sicotte, N., Reich, D., & Shinohara, R. (2019). An Automated Statistical Technique for Counting Distinct Multiple Sclerosis Lesions. *American Journal of Neuroradiology*, 39(4), 626–633. <https://doi.org/10.3174/ajnr.A5556>
- Dworkin, J., Sati, P., Solomon, A., Pham, D., Watts, R., Martin, M., Ontaneda, D., Schindler, M., Reich, D., & Shinohara, R. (2018). Automated Integration of Multimodal MRI for the Probabilistic Detection of the Central Vein Sign in White Matter Lesions. *American Journal of Neuroradiology*, 39(10), 1806–1813. <https://doi.org/10.3174/ajnr.A5765>
- Eitel, F., Soehler, E., Bellmann-Strobl, J., Brandt, A. U., Ruprecht, K., Giess, R. M., Kuchling, J., Asseyer, S., Weygandt, M., Haynes, J.-D., Scheel, M., Paul, F., & Ritter, K. (2019). Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. *NeuroImage: Clinical*, 24, 102003. <https://doi.org/10.1016/j.nicl.2019.102003>
- Elliot, C., & Arbel, T. (2016). *A Bayesian framework for 4-D segmentation of Multiple Sclerosis lesions in serial MRI in the brain* (Doctoral dissertation). McGill University. Retrieved October 13, 2021, from <https://escholarship.mcgill.ca/concern/theses/jd473013b>
- Fartaria, M. J., Bonnier, G., Roche, A., Kober, T., Meuli, R., Rotzinger, D., Frackowiak, R., Schlupe, M., Du Pasquier, R., Thiran, J.-P., Krueger, G., Bach Cuadra, M., & Granziera, C. (2016). Automated detection of white matter and cortical lesions in early stages of multiple sclerosis: Automated

## Bibliography

---

- MS Lesion Segmentation. *Journal of Magnetic Resonance Imaging*, 43(6), 1445–1454. <https://doi.org/10.1002/jmri.25095>
- Fartaria, M. J., Roche, A., Meuli, R., Granziera, C., Kober, T., & Bach Cuadra, M. (2017). Segmentation of Cortical and Subcortical Multiple Sclerosis Lesions Based on Constrained Partial Volume Modeling [Series Title: Lecture Notes in Computer Science]. *Medical Image Computing and Computer Assisted Intervention MICCAI 2017* (pp. 142–149). Springer International Publishing. [https://doi.org/10.1007/978-3-319-66179-7\\_17](https://doi.org/10.1007/978-3-319-66179-7_17)
- Fartaria, M. J., Sati, P., Todea, A., Radue, E.-W., Rahmzadeh, R., O'Brien, K., Reich, D. S., Bach Cuadra, M., Kober, T., & Granziera, C. (2019). Automated Detection and Segmentation of Multiple Sclerosis Lesions Using Ultra-High-Field MP2RAGE. *Investigative Radiology*, 54(6), 356–364. <https://doi.org/10.1097/RLI.0000000000000551>
- Favaretto, A., Lazzarotto, A., Poggiali, D., Rolma, G., Causin, F., Rinaldi, E., Perini, P., & Gallo, P. (2016). MRI-detectable cortical lesions in the cerebellum and their clinical relevance in multiple sclerosis. *Multiple Sclerosis Journal*, 22(4), 494–501. <https://doi.org/10.1177/1352458515594043>
- Feng, Y., Pan, H., Meyer, C., & Feng, X. (2018). A Self-Adaptive Network For Multiple Sclerosis Lesion Segmentation From Multi-Contrast MRI With Various Imaging Protocols [arXiv: 1811.07491]. *arXiv:1811.07491 [cs]*. Retrieved May 14, 2020, from <http://arxiv.org/abs/1811.07491>
- Filippi, M., Preziosa, P., Banwell, B. L., Barkhof, F., Ciccarelli, O., De Stefano, N., Geurts, J. J. G., Paul, F., Reich, D. S., Toosy, A. T., Traboulsee, A., Wattjes, M. P., Yousry, T. A., Gass, A., Lubetzki, C., Weinshenker, B. G., & Rocca, M. A. (2019). Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. *Brain*, 142(7), 1858–1875. <https://doi.org/10.1093/brain/awz144>
- Filippi, M., Rocca, M. A., Ciccarelli, O., De Stefano, N., Evangelou, N., Kappos, L., Rovira, A., Sastre-Garriga, J., Tintorè, M., Frederiksen, J. L., Gasperini, C., Palace, J., Reich, D. S., Banwell, B., Montalban, X., & Barkhof, F. (2016). MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. *The Lancet Neurology*, 15(3), 292–303. [https://doi.org/10.1016/S1474-4422\(15\)00393-2](https://doi.org/10.1016/S1474-4422(15)00393-2)
- Finck, T., Li, H., Grundl, L., Eichinger, P., Bussas, M., Mühlau, M., Menze, B., & Wiestler, B. (2020). Deep-Learning Generated Synthetic Double Inversion Recovery Images Improve Multiple Sclerosis Lesion Detection. *Investigative Radiology*, 55(5), 318–323. <https://doi.org/10.1097/RLI.0000000000000640>
- Fischl, B. (2012). FreeSurfer. *NeuroImage*, 62(2), 774–781. <https://doi.org/10.1016/j.neuroimage.2012.01.021>
- Fox, R. J., Beall, E., Bhattacharyya, P., Chen, J. T., & Sakaie, K. (2011). Advanced MRI in Multiple Sclerosis: Current Status and Future Challenges. *Neurologic Clinics*, 29(2), 357–380. <https://doi.org/10.1016/j.ncl.2010.12.011>
- Frangi, A. F., Niessen, W. J., Vincken, K. L., & Viergever, M. A. (1998). Multiscale vessel enhancement filtering [Series Title: Lecture Notes in Computer Science]. *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98* (pp. 130–137). Springer Berlin Heidelberg. <https://doi.org/10.1007/BFb0056195>

- Frischer, J. M., Weigand, S. D., Guo, Y., Kale, N., Parisi, J. E., Pirko, I., Mandrekar, J., Bramow, S., Metz, I., Brück, W., Lassmann, H., & Lucchinetti, C. F. (2015). Clinical and pathological insights into the dynamic nature of the white matter multiple sclerosis plaque: Dynamic Nature of MS Plaque. *Annals of Neurology*, 78(5), 710–721. <https://doi.org/10.1002/ana.24497>
- Gabr, R. E., Coronado, I., Robinson, M., Sujit, S. J., Datta, S., Sun, X., Allen, W. J., Lublin, F. D., Wolinsky, J. S., & Narayana, P. A. (2020). Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: A large-scale study. *Multiple Sclerosis Journal*, 26(10), 1217–1226. <https://doi.org/10.1177/1352458519856843>
- Geurts, J. J. G., Bo, L., Pouwels, P. J. W., Castelijns, J. A., Polman, C. H., & Barkhof, F. (2005). Cortical Lesions in Multiple Sclerosis: Combined Postmortem MR Imaging and Histopathology, 6.
- Geurts, J. J. G., Pouwels, P. J. W., Uitdehaag, B. M. J., Polman, C. H., Barkhof, F., & Castelijns, J. A. (2005). Intracortical Lesions in Multiple Sclerosis: Improved Detection with 3D Double Inversion-Recovery MR Imaging. *Radiology*, 236(1), 254–260. <https://doi.org/10.1148/radiol.2361040450>
- Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D. I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., Whyntie, T., Nachev, P., Modat, M., Barratt, D. C., Ourselin, S., Cardoso, M. J., & Vercauteren, T. (2018). NiftyNet: a deep-learning platform for medical imaging. *Computer Methods and Programs in Biomedicine*, 158, 113–122. <https://doi.org/10.1016/j.cmpb.2018.01.025>
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks [arXiv: 1406.2661]. *arXiv:1406.2661 [cs, stat]*. Retrieved May 5, 2020, from <http://arxiv.org/abs/1406.2661>
- Grahl, S., Pongratz, V., Schmidt, P., Engl, C., Bussas, M., Radetz, A., Gonzalez-Escamilla, G., Groppa, S., Zipp, F., Lukas, C., Kirschke, J., Zimmer, C., Hoshi, M., Berthele, A., Hemmer, B., & Mühlau, M. (2019). Evidence for a white matter lesion size threshold to support the diagnosis of relapsing remitting multiple sclerosis. *Multiple Sclerosis and Related Disorders*, 29, 124–129. <https://doi.org/10.1016/j.msard.2019.01.042>
- Haacke, E., Mittal, S., Wu, Z., Neelavalli, J., & Cheng, Y.-C. (2009). Susceptibility-Weighted Imaging: Technical Aspects and Clinical Applications, Part 1. *American Journal of Neuroradiology*, 30(1), 19–30. <https://doi.org/10.3174/ajnr.A1400>
- Hagiwara, A., Kamagata, K., & Aoki, S. (2020). Image Domain Transfer by Deep Learning is Feasible in Multiple Sclerosis Clinical Practice: *Investigative Radiology*, 55(5), 324–325. <https://doi.org/10.1097/RLI.0000000000000665>
- Harrison, D. M., Roy, S., Oh, J., Izbudak, I., Pham, D., Courtney, S., Caffo, B., Jones, C. K., van Zijl, P., & Calabresi, P. A. (2015). Association of Cortical Lesion Burden on 7-T Magnetic Resonance Imaging With Cognition and Disability in Multiple Sclerosis. *JAMA Neurology*, 72(9), 1004. <https://doi.org/10.1001/jamaneurol.2015.1241>
- Hashemi, S. R., Mohseni Salehi, S. S., Erdogmus, D., Prabhu, S. P., Warfield, S. K., & Gholipour, A. (2019). Asymmetric Loss Functions and Deep Densely-Connected Networks for Highly-Imbalanced Medical Image Segmentation: Application to Multiple Sclerosis Lesion Detection. *IEEE Access*, 7, 1721–1735. <https://doi.org/10.1109/ACCESS.2018.2886371>



## Bibliography

---

- Hemond, C. C., & Bakshi, R. (2018). Magnetic Resonance Imaging in Multiple Sclerosis. *Cold Spring Harbor Perspectives in Medicine*, 8(5), a028969. <https://doi.org/10.1101/cshperspect.a028969>
- hu, C., Kang, G., Hou, B., Ma, Y., Labeau, F., & Su, Z. (2020). Acu-Net: A 3D Attention Context U-Net for Multiple Sclerosis Lesion Segmentation. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1384–1388. <https://doi.org/10.1109/ICASSP40776.2020.9054616>
- Ineichen, B. V., Beck, E. S., Piccirelli, M., & Reich, D. S. (2021). New Prospects for Ultra-High-Field Magnetic Resonance Imaging in Multiple Sclerosis. *Investigative Radiology*, 56(11), 773–784. <https://doi.org/10.1097/RLI.0000000000000804>
- Interrater reliability: the kappa statistic. (2012). Retrieved October 5, 2021, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/>
- Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2), 203–211. <https://doi.org/10.1038/s41592-020-01008-z>
- Isensee, F., & Maier-Hein, K. H. (2019). An attempt at beating the 3D U-Net [arXiv: 1908.02182]. *arXiv:1908.02182 [cs, eess]*. Retrieved May 14, 2020, from <http://arxiv.org/abs/1908.02182>
- Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2018). Image-to-Image Translation with Conditional Adversarial Networks [arXiv: 1611.07004]. *arXiv:1611.07004 [cs]*. Retrieved May 5, 2020, from <http://arxiv.org/abs/1611.07004>
- Jacobs, B. M., Giovannoni, G., Cuzick, J., & Dobson, R. (2020). Systematic review and meta-analysis of the association between Epstein–Barr virus, multiple sclerosis and other risk factors. *Multiple Sclerosis Journal*, 26(11), 1281–1297. <https://doi.org/10.1177/1352458520907901>
- Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, 62(2), 782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual Losses for Real-Time Style Transfer and Super-Resolution [arXiv: 1603.08155]. *arXiv:1603.08155 [cs]*. Retrieved May 28, 2020, from <http://arxiv.org/abs/1603.08155>
- Kau, T., Taschwer, M., Deutschmann, H., Schönfelder, M., Weber, J. R., & Hausegger, K. A. (2013). The “central vein sign”: is there a place for susceptibility weighted imaging in possible multiple sclerosis? *European Radiology*, 23(7), 1956–1962. <https://doi.org/10.1007/s00330-013-2791-4>
- Kaunzner, U. W., Kang, Y., Zhang, S., Morris, E., Yao, Y., Pandya, S., Hurtado Rua, S. M., Park, C., Gillen, K. M., Nguyen, T. D., Wang, Y., Pitt, D., & Gauthier, S. A. (2019). Quantitative susceptibility mapping identifies inflammation in a subset of chronic multiple sclerosis lesions. *Brain*, 142(1), 133–145. <https://doi.org/10.1093/brain/awy296>
- Kaur, A., Kaur, L., & Singh, A. (2020). State-of-the-Art Segmentation Techniques and Future Directions for Multiple Sclerosis Brain Lesions. *Archives of Computational Methods in Engineering*. <https://doi.org/10.1007/s11831-020-09403-7>
- Kaur, A., Kaur, L., & Singh, A. (2021). State-of-the-Art Segmentation Techniques and Future Directions for Multiple Sclerosis Brain Lesions. *Archives of Computational Methods in Engineering*, 28(3), 951–977. <https://doi.org/10.1007/s11831-020-09403-7>

- Kazemini, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., & Mukhopadhyay, A. (2019). GANs for Medical Image Analysis [arXiv: 1809.06222]. *arXiv:1809.06222 [cs, stat]*. Retrieved May 5, 2020, from <http://arxiv.org/abs/1809.06222>
- Keough, M. B., & Yong, V. W. (2013). Remyelination Therapy for Multiple Sclerosis. *Neurotherapeutics*, 10(1), 44–54. <https://doi.org/10.1007/s13311-012-0152-7>
- Kidd, D., Barkhof, F., McConnell, R., Algra, P. R., Allen, I. V., & Revesz, T. (1999). Cortical lesions in multiple sclerosis. *Brain*, 122(1), 17–26. <https://doi.org/10.1093/brain/122.1.17>
- Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization [arXiv: 1412.6980]. *arXiv:1412.6980 [cs]*. Retrieved July 27, 2020, from <http://arxiv.org/abs/1412.6980>
- Klein, S., Staring, M., Murphy, K., Viergever, M., & Pluim, J. (2010). Elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging*, 29(1), 196–205. <https://doi.org/10.1109/TMI.2009.2035616>
- Kober, T., Granziera, C., Ribes, D., Browaeys, P., Schluep, M., Meuli, R., Frackowiak, R., Gruetter, R., & Krueger, G. (2012). MP2RAGE multiple sclerosis magnetic resonance imaging at 3 T. *Investigative Radiology*, 47(6), 346–352. <https://doi.org/10.1097/RLI.0b013e31824600e9>
- Koch-Henriksen, N., & Sørensen, P. S. (2010). The changing demographic pattern of multiple sclerosis epidemiology. *The Lancet Neurology*, 9(5), 520–532. [https://doi.org/10.1016/S1474-4422\(10\)70064-8](https://doi.org/10.1016/S1474-4422(10)70064-8)
- Kolb, H., Absinta, M., Beck, E. S., Ha, S.-K., Song, Y., Norato, G., Cortese, I., Sati, P., Nair, G., & Reich, D. S. (2021). 7T MRI Differentiates Remyelinated from Demyelinated Multiple Sclerosis Lesions. *Annals of Neurology*, 90(4), 612–626. <https://doi.org/10.1002/ana.26194>
- Kuhlmann, T., Ludwin, S., Prat, A., Antel, J., Brück, W., & Lassmann, H. (2017). An updated histological classification system for multiple sclerosis lesions. *Acta Neuropathologica*, 133(1), 13–24. <https://doi.org/10.1007/s00401-016-1653-y>
- Kumar, A., Murthy, O. N., Shrish, Ghosal, P., Mukherjee, A., & Nandi, D. (2019). A Dense U-Net Architecture for Multiple Sclerosis Lesion Segmentation [ISSN: 2159-3450]. *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, 662–667. <https://doi.org/10.1109/TENCON.2019.8929615>
- Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., & Matas, J. (2018). DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks [arXiv: 1711.07064]. *arXiv:1711.07064 [cs]*. Retrieved May 7, 2020, from <http://arxiv.org/abs/1711.07064>
- Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, 33(11), 1444–1444. <https://doi.org/10.1212/WNL.33.11.1444>
- La Rosa, F., Abdulkadir, A., Fartaria, M. J., Rahmanzadeh, R., Lu, P.-J., Galbusera, R., Barakovic, M., Thiran, J.-P., Granziera, C., & Bach Cuadra, M. (2020). Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE. *NeuroImage: Clinical*, 102335. <https://doi.org/10.1016/j.nicl.2020.102335>
- La Rosa, F., Beck, E. S., Abdulkadir, A., Thiran, J.-P., Reich, D. S., Sati, P., & Bach Cuadra, M. (2020). Automated Detection of Cortical Lesions in Multiple Sclerosis Patients with 7T MRI [Series Title: Lecture Notes in Computer Science]. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, 12264, 584–593. [https://doi.org/10.1007/978-3-030-59719-1\\_57](https://doi.org/10.1007/978-3-030-59719-1_57)



## Bibliography

---

- La Rosa, F., Fartaria, M. J., Abdulkadir, A., Granziera, C., Thiran, J.-P., & Cuadra, M. B. (2019). Deep learning-based detection of cortical lesions in multiple sclerosis patients with FLAIR, DIR, and MP2RAGE MRI sequences. Retrieved October 31, 2021, from <https://onlinelibrary.ectrims-congress.eu/ectrims/2019/stockholm/278829/francesco.la.rosa.deep.learning-based.detection.of.cortical.lesions.in.html>
- La Rosa, F., Yu, T., Barquero, G., Thiran, J.-P., Granziera, C., & Bach Cuadra, M. (2021). MPRAGE to MP2RAGE UNI translation via generative adversarial network improves the automatic tissue and lesion segmentation in multiple sclerosis patients. *Computers in Biology and Medicine*, 132, 104297. <https://doi.org/10.1016/j.compbimed.2021.104297>
- Lee, D., Moon, W.-J., & Ye, J. C. (2020). Assessing the importance of magnetic resonance contrasts using collaborative generative adversarial networks. *Nature Machine Intelligence*, 2(1), 34–42. <https://doi.org/10.1038/s42256-019-0137-x>
- Lei, Y., Harms, J., Wang, T., Liu, Y., Shu, H.-K., Jani, A. B., Curran, W. J., Mao, H., Liu, T., & Yang, X. (2019). MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Medical Physics*, mp.13617. <https://doi.org/10.1002/mp.13617>
- Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Magazine*, 37(3), 50–60. <https://doi.org/10.1109/MSP.2020.2975749>
- Lin, L. I.-K. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, 45(1), 255. <https://doi.org/10.2307/2532051>
- Lladó, X., Oliver, A., Cabezas, M., Freixenet, J., Vilanova, J. C., Quiles, A., Valls, L., Ramió-Torrentà, L., & Rovira, À. (2012). Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. *Information Sciences*, 186(1), 164–185. <https://doi.org/10.1016/j.ins.2011.10.011>
- Lou, C., Sati, P., Absinta, M., Clark, K., Dworkin, J. D., Valcarcel, A. M., Schindler, M. K., Reich, D. S., Sweeney, E. M., & Shinohara, R. T. (2021). Fully Automated Detection of Paramagnetic Rims in Multiple Sclerosis Lesions on 3T Susceptibility-Based MR Imaging. *NeuroImage: Clinical*, 102796. <https://doi.org/10.1016/j.nicl.2021.102796>
- LST – Lesion segmentation for SPM | Paul Schmidt – freelance statistician. (2018). Retrieved May 14, 2020, from <https://www.applied-statistics.de/lst.html>
- Lu, P.-J., Barakovic, M., Weigel, M., Rahmanzadeh, R., Galbusera, R., Schiavi, S., Daducci, A., La Rosa, F., Bach Cuadra, M., Sandkühler, R., Kuhle, J., Kappos, L., Cattin, P., & Granziera, C. (2021). GAMER-MRI in Multiple Sclerosis Identifies the Diffusion-Based Microstructural Measures That Are Most Sensitive to Focal Damage: A Deep-Learning-Based Analysis and Clinico-Biological Validation. *Frontiers in Neuroscience*, 15, 647535. <https://doi.org/10.3389/fnins.2021.647535>
- Lu, P.-J., Yoo, Y., Rahmanzadeh, R., Galbusera, R., Weigel, M., Ceccaldi, P., Nguyen, T. D., Spincemaille, P., Wang, Y., Daducci, A., La Rosa, F., Bach Cuadra, M., Sandkühler, R., Nael, K., Doshi, A., Fayad, Z. A., Kuhle, J., Kappos, L., Odry, B., ... Granziera, C. (2021). GAMER MRI: Gated-attention mechanism ranking of multi-contrast MRI in brain pathology. *NeuroImage: Clinical*, 29, 102522. <https://doi.org/10.1016/j.nicl.2020.102522>

- Lublin, F. D., Reingold, S. C., Cohen, J. A., Cutter, G. R., Sorensen, P. S., Thompson, A. J., Wolinsky, J. S., Balcer, L. J., Banwell, B., Barkhof, F., Bebo, B., Calabresi, P. A., Clanet, M., Comi, G., Fox, R. J., Freedman, M. S., Goodman, A. D., Inglese, M., Kappos, L., ... Polman, C. H. (2014). Defining the clinical course of multiple sclerosis: The 2013 revisions. *Neurology*, 83(3), 278–286. <https://doi.org/10.1212/WNL.0000000000000560>
- Luchetti, S., Fransen, N. L., van Eden, C. G., Ramaglia, V., Mason, M., & Huitinga, I. (2018). Progressive multiple sclerosis patients show substantial lesion activity that correlates with clinical disease severity and sex: a retrospective autopsy cohort analysis. *Acta Neuropathologica*, 135(4), 511–528. <https://doi.org/10.1007/s00401-018-1818-y>
- Madsen, M. A., Wiggermann, V., Bramow, S., Christensen, J. R., Sellebjerg, F., & Siebner, H. R. (2021). *Imaging cortical multiple sclerosis lesions with ultra-high field MRI* (preprint). Radiology and Imaging. <https://doi.org/10.1101/2021.06.25.21259363>
- Maggi, P., Absinta, M., Grammatico, M., Vuolo, L., Emmi, G., Carlucci, G., Spagni, G., Barilaro, A., Repice, A. M., Emmi, L., Prisco, D., Martinelli, V., Scotti, R., Sadeghi, N., Perrotta, G., Sati, P., Dachy, B., Reich, D. S., Filippi, M., & Massacesi, L. (2018). Central vein sign differentiates Multiple Sclerosis from central nervous system inflammatory vasculopathies: Central Vein Sign. *Annals of Neurology*, 83(2), 283–294. <https://doi.org/10.1002/ana.25146>
- Maggi, P., Fartaria, M. J., Jorge, J., La Rosa, F., Absinta, M., Sati, P., Meuli, R., Du Pasquier, R., Reich, D. S., Cuadra, M. B., Granziera, C., Richiardi, J., & Kober, T. (2020). CVSnet: A machine learning approach for automated central vein sign assessment in multiple sclerosis. *NMR in Biomedicine*, 33(5). <https://doi.org/10.1002/nbm.4283>
- Maggi, P., Kuhle, J., Schädelin, S., van der Meer, F., Weigel, M., Galbusera, R., Mathias, A., Lu, P.-J., Rahmzadeh, R., Benkert, P., La Rosa, F., Bach Cuadra, M., Sati, P., Théaudin, M., Pot, C., van Pesch, V., Leppert, D., Stadelmann, C., Kappos, L., ... Granziera, C. (2021). Chronic White Matter Inflammation and Serum Neurofilament Levels in Multiple Sclerosis. *Neurology*, 97(6), e543–e553. <https://doi.org/10.1212/WNL.00000000000012326>
- Maggi, P., Sati, P., Nair, G., Cortese, I. C. M., Jacobson, S., Smith, B. R., Nath, A., Ohayon, J., Pesch, V., Perrotta, G., Pot, C., Théaudin, M., Martinelli, V., Scotti, R., Wu, T., Du Pasquier, R., Calabresi, P. A., Filippi, M., Reich, D. S., & Absinta, M. (2020). Paramagnetic Rim Lesions are Specific to Multiple Sclerosis: An International Multicenter 3T MRI Study. *Annals of Neurology*, 88(5), 1034–1042. <https://doi.org/10.1002/ana.25877>
- Magliozzi, R., Reynolds, R., & Calabrese, M. (2018). MRI of cortical lesions and its use in studying their role in MS pathogenesis and disease course. *Brain Pathology*, 28(5), 735–742. <https://doi.org/10.1111/bpa.12642>
- Maranzano, J., Dadar, M., Rudko, D., De Nigris, D., Elliott, C., Gati, J., Morrow, S., Menon, R., Collins, D., Arnold, D., & Narayanan, S. (2019). Comparison of Multiple Sclerosis Cortical Lesion Types Detected by Multicontrast 3T and 7T MRI. *American Journal of Neuroradiology*, 40(7), 1162–1169. <https://doi.org/10.3174/ajnr.A6099>

## Bibliography

---

- Marques, J. P., Kober, T., Krueger, G., van der Zwaag, W., Van de Moortele, P.-F., & Gruetter, R. (2010). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *NeuroImage*, 49(2), 1271–1281. <https://doi.org/10.1016/j.neuroimage.2009.10.002>
- Marstal, K., Berendsen, F., Staring, M., & Klein, S. (2016). SimpleElastix: A User-Friendly, Multi-lingual Library for Medical Image Registration [ISSN: 2160-7516]. *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 574–582. <https://doi.org/10.1109/CVPRW.2016.78>
- Mattes, D., Haynor, D. R., Vesselle, H., Lewellyn, T. K., & Eubank, W. (2001). Nonrigid multimodality image registration. In M. Sonka & K. M. Hanson (Eds.). <https://doi.org/10.1117/12.431046>
- Mazurowski, M. A., Buda, M., Saha, A., & Bashir, M. R. (2019). Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *Journal of Magnetic Resonance Imaging*, 49(4), 939–954. <https://doi.org/10.1002/jmri.26534>
- McDonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H.-P., Lublin, F. D., McFarland, H. F., Paty, D. W., Polman, C. H., Reingold, S. C., Sandberg-Wollheim, M., Sibley, W., Thompson, A., Van Den Noort, S., Weinshenker, B. Y., & Wolinsky, J. S. (2001). Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Annals of Neurology*, 50(1), 121–127. <https://doi.org/10.1002/ana.1032>
- McKinley, R., Wepfer, R., Aschwanden, F., Grunder, L., Muri, R., Rummel, C., Verma, R., Weisstanner, C., Reyes, M., Salmen, A., Chan, A., Wagner, F., & Wiest, R. (2021). Simultaneous lesion and brain segmentation in multiple sclerosis using deep neural networks. *Scientific Reports*, 11(1), 1087. <https://doi.org/10.1038/s41598-020-79925-4>
- Mistry, N., Abdel-Fahim, R., Samaraweera, A., Mougin, O., Tallantyre, E., Tench, C., Jaspan, T., Morris, P., Morgan, P. S., & Evangelou, N. (2016). Imaging central veins in brain lesions with 3-T T2\*-weighted magnetic resonance imaging differentiates multiple sclerosis from microangiopathic brain lesions. *Multiple Sclerosis Journal*, 22(10), 1289–1296. <https://doi.org/10.1177/1352458515616700>
- Montavon, G., Samek, W., & Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73, 1–15. <https://doi.org/10.1016/j.dsp.2017.10.011>
- Mugler, J. P., & Brookeman, J. R. (1990). Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magnetic Resonance in Medicine*, 15(1), 152–157. <https://doi.org/10.1002/mrm.1910150117>
- Mukhoti, J., Kulharia, V., Sanyal, A., Golodetz, S., Torr, P. H. S., & Dokania, P. K. (2020). Calibrating Deep Neural Networks using Focal Loss [arXiv: 2002.09437]. *arXiv:2002.09437 [cs, stat]*. Retrieved September 14, 2021, from <http://arxiv.org/abs/2002.09437>
- Mussard, E., Hilbert, T., Forman, C., Meuli, R., Thiran, J.-P., & Kober, T. (2020). Accelerated MP2RAGE imaging using Cartesian phyllotaxis readout and compressed sensing reconstruction. *Magnetic Resonance in Medicine*, 84(4), 1881–1894. <https://doi.org/10.1002/mrm.28244>
- Myronenko, A. (2018). 3D MRI brain tumor segmentation using autoencoder regularization [arXiv: 1810.11654]. *arXiv:1810.11654 [cs, q-bio]*. Retrieved May 15, 2020, from <http://arxiv.org/abs/1810.11654>

- Nair, T., Precup, D., Arnold, D. L., & Arbel, T. (2020). Exploring uncertainty measures in deep networks for Multiple sclerosis lesion detection and segmentation. *Medical Image Analysis*, 59, 101557. <https://doi.org/10.1016/j.media.2019.101557>
- Ng Kee Kwong, K. C., Mollison, D., Meijboom, R., York, E. N., Kampaite, A., Thrippleton, M. J., Chandran, S., & Waldman, A. D. (2021). The prevalence of paramagnetic rim lesions in multiple sclerosis: A systematic review and meta-analysis (Q. Jiang, Ed.). *PLOS ONE*, 16(9), e0256845. <https://doi.org/10.1371/journal.pone.0256845>
- Nielsen, A. S., Kinkel, R. P., Madigan, N., Tinelli, E., Benner, T., & Mainero, C. (2013). Contribution of cortical lesion subtypes at 7T MRI to physical and cognitive performance in MS. *Neurology*, 81(7), 641–649. <https://doi.org/10.1212/WNL.0b013e3182a08ce8>
- Nielsen, A. S., Kinkel, R. P., Tinelli, E., Benner, T., Cohen-Adad, J., & Mainero, C. (2012). Focal cortical lesion detection in multiple sclerosis: 3 tesla DIR versus 7 tesla FLASH-T2\*. *Journal of Magnetic Resonance Imaging*, 35(3), 537–542. <https://doi.org/10.1002/jmri.22847>
- Oh, J., Vidal-Jordana, A., & Montalban, X. (2018). Multiple sclerosis: clinical aspects. *Current Opinion in Neurology*, 31(6), 752–759. <https://doi.org/10.1097/WCO.0000000000000622>
- Okubo, G., Okada, T., Yamamoto, A., Kanagaki, M., Fushimi, Y., Okada, T., Murata, K., & Togashi, K. (2016). MP2RAGE for deep gray matter measurement of the brain: A comparative study with MPRAGE: MP2RAGE for Deep Gray Matter Measurement. *Journal of Magnetic Resonance Imaging*, 43(1), 55–62. <https://doi.org/10.1002/jmri.24960>
- Okuda, D. T., Mowry, E. M., Beheshtian, A., Waubant, E., Baranzini, S. E., Goodin, D. S., Hauser, S. L., & Pelletier, D. (2009). Incidental MRI anomalies suggestive of multiple sclerosis: The radiologically isolated syndrome. *Neurology*, 72(9), 800–805. <https://doi.org/10.1212/01.wnl.0000335764.14513.1a>
- Pearce, J. (2005). Historical Descriptions of Multiple Sclerosis. *European Neurology*, 54(1), 49–53. <https://doi.org/10.1159/000087387>
- Pirko, I., Lucchinetti, C. F., Sriram, S., & Bakshi, R. (2007). Gray matter involvement in multiple sclerosis. *Neurology*, 68(9), 634–642. <https://doi.org/10.1212/01.wnl.0000250267.85698.7a>
- Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F. D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A. J., Waubant, E., Weinshenker, B., & Wolinsky, J. S. (2011). Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Annals of Neurology*, 69(2), 292–302. <https://doi.org/10.1002/ana.22366>
- Poser, C. M., & Brinar, V. V. (2004). Diagnostic criteria for multiple sclerosis: an historical review. *Clinical Neurology and Neurosurgery*, 106(3), 147–158. <https://doi.org/10.1016/j.clineuro.2004.02.004>
- Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks [arXiv: 1511.06434]. *arXiv:1511.06434 [cs]*. Retrieved June 3, 2020, from <http://arxiv.org/abs/1511.06434>
- Rahmanzadeh, R., Lu, P.-J., Barakovic, M., Weigel, M., Maggi, P., Nguyen, T. D., Schiavi, S., Daducci, A., La Rosa, F., Schaedel, S., Absinta, M., Reich, D. S., Sati, P., Wang, Y., Bach Cuadra, M., Radue, E.-W., Kuhle, J., Kappos, L., & Granziera, C. (2021). Myelin and axon pathology in multiple

## Bibliography

---

- sclerosis assessed by myelin water and multi-shell diffusion imaging. *Brain*, 144(6), 1684–1696. <https://doi.org/10.1093/brain/awab088>
- Reich, D. S., Lucchinetti, C. F., & Calabresi, P. A. (2018). Multiple Sclerosis (D. L. Longo, Ed.). *New England Journal of Medicine*, 378(2), 169–180. <https://doi.org/10.1056/NEJMra1401483>
- Reyes, M., Meier, R., Pereira, S., Silva, C. A., Dahlweid, F.-M., Tengg-Kobligk, H. v., Summers, R. M., & Wiest, R. (2020). On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. *Radiology: Artificial Intelligence*, 2(3), e190043. <https://doi.org/10.1148/ryai.2020190043>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation [Series Title: Lecture Notes in Computer Science]. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer International Publishing. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- Roy, S., Butman, J. A., Reich, D. S., Calabresi, P. A., & Pham, D. L. (2018). Multiple Sclerosis Lesion Segmentation from Brain MRI via Fully Convolutional Neural Networks [arXiv: 1803.09172]. *arXiv:1803.09172 [cs]*. Retrieved May 14, 2020, from <http://arxiv.org/abs/1803.09172>
- Sajjadi, M. S. M., Schölkopf, B., & Hirsch, M. (2017). EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis [arXiv: 1612.07919]. *arXiv:1612.07919 [cs, stat]*. Retrieved May 28, 2020, from <http://arxiv.org/abs/1612.07919>
- Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks [Series Title: Lecture Notes in Computer Science]. *Machine Learning in Medical Imaging* (pp. 379–387). Springer International Publishing. [https://doi.org/10.1007/978-3-319-67389-9\\_44](https://doi.org/10.1007/978-3-319-67389-9_44)
- Salem, M., Valverde, S., Cabezas, M., Pareto, D., Oliver, A., Salvi, J., Rovira, À., & Lladó, X. (2019). Multiple Sclerosis Lesion Synthesis in MRI using an encoder-decoder U-NET [arXiv: 1901.05733]. *arXiv:1901.05733 [cs]*. Retrieved May 5, 2020, from <http://arxiv.org/abs/1901.05733>
- Sati, P., Thomasson, D., Li, N., Pham, D., Biassou, N., Reich, D., & Butman, J. (2014). Rapid, high-resolution, whole-brain, susceptibility-based MRI of multiple sclerosis. *Multiple Sclerosis Journal*, 20(11), 1464–1470. <https://doi.org/10.1177/1352458514525868>
- Sati, P., George, I. C., Shea, C. D., Gaitán, M. I., & Reich, D. S. (2012). FLAIR\*: A Combined MR Contrast Technique for Visualizing White Matter Lesions and Parenchymal Veins. *Radiology*, 265(3), 926–932. <https://doi.org/10.1148/radiol.12120208>
- Sati, P., Oh, J., Constable, R. T., Evangelou, N., Guttmann, C. R. G., Henry, R. G., Klawiter, E. C., Mainero, C., Massacesi, L., McFarland, H., Nelson, E., Ontaneda, D., Rauscher, A., Rooney, W. D., Samaraweera, A. P. R., Shinohara, R. T., Sobel, R. A., Solomon, A. J., Treaba, C. A., ... Reich, D. S. (2016). The central vein sign and its clinical evaluation for the diagnosis of multiple sclerosis: a consensus statement from the North American Imaging in Multiple Sclerosis Cooperative. *Nature Reviews Neurology*, 12(12), 714–722. <https://doi.org/10.1038/nrneurol.2016.166>

- Sati, P., & Reich, D. S. (2021). 7 Tesla MRI will soon be helpful to guide clinical practice in multiple sclerosis centers – Commentary. *Multiple Sclerosis Journal*, 27(3), 364–365. <https://doi.org/10.1177/1352458520981738>
- Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V. J., Zimmer, C., Hemmer, B., & Mührlau, M. (2012). An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *NeuroImage*, 59(4), 3774–3783. <https://doi.org/10.1016/j.neuroimage.2011.11.032>
- SciPy 1.0 Contributors, Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., ... van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Shende, P., Pawar, M., & Kakde, S. (2019). A Brief Review on: MRI Images Reconstruction using GAN. *2019 International Conference on Communication and Signal Processing (ICCSP)*, 0139–0142. <https://doi.org/10.1109/ICCSP2019.8698083>
- Shiee, N., Bazin, P.-L., & Pham, D. L. (2008). Multiple Sclerosis Lesion Segmentation Using Statistical and Topological Atlases, 10.
- Shin, H.-C., Tenenholtz, N. A., Rogers, J. K., Schwarz, C. G., Senjem, M. L., Gunter, J. L., Andriole, K., & Michalski, M. (2018). Medical Image Synthesis for Data Augmentation and Anonymization using Generative Adversarial Networks [arXiv: 1807.10225]. *arXiv:1807.10225 [cs, stat]*. Retrieved May 5, 2020, from <http://arxiv.org/abs/1807.10225>
- Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition [arXiv: 1409.1556]. *arXiv:1409.1556 [cs]*. Retrieved May 7, 2020, from <http://arxiv.org/abs/1409.1556>
- Solomon, A. J., & Corboy, J. R. (2017). The tension between early diagnosis and misdiagnosis of multiple sclerosis. *Nature Reviews Neurology*, 13(9), 567–572. <https://doi.org/10.1038/nrneurol.2017.106>
- Solomon, A. J., Naismith, R. T., & Cross, A. H. (2019). Misdiagnosis of multiple sclerosis: Impact of the 2017 McDonald criteria on clinical practice. *Neurology*, 92(1), 26–33. <https://doi.org/10.1212/WNL.0000000000006583>
- Solomon, A. J., Schindler, M. K., Howard, D. B., Watts, R., Sati, P., Nickerson, J. P., & Reich, D. S. (2016). Central vessel sign” on 3T FLAIR\* for the differentiation of multiple sclerosis from migraine. *Annals of Clinical and Translational Neurology*, 3(2), 82–87. <https://doi.org/10.1002/acn3.273>
- Solomon, A. J., Watts, R., Ontaneda, D., Absinta, M., Sati, P., & Reich, D. S. (2018). Diagnostic performance of central vein sign for multiple sclerosis with a simplified three-lesion algorithm. *Multiple Sclerosis Journal*, 24(6), 750–757. <https://doi.org/10.1177/1352458517726383>
- Souplet, J.-C., Lebrun, C., Ayache, N., & Malandain, G. (2010). An Automatic Segmentation of T2-FLAIR Multiple Sclerosis Lesions. *10/2006*, 12.



## Bibliography

---

- Springer, E., Dymerska, B., Cardoso, P. L., Robinson, S. D., Weisstanner, C., Wiest, R., Schmitt, B., & Tractnig, S. (2016). Comparison of Routine Brain Imaging at 3 T and 7 T. *Investigative Radiology*, 51(8), 469–482. <https://doi.org/10.1097/RLI.0000000000000256>
- Styner, M., Lee, J., Chin, B., Chin, M. S., Commowick, O., Tran, H.-H., Jewells, V., & Warfield, S. (2009). 3D Segmentation in the Clinic: A Grand Challenge II: MS lesion segmentation, 6.
- Suh, C. H., Kim, S. J., Jung, S. C., Choi, C. G., & Kim, H. S. (2019). The “Central Vein Sign” on T2\*-weighted Images as a Diagnostic Tool in Multiple Sclerosis: A Systematic Review and Meta-analysis using Individual Patient Data. *Scientific Reports*, 9(1), 18188. <https://doi.org/10.1038/s41598-019-54583-3>
- Tallantyre, E. C., Dixon, J. E., Donaldson, I., Owens, T., Morgan, P. S., Morris, P. G., & Evangelou, N. (2011). Ultra-high-field imaging distinguishes MS lesions from asymptomatic white matter lesions. *Neurology*, 76(6), 534–539. <https://doi.org/10.1212/WNL.0b013e31820b7630>
- Tallantyre, E. C., Morgan, P. S., Dixon, J. E., Al-Radaideh, A., Brookes, M. J., Evangelou, N., & Morris, P. G. (2009). A Comparison of 3T and 7T in the Detection of Small Parenchymal Veins Within MS Lesions: *Investigative Radiology*, 44(9), 491–494. <https://doi.org/10.1097/RLI.0b013e3181b4c144>
- Tam, R. C., Traboulsee, A., Riddehough, A., & Li, D. K. (2012). Improving the clinical correlation of multiple sclerosis black hole volume change by paired-scan analysis. *NeuroImage: Clinical*, 1(1), 29–36. <https://doi.org/10.1016/j.nicl.2012.08.004>
- Tanner, M., Gambarota, G., Kober, T., Krueger, G., Erritzoe, D., Marques, J. P., & Newbould, R. (2012). Fluid and white matter suppression with the MP2RAGE sequence. *Journal of Magnetic Resonance Imaging*, 35(5), 1063–1070. <https://doi.org/10.1002/jmri.23532>
- Tardif, C. L., Collins, D. L., Eskildsen, S. E., Richardson, J. B., & Pike, G. B. (2010). Segmentation of Cortical MS Lesions on MRI Using Automated Lamina Profile Shape Analysis [Series Title: Lecture Notes in Computer Science]. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010* (pp. 181–188). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-15711-0\\_23](https://doi.org/10.1007/978-3-642-15711-0_23)
- Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., ... Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, 17(2), 162–173. [https://doi.org/10.1016/S1474-4422\(17\)30470-2](https://doi.org/10.1016/S1474-4422(17)30470-2)
- Thompson, A. J., Baranzini, S. E., Geurts, J., Hemmer, B., & Ciccarelli, O. (2018). Multiple sclerosis. *The Lancet*, 391(10130), 1622–1636. [https://doi.org/10.1016/S0140-6736\(18\)30481-1](https://doi.org/10.1016/S0140-6736(18)30481-1)
- Tomas-Fernandez, X., & Warfield, S. K. (2015). A Model of Population and Subject (MOPS) Intensities With Application to Multiple Sclerosis Lesion Segmentation. *IEEE Transactions on Medical Imaging*, 34(6), 1349–1361. <https://doi.org/10.1109/TMI.2015.2393853>
- Trapp, B. D., Ransohoff, R., & Rudick, R. (1999). Axonal pathology in multiple sclerosis: relationship to neurologic disability. *Current Opinion in Neurology*, 12(3), 295–302. <https://doi.org/10.1097/00019052-199906000-00008>

- Treaba, C. A., Conti, A., Klawiter, E. C., Barletta, V. T., Herranz, E., Mehndiratta, A., Russo, A. W., Sloane, J. A., Kinkel, R. P., Toschi, N., & Mainero, C. (2021). Cortical and phase rim lesions on 7 T MRI as markers of multiple sclerosis disease progression. *Brain Communications*, 3(3), fcab134. <https://doi.org/10.1093/braincomms/fcab134>
- Treaba, C. A., Granberg, T. E., Sormani, M. P., Herranz, E., Ouellette, R. A., Louapre, C., Sloane, J. A., Kinkel, R. P., & Mainero, C. (2019). Longitudinal Characterization of Cortical Lesion Development and Evolution in Multiple Sclerosis with 7.0-T MRI. *Radiology*, 291(3), 740–749. <https://doi.org/10.1148/radiol.2019181719>
- Valcarcel, A. M., Linn, K. A., Khalid, F., Vandekar, S. N., Tauhid, S., Satterthwaite, T. D., Muschelli, J., Martin, M. L., Bakshi, R., & Shinohara, R. T. (2018). A dual modeling approach to automatic segmentation of cerebral T2 hyperintensities and T1 black holes in multiple sclerosis. *NeuroImage: Clinical*, 20, 1211–1221. <https://doi.org/10.1016/j.nicl.2018.10.013>
- Valcarcel, A. M., Linn, K. A., Vandekar, S. N., Satterthwaite, T. D., Muschelli, J., Calabresi, P. A., Pham, D. L., Martin, M. L., & Shinohara, R. T. (2018). MIMoSA: An Automated Method for Intermodal Segmentation Analysis of Multiple Sclerosis Brain Lesions: Method For Inter-Modal Segmentation Analysis. *Journal of Neuroimaging*, 28(4), 389–398. <https://doi.org/10.1111/jon.12506>
- Valverde, S., Cabezas, M., Roura, E., González-Villà, S., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À., Oliver, A., & Lladó, X. (2017). Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *NeuroImage*, 155, 159–168. <https://doi.org/10.1016/j.neuroimage.2017.04.034>
- Valverde, S., Salem, M., Cabezas, M., Pareto, D., Vilanova, J. C., Ramió-Torrentà, L., Rovira, À., Salvi, J., Oliver, A., & Lladó, X. (2019). One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks. *NeuroImage: Clinical*, 21, 101638. <https://doi.org/10.1016/j.nicl.2018.101638>
- Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., & Suetens, P. (2001). Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Transactions on Medical Imaging*, 20(8), 677–688. <https://doi.org/10.1109/42.938237>
- Vrenken, H., Jenkinson, M., Pham, D., Guttman, C. R., Pareto, D., Paardekooper, M., de Sitter, A., Rocca, M. A., Wottschel, V., Cardoso, M. J., Barkhof, F., & on behalf of the MAGNIMS Study Group. (2021). Opportunities for Understanding MS Mechanisms and Progression With MRI Using Large-Scale Data Sharing and Artificial Intelligence. *Neurology*, 10.1212/WNL.0000000000012884. <https://doi.org/10.1212/WNL.0000000000012884>
- Walsh, A. J., Lebel, R. M., Eissa, A., Blevins, G., Catz, I., Lu, J.-Q., Resch, L., Johnson, E. S., Emery, D. J., Warren, K. G., & Wilman, A. H. (2013). Multiple Sclerosis: Validation of MR Imaging for Quantification and Detection of Iron. *Radiology*, 267(2), 531–542. <https://doi.org/10.1148/radiol.12120863>
- Walton, C., King, R., Rechtman, L., Kaye, W., Leray, E., Marrie, R. A., Robertson, N., La Rocca, N., Uitendhaag, B., van der Mei, I., Wallin, M., Helme, A., Angood Napier, C., Rijke, N., & Baneke, P. (2020). Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Multiple Sclerosis Journal*, 26(14), 1816–1821. <https://doi.org/10.1177/1352458520970841>



## Bibliography

---

- Wattjes, M. P., Lutterbey, G. G., Gieseke, J., Träber, F., Klotz, L., Schmidt, S., & Schild, H. H. (2007). Double inversion recovery brain imaging at 3T: diagnostic value in the detection of multiple sclerosis lesions. *AJNR. American journal of neuroradiology*, 28(1), 54–59.
- Wattjes, M. P., Ciccarelli, O., Reich, D. S., Banwell, B., de Stefano, N., Enzinger, C., Fazekas, F., Filippi, M., Frederiksen, J., Gasperini, C., Hachohen, Y., Kappos, L., Li, D. K. B., Mankad, K., Montalban, X., Newsome, S. D., Oh, J., Palace, J., Rocca, M. A., ... Oh, J. (2021). 2021 MAGNIMS–CMSC–NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. *The Lancet Neurology*, 20(8), 653–670. [https://doi.org/10.1016/S1474-4422\(21\)00095-8](https://doi.org/10.1016/S1474-4422(21)00095-8)
- Weeda, M., Brouwer, I., de Vos, M., de Vries, M., Barkhof, F., Pouwels, P., & Vrenken, H. (2019). Comparing lesion segmentation methods in multiple sclerosis: Input from one manually delineated subject is sufficient for accurate lesion segmentation. *NeuroImage: Clinical*, 24, 102074. <https://doi.org/10.1016/j.nicl.2019.102074>
- Wisnieff, C., Ramanan, S., Olesik, J., Gauthier, S., Wang, Y., & Pitt, D. (2015). Quantitative susceptibility mapping (QSM) of white matter multiple sclerosis lesions: Interpreting positive susceptibility and the presence of iron: Iron and Myelin Content of MS Lesions with MRI. *Magnetic Resonance in Medicine*, 74(2), 564–570. <https://doi.org/10.1002/mrm.25420>
- Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P. L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., & Firmin, D. (2018). DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction [Conference Name: IEEE Transactions on Medical Imaging]. *IEEE Transactions on Medical Imaging*, 37(6), 1310–1321. <https://doi.org/10.1109/TMI.2017.2785879>
- Yao, B., Bagnato, F., Matsuura, E., Merkle, H., van Gelderen, P., Cantor, F. K., & Duyn, J. H. (2012). Chronic Multiple Sclerosis Lesions: Characterization with High-Field-Strength MR Imaging. *Radiology*, 262(1), 206–215. <https://doi.org/10.1148/radiol.11110601>
- Young, I. R., Randell, C. P., Kaplan, P. W., James, A., Bydder, G. M., & Steiner, R. E. (1983). Nuclear Magnetic Resonance (NMR) Imaging in White Matter Disease of the Brain Using Spin-Echo Sequences: *Journal of Computer Assisted Tomography*, 7(2), 290–294. <https://doi.org/10.1097/00004728-198304000-00016>
- Yu, T., Canales-Rodríguez, E. J., Pizzolato, M., Piredda, G. F., Hilbert, T., Fisch-Gomez, E., Weigel, M., Barakovic, M., Bach Cuadra, M., Granziera, C., Kober, T., & Thiran, J.-P. (2021). Model-informed machine learning for multi-component T2 relaxometry. *Medical Image Analysis*, 69, 101940. <https://doi.org/10.1016/j.media.2020.101940>
- Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31(3), 1116–1128. <https://doi.org/10.1016/j.neuroimage.2006.01.015>
- Zeng, C., Gu, L., Liu, Z., & Zhao, S. (2020). Review of Deep Learning Approaches for the Segmentation of Multiple Sclerosis Lesions on Brain MRI. *Frontiers in Neuroinformatics*, 14, 610967. <https://doi.org/10.3389/fninf.2020.610967>

- Zhang, Y.-D., Govindaraj, V. V., Tang, C., Zhu, W., & Sun, J. (2019). High Performance Multiple Sclerosis Classification by Data Augmentation and AlexNet Transfer Learning Model. *Journal of Medical Imaging and Health Informatics*, 9(9), 2012–2021. <https://doi.org/10.1166/jmihi.2019.2692>

# Curriculum Vitae

## Francesco La Rosa

<b>Nationality</b>	Italian	<b>Date of Birth</b>	4 <sup>th</sup> January 1993
<b>Address</b>	Chemin de Chandieu 1, 1006 Lausanne, Switzerland	<b>Mobile Phone</b>	+41 77 511 52 45
		<b>Email</b>	<a href="mailto:francescolaros@gmail.com">francescolaros@gmail.com</a>
		<b>Linkedin</b>	<a href="https://www.linkedin.com/in/francesco-la-rosa">linkedin/francesco-la-rosa</a>

## Education and training

### Apr. 2018 - **PhD in Electrical Engineering**

Present *École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.*  
*Supervisor:* Prof. Jean-Philippe Thiran. *Co-supervisor:* Dr. Meritxell Bach Cuadra.  
*Thesis:* "Deep learning-based analysis of multiple sclerosis lesions with high and ultra-high field MRI".  
*Publications:* 8 journal articles, 3 conference article, 22 conference abstracts.  
Supervision of 2 master students and 2 bachelor students.  
Teaching assistant of 2 bachelor courses at EPFL.

### Sept. 2019 - **Visiting PhD student**

Jan. 2020 *The National Institutes of Health (NIH), Bethesda, Maryland, US.*  
*Supervisor:* Dr. Daniel Reich.  
*Project:* applying machine learning techniques previously developed to the automated detection of multiple sclerosis cortical lesions with 7T MRI.

### Sept. 2015 - **MSc in Applied Physics**

Oct. 2017 *Università di Bologna, Bologna, Italy.*  
*Thesis:* "A deep learning approach to bone segmentation in CT scans". *Supervisor:* Prof. Renato Campanini. *Co-supervisor:* Dr. Markus Wenzel. Available at: [ams.laura.unibo.it/14561](https://ams.laura.unibo.it/14561) *Final Grade:* 110/110 cum laude.

### Sept. 2016 - **Master thesis research period**

June. 2017 *Fraunhofer MEVIS: Institute for Medical Image Computing, Bremen, Germany.*  
Website: [www.mevis.fraunhofer.de](http://www.mevis.fraunhofer.de)

---

Sept. 2012 - **BSc in Physics**

Dic. 2015 *Università di Bologna, Bologna, Italy.*

*Thesis: "Studies on ToF-PET using Cherenkov radiation". Supervisor: Prof. Tiziano Rovelli. Available at: [amslaurea.unibo.it/9543](https://amslaurea.unibo.it/9543) Final Grade: 110/110.*

Sept. 2014 - **Erasmus exchange student**

June 2015 *University of Bristol, Bristol, UK.*

## Work experience

Sept. 2016 - **Internship**

Oct. 2016 *Bioretics, Cesena, Italy.*

Website: [www.bioretics.com](http://www.bioretics.com)

## Honours and awards

2021 **Honorable reviewer**

International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI).

2019 **Shortlisted in the Poster Awards**

European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS) conference. *Abstract:: "Deep learning-based detection of cortical lesions in multiple sclerosis patients with FLAIR, DIR and MP2RAGE MRI sequences".*

## Fundings

2021 **SNF Postdoc Mobility fellowship**

Fellowship for a 24 months stay abroad awarded by the Swiss National Science Foundation (SNF).

2021 **Education grant ECTRIMS**

Education grant to participate to the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS) 2021 conference.

2021 **Education stipend ISMRM**

Education stipend to participate to the International Society for Magnetic Resonance in Medicine (ISMRM) 2021 conference.

---

2020	<b>Education grant ECTRIMS</b> Education grant to participate to the European Committee for Treatment and Research in Multiple Sclerosis (ECTRIMS) 2020 conference.
2020	<b>Education stipend ISMRM</b> Education stipend to participate to the International Society for Magnetic Resonance in Medicine (ISMRM) 2020 conference.

## Skills and Competences

- **Programming**

Python, MATLAB, C/C++.

- **Software / Libraries**

TensorFlow, Keras, Pytorch, MONAI, Microsoft Office Suite, ITK, Elastix, ANTs, FSL, FreeSurfer, SPM,  $\LaTeX$ .

- **Languages**

*Italian*: Native. *English*: Full professional proficiency. *French*: Basic knowledge. *German*: Basic knowledge.



# List of publications

## Articles in peer-reviewed journals

1. **La Rosa F**, Abdulkadir A, Fartaria MJ, Rahmanzadeh R, Lu PJ, Galbusera R, Barakovic M, Thiran JP, Granziera C, Bach Cuadra M. Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE. *NeuroImage: Clinical*, 2020; 27, p.102335. doi: [10.1016/j.nicl.2020.102335](https://doi.org/10.1016/j.nicl.2020.102335)
2. **La Rosa F**, Yu T, Barquero G, Thiran JP, Granziera C, Bach Cuadra M. MPRAGE to MP2RAGE UNI translation via generative adversarial network improves the automatic tissue and lesion segmentation in multiple sclerosis patients. *Computers in Biology and Medicine*, 2021, 132, p.104297. doi: [10.1016/j.combiomed.2021.104297](https://doi.org/10.1016/j.combiomed.2021.104297)
3. Barquero G, **La Rosa F**, Kebiri H, Lu PJ, Rahmanzadeh R, Weigel M, Fartaria MJ, Kober T, Théaudin M, Du Pasquier R, Sati P, Reich DS, Absinta M, Granziera C, Maggi P, Bach Cuadra M. RimNet: A deep 3D multimodal MRI architecture for paramagnetic rim lesion assessment in multiple sclerosis. *Neuroimage: Clinical*, 2020, 28, p.102412. doi: [10.1016/j.nicl.2020.102412](https://doi.org/10.1016/j.nicl.2020.102412)
4. Müller J, **La Rosa F**, Beaumont J, Tsagkas C, Rahmanzadeh R, Weigel M, Bach Cuadra M, Gambarota G, Granziera C. FLAWS: New Sensitive 3T MRI Contrasts for Cortical Lesion Detection in Multiple Sclerosis. Accepted for publication in *Investigative Radiology*.
5. Maggi P, Fartaria MJ, Jorge J, **La Rosa F**, Absinta M, Sati P, Meuli R, Du Pasquier R, Reich DS, Bach Cuadra M, Granziera C, Richiardi J., Kober T. CVSnet: a machine learning approach for automated central vein sign assessment in multiple sclerosis. *NMR in Biomedicine*, 2020, 33(5), p.e4283. doi: [10.1002/nbm.4283](https://doi.org/10.1002/nbm.4283)
6. Rahmanzadeh R, Lu PJ, Barakovic M, Weigel M, Maggi P, Nguyen TD, Schiavi S, Daducci A, **La Rosa F**, Schaedelin S, Absinta M, Reich SR, Sati P, Wang Y, Bach Cuadra M, Radue EW,

---

Kuhle J, Kappos L, Granziera C. Myelin and axon pathology in multiple sclerosis assessed by myelin water and multi-shell diffusion imaging. *Brain*, 2021. <https://doi.org/10.1093/brain/awab088>

7. Maggi P, Kuhle J, Schädelin S, van der Meer F, Weigel M, Galbusera R, Mathias A, Lu PJ, Rahmanzadeh R, Benkert P, **La Rosa F**, Bach Cuadra M, Sati P, Théaudin M, Pot C, van Pesch V, Leppert D, Stadelmann C, Kappos L, Du Pasquier R, Reich SR, Absinta M, Granziera C. Chronic White Matter Inflammation and Serum Neurofilament Levels in Multiple Sclerosis. *Neurology*, 2021. <https://doi.org/10.1212/WNL.0000000000012326>
8. Lu PJ, Barakovic M, Weigel M, Rahmanzadeh R, Galbusera R, Schiavi S, Daducci A, **La Rosa F**, Bach Cuadra M, Sandkühler R, Kuhle J, Kappos L, Cattin P, Granziera C. GAMER-MRI in Multiple Sclerosis Identifies the Diffusion-Based Microstructural Measures That Are Most Sensitive to Focal Damage: A Deep-Learning-Based Analysis and Clinico-Biological Validation. *Frontiers in Neuroscience*, 2021. <https://doi.org/10.3389/fnins.2021.647535>
9. Lu PJ, Yoo Y, Rahmanzadeh R, Galbusera R, Weigel M, Ceccaldi P, Nguyen TD, Spincemaille P, Wang W, Daducci A, **La Rosa F**, Bach Cuadra M, Sandkühler R, Nael K, Doshi A, Fayad ZA, Kuhle J, Kappos L, Granziera C. GAMER-MRI: Gated-attention mechanism ranking of multi-contrast MRI in brain pathology. *Neuroimage Clinical*, 2020. <https://doi.org/10.1016/j.nicl.2020.102522>

### Journal articles under revision

1. **La Rosa F**, Beck ES, Maranzano J, Todea RA, Gelderen Pv, de Zwart J, Luciano NJ, Duyn JH, Thiran JP, Granziera C, Reich DS, Sati P, Bach Cuadra M.  
Multiple sclerosis cortical lesion detection with deep learning at ultra-high-field MRI.  
Currently under revision in *NMR in biomedicine*.
2. **La Rosa F**, Wynen M, Al-Louzi O, Beck ES, Huelnhagen T, Maggi P, Thiran JP, Kober T, Shinohara RT, Sati P, Reich SD, Granziera C, Absinta M, Bach Cuadra M.  
Cortical lesions, central vein sign, and paramagnetic rim lesions in multiple sclerosis: emerging machine learning techniques and future avenues.  
Currently under revision in *Neuroimage: Clinical*.
3. Rahmanzadeh R, Weigel M, Lu PJ, Melie-Garcia L, Nguyen TD, Cagol A, **La Rosa F**, Barakovic M, Wang Y, Bach Cuadra M, Radue EW, Gaetano L, Kappos L, Kuhle J, Magon S, Granziera C.  
A comparative assessment of sensitivity to multiple sclerosis pathology and intrascanner



---

reproducibility for myelin-sensitive MRI measures.

Currently under revision in *Neuroimage: Clinical*.

4. Rahmanzadeh R, Galbusera R, Lu PJ, Bahn E, Weigel M, Barakovic M, Franz J, Nguyen TD, Cagol A, Spincemaille P, Schiavi S, Daducci A, **La Rosa F**, Absinta M, Sati P, Bach Cuadra M, Radue EW, Leppert D, Kuhle J, Kappos L, Bruck W, Reich DS, Stadelmann C, Wang Y, Granziera C. A new advanced MRI biomarker identifies fully remyelinated lesions in Multiple Sclerosis: An in vivo cross-sectional and longitudinal study with double-blind postmortem validation.

Currently under revision in *Annals of Neurology*.

## Peer-reviewed conference papers

1. **La Rosa F**, Beck ES, Abdulkadir A, Thiran JP, Reich DS, Sati P, Bach Cuadra M. Automated Detection of Cortical Lesions in Multiple Sclerosis Patients with 7T MRI. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 584-593), 2020. Springer, Cham. [https://doi.org/10.1007/978-3-030-59719-1\\_57](https://doi.org/10.1007/978-3-030-59719-1_57)
2. **La Rosa F**, Fartaria MJ, Kober T, Richiardi J, Granziera C, Thiran JP, Bach Cuadra M. Shallow vs deep learning architectures for white matter lesion segmentation in the early stages of multiple sclerosis. In International MICCAI Brainlesion Workshop (pp. 142-151), 2018. Springer, Cham. [https://doi.org/10.1007/978-3-030-11723-8\\_14](https://doi.org/10.1007/978-3-030-11723-8_14)
3. Nguyen HG, Pica A, Hrbacek J, Weber DC, **La Rosa F**, Schalenbourg A, Sznitman R, Bach Cuadra M. A novel segmentation framework for uveal melanoma in magnetic resonance imaging based on class activation maps. In International Conference on Medical Imaging with Deep Learning (pp. 370-379), 2019. PMLR. <http://proceedings.mlr.press/v102/nguyen19a.html>

## Abstracts in proceedings of international conferences

### ECTRIMS 2021

#### *Joint ACTRIMS-ECTRIMS Meeting, Virtual Event*

1. **La Rosa F**, Lu PJ, Rahmanzadeh R, Cagol A, Barakovic M, Thiran JP, Granziera C, Bach Cuadra M. CLaiMS-Net: cortical lesions assessment in multiple sclerosis patients via a convolutional neural network and a single 3T MRI acquisition. Abstract number: 1401.

- 
2. **La Rosa F**, Beck ES, Maranzano J, Thiran JP, Granziera C, Reich DS, Sati P, Bach Cuadra M. Automated detection of cortical lesions with single and multi-contrast 7T MRI. Abstract number: 295.
  3. Müller J, **La Rosa F**, Beaumont J, Tsagkas C, Rahmanzadeh R, Barakovic M, Yaldizli O, Weigel M, Bach Cuadra M, Kappos L, Kuhle J, Gambarota G, Granziera C. Cortical Lesion Detection using FLAWS in Multiple Sclerosis. Abstract number: 779.
  4. Rahmanzadeh R, Weigel M, Lu PJ, Melie-Garcia L, Nguyen TD, Barakovic M, **La Rosa F**, Wang Y, Bach Cuadra M, Radue EW, Gaetano L, Kuhle J, Kappos L, Magon S, Granziera C. An assessment of intra-scanner reproducibility of myelin-sensitive MRI measures.

## **ISMRM 2021**

### ***ISMRM & SMRT Annual Meeting, Virtual Event***

5. **La Rosa F**, Barquero G, Al-Louzi O, Maréchal BM, Kober T, Thiran JP, Sati P, Reich DS, Maggi P, Absinta M, Bach Cuadra M. Automated assessment of paramagnetic rim lesions in multiple sclerosis patients with 3T and 7T MP2RAGE. Abstract number: 3507.
6. Wynen M, **La Rosa F**, Sellimi A, Barquero G, Perrotta G, Lolli V, Van Pesch V, Granziera C, Kober T, Sati P, Macq B, Reich DS, Absinta M, Bach Cuadra M, Maggi P. Longitudinal automated assessment of paramagnetic rim lesions in multiple sclerosis using RimNet. Abstract number: 2814.
7. Rahmanzadeh R, Lu PJ, Barakovic M, Weigel M, Gaetano L, Galbusera R, Nguyen TD, **La Rosa F**, Reich DS, Sati P, Wang Y, Bach Cuadra M, Radue EW, Kuhle J, Kappos L, Magon S, Granziera C. Quantitative myelin-sensitive MRIs exhibit differential sensitivity to multiple sclerosis pathology in distinct brain lesions and regions. Abstract number: P1232.
8. Lu PJ, Barakovic M, Weigel M, Rahmanzadeh R, Galbusera R, Schiavi S, Daducci A, **La Rosa F**, Bach Cuadra M, Sandkühle R, Kuhle J, Kappos L, Cattin P, Granziera C. Identification of diffusion-based micro-structural measures most sensitive to multiple sclerosis focal damage using GAMER-MRI. Abstract number: P606.

## **MSVirtual2020**

### ***Joint ACTRIMS-ECTRIMS Meeting, Virtual Event***

9. **La Rosa F**, Yu T, Barquero G, Thiran JP, Granziera C, Bach Cuadra M. Translating mprage to mp2rage improves the automatic tissue and lesion segmentation in multiple sclerosis

---

patients. Abstract number: P0017.

10. Barquero G, **La Rosa F**, Kebiri H, Rahmanzadeh R, Weigel M, Fartaria MJ, Kober T, Theaudin M, Du Pasquier R, Sati P, Reich DS, Absinta M, Granziera C, Maggi P, Bach Cuadra M. Rimnet: a deep 3D multimodal MRI architecture for paramagnetic rim lesions assessment in multiple sclerosis. PS16.04
11. Barakovic M, Rahmanzadeh R, Lu PJ, Maggi P, Absinta M, **La Rosa F**, Bach Cuadra M, Schiavi S, Daducci A, Sati P, Reich DS, Kuhle J, Weigel M, Kappos J, Granziera C. Studying intralesional axonal damage in MS white matter lesions with diffusion mri biophysical models. Abstract number: P0647.
12. Rahmanzadeh R, Lu PJ, Barakovic M, Weigel M, Schaedelin S, Nguyen TD, Schiavi S, Daducci A, **La Rosa F**, Wang Y, Bach Cuadra M, Radue EW, Kuhle J, Kappos J, Granziera C. Advanced magnetic resonance imaging for myelin and axonal density in MS: correlation with clinical disability and serum neurofilament levels. Abstract number: P0534.

## **ISMRM 2020**

### ***ISMRM & SMRT Virtual Conference & Exhibition***

13. **La Rosa F**, Abdulkadir A, Fartaria MJ, Rahmanzadeh R, Galbusera R, Thiran JP, Granziera C, Bach Cuadra M. Automatic detection of Multiple Sclerosis cortical lesions based on 3D-FLAIR and MP2RAGE sequences. Abstract number: 1862.
14. Barakovic M, Rahmanzadeh R, Lu PJ, Maggi P, Sati P, Reich D, **La Rosa F**, Bach Cuadra M, Schiavi S, Daducci A, Kuhle J, Weigel M, Kappos L, Granziera C. A comparison of seven white matter diffusion microstructure models to study intralesional damage in Multiple Sclerosis. Abstract number: 1390.
15. Rahmanzadeh R, Lu PJ, Barakovic M, Galbusera R, Weigel M, Schaedelin S, Maggi P, Nguyen TD, Schiavi S, **La Rosa F**, Reich DS, Sati D, Wang Y, Bach Cuadra M, Radue EW, Kuhle J, Kappos J, Granziera C. A quantification of myelin and axonal damage across multiple sclerosis lesions and clinical subtypes with myelin and diffusion MRI. Abstract number: P0058.

## **ECTRIMS 2019**

### ***Joint ACTRIMS-ECTRIMS Meeting, Stockholm, Sweden***

16. **La Rosa, F**, Fartaria, M.J., Abdulkadir, A., Rahmanzadeh, R., Lu, P.J., Galbusera, R., Granziera,

---

C., Thiran, J.P. and Bach Cuadra, M. Deep learning-based detection of cortical lesions in multiple sclerosis patients with FLAIR, DIR, and MP2RAGE MRI sequences. Abstract number: P469. Shortlisted in the ECTRIMS Poster Awards.

17. Richiardi J, Maggi P, Fartaria MJ, **La Rosa, F.**, Jorge J, Sati P, Reich DS, Du Pasquier R, Meuli R, Bach Cuadra M, Kober T. Differential diagnosis of multiple sclerosis with machine learning-based central vein sign recognition. Abstract number: P971.
18. Rahmanzadeh R, Lu PJ, Weigel M, Galbusera R, **La Rosa F**, Bach Cuadra M, Schiavi S, Radue EW, Kuhle J, Kappos J, Granziera C. Axonal damage explains part of and extends beyond the diffuse pathology evidenced by T1 mapping in normal-appearing white matter of multiple sclerosis patients. Abstract number: P1259.
19. Lu PJ, Rahmanzadeh R, Galbusera R, Odry B, Weigel M, **La Rosa F**, Bach Cuadra M, Nguyen T, Wang Y, Daducci A, Kuhle J, Cattin P, Kappos L, Granziera C. Deep learning analysis applied to multi-parametric advanced MRI shows higher myelin content and neurite density in juxtacortical lesions compared to periventricular lesions. Abstract number: P529.
20. Rahmanzadeh R, Lu PJ, Weigel M, Galbusera R, Nguyen TD, Schiavi S, Wang Y, **La Rosa F**, Bach Cuadra M, Radue EW, Kuhle J, Kappos J, Granziera C. Axonal and myelin injury in white matter and cortex of relapsing-remitting and progressive multiple sclerosis patients: a combined myelin water and multi-shell diffusion MRI study. Abstract number: P1258.

### Other conferences

21. **La Rosa F**, Thiran JP, Bach Cuadra M. A subtraction image-based method to detect new appearing multiple sclerosis lesions on single-contrast FLAIR MRI. MSSEG-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure.
22. Muller J, **La Rosa F**, Beaumont J, Tsagkas C, Rahmanzadeh R, Barakovic M, Yaldizli O, Weigel M, Bach Cuadra M, Kappos L, Kuhle J, Gambarota G, Granziera C. FLAWS: A New Sensitive Method for Cortical Lesion Detection in Multiple Sclerosis at 3T MRI. Swiss Society of Neurology (SNG) conference.