

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



Analysis of Magnetic Resonance Images in the Study of Multiple Sclerosis Disease

Joana Esteves Fonseca

Mestrado em Engenharia Biomédica

Supervisor: João Manuel R.S. Tavares, Ph.D.
Faculdade de Engenharia da Universidade do Porto

July 19, 2021

Analysis of Magnetic Resonance Images in the Study of Multiple Sclerosis Disease

Joana Esteves Fonseca

Mestrado em Engenharia Biomédica

Faculdade de Engenharia da Universidade do Porto

July 19, 2021

Resumo

A esclerose múltipla é a doença desmielinizante mais comum com maior prevalência em adultos jovens. Apesar de todos os conhecimentos existentes relativamente à doença, ainda há muito que se desconhece, o que sustenta a busca contínua por respostas. Atendendo à sua natureza progressiva e degenerativa, esta acarreta, entre outras, uma degradação das capacidades motoras e cognitivas para os seus doentes diminuindo significativamente a sua qualidade de vida. Atualmente, ainda não foi descoberta uma cura o que limita a resposta providenciada a estes pacientes cujo tratamento se concentra em retardar a progressão e o processo inflamatório assim como a gestão dos sintomas.

O diagnóstico é realizado principalmente por meio de técnicas de imagiologia, como ressonância magnética e da avaliação visual realizada pelos profissionais de saúde. Este processo é repetitivo e desgastante para estes profissionais, por isso as técnicas de análise e processamento de imagens têm se tornado uma ferramenta necessária para auxiliar o diagnóstico, resultando no prolongamento das condições de vida desses pacientes.

Este projeto teve como objetivo estudar diferentes técnicas de processamento e análise de imagens e as etapas necessárias para implementar um algoritmo de segmentação de lesões de esclerose múltipla automatizado utilizando imagens de ressonância magnética. Paralelamente, uma exploração de métodos de pré-processamento, segmentação e classificação, apoiados pelo estado da arte e conceitos fundamentais relacionados à doença de esclerose múltipla, técnica de ressonância magnética, noções de *machine learning*, também são apresentados neste trabalho.

A metodologia proposta foi baseada num modelo de cascata cuja entrada são *patches*. Este algoritmo é então composto por duas redes neurais convolucionais idênticas que diferenciam os dados de entrada entre *patches* lesão e não lesão, sendo o resultado uma máscara binária das lesões estimadas.

A metodologia foi avaliada num conjunto de dados disponíveis publicamente, o *MICCAI 2008 MS Lesion Segmentation Challenge*. Os resultados obtidos demonstraram a relevância das técnicas de pré-processamento, visto que melhoram a qualidade da imagem e, conseqüentemente, os resultados subsequentes. Embora a amostra com informação da segmentação realizada pelos especialistas fosse pequena, e se tenha sido parcialmente ultrapassado com a aplicação da técnica de aumento de dados, os resultados em valores de performance foram de um modo geral promissores.

Abstract

Multiple sclerosis is the most common demyelinating disease with a higher diagnosis prevalence in young adults. Despite all possessed knowledge related to the condition, there is yet a lot unknown, which fuels the continuous seek for answers. Considering its progressive and degenerative nature, the disability caused on its patients decreases their quality of life significantly. There is no cure, limiting the response provided to these patients whose treatment focuses on retarding the progression of the inflammatory process and managing the symptoms. The diagnosis is mainly performed on imaging techniques, like magnetic resonance and the visual assessment of health professionals. This process is repetitive and exhausting to this professional, so image analysis and processing techniques have become a required tool and ally to assist in the diagnosis and treatment by the experts, resulting in prolonging the life conditions of these patients.

This project aimed to study different image processing and analysis techniques and the required steps to implement an automated MS lesion segmentation algorithm using MR images. Adjacent, an exploration of pre-processing, segmentation, and classification methods, supported by the state-of-the-art and fundamental concepts related to multiple sclerosis diseases, magnetic resonance imaging technique, notions of machine learning, too, presented in this work.

The proposed methodology was based on a patch-wise cascade model composed of two identical convolutional neural networks that differentiate between the lesion and non-lesion patches, resulting in a binary mask of the estimated lesions. The methodology was evaluated on the public dataset MICCAI 2008 MS Lesion Segmentation Challenge.

The results obtained demonstrated the pre-processing techniques' relevance since they improve the image quality and, therefore, the subsequent results. Although the labeled sample was small and partially overcome with the application of the data augmentation technique, the overall segmentation results were promising.

Contents

1	Introduction	1
1.1	Context	1
1.2	Objective and Methodology	3
1.3	Contributions	4
1.4	Dissertation Structure	4
2	Background	7
2.1	Multiple Sclerosis	7
2.1.1	Description of the Multiple Sclerosis Disease	7
2.1.2	Diagnosis Criteria	10
2.1.3	Treatment	10
2.2	Magnetic Resonance Imaging	11
2.2.1	MRI Principles	11
2.2.2	Image Acquisition	13
2.2.3	Application in Multiple Sclerosis	15
2.2.4	Image Artifacts	16
2.3	Summary	16
3	Machine Learning and Deep Learning	19
3.1	Machine Learning	19
3.2	Learning Scenarios	20
3.3	Deep Learning	21
3.4	Machine Learning Algorithms	21
3.4.1	Support Vector Machine	21
3.4.2	Logistic Regression	21
3.4.3	Decision Tree	21
3.4.4	Random Forest	22
3.4.5	Linear Discriminate Analysis	22
3.4.6	K-Nearest Neighbour	22
3.4.7	K-Means	22
3.5	Deep Learning Algorithms	23
3.5.1	Deep Neural Network	23
3.5.2	Restricted Boltzmann Machine and Deep Belief Network	23
3.5.3	Convolutional Neural Networks	23
3.5.4	Generative Adversarial Network	23
3.6	Convolutional Neural Networks	23
3.6.1	Architecture Overview	24
3.6.2	Type of Layers	24

3.6.3	Filter Hyperparameters	25
3.6.4	Activation Function	26
3.6.5	Network Training	26
3.6.6	Optimization	27
3.7	Summary	28
4	Image Processing and Analysis - State of the Art	29
4.1	Overview	29
4.2	Method	30
4.3	Related work	30
4.3.1	Preprocessing	32
4.3.2	Multiple Sclerosis - Lesion Segmentation	33
4.3.3	Multiple Sclerosis - Classification	40
4.4	Evaluation Metrics	47
4.5	Summary	48
5	Methodology	51
5.1	Overview	51
5.2	Dataset	52
5.3	Image Preprocessing	53
5.3.1	Registration	53
5.3.2	Extraction of Non cerebral Structures - Skull stripping	53
5.3.3	Bias Field Correction	56
5.3.4	Downsample, Crop and Intensity Normalization	56
5.4	Lesion Segmentation	57
5.4.1	Model Inputs and the Cascade Configuration	57
5.4.2	Model Architecture	58
5.4.3	CNN Training	59
5.4.4	Data Augmentation	59
5.4.5	CNN Testing	60
5.4.6	Evaluation	60
5.4.7	Comparison with other methods	61
5.5	Lesion Evaluation	62
5.6	Adopted Resources	62
5.6.1	ITK-SNAP	63
5.6.2	Anaconda - Jupyter Notebook	63
5.6.3	Tensorflow and Keras	64
5.6.4	Matlab, SPM and LST	64
5.7	Summary	64
6	Results and Discussion	67
6.1	Implementation	67
6.1.1	Computational Infrastructure	67
6.1.2	Image Formats	67
6.2	Image Preprocessing	68
6.3	Lesion Segmentation	70
6.4	Lesion Evaluation	74
6.5	Summary	75

7	Conclusion and Future Work	77
7.1	Conclusion	77
7.2	Future Work	79
	References	81

List of Figures

1.1	Lesion progression of a patient with multiple sclerosis obtained at 1.1a baseline, and 1.1b 1 year, 1.1c 2 years and 1.1d 3 years later.	2
2.1	Multiple sclerosis disease courses of RRMS, SPMS and PPMS.	8
2.2	Multiple sclerosis prevalence per 100,000 population by country.	9
2.3	Simplification of the protons without and with an external field. 2.3a without the external magnetic field the protons take a aleatory orientation. 2.3b is represented the alignment of the protons when passed by an external magnetic field, assuming either a parallel or anti-parallel orientation.	12
2.4	MRI system with operator controls and visualization panels, and the various sub-systems	14
3.1	Comparison between a CNN with the human brain functions.	24
3.2	Example of the application of different pooling methods.	25
3.3	Highlight the difference of the application of the dropout technique.	26
3.4	Scheme with different types of ReLU functions with mathematical equations and graphs.	27
4.1	Usual pipeline followed in image analysis and processing studies, divided into four main steps: acquisition, preprocessing, segmentation and quantification. . .	30
4.2	PRISMA diagram describing the results of the executed literature search.	31
5.1	Pipeline of the proposed methodology.	52
5.2	Flowchart of the BET algorithm.	54
5.3	Deepbrain architecture schematic.	55
5.4	Flowchart of the N4 procedure, where B represents the bias field.	57
5.5	Training Pipeline. Input MRI sequences are processed to create an array of patches, which is included a balanced number of patches from each of the classes. This array is inputted to CNN ₁ to be trained. The originally unbalanced array is then tested to this network, and its results are threshold, the classes are balanced, and used as input to CNN ₂ , concluding the training process.	58
5.6	Proposed model.	59
5.7	Flowchart of the testing procedure of the proposed cascade model. Similar to training, an array of patches is formed and applied sequentially to each network to be evaluated.	60
5.8	ITK-SNAP interface with an example from the MICCAI 2008 dataset. The four black panels allow the visualization of the three anatomical planes of the medical image, and the fourth, the bottom left panel, permits a 3D visualization of the segmentation information.	63

5.9	The report retrieves from the LST software. The left part presents information related to the images used and algorithm name and parameters, like threshold and the maximum number of iterations if the selected algorithm is the LGA. Furthermore, it presents an illustration of the lesion location in the left bottom section. The right section exhibits the outcomes and a slide frame to visualize the overlay obtained lesion segmentation annotations.	65
6.1	Results from the application of the BET algorithm on the MICCAI dataset, the images were obtained from the same axial slice and patient.	68
6.2	Results from the application of the Deepbrain algorithm on the MICCAI dataset, the images were obtained from the same axial slice and patient.	69
6.4	Parameter optimization for the cascaded CNN, tbin and lmin, versus DSC coefficient, obtained for the UNC model, using T1-w and FLAIR images as input sequences.	73
6.5	Obtained results for the joined masks visualized in ITK-Snap. The red color represents the lesion segmented by the expert, the green estimated lesion, and the blue intersection between the previous two. This example was obtained with the same subjects reported in the previous Table. Subject <i>CHB Case06</i> resultant masks for two center data model, 6.5a, and the CHB data trained model, 6.5b. Subject <i>UNC Case10</i> retrieved for the model using both center 6.5c, and the UNC data 6.5d. The presented slices are not the same in both subjects for the different models.	75

List of Tables

2.1	MS phenotype and characterization.	8
2.2	McDonald Criteria for multiple sclerosis diagnosis.	11
2.3	T1 and T2 relaxation time constant for different tissues.	13
3.1	Differences between supervised and unsupervised learning.	20
4.1	Articles with methodologies related to multiple sclerosis lesions segmentation using machine/deep learning	36
4.2	Model characteristics, contributions, benefits, and limitations of the segmentation articles analyzed.	38
4.3	Articles related to MS classification, diagnosis, or characterization on MRI images using ML and DL methods. The column regarding the multi-modality refers to the studies utilizing more than conventional MRI sequence.	44
4.4	Model, contributions, benefits, and limitations of the articles related to the MS classification.	45
4.5	Evaluation metrics for algorithms	48
6.2	Initial min threshold variations. The results were obtained working with the same model trained with the same setting, modifying the threshold values, and testing the same subjects. Mean DSC, VD, TPF, and FPF are presented for each threshold value.	71
6.3	Results obtained for the different models, trained on T1 and FLAIR modalities. The two models from the UNC center were trained one with the annotations from the in-house expert and the other with label from the CHB center expert. Mean DSC, VD, TPF, and FPF are presented for each model.	72
6.4	Results obtained for the different models without applying data augmentation. The two models from the UNC center were trained one with the annotations from the in-house expert and the other with label from the CHB center expert. Mean DSC, VD, TPF, and FPF are presented for each threshold value.	73
6.5	Segmentation results for each of the evaluated methods on the two MS datasets. Results are shown for LST algorithms, LGA and LPA and the proposed approach trained with the two center MICCAI dataset. Mean DSC, VD, TPF, and FPF are presented for each method and database.	74
6.6	Characterization of the segmentation models for the different models. The same subjects are compared. The lesion percentage column presents the ratio of lesions with brain volume.	75

Abbreviations and Symbols

List of Abbreviations

ABC	Artificial Bee Colony
BET	Brain Extraction Tool
BSGLMM	Bayesian Spataial Generalized Linear Mixed Model
CDMS	Clinically-Defined Multiple Sclerosis
CEN	Convolutional Encoder Network
CIS	Clinical Isolated Syndrome
cMRI	conventional MRI
CNN	Convolutional Neural Network
CNS	Central Nervous System
CSV	Central Vein Sign
CT	Computed Tomography
DBN	Deep Belief Network
DF	Deformation fields
DL	Deep Learning
DNN	Deep Neural Network
DOST	Discrete Orthogonal Stockwell Transform
DSBI	Diffusion Basis Spectrum Imaging
DSC	Dice Similarity Coefficient
DT	Decision Tree
DTI	Diffusion-tensor Imaging
DWI	Diffusion-weighted Images
DWT	Discrete Wavelet Transform
EDSS	Expanded Disability Status Scale
EEG	Electroencephalography
ELM	Extreme Learning Machine
FLAIR	Fluid Attenuated Inversion Recovery
fMRI	functional Magnetic Resonance Imaging
FN	False Negative
FP	False Positive
FPR	False Positive Rate
F _s	F-score
FSL	FMRIB Software Library
GAN	Generative Adversarial Network
GM	Grey Matter
HDLS	Hereditary Diffuse Leukoencephalopathy with Spheroids
HWT	Haar Wavelet Transform

ITK	Insight Toolkit software
KNN	K-Nearest Neighbours
LDA	Linear Discriminate Analysis
LGCP	Log Gaussian Cox Process
LR	Logistic Regression
LRP	Layer-wise Relevance Propagation
ML	Machine Learning
MPRAGE	Magnetization-Prepared Acquisition with Gradient Echo
MRI	Magnetic Resonance Imaging
MRSI	Magnetic Resonance Spectroscopic Imaging
MS	Multiple Sclerosis
MSIF	Multiple Sclerosis International Federation
MT	Magnetization Transfer Images
NAGM	Normal Appearing Grey Matter
NAWM	Normal Appearing White Matter
NMR	Nuclear Magnetic Resonance
NPV	Negative Predictive Value
PD-W	Proton Density Weighted images
Pixel	Picture element
PPMS	Primary Progressive Multiple Sclerosis
PPV	Positive Predictive Value
qMRI	quantitative MRI
RBM	Restricted Boltzman Machines
RF	Radio Frequency
RF	Random Forest
ROC	Receiver Operating Characteristic curve
ROI	Region of Interest
RPMS	Remitting Progressive Multiple Sclerosis
RRMS	Relapsing-Remitting Multiple Sclerosis
SNR	Signal-to-Noise Ratio
SPMS	Secondary Progressive Multiple Sclerosis
STIR	Short Tau Inversion Recovery
SVM	Support Vector Machine
SWE	Stationary Wavelet Entropy
T1-W	T1 weighted images
T2-W	T2 weighted images
TE	Time of Echo
TI	Time of Inversion
TN	True Negative
TP	True Positive
TPR	True Positive Rate
TR	Time of Repetition
US	Ultrasounds
VD	Volume Difference
Voxel	Volume element
WM	White Matter

List of Symbols

β_0	External magnetic field
γ	Gyromagnetic ratio
ω_0	Angular precessional or Lamor frequency

Chapter 1

Introduction

This chapter presents a small overview of the dissertation. Starting with a brief contextualization regarding the pathology, followed by medical imaging modality and goals for the study. Afterward, it is drawn a summary of the structure of this report with a short description of each chapter. Ultimately, a description of the contributions achieved during the development of the project.

1.1 Context

Humans are continually processing signals either from the outside or within the body. All these are processed and managed by the central nervous system (CNS) that regulates all our actions (voluntary or involuntary). The brain and spinal cord compose the CNS, representing a portion of the nervous system. Its comprehension and behavior are still a fevering study area due to its complexity and relevancy. It also considers that alterations could lead to significant impairments of the subjects.

Multiple sclerosis (MS) is an auto-immune inflammatory disease known for its demyelinated effect and progression. According to the Multiple Sclerosis International Federations (MSIF) most recent number, about 2.8 million people worldwide are affected by multiple sclerosis, with an estimated minimum of 300 people being diagnosed each day [1]. MS has no cure and is characterized by numerous plaques, lesions that can evolve and deteriorate the condition of the patient. These lesions occur due to the attack on the myelin of nerve cells by the immunologic system, which recognized it as intrusion cells, befalling an inflammatory process and consequent presentation of the lesions.

The first reference to MS remote to the beginning of the XIX century, on a personal journal owned by Sir Augustus D'Este, meticulously detailing his symptoms. Although he was never formally diagnosed, the similarity to MS signs leads to that conclusion posthumous. McKenzie executed the first medical acknowledgment of MS regarding a patient who showed progressive confusion of vision. Years later, Martin Charcot, a French neurologist, presented the first delin-eation diagnosis criteria for the pathology and a further description of the condition attributing the name of "Sclerosis plaques" [2].

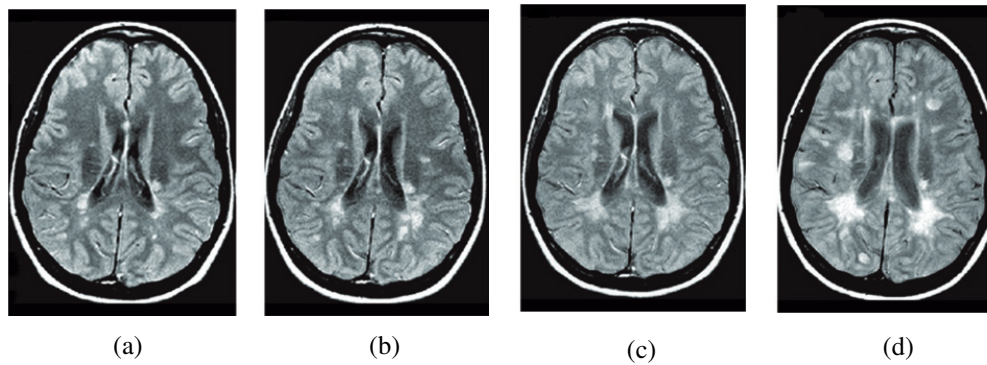


Figure 1.1: Lesion progression of a patient with multiple sclerosis obtained at 1.1a baseline, and 1.1b 1 year, 1.1c 2 years and 1.1d 3 years later (From [3]).

The MS diagnosis is challenging because the episodes occur in indefinite periods, especially in the early stages. Nonetheless, the process became more manageable after medical imaging techniques such as MRI were adopted into the clinic, ceasing the diagnosis to relying merely on clinical analysis to be further supported with MR images.

Bloch and Purcell reported the nuclear magnetic resonance (NMR) phenomenon in 1946, giving them a Nobel prize. However, the first clinical magnetic resonance imaging (MRI) based on the physical event was only acquired in 1980. The technique rapidly evolved, and it is nowadays one of the most performed medical imaging modalities [4]. It is a non-invasive imaging technique considering that, unlike similar systems, it does not require ionizing radiation. Initially, it was considered a method merely for anatomical representation, but the researchers were able to create new ways to acquire functional images and spatially localized spectra. MRI is a versatile and powerful tool in the clinical field with a vast range of applications such as neurological, cardiovascular, and musculoskeletal [5].

The MR images result from the strong magnetic field and radiofrequency (RF) waves while interacting with biological tissues. They are distinguished by their high contrast and detail of the anatomical structures permitting a better evaluation concerning irregularities.

These medical images may require additional examination and adjustments that could subject the patients to the second series of scans or the professional spending significant amounts of time, which ultimately means consuming valuable resources. With imaging and engineering efforts, some methods were developed that assist in image processing and analysis. These techniques have been highlighted as a helpful mechanism since it facilitates the extraction of information that would be difficult to acquire, consequently removing some subjectivity of the image analysis procedures. Furthermore, it reduces the realization of tedious and lengthy tasks by medical professionals [6]. In the medical area, the implementation of these techniques mainly increased for disease diagnostic and characterization.

1.2 Objective and Methodology

Multiple sclerosis impacts the lives and well-being of their patients, combined with a predominant incidence in young adults and its uncertain causes, make MS an intense research area joining the various fields, particularly medical and engineering.

So far, MS can not be cured or prevented due to its unknown origin. Therefore can merely be handled with medication for retarding the progression and severity due to the accumulation of lesions and managing the symptoms like visions and sensory-motor problems.

The issue and weighing the affected by multiple sclerosis and the conscientious of the current reality for the diagnosis chain in some medical facilities, where the assessment is mainly or entirely performed by visual evaluation of the health professional, gave the motivation to conduct the study described in this report. Therefore, it has become necessary to rely on the image modalities and analyses to advance the diagnosis, preferentially in an early stage, so that physicians could expedite the treatment prescription. The following points summarize the objectives established for the development of this dissertation:

- Preparation and implementation of a methodology to process and analyses magnetic resonance images of multiple sclerosis patients;
- Exploration and selection of preprocessing, segmentation, and classification processes culminate in an algorithm for assisting the medical experts in identifying multiple sclerosis lesions.

The previous points can be further elaborated, and so more specifically, the purposes for this dissertation are:

- Recognize anatomical, epidemiological, and clinical characteristics of the pathology;
- Investigate physical concepts, conditional factors of the magnetic resonance image technique in neurological studies;
- Describe the main steps of the image processing and analysis methodologies and techniques used in the literature;
- The search of metrics for evaluating the obtained results;
- Build guidelines to promote the implementation of automatic MS lesion segmentation in the clinical environment;
- Illustrate the viability of adopting segmentation techniques to distinguish and classify MS lesions.

The implemented methodology is supported through the experimental study to segmentation MS lesions using publicly available MR images. The information assembled during the research in the existing literature supported the decision for this approach and the techniques used to accomplish it.

1.3 Contributions

The principal contributions that emerged from this project related, on the one hand, to a theoretical plan, where the background analyzes of the literature centering on the approached thematic. On the other hand, in a practical field, where there are included all the experiments performed. Therefore, the contributions reach with this work could be outlined with:

- The contextualization of the stages and the methodology for the analysis of MR images in patients with multiple sclerosis;
- A search for the computational solution in this field, to establish the state-of-the-art;
- The experiment, application, and comparison of preprocessing techniques in order to enhance the images;
- The development of a methodology to accomplish the segmentation of MS lesion from MR images;
- The submission of an article in the *Neurocomputing Journal*, based on literature review, perform to assess the state-of-the-art.

1.4 Dissertation Structure

This report is divided into seven chapters, closing with the bibliographic references. The current chapter introduces the study theme, further explored in the following two chapters - multiple sclerosis characterization and specification, magnetic resonance imaging principles, and machine learning concepts and algorithms. These constitute the theoretical groundwork for the fourth chapter, which describes the methodology pursued to retrieve the state-of-the-art studies using learning algorithms for processing MRI images related to MS subjects. The methodology was developed and implemented, attending to the knowledge acquired in these former chapters, explained in the fifth chapter, and the resources used in the process. The collected results are displayed and examined in chapter six. Finally, the seventh chapter presents the overall conclusions and future work tasks and prospects. In the following points, each of the later chapters is succinctly summarized:

- **Chapter II - Background**

This chapter approaches the basic fundamentals of multiple sclerosis to expand the understanding of the disease. It is organized into five sections where the pathology is characterized, both from a medical and epidemiological perspective, the diagnosis procedures are detailed, and the current possible treatment courses to help patients prolong their life quality.

The MRI principles, hardware, image acquisition parameters, and pulse sequences, which particularize some scans settings, represent the subjects reported in this chapter organized in five sections. Furthermore, the MRI technique application in multiple sclerosis primarily focusing on the lesion characterization and the quality of the images

- **Chapter III - Machine Learning and Deep Learning**

In this chapter, the essential concepts of machine learning and deep learning are explored alongside a brief description of the standard algorithms of both computed methods. Moreover, different kinds of learning are analyzed in order to facilitate the comprehension of algorithms presented in the subsequent chapter. Furthermore, Convolutional Neural Networks are more extensively examined as they are at the core of the proposed methodology.

- **Chapter IV - Image Processing and Analysis - State-of-the-Art**

State-of-the-art is obtained following a structured approach that is specified in this chapter. Prior to carefully analyzing the discover articles, central insights associated with image processing and analysis methods are introduced. Then, the studies were split into two points respecting their objectives. Finally, the main ideas are retrieved and highlighted in the last of the five sections that comprise this chapter.

- **Chapter V - Methodology**

The adopted methodology is presented in this chapter, divided into five sections, describes the preprocessing steps, the lesion segmentation, which includes the model architecture training, and the evaluation metrics used. Moreover, the public datasets adopted to assess the presented methodology are reported, particularly concerning the acquisition protocols, modalities provided, and any processing performed to the released images.

- **Chapter VI - Results and Discussion**

This chapter has described the outcomes of the application of the previously disclosed methodology. Different experiments are conducted and analyzed in an attempt to reach better results and observe the performance of the model under different conditions. Moreover, the results are compared with those obtained with existing algorithms.

- **Chapter VII - Conclusions and Future Work**

The last chapter presents a final summary of the work, highlighting the significance of the pathology, methodology, and results. Additionally, a future work section outlining some experiments or adjustments that could be done in the prospect.

Chapter 2

Background

Multiple sclerosis is a neurological condition that can affect the brain and the spinal cord, distinguished by lesions resulting from the demyelination of the nerve cells, also designated as plaque. The demyelinating plaques lead to neurological signs and symptoms, which worsen with the evolution of the disease.

Medical imaging was one of the most significant technological advances in the medical field. Medical decisions such as disease diagnosis and course of treatment or clinical procedures like planning surgery are supported through these techniques. The acquisition of images using MRI has been growing since this modality does not admit ionizing radiation (X-rays or gamma rays) as a means for image acquisition. The results present an elevated sensitivity for tissue differentiation.

The current chapter explores this neurological disease, with a brief MS filling in the medical history, accompanied by the description of the condition, such as evolution, symptomology, and treatment. Moreover, the elementary principles and fundamentals behind the MRI technique and image acquisition are described and the application of this technique in MS and its specificities. Conclusively, few considerations and inferences are drawn.

2.1 Multiple Sclerosis

2.1.1 Description of the Multiple Sclerosis Disease

Multiple sclerosis is a chronic inflammatory disease of the central nervous system, demyelinating, with an unknown etymology. It is the most usual nontraumatic disabling neurologic illness, affecting people between the ages of 20 to 40, and has a higher predominant incidence in women with a ratio of 2:1 between genders [7]. The lesions, a product of the demyelinating process, appear in CNS, with more frequency in the white matter, spinal cord, and optical nerve. Due to the attack on the myelin, MS can further be considered as an autoimmune disease, given that T helper 1 (Th1), specific cells of the immune system, induce macrophage and microglial activation production responsible for the demyelination process, and axonal loss [8].

The evolution of MS regarding the progression of the neurologic symptoms can be arranged in the following forms described in Table 2.1. These phenotypes were originally established in 1996

Table 2.1: MS phenotype and characterization (Adapted from [9]).

MS phenotype	Description
Clinically Isolated Syndrome (CIS)	A first episode that could become MS if additional activity happen.
Relapsing Remitting MS (RRMS)	Episodes of worsening function followed with total or partial recovery, no apparent progression.
Secondary Progressive MS (SPMS)	Following the RRMS; steadily progressive, with or without relapses.
Primary Progressive MS (PPMS)	Steadily worsening from onset of symptoms without initial relapses or remissions.
Relapsing Progressive MS (RPMS)	Course eliminated in the 2013 revision, patients who were previously diagnosed with RPMS fall into the PPMS type.

and were reexamined in 2013 [9]. This revision included a new course, Clinical Isolated Syndrome (CIS), which refers to a first episode of demyelination that could turn into MS if further activity occurs. Relapsing-Remitting MS (RRMS) is distinguished by alternating cycles of neurological dysfunction, relapses, and clinical stability intervals, remission. These intervals have no definite time varying from individual to individual, Figure 2.1a. The disease course eventually evolves to the secondary progressive MS (SPMS) due to the worsening of neurologic function over time. The preponderance of patients presenting this MS phenotype result from unattended RRMS. SPMS does not describe a steady progression, instead shows intervals with neurological activity similar to RRMS, i.e., with both periods of relapses and remission, as illustrated in Figure 2.1b. Lastly, primary progressive MS (PPMS) is defined by worsening from the origin of symptoms without early relapses or remissions, with a progression also not uniform with some periods of relapses as shown in Figure 2.1c. RRMS is the most common of the present types, discover in approximately 85% of the MS patients [9].

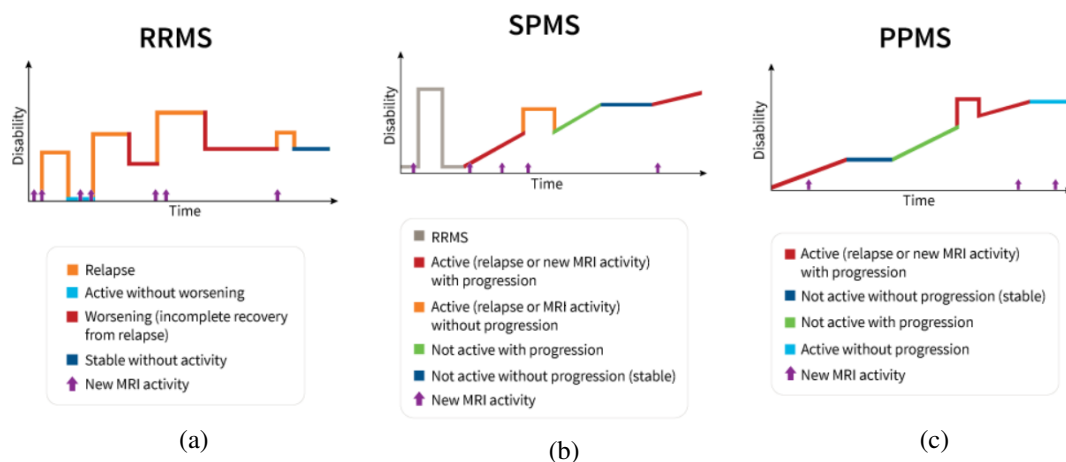


Figure 2.1: Multiple sclerosis disease courses of RRMS, SPMS and PPMS (From [10]).

Although there is no conclusive explanation of the causes that trigger the immune system to attack myelin, studies steer to a combination of genetic and environmental factors.

These epidemiological studies are essential for a better understanding of the characteristics of the disease regarding, for instance, age and gender. This disease, as formerly referred, is twice more common in women than men, with a higher prevalence in young adults. Nonetheless, it can also appear in children, yet the incidence rate is lower (4% to 5% with ages under sixteen), and of then around 97% to 99% have RRMS [11]. These values increase after 18, reaching a maximum of people between 25 and 35 years. Estimations are that 2.8 million persons live with MS worldwide [12]. According to later epidemiological studies, the predominance and frequency of the disease have been rising. While comparing the values from the records of 2013 with those recently acquired shows that the numbers of patients increase by 300 000 thousand globally [12].

Countries with more significant latitude values, particularly more expressive in North American countries, Europe, Australia, and New Zealand, presented higher prevalence rates. Contrastly to equatorial and tropical regions, such as Asia, where the predominance levels decrease, illustrated by Figure 2.2. This information leads research to investigate the influence of sun exposure and vitamin D production in the body with MS patients.

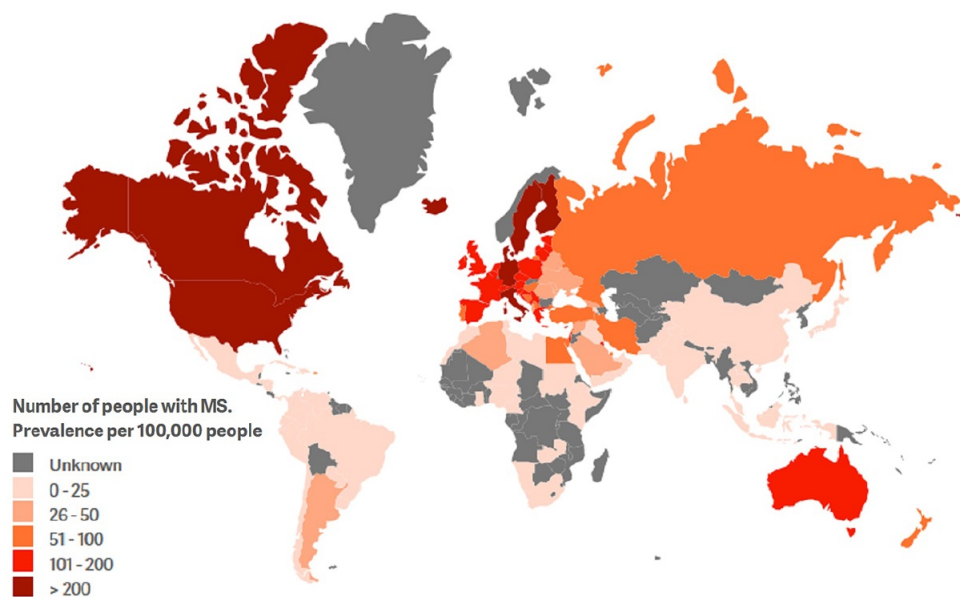


Figure 2.2: Multiple Sclerosis prevalence per 100,000 population by country (From [9])

Besides the environmental and geographic factors, genetics play an essential part, considering that the risk is higher for those with relatives who possess the disease (approximately 13% for all phenotypes) than otherwise [13].

As mentioned earlier, the implications of different factors are not entirely understood. Exposure to sunlight, infectious agents, and lifestyle are suggested as triggering factors [13], yet no concrete conclusions have been made.

2.1.2 Diagnosis Criteria

The MS diagnosis criteria have mutated over time based on clinical features, such as symptoms, signs, evolution, and remission, initially delineated by Charcot in 1868. Only until 1954, the first accurate diagnosis scheme appeared, published by Allison and Millar reporting three categories early, probable, and possible [14].

The fact that disease evolution has various time intervals between manifestations causes early identification and diagnosis challenging. The diagnosis is performed by clinical evaluation with the support of complementary data such as imaging using magnetic resonance, histology with the study of the cerebral fluid (CSF) by lumbar puncture, and EEG with the evoke potentials.

As a result of the technological advances, especially MRI quality, in 2001, the McDonald criteria were incorporated to consider the spatial and temporal dissemination from the MR images. These criteria had subsequent revisions in order to specialize further and increase sensibility in diagnose assessment. The current McDonald criteria reexamined in 2017 [15], is succinctly presented in Table 2.2.

The MR images have been employed in the MS diagnosis, considering a non-ionizing image modality with high sensitivity for tissue differentiation.

2.1.3 Treatment

Despite the numerous breakthrough related to medical treatments, there is no cure for MS. Instead, therapeutics can be split into MS-specific and symptomatic therapies frequently employed in various disease areas. The MS-specific treatments regard therapies that modify the disease by slowing the progression and the relapses. Meanwhile, symptomatic therapies assist in managing the inherent symptoms like weakness, fatigue, vision problems, and numbness or tingling [16].

The present medication has proven to increase efficacy, particularly in the early stages, as the RRMS type, because of the reduced severity of relapses [16]. Medication such as corticosteroids, cyclophosphamide, a cell cycle inhibitor, and intravenous immunoglobulin can all be administered to aid blunt the autoimmune process. Additionally, plasmapheresis can be useful considering that while plasma is filtered, disease-causing antibodies are excluded. Immunosuppressants are also prescribed to decrease the level of inflammatory cytokines in the brain, increase T regulatory cell function, or even block T cells from getting into the brain.

On the other hand, there are not many options available for progressive MS treatment. Essentially the available therapies focus on symptoms management: physical and cognitive recovery therapy including sensory, motor, and cognitive symptoms[17]. Finally, and attending to the epidemiological factors referred above, the vitamin D role has had an increasing interest as a potential alternative course of treatment. Moreover, complementary alternative therapies are dietary supplements like omega 3 and lipoic acid [18, 19].

¹The CNS areas: periventricular, cortical, juxtacortical, infratentorial or spinal cord.

Table 2.2: McDonald Criteria for multiple sclerosis diagnosis (Adapted from [15])

Clinical Presentation	Additional Data required to Diagnose MS
2 or more attacks and clinical evidence of 2 or more lesions;	None
2 or more attacks and clinical evidence of 1 lesion with historical evidence of prior attack involving lesion in different location	None
2 or more attacks and clinical evidence of 1 lesion	DIS shown by one of these criteria: - additional clinical attack implicating different CNS site - 1 or more MS-typical T2 lesions in 2 or more areas of CNS ¹
1 attack and clinical evidence of 2 or more lesions	DIT displayed by one of these criteria: - additional clinical attack - simultaneous presence of both enhancing and non-enhancing MS-typical MRI lesions, or new T2 or enhancing MRI lesion compared to baseline scan (without regard to timing of baseline scan) - CSF oligoclonal bands
1 attack and clinical evidence of 1 lesion	Proceed as for 2 or more attacks and clinical evidence of 1 lesion AND Proceed as for 1 attack and clinical evidence of 2 or more lesions
1 year of disease progression (retrospective or prospective)	DIS shown by at least two of these criteria: - 1 or more MS-typical T2 lesions (periventricular, cortical, juxtacortical or infratentorial) - 2 or more T2 spinal cord lesions - CSF oligoclonal bands

DIT= Dissemination in Time **DIS**= Dissemination in Space

2.2 Magnetic Resonance Imaging

2.2.1 MRI Principles

Originated from the nuclear magnetic resonance (NMR) phenomenon experimentally reported by Bloch and Purcell in 1946, it rapidly evolved, and in 1980 the first MR image was produced [4].

MRI explores the happening of nuclear MR, whereby atomic nuclei exposed to a strong magnetic field absorb and reemit electromagnetic waves at the “resonant” frequency, the characteristic frequency, settling into the radiofrequency spectrum [5].

2.2.1.1 Magnetic Characteristic of Proton

Some atomic nuclei, such as hydrogen or phosphorus, possess a property designated as "spin" that describes the spinning around the own nucleus axis, not in the literal sense. Only through its integral parts produce a magnetic moment, generating a magnetic field with the respective north and south poles [9]. The magnetic moment of a single proton is not detectable. However, more significant numbers, like hydrogen atoms existing in water, fat, and complex macromolecules

within tissues, have an arbitrary orientation of their protons, Figure 2.3a, which is not sufficient to observe a magnetization of the sample.

When placed in a strong magnetic field B_0 , the protons align with the field in a parallel, a low-energy state, and anti-parallel, a high-energy state, direction, as illustrated in Figure 2.3b. The thermal energy that causes the random alignment of the protons also interferes with the new distribution. Furthermore, at equilibrium, there is a fraction with more protons in the parallel orientation [20].

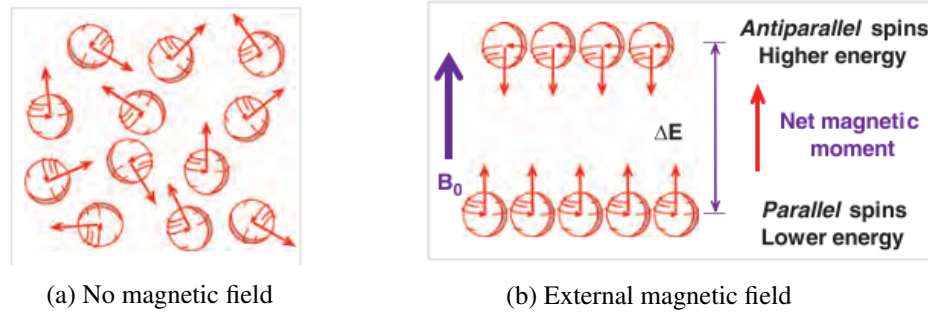


Figure 2.3: Simplification of the protons without and with an external field. 2.3a without the external magnetic field the protons take a aleatory orientation. 2.3b illustrate the alignment of the protons when passed by an external magnetic field, assuming either a parallel or anti-parallel orientation (From [20]).

In addition to the energy separation between the two directions, the protons undergo a torque perpendicular to the B_0 , making them rotate or precess around the magnetic field axis. The velocity associated with this rotation is designated as the Larmor frequency, describe in the equation [9].

$$\omega_0 = \gamma * B_0 \quad (2.1)$$

Where ω_0 represents the angular precessional or Larmor frequency (HZ or MHz), and γ the gyromagnetic ratio singular to each element.

2.2.1.2 Relaxation Times

Relaxation is the process in which spins release the energy gained from a RF pulse[21]. The MRI signal is affected mostly by the various relaxations:

- T1 relaxation time (spin-lattice or longitudinal relaxation) - is the measured recovery time of the net magnetization vector to its base state in the direction of the main magnetic field B_0 subsequent to the RF pulse excitation[20]. T1 is also defined as the interval needed for the system to recover to 63% of its equilibrium value follow excitation [8];
- T2 relaxation time (spin-spin or transverse relaxation) - T2 relaxation time (spin-spin or transverse relaxation) - refers to the continued dephasing of spinning dipoles after the RF

excitation leading to the decay in the magnetization in the transverse plane. The T2 relaxation time is the period it takes from dephasing to decay the signal to 37% of its initial value[8, 22].

Both relaxation times are affected according to the type of tissue, molecular motion, and size. When comparing T1 with T2 values, the first is longer than the second, with T1 being 5 to 10 times bigger than T2 [20], as demonstrated by the Table 2.3. For instance, the T1 value of fat tissue is smaller than CSF, which similarly occurs to T2.

Table 2.3: T1 and T2 relaxation time constant for different tissues (Adapted from [20]).

Tissue	T1, 0.5T (ms)	T1, 1.5T (ms)	T2 (ms)
Fat	210	260	80
White matter	500	780	45
Gray matter	650	900	100
Cerebrospinal fluid	1800	2400	160

2.2.2 Image Acquisition

2.2.2.1 Hardware

Like its imaging modalities peers, the MRI scanner possesses a computer to program the acquisitions, store and visualize the images. Additionally, at a lower level, primary magnets, the gradient system, and the RF coils. Primary magnets are responsible for creating a uniform magnetic field. The gradient system is composed of a gradient coil that produces linear-gradient magnetic fields superimposed over the field produced by the magnets. The RF coils attend as transmitter and receiver of the signal [5].

There is a variety of MRI scanners regarding the field strengths. For example, with clinical purposes, the most common range goes from 1.5T to 3T scanners, whereas the 7T scanners are used for research [5].

2.2.2.2 Basic Acquisition Parameters

Differences in the T1 and T2 relaxation times along with proton density of the tissue are huge influencing factors of the MRI contrast sensitivity images[20]. The following points describe machine-based parameters:

- **Time of Repetition, TR** - the time interval between an excitation pulse and the next pulse. In this interval, T2 decay and T1 restoration, i.e., TR, establish how much longitudinal magnetization recovers between pulses, measured in milliseconds.
- **Time of Echo, TE** - defines the period between the excitation pulse and the peak of the signal induced in the coil.

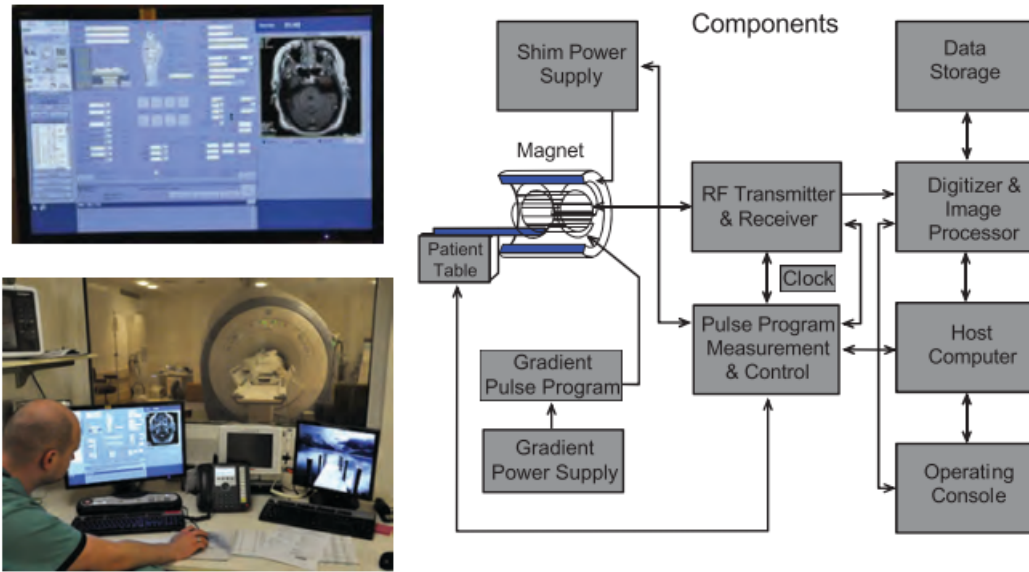


Figure 2.4: MRI system with operator controls and visualization panels, and the various subsystems [20]

- **Time of Inversion, TI** - time elapsed between the initial 180° pulse and the 90° readout pulse. This 180° pulse creates maximum tissue saturation.

2.2.2.3 Pulse Sequences

Pulse sequences are programmed combinations of signal acquisition, repeated RF pulses, and magnetic gradients [23, 5]. These pulse sequences are designed to obtain optimal signal contrast for specific purpose. Exist three significant sequence types gradient-echo, spin-echo, and inversion recovery. Moreover, joining these models with the spatial location approaches creates "contrast-weighted" images [20].

A. Spin-Echo Sequences

A Spin-Echo (SE) sequence uses two RF pulses in a TR interval, a 90° RF pulse and a production signal, followed by a refocusing 180° RF pulse to create an echo.

Implementing the spin-echo approach exist some specific pulse sequences like T1-weighted (T1-w), proton density (PD), T2-weighted (T2-w).

The difference between approaches is the emphasis given to the parameter that names the method—for instance, T1-w is devised to produce a contrast mainly using the T1 properties of the tissues. For that, short TR is managed in combination with small TE to limit T2 decay while acquiring the signal. The PD depends on variations in the number of magnetized protons per unit volume of tissue. Therefore this difference is performed by reducing the influence of T1 recovery and T2 decay, consequently selecting long TR and short TE values to minimize T1 and T2 differences, respectively. Similarly, T2-w operated by reducing T1 with long TR and highlight

T2 with long TE. Examining the approaches, PD performs the highest signal intensity and largest Signal-to-Noise Ratio (SNR). However, image contrast is comparatively low, opposite to T2-w images that present high tissue contrast.

B. Inversion Recovery

Inversion Recovery (IR) highlights T1 of the tissues by prolonging the longitudinal recovery amplitude by a factor of two. A primary 180° excitation pulse inverts the longitudinal magnetization, following a control delay, the time inversion, a 90° pulse rotates the recovered portion of magnetization. Subsequently, a second 180° pulse is executed at half of the TE, which refocuses the transverse magnetization as an echo at the end of the TE time.

Applications of this type of pulse are the Short Tau Inversion Recovery (STIR) and the Fluid Attenuated Inversion Recovery (FLAIR).

STIR uses short T1, and the magnitude signal is positive as a result of signal processing. The acquired images have a low intensity for material with short T1, which might be useful for suppressing tissue signals. For instance, with a slightly smaller T1 than other tissues, as seen in Table 2.3, fat will appear darker. Contrastly, considering the longer T1 of tissues like CSF, the signal levels could ruin the inversion recovery images. In FLAIR sequences, these water-bound tissues are lessened by selecting T1 close to the bounce point of CSF, permitting an enhanced visualization of the neighboring structures.

C. Gradient-Echo Sequences

Gradient-Echo (GE) uses a magnetic field gradient employed in one direction and then reversed to cause an echo, rather than the 180° inversion pulse as in the SE sequences.

Specific applications of this type of pulse sequences are the gradient echo sequence with long and short TR, and coherent GE.

The coherent GE presents the timing of the RF pulse with the dephasing and rephasing by reversal of gradient polarity to create an echo at a controllable time TE for the frequency encode gradient, where identification of proton location based on the frequency is performed.

2.2.3 Application in Multiple Sclerosis

No individual test exists that diagnoses MS, not even MRI. Nonetheless, as reported in the previous section, MR images are crucial for multiple sclerosis diagnoses and monitoring progression. Its relevance is even highlighted in the McDonald diagnosis criteria since it requires these medical images to evaluate the disease better, significantly based on visualizing the dissemination of lesions in space and time [24].

Since evaluating lesions, either the evolution of lesions or the appearance of new ones is an integral part of assessing the disease, selecting different MRI sequence types is needed to improve the sensitivity contrast for lesion characterization, such as T1-w, T2-w, and FLAIR are strong examples.

MS lesions typically have a round to oval shape and diameter size that vary from millimeters to one or two centimeters. They develop in both hemispheres, yet the pattern is not symmetric in the early stages. The lesion can occur in any CNS regions. Multiple sclerosis alters white matter regions like the periventricular, juxtacortical, and spinal cord.

FLAIR, especially 3D images, is reported as the most sensible sequence type since the MS lesions in white matter are T2-hyperintense, i.e., lesion intensity values in this type of image higher for most CNS regions. Additionally, to confidence confirmed T2-w, PD-weighted, or T1-weighted magnetization-prepared acquisition with gradient echo (MPRAGE) sequences could be also be considered [25]

2.2.4 Image Artifacts

Artifacts present as positive or negative signal intensities that do not correctly represent imaged anatomical structures. Some of these could be insignificant, yet, others could lead to incorrect assessment through the obscuring, mimicking anatomy, or illness [20]. MRI artifacts are categorized into three classes - caused by the machine, the patient, and signal processing.

Machine artifacts are related to deformity or misplacement of anatomy befalling if the magnetic field is not entirely uniform. Field distortions are produced by the presence of ferromagnetic objects like implants, prostheses, or items transported by the patients, such as coins and keys. The RF pulses and other frequencies of the machine work in the same bandwidth that standard equipment, TV, or computers, since the precise position and spatial extent rely on the resonant frequency of the scanner, these RF interferences cause spatial variations with contrast-reducing artifacts [20].

Of the artifacts related to the patient, the most prevalent is motion resulting from body function such as breathing, cardiac movement, blood flow, and CSF pulsation or movements of the patients, voluntary movements [26]. Moreover, chemical shift artifacts seem as dark or bright intensities at the liquid-water, i.e., fat versus water, interfaces caused by the Larmor frequencies differences of the particles [20, 26].

Furthermore, close to definite edges and elevated contrast transitions appear as multiple alternating bands of bright and dark signal, called ringing, that gradually weakens with distance. These artifacts are denominated as ringing artifacts, and they result from a short sampling of high frequencies.[26].

2.3 Summary

This chapter aims to facilitate the comprehension of both the pathology and the medical imaging technique.

Multiple sclerosis is a cause for concern, considering the degradation and discomfort that it causes in its patients. Highlight that the disease has a higher incidence in young adults affecting more the female gender. Regions with more significant latitudes, i.e., redrawn from the equator and

having direct kin diagnosed, show higher prevalence rates, forcing genetic and environmental risk factors. It is defying to ascertain an early diagnosis due to the progressive nature of this condition. This process is carried with defined criteria that included additional medical information such as MRI scans. The analysis of these images is a crucial step to a fast evaluation to be adequately treated. The current treatments for MS are MS-Specific, assisting in the progression and worsening of the disease and symptom management.

MRI is one of the most versatile medical image techniques, either from an application perspective, practiced in children, pregnant women, and for all body scans, or from an image acquisition view with a variety of sequences chosen in a way that best fits the application objectives. The images are acquired based on signal production from magnetic fields and radiofrequency pulses interacting with protons from the anatomic structures, like hydrogen.

Regarding MS, this technique is fundamental to determine or evaluate the disease. Its progression, with T2 and FLAIR, was reported as having a higher contrast sensitivity for the MS lesions than other sequences.

MR images can be deemed unusable due to the presence of some artifacts. It is essential to understand its causes to correct or attenuate, permitting the medical professionals to perform their assessments with high confidence and efficiency.

Chapter 3

Machine Learning and Deep Learning

Machine learning is a field of research that gives computers the capacity to learn without being specifically programmed. It has evolved and is applied for numerous purposes in most various areas. Deep learning is a subfield of machine learning that further explores methods that approximate the objective of machine intelligence.

This chapter explores machine learning, the concept, learning types, algorithms, and applications. Additionally, deep learning is discussed, singularities of each strategy are explored, and the crescendo of studies with this approach in the various fields.

3.1 Machine Learning

Machine learning (ML) is a subfield of artificial intelligence (AI) focusing on analyzing methods of collecting information from data [27]. ML aims to facilitate the design of intelligent systems whose implementation can be effective to solve real-life situations [28].

ML first appeared in 1950 and was developed for playing the game of chess. Afterward, it has evolved and has numerous applications in most various areas, like finance and entertainment.

Although the versatility of ML algorithms has been tested and improved over the years, its complex nature and the dependency of data produce some challenges [28]. Among concerns of machine learning are:

- Data acquisition and processing, which are time-consuming procedures.
- Data quality leads to problems with processing and feature extraction.
- The lack of sufficiently large datasets for training the algorithms.
- Overfitting and underfitting reflect precarious settings of the model and producing a high number of features - the curse of dimensionality.

3.2 Learning Scenarios

According to the approach of the models, data types, data access, and flow of processing training data, there are different strategies related to the learning process.

The two major groups are supervised and unsupervised learning. Overall, the significant difference is in handled data type. Supervised uses known and labeled data, contrary to unsupervised learning that learns from unlabeled data [27].

Iteratively, supervised learning algorithms predict the training dataset and are correct by the supervisor before achieving an appropriate performance level. On the other hand, unsupervised learning has no supervisor correcting the data, deriving structures from the input data to obtain results.

Given all these variations, supervised learning is a tool for addressing classification and regression problems, whereas unsupervised is mainly used for data clustering, as presented in Table 3.1, with a resume of some significant differences between the two learning standards.

Table 3.1: Differences between supervised and unsupervised learning (Adapted from [27]).

	Supervised	Unsupervised
Input data	Known and with labels	no labels
Computational complexity	High	Medium
Accuracy and reliability of the results	High	Medium
Problem ask	Classification and regression	Clustering and dimensionality reduction
Algorithms	KNN, SVM, Logistic Regression, Random Forest, among others	K-means

Furthermore, there is a compound of these previous examples, semisupervised learning, where the training set includes both label and unlabeled data. This attempts to build a model that produces more reliable performance than with an entirely labeled data configuration.

More recently, a new kind appears that enables the model to learn in a dynamic environment, designated as reinforcement learning. The learner gathers information from the environment, which is influenced and gives a reward for each interaction. The objective is for the model to optimize the rewards after some iterations [28]. This configuration offers a more accurate scenario of the natural learning process.

3.3 Deep Learning

Even though ML algorithms evolve to process vast data, they were still constrained by relying heavily on data features that require human expertise to navigate the features which the algorithm should be searching for [29]. All these reasons, combined to bring the ML technology closer to "creating intelligent machines", led to the Deep Learning (DL) notion designed [28]. DL is spurred by the human brain architectures, particularly the connection between neurons mimicked by the neural networks.

DL algorithms are characterized by adopting a hierarchical layered architecture, where the data is learned in distinct layers. It attempts to learn between the original data and the output while working to understand the representation of the data. Consequently, converting DL into a unique system compared with ML, considering that it does not require extensive expertise in areas such as feature extraction and efficiency concerning performance and scalability [28].

3.4 Machine Learning Algorithms

3.4.1 Support Vector Machine

Support Vector Machine (SVM) is one of the principal machine learning algorithms recognized for its accuracy and robustness. SVM is a supervised learning model, and it is applied in classification problems.

It aims to create an optimal hyperplane separation that maximizes margin, decision boundary and reduces the number of incorrectly classified samples [27]. The margin is the smallest space between the hyperplane point and the near located data points, the support vectors [28]. This algorithm is versatile because, in some non-separable cases, it is still possible to overcome that difficulty using a linear classifier that projects the dataset into a higher dimension where SVM can separate it. Nonetheless, these projections require caution because, with the increase of dimensions, the amount of data augments leading to the risk of overfitting and higher computational complexity [27].

3.4.2 Logistic Regression

As previously demonstrated in Table 3.1, Logist Regression (LR) is associated with classification problems, regardless of what the name might suggest LR works with the probabilities of a given event/class occur considering the input values [30]. This algorithm presents some advantages, such as simplicity of implementation and computational efficiency. However, it presents difficulty in solving non-linear problems and tends to overfit, mainly if independent variables are unknown.

3.4.3 Decision Tree

The decision tree (DT) is a supervised ML algorithm that tackles the problem by continuously dividing data based on unique parameters. The evaluation and implementation of the parameters

are carried out at the "leaf" stage, followed by the differentiation at the node. This approach is easy to define and handle categorical and quantitative values. However, it is difficult to monitor the size of the tree, which can lead to a mistake in sampling or even overfitting [30].

3.4.4 Random Forest

Random Forest (RF) is an ensemble of DT algorithms, the "forest". The ensemble learning scheme improves the overall outcome by combining different algorithms to address the equivalent problem [30, 31]. The algorithm requires two initial parameters the number of trees and the number of randomly selected features. If the number of trees is expanded, the RF creates high variance and low bias.[31]. In a general view, an understandable decision-making process, easy implementation, and robustness to outliers and noise are some of the highlighted advantages of this method.

3.4.5 Linear Discriminate Analysis

Linear Discriminate Analysis (LDA) is used as a dimension reduction agent and a sturdy classifier. This algorithm discovers a linear combination of features that differentiate the classes in the study, which are presumed to be normally distributed [28].

3.4.6 K-Nearest Neighbour

KNN is a classification algorithm that groups a given dataset into classes. For that uses the similarity between data, which is evaluated considering the K neighbors majority classification. The neighbors are established based on distance metrics which determine the closeness, for instance, Euclidean distance[28]. Its implementation and flexibility regarding the classification process are some of the advantages. On the other hand, it depends on the distance between the k-nearest neighbors, which for bigger datasets would lead to high computational times and overfitting, and requires the value of k neighbors in advance [30, 32].

3.4.7 K-Means

K-means is a clustering unsupervised ML algorithm, where K represents the number of clusters. Its objective is to obtain the k-cluster in data by iteratively minimizing the measure between the cluster center of the class and the given observation [27].

One issue for this algorithm is that they require K as an initialization, which influences the convergence speed and performance, being computationally expensive, especially for large values of K.

3.5 Deep Learning Algorithms

3.5.1 Deep Neural Network

Deep Neural Network (DNN) is similar to conventional neural networks, except these have many hidden layers, oppose to the one or two layers presented by the traditional networks [28].

3.5.2 Restricted Boltzmann Machine and Deep Belief Network

Restricted Boltzmann Machine (RBM) is a generative stochastic artificial neural network that learns how to represent input data by approximately estimating its likely distribution. By stacking various RBMs, a Deep Belief Network (DBN) is designed. Its key component is a greedy layer-wise learning approach that optimizes the weights depending on the size and extent of the networks [33, 34].

3.5.3 Convolutional Neural Networks

Convolutional Neural Networks (CNN) focuses on mirroring the functionality and operation of our visual cortex [28]. CNN is mostly used in image processing for recognizing spatial and temporal features from images. Its structure is characterized as a connected collection of layers running convolutional filters that attribute weights, and biases to single features [33, 35]. Besides the convolutional layers, CNN possesses other layer types and specificities, further detailed in Section 3.6.

An example of a CNN-based architecture is U-Net, composed of two-part, convolutional, and decoder, shaped in a U form; it is commonly used for segmentation problems. The first part extracts features, and the second part reconstructs the segmented image [36].

3.5.4 Generative Adversarial Network

Generative Adversarial Network (GAN) is a model capable of synthesizing new data by making a generative and discriminatory network works collectively or against each other. The generative network produces artificial data, while the discriminatory network seeks to differentiate between synthetic and real data [35].

3.6 Convolutional Neural Networks

Convolutional Neural Networks (CNN), as formerly referenced, are very comparable to conventional Neural Networks (NN), with neurons that possess learnable weights and bias, mirroring the functionality and operation of our visual cortex [28]. They are mainly adopted for image processing to recognize spatial and temporal features.

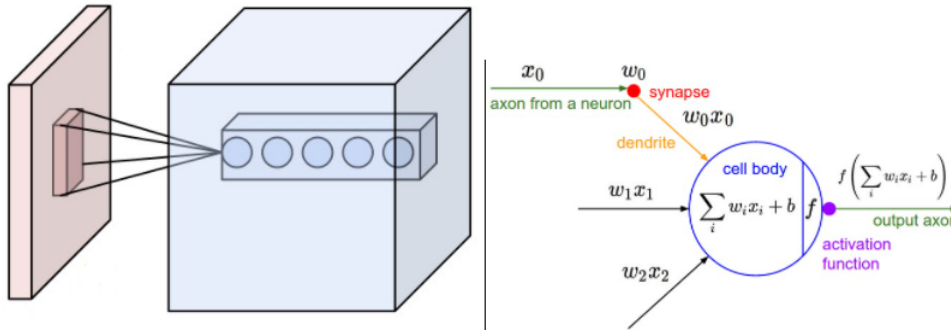


Figure 3.1: Comparison between a CNN with the human brain functions (Adapted from [37]).

3.6.1 Architecture Overview

Unlike a regular NN, the layers of a CNN have neurons arranged in various dimensions, for instance, width, height, length, and depth, to process 3D images. Its structure is characterized as a connected collection of layers running convolutional filters that attribute weights and biases to single features [35, 33]. Each layer comprises neurons that are fully connected to the entire neurons in the preceding layer, and in a single layer, they operate autonomously, not sharing any connection. Besides the convolutional layers CNN might possess pooling layers that diminish the image size and additional types further explored in the following subsection.

The CNN architecture is very versatile and has been numerous CNN-based systems develop in the past years, like the formerly presented U-Net model [36].

3.6.2 Type of Layers

A. Convolution Layers

The convolution layer (CONV) employs filters to execute convolution operations while scanning the input in terms of dimensions. The filter size and stride are two of its hyperparameters that can be manipulated to obtained different outcomes. The result of these operations is nominated as a feature map, or activation map [38].

B. Pooling Layer

Pooling layers are usually embedded within consecutive Conv layers to lessen the image dimensions, which consequently reduces the number of parameters and computation in the network [37]. This reduction is accomplished since pooling layers blended adjacent units, pixels or voxels, of an image region into a unique value [28]. The neighboring pixels are generally chosen from the square matrix, the number of units merged and the output varies depending on the task. These tasks are primarily Max and the Mean(average) pooling. Max pooling involves using the largest value, rather than the average, as in mean pooling [39]. The Figure 3.2 illustrates the pooling process for both mean and max polling, using a 2x2 matrix on the input.

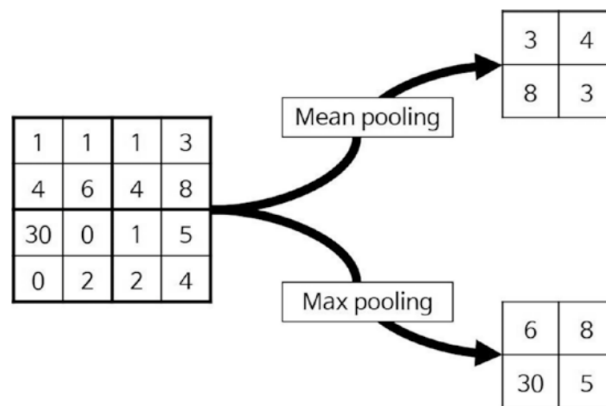


Figure 3.2: Example of the application of different pooling methods.

C. Fully Connected

The fully connected layer (FC) works with a flattened input, with each input connected to all neurons. FC layers, when present, are typically found near the end of CNN architectures and can be used to optimize objectives such as class scores [38].

D. Batch Normalization

They commonly included immediately after a Conv layer and before non-linearities, is based on a technique developed to suitably initialize the network by explicitly forcing the activation functions during a network to practice a unit gaussian distribution at the start of the training [40].

E. Dropout

Overfitting is a constant issue in machine learning tasks. The dropout technique was created to prevent it, where the term expresses the dropping of units in the network [41]. In practice, the units are not drop out but momentarily detached from the network, along with their income and outgoing connections, as shown in the Figure 3.3. The detached units during the process are selected randomly, with a probability p that can be defined based on the validation set or just a fixed value. Dropout is only used during the network training since it would not be reasonable to average the predictions from many smaller models resultant from its application.

3.6.3 Filter Hyperparameters

As mentioned previously, the filter used in the convolutional layer possesses some parameter that impacts both the output and the performance of the layer and further details the following points.

- **Dimensions of the filter** - also referred to as Kernel Size, is the size of the filter.
- **Stride** - The represents the number of image units the window moves after each operation.

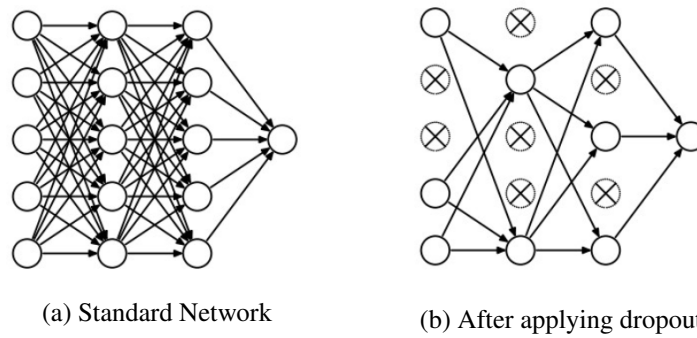


Figure 3.3: Highlight the difference of the application of the dropout technique. (From [41])

- **Padding** - expresses the method of adding a value, typically zeros, of units to each side of the input boundaries [38], guaranteeing that the filter process the image edges. These values can be specified or automatic define by one of the following modes:
 - **Valid:** no padding is made, and drops the last convolution if the dimensions do not match.
 - **Same:** padding is added, which allows the output image to has the exact dimensions as the input.

3.6.4 Activation Function

The activation function is used to calculate the output response of a neuron. To obtain that reply, the sum of the incoming signal is applied with an activation, as illustrated in Figure 3.1. This function is implemented throughout the neurons inside of the same layer. The activation function can have several representations, linear, nonlinear, the most well-known: Sigmoid function, Rectified Linear Unit (ReLU), and Softmax. The sigmoid is defined by following the mathematical equations:

$$f(x) = \frac{1}{1 + e^{-1}} \quad (3.1)$$

Furthermore, ReLU has some variants that can best fit the solution intended, as presented in the Figure 3.4.

The Softmax receives as input a vector of scores and outputs the probability of that same vector. For that particularity, it is seldom perceived as a generalized logistic function.

3.6.5 Network Training

Keeping learning in a particular environment is one of the most significant capabilities of an NN. The learning process is an interactive procedure that continuously adjusts the weights and bias estimated, concluding when the network finds a generalized solution for the task. The environment

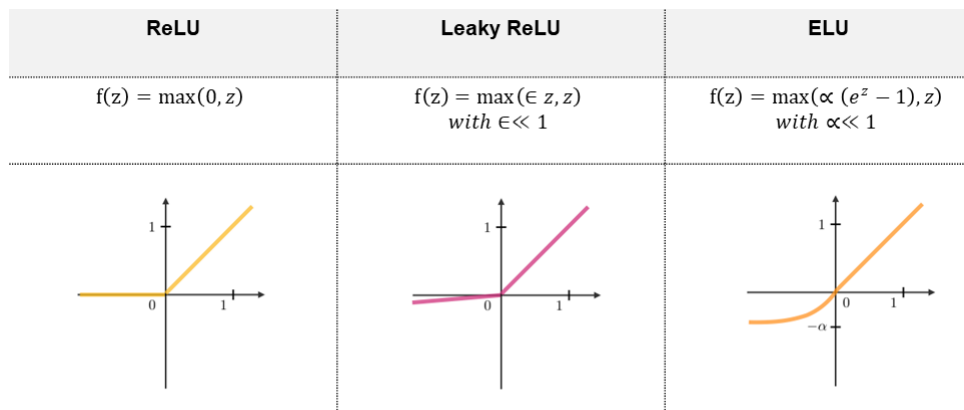


Figure 3.4: Scheme with different types of ReLU functions with mathematical equations and graphs (Adapted from [38]).

reflects the combination selected between the training parameters describe in the following points [42]:

- **Learning Rate** - coordinates the update of the weight at the completion of each batch.
- **Momentum** - controls the value to let the previous update influence the current weight update.
- **Number of epochs** - the iterations of the entire training dataset to the network during training
- **Batch-size** - the number of patterns given to the network before the weights are updated

3.6.6 Optimization

Computed learning, either machine or deep learning, aims to reduce the difference between the estimated and the actual outputs, signifying that the intention is to minimize the loss function by optimizing the weight of the network. This minimization is accomplished with the assistant of optimizer algorithms. In addition, there are various methods available, like gradient descent, one of the most known and more straightforward, being its principles at the core of the majority of the optimizer, updating the network parameter moving in the gradient direction that leads towards a minimum. Furthermore, exist more complex algorithms, like *Adadelata* or *Adam*, which are further explored in the following points.

- **Adadelata**

Adadelata, method was derived from a pre-existing one, ADAGRAD, to succeed two of its principal barriers, the rapid decay of learning rates during training and the manual picked global learning rate [43]. Therefore, the concept behind this is to use a restricted window size to add the previously squared gradients, rather than from intervals ranging from 1 to t (time steps), which

led to a decrease in the learning due to the increase of addition of the gradients. This process is accomplished using exponentially weighted averages.

- **Adam**

Adam represents a short notation for Adaptive Moment Estimation, which Diederik Kingman and Jimmy Lei Ba developed [44]. Similar to the previous optimizer, Adam also uses exponentially weighted averages. It is a gradient-based optimization of stochastic objective functions supported on adaptative estimates of lower-order moments. i.e., Adam determines the specific adaptive learning rate for each parameter from estimates of the first and second moments of the gradients. Its implementation is straightforward and demands low memory.

3.7 Summary

This chapter presents an overview of machine learning and deep learning, two growing methods widely used in various areas, particularly health. There are two significant groups: supervised and unsupervised learning regarding the learning scenarios, being the main difference between them the type of data used. Furthermore, reinforcement learning is a more recent approach that grows more similar to the actual learning environment. The machine learning algorithm can have many purposes, such as regression or classification. However, they require human expertise for a task like features extraction, unlike deep learning algorithms that are more automated. Overall, both approaches present robust and accurate algorithms.

Moreover, the CNN algorithm is a versatile algorithm specifically designed for the image analysis task. The convolutional layers, notably this algorithm, are the main ones responsible for analyzing the features. These layers have some parameters such as padding, stride, and kernel size that influence the output results. As for training, standard throughout neural networks, learning rate, number of epochs, and batch size are some of the parameters that could be tuned in order to achieve a better network performance.

Chapter 4

Image Processing and Analysis - State of the Art

This chapter discusses the state-of-the-art using image analysis techniques on neurological magnetic resonance images of multiple sclerosis patients. It starts with a presentation of a few common concepts for image analysis methods. A description of the search methodology, databases, and keywords used, followed by a report of the algorithms and their evaluation metrics. Lastly, an exposition of the methods used in the various studies and concluding considerations.

4.1 Overview

Each study presents its approach with fewer or more particularities in order to achieve its goals. Nonetheless, all of them show some resemblance in the pipeline or organization of the methodologies, divided into the following steps: image acquisition, pre-processing, segmentation, and quantification or classification. All these stages are illustrated in the scheme of the Figure [4.1](#).

The acquisition represents the obtaining of the images. The acquisition of medical images is the process to obtain images of the human body through a medical imaging technique like Computed Tomography (CT), Ultra-Sounds (US), x-ray, or in this case, magnetic resonance machine. Each imaging technique has distinctive specifications and hardware.

Medical images are regularly degenerated, for instance, over noise caused by intervention from the sources during the acquisition process. Attenuation or correction of these defects can be accomplished with image processing techniques. The images obtained from the previous step are processed to enhance the quality and suppress irrelevant information, forming the pre-processing step. The main objective of this step is to obtain clear images facilitating the subsequent steps [\[6\]](#).

Segmentation represents the separation of structures of interest from others and the background. In medical imaging, segmentation is used to analyze anatomical structures and diseased regions. It is also adopted as a primary step for visualization and condensation. Segmentation is

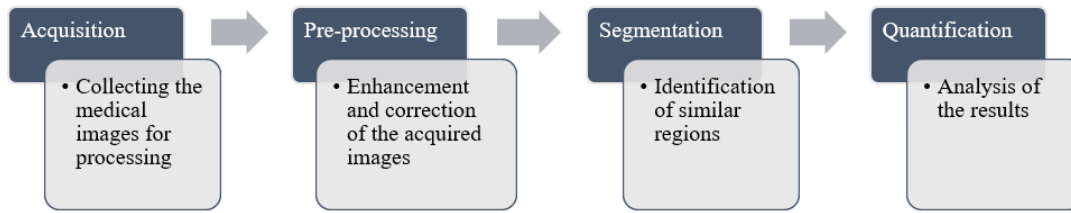


Figure 4.1: Usual pipeline followed in image analysis and processing studies, divided into four main steps: acquisition, preprocessing, segmentation and quantification.

achieved by identifying all image units of the structure or by delineating the boundary of the structure with surrounding tissues. The algorithms used in this step work on the intensity or texture changes of the images [6].

An interpretation of the results is essential to understand the performance of the algorithm, and in light of the study theme, examine the results. These analyses are of broader interest while investigating medical images, which profit with accurate, agile, and unbiased computational resource measures. These quantitative measurements provide the analysis of structure and function in healthy and unhealthy situations using many aspects of the data, such as tissue shape, size, and texture [6].

4.2 Method

The search was executed in Pubmed and Scopus databases with the subsequent keywords "multiple sclerosis", "machine learning", "deep learning", "MRI", "magnetic resonance", "diagnosis", "classification" and "segmentation". The previous keywords were organized in the following query, ("multiple sclerosis") and ("machine learning" or "deep learning") and ("MRI" or "magnetic resonance") and ("diagnosis" or "classification"). Consequently, 127 documents were obtained at the time of the search, of which only 65 were considered relevant for additional analysis. The remaining 62 documents were excluded based on the following criteria: MRI was not used as the medical imaging modality; MS was not one of the pathologies in the study; the document type was a journal editorial or a conference meeting review. Of the papers initially chosen, 8 were review reports and are further analyzed in the following section. Of the 57 original studies selected based on abstract, 48 articles were preserved after due consideration. They are focused on MS and MR imaging analysis using either machine or deep learning algorithms in the process. The entire methodology is summarized in the PRISMA diagram presented in Figure 4.2.

4.3 Related work

Multiple sclerosis identification heavily depends on lesion segmentation. Therefore of the selected 48 papers, it was possible to further separate them according to their methodology objectives,

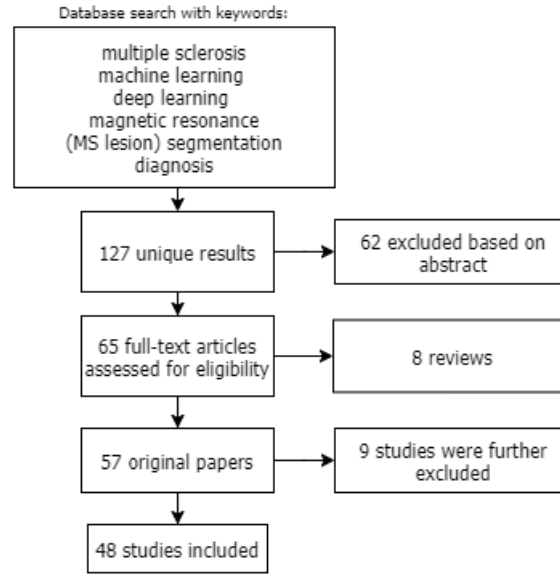


Figure 4.2: PRISMA diagram describing the results of the executed literature search.

which resulted in two main topics, lesion segmentation, and disease diagnosis or classification. Consequently, 24 documents had as a goal the lesion segmentation and 24 the disease classification, resulted from that division.

Furthermore, eight reviews were found in the performed search; a few of them include both the segmentation and classification of the chosen works. In essence, most of these reviews focus on the role of ML to develop tools to assist in diagnosing multiple diseases, the designated CAD (Computed-Aided Diagnosis) systems. Either focusing on investigating various neurological diseases [45, 46, 47] such as Alzheimer’s disease and Parkinson’s, or merely concentrating in MS [48, 49], the authors outline the importance of the data and its quality. The scarcity of large datasets, particularly with multidimensional patient data, represents a limitation to implementing machine learning methods. Moreover, data are frequently biased against specific demographic groups, restricting the generalization of results [50], being one of the reasons missing data, an occurrent issue when dealing with medical data [45]. Beyond a diagnosis, there is an urgency for an early diagnosis, especially for neurodegenerative disease, in which the neuronal degeneration process complexity bestows a challenge in generating the robust tool to assist [50]. Some of the most common AI techniques, such as SVMs, Logistic Regression, Random Forest, and CNNs, have been extensively adopted in the MS studies, either for the detection and segmentation of both lesions and tissues, as for course prediction and pathology classification [49].

Ultimately, all authors agree that applying these automated learning methods will facilitate the prediction of the disease for experts, leading to improved patient outcomes.

With a different aim, in [51] the authors describe MS neuropathology and the MR images for assessing disease severity and treatment response. MS affects both grey and white matter, with higher expression in the last, which results in a further categorization acute, active/inactive, and

chronic according to type and intensity of inflammatory response. For MRI, the authors discuss the higher sensitivity presented by the technique in disclosing disease activity, denoting either a clinical sign of relapses or disability progression. The versatility of MRI is impressive, allowing the collection of diverse information. With a variety of sequence type beyond the conventional MRI, for instance, the authors [52] investigate proton magnetic resonance spectroscopy (H-MRS), which has been utilized to extract biomarkers to understand MS, especially the numerous metabolic alterations of its lesions.

From publication dates, it is observed that ML was widespread in the last decade. Nevertheless, deep-learning has been more implemented in the previous 3 to 4 years.

The dataset source is distinguished as private studies that either acquire their data or acquired from a clinical trial that is not publicly available, obtained from challenges or databases such as ISBI 2015 [53], MICCAI 2008 [54], and CombiRX [55].

Some articles mention either ground truth was obtained from experts (radiologist or neurologist) or using existing software regarding the lesion segmentation. The ground truth source is quite relevant for comparing the obtained results with those from reliable origins.

4.3.1 Preprocessing

Despite the learning method employed, input data might require preprocessing, especially when the input is medical imaging. Nearly every selected paper reported a pre-processing step in the methodology. Skull-stripping, image registration, bias field correction, intensity normalization, and noise reduction are the most reported current processing techniques, which are further detailed in the following points:

- **Registration:** the MRI sequence of a patient is registered to the same space. The objective is to geometrically align the different sequences so that the images could be properly analyzed.
- **Skull-stripping:** the skull is removed from the images reducing unnecessary information, an error due to the intensity of the skull in some sequences type. Apart from the skull, some other structures like eyes orbits are excluded.
- **Bias field Correction:** also designated as inhomogeneities intensity correction refer to variations in the same tissue caused by non-homogeneity of static the magnetic field or applied in the scanner.
- **Noise Reduction:** during the acquisition process, noise could be introduced in the images, to attenuated it, filter methods are applied;
- **Intensity Normalization:** the normalization of the intensity values is essential to guarantee that all datasets, regardless of the hardware used to acquire the images, are on the same range of values.

In most instances, these procedures were implemented with the support of existing software as ITK [56], FSL toolbox [57, 58], or FreeSurfer [58].

Furthermore, post-processing is employed in some works to prevent false positives [59, 60]. Essentially growing algorithms [61] to improve the delineation of actual lesions, smoothing [59], and more specific handling Gaussian kernels in the output of the training models presented [56], and domain-specific segmentation like erosion and closing [62].

4.3.2 Multiple Sclerosis - Lesion Segmentation

Brain lesions due to the demyelination of the nerve cells are one of the known characteristics of MS. Among other metrics, they support the medical experts in diagnosing the disease, determining the current stage, and assessing the condition progression, particularly concerning the appearance of new lesions.

The presented reasons make lesion segmentation a required step in the study of multiple sclerosis. The different algorithms in the selected articles used various approaches regarding their input pixel or voxel, depending on data input dimensions, which could be 2D and 3D, respectively. Moreover, the input could be patches or even whole images, which can also have the same dimension differentiation as pixels and voxels.

The majority of articles in this segmentation subsection use CNN or CNN-based architectures, a deep learning algorithm commonly used in image analysis. Starting with a voxel-wise approach, Tanya Nair and colleagues [63] present a study that, besides the segmentation, explores the simulation of introducing uncertainties, using the Monte Carlo algorithm and analyzing its performance. Two types of patches, 2D or 3D (matrix of pixels or voxels, respectively), are found within a patch input approach. With a 2D patch approach, in [62] the authors use a supplementary source of anatomical information to test two experiments where they investigate the impact of knowing a position of large-scale anatomical structures the weight of identified tissue nature on the disclosure of demyelinating plaques. Another application of the same approach [59] studies four models where input information from the contrast images and time points varies among single or multiple, meaning that combinations of single and multiple contrasts and time points are used in the algorithms. After comparing the four designs with data collected from two raters, the multiple contrast images and time points model presented better results. Furthermore, two versions of this model were submitted to the challenge, one using post-processing and the other without post-processing with the first received a higher score.

CNN allows different configurations from a simple architecture to the inclusion of lesion information and a cascade-like structure. A multi-class segmentation with two cascades CNNs is described in [64] obtaining the best results using a 3-class cascade method, using only two types of MRI sequences. However, adopting a similar cascade fashion using FLAIR sequences as input produced better results. Additionally, a modified Region-based CNN (R-CNN) is proposed in [57], permitting the input classification with one layer's addition. This modification eliminated the use of post-processing and classification through SVM used in the original R-CNN. Proceeding with the 3D patch approach, [58] focus on rim lesions, chronic active lesions enhanced with a

paramagnetic RIM, which are considered a biomarker for MS. In a CNN-based algorithm designated as RimNet, lesions are classified as Rim+ or Rim- lesions and compared with the experts ground truth.

A convolutional encoder network (CEN) is divided into two pathways, convolutional and deconvolutional, interconnected. CEN is a CNN-based architecture model [65]. Tom Brosch and colleagues [65] report a simple structure CEN with just three layers. The first layer, the input, is composed of the image voxels, accompanied by the convolutional layer, which automatically learns features from the data, and lastly, the deconvolutional layer uses the learned features to compute the probabilistic lesion mask.

Also, [66] present a three-layer and also a robust seven-layer CEN study. In both algorithms, the authors pre-trained their CENs with a stack of convRBMs (restricted Boltzmann machines). The first is trained on input images, whereas the following convRBMs are executed on hidden activations of the first one. When comparing both CENs with the existing methodologies, they displayed generally better outcomes. In both studies, the probabilistic output maps are thresholded to obtain the binary segmentation mask.

Proceeding amidst deep-learning procedures, U-Net is widely used for segmentation purposes in medical imaging. Like the previous algorithms, U-Net is a specification of the CNN architecture. This algorithm is also utilized for multi-class segmentation of different tissue samples, such as white and grey matter, instead of only focusing on lesion segmentation. A 2-dimensional U-Net is implemented in [67], which beyond the segmentation, this study intended to investigate diverse contrast input MRI sequences in the segmentation process. Identically, [68] describes a 3D U-Net for a multi-class segmentation process. Additionally, using data-augmentation techniques, tests to the convergence of the algorithm were executed, which resulted in a high convergency rate for all classes. The overall results were promising, yet overfitting was reported for the lesion class that decreased after the augmented data training. Finally, when comparing the results with ground truth masks collected with MRIAP software, the U-Net masks demonstrate a high degree of agreement. A U-Net Fully-CNN [69] evaluates the lesion segmentation and the detection according to lesion size. Lastly, [70] focuses on new lesion detection segmentation, implementing a U-Net Fully-CNN patch approach, using, among others, deformation fields.

A distinct approach implemented with a generative adversarial network (GAN) is described in [71]. One component of the GAN, a generator, is used to create a mapping model of the input images. This part is composed of multiple encoding and one decoding path. Meanwhile, based on a patch procedure, the discriminators detected the difference between the real ground truth and fake lesion masks provided through the generator versus the generator's adversarial learning process.

Deep learning can also be used in segmentation studies, yet not for the segmentation process task. Instead, it helps uncover features to help discriminate classes. These models are denoted as hybrid approaches, given that they combine DL and ML algorithms. The authors [72] propose the first step to learn features from unlabeled images using restricted Boltzmann machines. Two RBM models were designed to process the two MRI sequences retrieving smaller-scale features.

Then, the large-scale features were obtained with a similar method using deep belief networks formed by stacking RBM models. After the feature vector is built, the labels are brought in, and the segmentation is completed using the RF, a supervised machine learning algorithm.

Fuzzy-KNN was chosen in [61] to study the detection and segmentation of white matter (WM) and cortical lesions. Information like voxel intensity, spatial location, and tissue type probability

Table 4.1: Articles with methodologies related to multiple sclerosis lesions segmentation using machine/deep learning

Author	Year	Dataset Size	Dataset Source	Ground truth by	Algorithm	Challenge Score	Evaluation
Ehab Essa et al. [57]	2020	45	public ¹	not mention	CNN	83.25%	DSC(UNC)=33% and DSC(CHB)=38%, R=61.8%(w/transfer learning).
Nils Gessert et al. [73]	2020	122	private	semi-automatic	CNN	-	DSC=62.2% (the higher DSC=65.6%), FPR=26.4% and TPR=74.2%.
Tanya Nair et al. [63]	2018	1064	public	semi-automatic and expert	CNN	-	no uncertainty: FDR of 20%, and TPR=47.5%; uncertainty: FDR=20%, and TPR=80%.
Germaín Barquero et al. [58]	2020	124	public	experts	CNN	-	single-modal: R=70.6%, TNR=94.9%, PPV=56.9%, and NPV=97.1%; inter-center: R=75.8%, TNR=95.1%, PPV=52.8%, and NPV=98.3%.
B. Stasiak et al. [62]	2018	96	private	not mention	CNN	-	Exp I (higher results): PPV=53%, R=55.6%, and Fs=54.29%; Exp II (higher results): PPV=59.81%, R=57.18%, and Fs=58.47% (DSC=58.47% compared w literature)
Ariel Birenbaum et al. [59]	2017	19	public ³	trained raters	CNN	1-91.267%, II=90.07%	DSC=62.7% for the two proposed
Sergi Valverde et al. [74]	2017	45 scans	public ¹	experts	CNN	-	DSC=56%, VD=27.5%, TPR=68.2%, FPR=33.6% and PPV=66.1%
Erol Kazancli et al. [64]	2018	59	no mention	experts	CNN	-	VD=22.8%, TPR=59.7% and FDR=42.5% for a patch(11x11) 3 class, cascade approach.
Tom Brosch et al. [66]	2016	238	mix ^{1,3}	partially semi-automated	CEN	Mic=84.07%	MICCAI: LTPR=51.6%, LFPR=51.3% and VD=57.8%; ISBI: DSC=68.3%, LTPR=74.5%, LFPR=54.5%; private: DSC=63.83%, LTPR=62.49%, LFPR=36.1% and VD=32.89% (all input modalities);
Tom Brosch et al. [65]	2015	520	mix ¹	not mention	CEN	-	DSC=57.38% (trained w/100 subjects); TPR=39.71%, PPV=41.38% and DSC=35.52%(smaller set).
Chaoyi Zhang et al. [71]	2019	126	private	experts	GAN	-	DSC=67.2%, R=69.2%, PPV=72.4% and Fs=70.8%;
Ponnada Narayana et al. [67]	2020	1008	public ²	experts	U-Net	-	only FLAIR: DSC=91%, LTPR=82% and LFPR=47%; all inputs: DSC=90%, LTPR=81%, and LFPR=28%.
Refatt E Gabr et al. [69]	2020	1003	public ²	expert	U-Net	-	DSC(lesion)=82%, DSC(WM)=95%, DSC(GM)=96%, and DSC(CSF)=99%; TPR=79%, and FPR=34% for all size lesion
Mostafa Salem et al. [70]	2020	60	private	not mention	U-Net	-	DSC(seg)=55%, DSC(det)=83%, TPF=83.09%, and FPF=9.36%.
Ponnada Narayana et al. [68]	2019	1008	public ²	expert	U-Net CNN	-	DSC values of 94% for WM and GM, 0.97% for CSF, and 85% for lesion; R ² values of 91.2%, 88.3%, 99.67%, and 99.6% for WM, GM, CSF and lesions.
Hrishikesh Deshpande et al. [75]	2015	14	private	expert	ADL	-	4D-D: DSC=49.8%, PPV=45.3% and R=63.2%;
Snehashis Roy et al. [76]	2015	122	mix	not mention	SDL	-	2D-D: DSC=43.5%, PPV=39.8%, and R=58%.
Mário João Faria et al. [61]	2016	39	private	expert	fuzzy-kNN	-	DSC=55.5%, LFPR=44.4%, VD=26.7% and LTPR=36.8%.
Mosfiata Salem et al. [56]	2018	60	private	expert	LR	-	DR(WM) 77% and DR(cortical)=62% (better results).
Zahra Karimghaloo et al. [77]	2016	1760	public	experts	AMCRF	-	Large lesion: DSC(seg)=59%, DSC(det)=93%, TPF=91.3% and FPF=5.88%;
Youngjin Yoo et al. [72]	2014	581	public	experts	RF	-	Small lesion: DSC(seg)=24%, DSC(det)=26%, TPF=34%, and FPF=24.09%;
Colm Elliott et al. [78]	2013	255	private	automatic, validated by expert	RF and Bayesian	-	Thres=0.3; TPF=74.3%, FPF=11.86% DSC(det)=77%, DSC(seg)=56%, and Fs =80.06%
Amnegreet Van Opbroek et al. [60]	2015	20	public	no mention	SVM	CHB=80% UNC=75%	Overall results R=91%, FDR=33%, and average FP=46% (counts per scan).
Ozdemir Cetin et al. [79]	2020	6	private	experts	supervised learning	-	DSC=38%, TPR=58%, and PPV=35%;
							target R=80%, an FDR=8%, target R=90%, an FDR=23%.
							Score(CHB)=80%;TPR=58.7% and FPR=70.3% w/UNC rater;
							Score(UNC)=75%; TPR=44.9%, and FPR=69.6% w/UNC rater;
							20 models 50/50split: A=90%, R=62%, and TNR=92%; DSC=62% (wall input sequences)
							6 LOO models: A=90%, R=65% and TNR=91%; DSC=71% (wall input sequences)

A=Accuracy, R= Recall, (L)TPR=(Lesion)True Positive Rate, TNR=True Negative Rate, NPV= Negative Predictive Value, PPV= Positive Predictive Value, (L)FPR=(Lesion) False Positive Rate, Fs= F-score. ¹MICCAI 2008 Challenge; ²CombiRX database; ³ISBI 2015Challenge.

was among the extracted features to distinguish between the six classes, WM, four different cortical, and Deep-GM lesions. This algorithm returns a lesion probability map rather than the binary segmentation, which is obtained by thresholding the output of the algorithm. The threshold was decided in order to maximize the Dice Score Coefficient (DSC) between the obtained results and manually segmented ground truth lesion, with a leave-one-out cross-validation method. Moreover, a post-processing set was executed to exclude the outlier voxel of the binary mask, decreasing the false-positive rate, followed by a region-growing technique to enhance the delineation of true lesions. Overall, this approach shows encouraging results, especially for small WM and cortical lesions.

Most papers perform segmentation of the brain tissues, followed by lesion segmentation is completed. Ozdemir and colleagues [79] describe a single step lesion classification running a supervised machine learning method. To the ground-truth obtained from the expert, 30 voxels sample intensity values were collected of each class from every subject. The voxel coordinates samples are used as the input strategy to extract the intensities voxels at the same location for the different MRI sequences. After the training, the centroids from the corresponding dataset were calculated, and a centroid matrix is generated. They are later recalculated based on the Euclidean distance to the nearest centroid for each iteration of the cross-validation process. The output is a 3D decision map of the different classes for each test subject on which an ellipsoid mask is applied to identify false-negative MS lesion events. Finally, the results are compared to the ground truth provided by the experts.

With the combination of sparse coding and dictionary learning, the authors [75, 76] describe two similar approaches to segment brain tissue. The MRI sequences are divided into patches, labeled using manual segmentation [75] or atlas prior information [76], and applied for training and test. Different size dictionaries, atoms learned (elements composing of the dictionary), and the number of classes is studied. This kind of learning is used to reduce redundancy, which could lead to overfitting and computational burden.

As stated earlier, new lesion detection is a necessary factor in MS evaluation. In clinical practice, follow-up imaging is performed and later visually compared to the baseline, a previous exam used as a temporal reference. Mofata Salem et al. [56] offer a supervised model to detect new T2-w lesions working on features from image intensities, subtraction values, and deformation fields (DF). The authors tested four different models based on logistic regression and combinations of managing with or without DF and baseline. The final assessments returned that the best results were accomplished by the model working on both DF operators and baseline.

In the same line, [78] present a two-stage classification process. In a first step, a generative Bayesian model is adopted to acquire a probabilistic brain tissue (CSF, GM, WM, and lesions) of each voxel in the baseline and follow-up images. Secondly, an RF algorithm provides the new lesion identification using the voxels recognized as new lesion candidates and grouped, refining the previous voxel classification.

Beyond lesion segmentation, a commonly reported difficulty is related to detecting small-sized lesions. In order to investigate this predicament, some of the above studies [67, 69, 72] performed

studies into distinguishing different size lesion, but even though the results were promising, the ratio between small and larger lesions is the principal justification for the results. Optimization learning methods are used to fit the model to the training data: Adam [69, 70, 68], Adadelta [59, 74, 65] are the most widely implemented, SGS [80] and RMSProp [80, 65] were also reported. Besides the optimization methods, some techniques are used, such as dropout [59] and data augmentation [71], which permits an expanded diversity in the dataset, avoiding overfitting the data, mainly applied in studies with small datasets. Moreover, transfer learning is implemented and compared with a standard supervised-learning classifier [60]. Implementing this technique allows learning from data that might be similar but not necessarily precisely illustrative for the target data. Four transfer learning models are matched to a conventional SVM algorithm, which outperformed the traditional model even with a small training sample data.

Table 4.1 presents a brief overview of the previously discussed studies, with info on the datasets used, which algorithms were chosen to perform the segmentation, and the achieved results. Moreover, the methods are further detailed concerning contributions and limitations in Table 4.2.

Table 4.2: Model characteristics, contributions, benefits, and limitations of the segmentation articles analyzed.

Authors	Model	Contributions\ Benefits	Limitations
Ehab Essa et al. [57]	3D patch-wise R-CNN	The probabilistic maps from different R-CNN of the various modalities are fused using an adaptive neuro-fuzzy inference system (ANFIS)	Requires a R-CNN for each image modality
Nils Gessert et al. [73]	Two-path CNN	Fusing the information of two-time points images using various processes (addition and concatenation) for segmenting new and enlarge lesions	Comparison between the multiple points is no simultaneous, straightforward with multiple pairwise comparisons
Tanya Nair et al. [63]	3D voxel-wise CNN	Analysis and deeper explanation of the uncertainty measures that affect the segmentation of lesions	The detection of small lesions contributes to higher uncertainty values due to their misclassification
Germán Barquero et al. [58]	3D patches CNN	The evaluation of paramagnetic rim lesion in MS patients using multi-center data	RimNet presents a low positive predictive value when compared to the experts due to its high number of false positives
B. Stasiak et al. [62]	2D patch CNN (with 6-layers)	Assessment of influence of known location of large-scale anatomical structures and tissue types	The introduction of information could compromise the generalization of the treatment of the usable data
Ariel Birenbaum et al. [59]	2D patches CNN	Longitudinal multi-view CNN tested with multiple and single contrast images and time points	The first time point is segmented with a single time model, and the remaining time point is acquired with a multiple time model
Sergi Valverde et al. [74]	Two 3D patch CNN designed in a "cascade" fashion	The first CNN is trained to expose viable candidates voxels, while the second reduces the number of misclassified voxels from the former network	The negative class is under-sampled in the first network, increasing the probability of miss predicting the voxels

Table 4.2 Continued.

Authors	Model	Contributions\ Benefits	Limitations
Erol Kazancli et al. [64]	3D patch-wise CNN "cascade" fashioned	The cascade fashion methodology is used to reduce false positive different multi-class models are tested	under-sampled of some classes in the multi-class analysis; neglecting the learning of some exceptions required in the domain knowledge adopted by the experts
Tom Brosch et al. [65]	3D CEN with 3 layers	Feature extraction and segmentation prediction in a single model; the features are tuned for a combination of image types and segmented for very unbalanced classes	Requires around 100 samples for training and testing to stabilize its performance
Tom Brosch et al. [66]	3D CEN with 3 and 7 layers	The methodology combines the advantages of CEN and U-Net exploring a higher number of layers and shortcut connections	with the increase of the number of layers, it requires a bigger dataset; challenges datasets was limited to the 3 layers CEN due to size
Chaoyi Zhang et al. [71]	GAN	The generator segment of the GAN allows and end-to-end learning integrating the various images modalities	Requires post-processing set, and uses optimization to attenuate the hard-to-convergence property of the GAN
Ponnada Narayana et al. [67]	2D U-Net	A multi-class method with the segmentation of various brain tissues and the MS lesions	2D architecture limited by the MRI protocol of the database adopted
Refatt E Gabr et al. [69]	U-Net FCNN	Multi-class method; FCNN was capable to detect small-sized lesions that were not presented in the ground truth	Small heterogeneity of the dataset, ground truth not entirely accurate due to the application of semi-automated methods
Mostafa Salem et al. [70]	Two 3D U-Net	One U-Net utilized to learn deformation fields (DFs) and registration of baseline and follow-up and then used as input to the second that performs the segmentation	Small dataset and deficient annotations of inflammation areas by the experts
Ponnada Narayana et al. [68]	2D U-Net	Multi-class segmentation of brain tissues and lesions	Exclusion of cortical lesion due to MRI protocol adopted
Hrishikesh Deshpande et al. [75]	Dictionary learning and sparsing code with 3D patches	Representation of high-dimensional images using few characteristic signals in a low-dimensional subspace	PCA inability to analyze the non-linearity in data when used with a higher dictionary size
Snehashis Roy et al. [76]	Patch-based sparse dictionary and atlas priors	Lower reconstruction error and the computational burden is reduced; uses priors to localize different tissue based on the atlas information	For a new dataset from a distinct scanner or sequence, a few atlases require to be manually delineated
Mário João Fartaria et al. [61]	K-NN using 3D img	Focuses on early disease and low disability patients, able to disclose small lesion of 3 voxels size	Low lesion sample of the training dataset
Mosfata Salem et al. [56]	Voxel-wise LR	Join intensity and deformation-bases strategies from baseline, follow-up, subtraction images, and DF operators to identify new lesions	In the training step, standard spatial and texture features were discarded; a small dataset
Zahra K. et al. [77]	AMCRF voxel-wise	Remove the false-positive lesion candidates, gather in a first stage, and refine the boundaries of the detected lesions, performed in the second stage	Adopts only Gad-enhancing lesions

Table 4.2 Continued.

Authors	Model	Contributions\ Benefits	Limitations
Youngjin Yoo et al. [72]	Patch-wise RF	Employs RBMs and deep learning to extract and learn the features	High-dimensionality of the feature vectors, poor characterization of the interaction of label and unlabeled data
Colm Elliot et al. [78]	Voxel-wise Bayesian and lesion-level RF	New lesion classifier allows for a trade-off of sensitivity and specificity by the use of a user-specified confidence threshold	Ground truth is no perfect
Annegreet Van Opbroek et al. [60]	Four Voxel-wise SVM-based models	Study the influence of transfer-learning technique in improving overall results	Restricted to voxel-wise classification; a small dataset
Ozdemir Cetin et al. [79]	2D supervised learning model	Classification of lesions in a single step, without relying on the segmentation of other tissue first	2D architecture limits the spatial information, and poor small dataset restricted to Euclidean distance parameter

4.3.3 Multiple Sclerosis - Classification

MS diagnosis is essential to slow the disease progression with the proper treatment prescription at the earliest stage possible. Distinguishing MS patients from healthy subjects is an important step, presented as the general purpose of the studies reviewed on this classification task.

The significance of an early diagnosis directs researchers to concentrate on Clinically Isolated Syndrome (CIS) patients, with several studies attempting to differentiate CIS from healthy subjects. It is considered as the beginning of MS because it might evolve for an MS phenotype. Furthermore, in the endeavor to predict its evolution, some studies investigate the possibility of converting CIS to a more advanced MS phenotype [81, 82, 83]. Moreover, the most common classification studies distinguish between healthy and MS patients, as seen in the articles. Ambitious researchers try to separate between multi-stages of MS. Some specific classifications are performed to separate normal-appearing white matter (NAWM) from normal-appearing grey matter (NAGM) or if the pathology is worsening.

Of the numerous machine learning algorithms employed in classification problems, SVM is one of the most used. It can have a variety of pipelines such as voxel, ROI, or using the entire image. With a voxel-wise approach, studies [84, 85, 82] present different classifications. Demographic and clinical features, such as the Expanded Disability Status Scale (EDSS), are handled to separate early/late, Benign/Non-Benign MS subjects and low/high lesion load [85]. Benign MS course is characterized as a state where the patient stays utterly functional in all neurological processes 15 years post clinical onset, a determination which relies principally on the EDSS. Additionally, using lesion count, medium length of lesions from the center of the brain, and CIS type, the authors in [82] investigated the conversion of CIS into clinically-defined multiple sclerosis (CDMS). Furthermore, in order to work with features that present a higher discriminative power, the authors performed a feature selection step using a forwards recursive feature elimination (fRFE) algorithm. Similar goals as the previous study, [81] presents an ROI approach, where the brain is segmented into regions and used to extract some features applied to the SVM

classifier. Additionally, various experiments are conducted to assess the influence of sample size, cross-validation partitioning, using single or multi-center data, i.e., from various centers, which most likely have no MRI scanners from the same manufacturer, consequently increasing the variety of the samples.

Sara C. Reitz and et al. [86] also employ an ROI approach to investigate parameters, like the mean and standard deviation value of the relaxation T1 and T2, that significantly discriminate normal-appearing WM from normal WM. If time course varies among patients with and without radiological disease activity, meaning emerging of new lesions. The findings showed that the standard deviation values differ considerably between patients and controls in all tissues examined. Regarding the time course evaluation, mean T2 time varied expressively among the two groups. Finally, mean magnetic transfer ratio and proton density were ranked as the two most relevant features in separating the classes from the studied parameters.

The appearance or absence of MS using quantitative MRI (qMRI) images is classified using KNN and SVM [87]. The performance of the model was evaluated on MRI sequences that were severely damaged from the movements of the patient, which represents a current problem in the clinic. This non-stillness is associated with the discomfort of being confined to a small space and the noise produced by the machine during the acquisition process. KNN was tested with various k values, and both algorithms were executed with and without motion-free data. As expected, the models tested with motion-free data had lower false prediction rates regardless of the qMRI-based features applied. Comparing the performance of both methods when using a dataset that combined motion and motion-free data reveals no notable difference. Nevertheless, testing solely with motion-free data KNN presents a lower false prediction rate than SVM.

Hereditary diffuse leukoencephalopathy with spheroids (HDLS) is a rare condition defined by demyelination, axonal loss, neurodegeneration, identical to MS, which leads to an underdiagnosis of HDLS. The similarities make it challenging to discriminate between both pathologies. Based on qMRI features, [88] presents a promising classification between the HDLS and MS. Quantitative maps averaged in GM and WM and volume fraction calculated of each tissue were utilized as features in a Linear Discriminant Analysis algorithm. With the rareness of the disease, which causes the unbalance between the classes, additional procedures were examined to increase the robustness of the design. A mean model was built and trained with augmented data from the correct classification models of the cross-validation average. This technique consisted of an arrangement amidst original features and measurement noise, resulting in a decrease of the true positive detection rate for either class with a noise level increase.

As it happened in the SVM algorithm approach, in the random forest from the selected articles, it is found a variety of practices regarding the input of the algorithms, patches [89], lesion-wise [83], and ROI [90]. In [89], the authors describe a methodology using a deep belief network (DBN), a deep learning method, to learn a hidden feature representation from image patches of NAWM and NAGM. These patches are chosen, so only normal-appearing image patches are used, i.e., patches overlaying with focal lesion are rejected. Then, to avoid overfitting the design, feature selection is completed (LASSO), and the resultant features are finally applied to the random forest

classifier.

V. Wottschel and colleagues [83] used a lesion-wise approach with an oblique random forest classifier to examine the evolution of CIS phenotype to MS, similar to earlier discussed articles. In opposition to the traditional RF, an oblique RF explicitly learns the optimal separation in two classes employing linear multivariate designs. Different classification models were designed according to the features specifications: based on intensity, shape, and a combination between the two natures of elements. Additionally, to verify the lesion segmentation procedure impact, lesions were segmented using two semi- and fully-automated methods. In essence, no significant variations were reported regarding the segmentation procedure. Contrastly, shape features models proved a high discriminative potential comparative to intensity models, which did not improve the classifier.

Furthermore, with an ROI approach, [90] aims to investigate the application of derived Stockwell transform, the polar Stockwell, and the discrete orthogonal Stockwell transforms for texture analysis in synthetic and MRI images and examines the significance of the associate features for MS classification. The ROIs were divided into confirmed lesions and NAWM to release storage space. The transforms were computed in a voxel-wise manner and then processed to agree to the ROI approach. An initial RF model was run to determine the feature ranking using a random set of features and cross-validation. This ranking was based on Gini index and SHAP method. Succeeding, a second RF model was executed based on the features selected by the previous model, ROI mean signal intensity, and the estimated probability of an ROI being a lesion. In general, outputs are shown no notable contrast between the features from the transforms. Nonetheless, DOST obtained higher results.

Identically, yet using the entire images used as input in CNN algorithms, the authors [80] present a study where the data is initially decomposed into four sub-bands using a two-dimensional Haar Wavelet Transform (HWT). One of these sub-bands is used as the input of the different CNN networks while the other three are merged as one layer with the network outputs. Moreover, different wavelets were tested, but Haar presented higher accuracy results.

As occurred in the segmentation, CNN is regularly implemented as a DL algorithm in selected disease classification articles. Fabian Eitel and colleagues [91] present a CNN, which is pre-trained on a large dataset of Alzheimer's disease patients and healthy subjects. Afterward, various CNN models were trained using MS and healthy controls data (with and without pre-training or lesion-filling). Additionally, layer-wise relevance propagation (LRP) was applied to create the heatmap, which assisted in further analysis. As a result, the pre-training model presented better results for most of the considered metrics. In a comparative approach, [92] adopts dropout to overcome an overfitting issue, batch normalization, and stochastic pooling to train the model. In complement, data augmentation, a technique applied to the original images in order to generate additional data, is included to increase the accuracy of the classification process. For this concrete case, rotation, scaling, noise injection, and random translation were chosen as data augmentation techniques.

In [93] the authors analyzed four clinical profiles with a multi-classification process utilizing graph structural connectivity information extracted from conventional MRI(cMRI) and diffusion-

tensor imaging (DTI). To conclude the CNN approaches and focusing on detecting a vein at the center of WM lesions, the central vein sign (CVS) is considered a biomarker for MS, Pietro Maggi et al. [94] presents two methods in CVS assessments in WM lesions. The first is based on the application of a vesselness filter previously proposed. The second with CVSnet, a three-layer CNN, using patches that considered the center of mass of each lesion the center of the respective patch. Overall, the final results exhibited slightly better outcomes for the CVSnet.

A deep-learning and multi-modal approach is described, adopting a DNN that combines cMRI sequence with diffusion-weighted images (DWI) and Magnetization Transfer (MT) images. From the modulation of DWI signals, diffusion basis spectrum imaging (DBSI) models are created, allowing metrics related to fiber fraction and water content to be obtained. The authors differentiated MS lesions according to tissue and MRI sequences used; for instance, WH on T1-weighted imaging lesions are designated as "black holes" or "gray holes" based on the level of hypointensity. Four different models were developed using normalized intensities of the conventional sequences in conjunction with various metrics: DBSI-DNN using DBSI metrics, DTI-DNN, which used DTI metrics, MTR-DNN from the MT sequences, and DNN model implemented with only the normalized intensities from the cMRI. The overall performances revealed that the DBSI-DNN model presented the best concordance with the expert evaluation for the different MS lesion subtypes.

A new innovated hybridized image classification procedure is presented in [95], based on discrete wavelet transform (DWT), artificial bee colony (ABC), and extreme learning machine (ELM) - MABC-ELM. DWT is frequently admitted as the most effective manner for feature extraction and image decomposition. ELM is a learning algorithm that is a single-layer feed-forward neural network with quick and efficient learning ability. ABC is utilized for optimization of weight and bias of ELM. The last component is modified in order to supply every solution with a reasonable possibility of being selected in the next iteration. This methodology outperformed the compared algorithms in every metric in regards to MS classification. Furthermore, by examining ELM and ABC-ELM results, the proposed MABC-ELM can produce a faster convergence rate. A similar application [96] explores stationary wavelet entropy (SWE) to extract features and then compare them in three machine learning algorithms: decision tree, KNN, and SVM. After all, the execution of the different models, the highest results were obtained for the combination of SWE and KNN, which was compared with reliable previous works. The proposed method still performed efficiently, outperforming some of the metrics like accuracy and precision.

Based on spatial information, [97] describes two peculiar algorithms: a Bayesian Spatial Generalized Linear Mixed Model (BSGLMM) and Log Gaussian Cox Process (LGCP). BSGLMM considers the two-fold lesion maps and the spatial relationship between neighboring voxels, while LGCP accounts for the random spatial change in lesion location. The authors added a multi-class SVM that was run to obtain the optimal feature set to use for comparison with the previous ones. Tested on T1 and T2 images, higher results are obtained with T2, and even higher overall accuracy and average accuracy values with LGCP.

In a different methodology, Marco Pota and colleagues [98] use a multivariate Linear Fuzzy Analysis (LFA) to integrate multiple sensitive neuroimaging markers, such as WM lesion load, to

Table 4.3: Articles related to MS classification, diagnosis, or characterization on MRI images using ML and DL methods. The column regarding the multi-modality refers to the studies utilizing more than conventional MRI sequence.

Author	Year	Dataset Size	Dataset Source	Multi Modality	Lesion Segmentation by	Algorithm	Classes	Evaluation
Fabian Eitel et al. [91]	2019	147	private	N	semi-automated	CNN	MSHC	Wo pre-train: A=71.23%, R=68.46%, TNR=74%, and AUC=85.46%; Wo pre-train: A=87.04%, R=93.08%, TNR=81%, and AUC=96.08%;
Pietro Maggi et al. [94]	2020	80	public	N	manually-segmented	CNN	CSV+CSV	CUSnet: R=83%, TNR=75%, and A=79%; Neckless filter: R=67%, TNR=79%, and A=73%;
Ali Aljamaat et al. [80]	2020	58	mix	N	no mention	CNN	MSHC	A=99.05%, PPV=99.43%, and R=99.14%;
Shui-Hua Wang et al. [92]	2018	64	public	N	experts	CNN	MSHC	Data augmentation: A=98.77%, TNR=98.76% and R=98.77%; No augmentation: A=98.2%, TNR=98.18%, R=98.22%;
Aldo Marzullo et al. [93]	2019	114	private	Y	not mention	CNN (graph)	MS phenotypes and HC	HCMS: unweighted - F=80%, PPV=81%, and R=80%; Weighted - F=92%, PPV=93% and R=93%;
Zezhong Ye et al. [99]	2020	38	private	Y	expert	DNN	Lesion differentiation	PBH: AUC=99.1%, R=95.9%, TNR=95% and F=92.3%; FGH: AUC=97.7%, R=92.9%, TNR=93.4%, F=82.3%; AGH: AUC=98.7%, R=95.1%, TNR=95%, and F=88.7%; NBH: AUC=98.1%, R=93.5%, TNR=93.1%, F=91.8%;
A. Reddy et al. [95]	2020	24 ¹	public	N	not mention	MABC-ELM	MS vs other diseases	NAWM: AUC=99.8%, R=99.1%, TNR=97.3%, and F=97.3%; (DBS1-DNN models performance)
A. Ion-Mărgineanu et al. [100]	2017	87	private	Y	-	LDA, SVM and RF	MS phenotypes	A=99%, TNR=97%, R=100% and F=98%;
Yijun Zhao et al. [101]	2017	1693	public ²	N	semi-automated	SVM and LR	MS phenotypes worsening/non-worsening	CISPR - AUC=77%, R=25%, TNR=100%; CIS/PP - AUC=95%, R=83%, and TNR=90%; RR/PP - AUC=68%, R=56%, and TNR=76%; RR/SP - AUC=73%, R=57%, and TNR=77%; (overall higher test w/SVM)
Sarah C. Reitz et al. [86]	2017	21	public	N	not mention	SVM	NAWM/NAWM	LR (Y, white) - A=84.9%, R=67%, and AUC=55%; SVM (coste LS) (Y, white) - A=65%, R=60%, and TNR=57%; (ZY, white) - A=72%, R=77%, and TNR=69%;
Viktor Wetschel et al. [81]	2019	400	public	N	semi-automated	SVM	CIS converts	A=66.7%
Kerstin Bendfeldt et al. [85]	2019	100	public	N	semi-automated	SVM	Various studies	Single-center: A=95% CI=92.2%, R=86%, and TNR=99.7%; Multi-center: A=95% CI=73.3%, R=73.2%, and TNR=73.3% (w/LOO-CV);
Mariana Zurita et al. [84]	2018	157	private	Y	not mention	SVM	RRMSHC	Early/late MS: A=85%, R=82.3%, and TNR=88.2%; Low/high T2-Lesion Volume - A=83%, R=80%, and TNR=85%; BMS/MS - A=77%, R=76.9%, and TNR=76.9%;
V. Wetschel et al. [82]	2015	74	private	N	expert w/semi-automated	SVM	CIS converts to CIDMS	rsMRI: A=85.9%, PPV=85.3%, R=86.8%, and TNR=85%; combine: A=87.7%, PPV=87.5%, R=88.1%, and TNR=87.4%; 3Y - A=68%, R=60%, TNR=76%, PPV=72%, and NPV=65%;
Gabriel Kucivar et al. [102]	2016	103	private	Y	no mention	SVM	MS phenotypes and HC	LY - A=71.4%, R=77%, TNR=56%, PPV=70%, and NPV=74%; 3Y - A=68%, R=60%, TNR=76%, PPV=72%, and NPV=65%;
Gabriel Mangear et al. [88]	2020	18	private	Y	no mention	LDA	HDL/MS	HC/CIS: PPV=92%, R=91.7%, and F=91.8%; CIS/RR: PPV=92%, R=91.7%, and F=91.8%; RR/PP: PPV: R: F=75.66%; RR/SP: F=68.8%; SP/PP: PPV=59.6%, R=61%, and F=60.3%; CIS/RR/SP: PPV=71.3%, R=70%, and F=70.6% (for all characteristics);
A. Ion-Mărgineanu et al. [103]	2017	105	private	Y	no mention	LDA, SVM and RF	MS phenotypes	R=97.2% and 99.6% for HDLS and MS
Youngjun Yoo et al. [89]	2018	99	private	N	semi-automated	RF	NAWM/NAWM (MS/HC)	wall feature and LDA: CIS/RR: F=65%; CIS/RR/RR/SP - F=65%; RR/PP - F=83%; RR/SP - F=87% (equal to SVM)
Haike Zhang et al. [83]	2019	84	private	N	semi-automated	RF (oblique)	CIS converts	Deep-learned: A=86.9%, R=85.5%, TNR=88.6%, and AUC=87%; (Wo LASSO); A=87.9%, R=87.3%, TNR=88.6%, and AUC=88% (w/LASSO); Deep-learned: A=74.7%, R=74.5%, TNR=75%, and AUC=74.8% (NAWM); A=73.7%, R=74.5%, TNR=72.7%, and AUC=73.6% (NAWM);
Glen Pridham et al. [90]	2020	6	private	N	semi-automated	RF	NAWM	Intensity: A=62%, R=62%, TNR=61%, PPV=85%, NPV=50%; Shape: A=85%, R=94%, TNR=50%, PPV=87%, NPV=69%; Shape (LST): A=82%, R=95%, TNR=43%, PPV=84%, NPV=67%; PST (all sequence): A=88%, TNR=87%, and R=91%;
Helko Neeb et al. [87]	2019	97	private	Y	no mention	KNN and SVM	MSHC	DOST (all sequence): A=89%, TNR=87%, and R=91%;
Yulong Zhang et al. [96]	2016	38	mix	N	expert	KNN, DT and SVM	MSHC	SVM: FPR=26.5% (all data), FPR=24.4% (motion-free); KNN (n=7): FPR=25.5% (all data), and FPR=20.4% (motion-free);
Bernad Tischler et al. [97]	2014	250	private	N	semi-automated	Bayesian Spatial	MS phenotypes	KNN: A=99.94%, TNR=99.33%, PPV=99.09% and R=96.15%;
Marcio Pota et al. [98]	2019	110	private	N	semi-automated	LFA	RRMSHC	BSGLME: AUC=78.3%, and AUC=82.3%; LCPE: AUC=75.3% and AUC=85.1%;

A=Accuracy, R= Recall, TPR=True Positive Rate, TNR= True Negative Rate, NPV= Negative Predictive Value, FPR= False Positive Rate, Fs= F-score; ¹ Represents number of scans instead of subjects; ² CLIMB dataset.

discriminate RRMS subjects from healthy participants.

This algorithm comprised two central elements the fuzzy sets correlated with the chosen variables of interest and the rule base. The abnormal WM (AWM) -fraction was verified as one of the most reliable markers among the adopted features. Nevertheless, according to the authors, this parameter alone was not sufficient since, among other causes, patients with low values of AWM-fraction could either be reduced lesion load MS patients or healthy controls showing WM hyperintensities due to noise. Therefore, this parameter was combined and tested with other markers. This procedure generated higher results with the GM-R2 union, which exhibited higher accuracy results in single feature tests.

Besides the classification algorithm and experimenting on a large dataset, Yijun Zhao and his colleagues [101] execute essays on agents that could influence algorithm performance, such as class imbalance and misclassification cost. Combinations of bagging/undersampling and different cost ratios models are tested using SVM models, which are then compared to logistic regression. Overall results, the use of misclassification cost in SVM models presents a more realistic model. The cost ratio of 1.5, the mean value, shows the best trade-off between predictive accuracy and false predictions.

As perceived in the previously detailed studies, some methodologies use MRI modalities besides the conventional sequences (T1, T2, and FLAIR). For instance, functional MRI (fMRI) together with Diffusion Tensor Imaging (DTI) in study [84] also used in [93]; quantitative MRI employed in [87, 88]; lastly, the MR spectroscopic imaging (MRSI) used in [103, 100].

Furthermore, it was noted that to separate the different MS phenotypes studies use different strategies. Either use binary classifications [93, 103, 100, 102], where the various types are distinguished between each other or healthy controls in a two-class problem. In contrast, a multi-class method [93, 102, 97], the algorithms differentiates between more than two classes.

Table 4.3 presents a brief overview of the previously discussed studies, with info on the datasets used, which algorithms were selected for the classification task, and the results reached. They are further detailed in Table 4.4.

Table 4.4: Model, contributions, benefits, and limitations of the articles related to the MS classification.

Authors	Model	Contributions\ Benefits	Limitations
Fabian Eitel et al. [91]	3D CNN and layer-wise relevant propagation (LRP)	LRP utilization to disclose relevant feature to assist the decision of the CNN	Limited sample size and computational burden of the architecture
Pietro Maggi et al. [94]	3D CNN ("CVS-net") evaluated lesion- and patient-wise	Focus on CVS a biomarker to detect MS	The need for an automated reproducible and nonoperator-dependent procedure that can define the frequency of CVS lesions in individual scans
Ali Alijamaat et al. [80]	2D discrete HWT and CNN	The HWT enables both spatial and spectral information of the input images to be used	Small dataset sample and downsampling that images undergo, reduction of possibly relevant information

Table 4.4 continued.

Authors	Model	Contributions\ Benefits	Limitations
Shui-Hua Wang et al. [92]	CNN (14-layers)	Combination of batch normalization, dropout, and stochastic pooling with CNN architecture	High computational power required, and limited dataset sample
Aldo Marzullo et al. [93]	Graph-based CNN	Differentiation between the MS phenotypes using brain connectivity information (local graph)	The small number of patients, and imbalance between classes
Zezhong Ye et al. [99]	Voxel-wise DNN	Application of advance MRI techniques (DTI and DSBI)	Limited heterogeneity of the data
A. Reddy et al. [95]	MABC-ELM feature extraction using 2D-DWT	Hybrid image classification methodology; ELM as the ability to learn the network quickly and efficiently	Reduce sample, higher hidden layers number increase the results but increases the computational burden
A. Ion-Mărgineanu et al. [100]	Voxel-wise LDA, RF, SVM and CNN	Adoption of Spectroscopic Imaging for discriminating between MS courses	MRSI features when classifying moderate from severe MS courses
Yijun Zhao et al. [101]	SVM (linear and non-linear kernels)	Assess disease progression while analyzing parameters such as class imbalance, and classification cost	concentrates mostly on clinical data points and quantitative MRI features; The progression of the disease is based on changes in EDSS over 5 years
Sarah C. Reitz et al. [86]	ROI-wise SVM	Application of qMRI to characterize NAWH and NWM, also the influence of parameter time-courses in patients with and without radiological disease activity	Small sample size, focus only one MS phenotype, and absence of longitudinal data of HC
Viktor Wottschel et al. [81]	ROI-wise SVM	Recursive feature elimination (RFE) and weight averaging combine with SVM to select the relevant features	Downsample, limited dataset
Kerstin Bendfeldt et al. [85]	Voxel-wise SVM	Multivariate SVM analysis for classification of GM patterns in MS, especially in early/late, low/high lesion load studies	Impact of medication in the results
Mariana Zurita et al. [84]	12 SVM models combining different input types	Fractional anisotropy maps, structural and functional connectivity were used as features	Low accuracy for EDSS below 1.5, i.e., low disease disability; DTI and rs-fMRI data were the only data used in one of the models
V. Wottschel et al. [82]	SVM using demographic and radiological features of two different time samples	Assess converting of CIS patients into MS; forward RFE adopted to select the relevant features	Bias in the accuracy results due to small dataset size; features are selected a priori and associated with WM lesions and are based on lesion masks
Gabriel Kocevar et al. [102]	Graph-based SVM combined with RBF kernel	Utilization of global graph-derived markers to differentiate between the various phenotypes and HC	Random choice of parcellation method for graph nodes definition; the diffusion acquisition was performed using 24 directions, which is comparatively low to current standards in diffusion imaging
Gabriel Mangeat et al. [88]	ROI-wise LDA based on tissue masks	Discriminate between HDLS and MS; mean model and noise analysis are implemented in the attempt to increase the robustness of the method	Class imbalance due to the rareness of the HDLS disease; significant differences of the 3T models used for the study

Table 4.4 continued.

Authors	Model	Contributions\ Benefits	Limitations
A. Ion Mărgineanu et al. [103]	2D LDA, SVM, and RF classifiers adopted for 9 binary classification tasks	Metabolic features (like Choline or Creatine) to the characterization of MS patients and the different phenotypes	Data was limited by only using a 1.5T acquire images, which reduces the sensitivity of lesion load segmentation, and influences the accuracy of metabolic quantification; The spectroscopic signal scales various among patients
Youngjin Yoo et al. [89]	patch-wise RF	Two modality-specific DBNs are used to learn a latent joint myelin-T1w features; LASSO adopted for feature selection	Reduce training sample size; Includes RRMS patients, not including progressive MS patients
Haike Zhang et al. [83]	RF based on lesion features (shape and intensity)	lesion features to evaluate converting from CIS to MS	shape features of small lesions can be influenced by the spatial resolution of a scan
Glen Pridham et al. [90]	ROI-wise RF	texture analysis ability of the DOST vs PST using both synthetic and MR images	small patient sample size, down-sampling approach
Heiko Neeb et al. [87]	KNN and SVM	Quantitative maps of tissue T1, T2*, total and myelin water content applies to differentiate between MS and HC; The influence of motion artifacts in the method robustness	Some lesions are lost due to their large T1, which results in the exclusion in one of the classes (NAGM+Lesions)
Yudong Zhang et al. [96]	DT, KNN, and SVM	Histogram stretching adopted to facilitate the comparability between data from different scanning; SWE for feature extraction	Reduce sample extent and only using histogram stretching in the intra-scans normalization analysis
Bernd Taschler et al. [97]	BSGLMM and LGCP	Manipulate spatial structures of MS lesion maps recognizing the binary nature of lesion data	In the LGCP not allowing the incorporation of lesion-specific features as supplementary marks
Marco Pota et al. [98]	ROI LFA	Calculation of R1, R2 and PD maps, and the creation of multi-feature Quantitative Magnetic Color Images (QMCI) to identify the position of normal brain clusters	Reduce heterogeneity of the data comprising the generalization

4.4 Evaluation Metrics

As previously discussed, the evaluation is essential since it permits retrieving information for the performance of the algorithms but also features that facilitate quantitative analysis of the structures.

The validation of an algorithm on a test set permits the numeric representation of its results due to its capacity to see how many predictions were correct, true positive (TP) and true negative (TN), and the incorrect, false negative (FN), and false positive (FP). From these numeric representations are derived some additional metrics such as accuracy, which measure the algorithm capacity to classify a sample considering the entire sample set correctly or sensibility that expose the ability of the algorithm to consider within all the correct data points positively. Moreover, it is conceivable to establish the weight applied towards sensitivity or specificity in some algorithms, designated as

the ROC curve (Receiver operating characteristic curve). This curve represents the variation of sensitivity, and its effect on specificity [104].

It was observed a slight difference in the chosen metrics. The majority of the segmentation articles adopted the dice similarity coefficient (DSC) as the primary evaluation metric; others such as true and false positive rates (TPR/FPR). Moreover, the volume difference (VD) measures the absolute difference between the obtained masks and the ones from the ground truth. On the other hand, most reports use metrics like accuracy, specificity, or even the area under the ROC curve, AUC, to evaluate the disease classification algorithms.

The table 4.5 describes some of the metrics used in both topics of the selected articles.

Table 4.5: Evaluation metrics for algorithms

Designation	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FN+FP}$
Specificity or True negative rate	$\frac{TN}{TN+FP}$
Sensitivity or Recall or True Positive Rate	$\frac{TP}{TP+FN}$
False Positive Rate	$\frac{FP}{TP+FP}$
Positive Predictive Value or Precision	$\frac{TP}{TP+FP}$
Negative Predictive Value	$\frac{TN}{TN+FN}$
F-score	$\frac{(1-\beta^2)*Precision*Recall}{\beta^2*(Precision+Recall)}$

4.5 Summary

Image processing and analysis techniques have become a useful tool in a vast range of applications. In the clinical environment, these procedures have increased for pathologies diagnosis and characterizations. Concerning the MS disease, it is widely implemented for lesion detection and segmentation, and diagnosis.

There are innumerable available articles on the thematic with a diversity of algorithms. Regarding the examined works in the pre-processing step, it is perceptible a transverse procedure described in most studies, being the enhancement techniques essential for improving analyses in the subsequent steps. For the segmentation step, it is exhibited a tendency for the adoption of supervised methods. In the classification step, constant revisions have been addressed from diagnosis criteria to disease diagnosis and treatments. In this stage, studies still rely on ML algorithms to distinguish different aspects of MS. However, in both steps, from the state-of-the-art is observed

a trend for the implementations of DL methods. Ultimately, studies report satisfactory and encouraging results. Still, there can not be a single generalized method for either lesion segmentation or disease classification.

Evaluation of any algorithms can be performed with a spectrum of metrics that permit examining parameters such as the efficacy and reliability of the generated results, which could be further handled to thoroughly compared with other similar articles discovered in the literature.

Chapter 5

Methodology

In the previous chapters, various notions were introduced concerning MS pathology, MR imaging, machine learning, deep learning, as well as image processing techniques. In addition, the different approaches related to MS computed image analysis studies found in the literature revealed two central topics lesion segmentation and disease classification. Since the latter subject depends on the lesions segmentation, the proposed approach focuses on the segmentation of MS lesions.

This chapter presents the proposed methodology, beginning with an overview of the pipeline, accompanied by a description of the main steps, with a theoretical explanation of the methods adopted in each step - preprocessing methodologies, segmentation, and evaluation. Moreover, some of the resources included in the process are presented, concluding with a review of the relevant information.

5.1 Overview

This project aims to explore pre-processing techniques and the segmentation of multiple sclerosis lesions from patient head MR images. Therefore the focus was on the pre-processing, segmentation, and evaluation of an image analysis algorithm. The Figure 4.1 illustrates the selected steps, which are going to be further detailed in the succeeding sections.

In the first instance, the images were visualize using the ITK Snap software, a graphical tool that allows to see the medical image in 3 plans of the 3D image, also permits the overlapping of the annotations and segmentation masks to the original image. Since the images were already registered to the same space, the first pre-processing step executed was the skull-stripping using the Brain Extraction Tool (BET) and the Deepbrain algorithm, discussed in subsection 5.3.2. After that, to attenuate the inhomogeneities intensity caused during the acquisition, the images were subjected to a bias field a correction method, N4 detailed in subsection 5.3.3. Lastly, seeking to reduce the computational power and time needed, data was downsampled to half of the cropped size, which reduce the number of slices, particularly the ones on the edges of the head where the portion of the background is significant compared to the brain info. To terminate the pre-processing step, the intensity of the images was normalized.

The images were then used in the segmentation algorithm, which results in an estimation of the lesion segmentation. Ultimately, the pipeline finishes with evaluating the acquired lesion mask concerning those provided by the challenge experts. The pipeline of the followed methodology is illustrated in Figure 5.1.

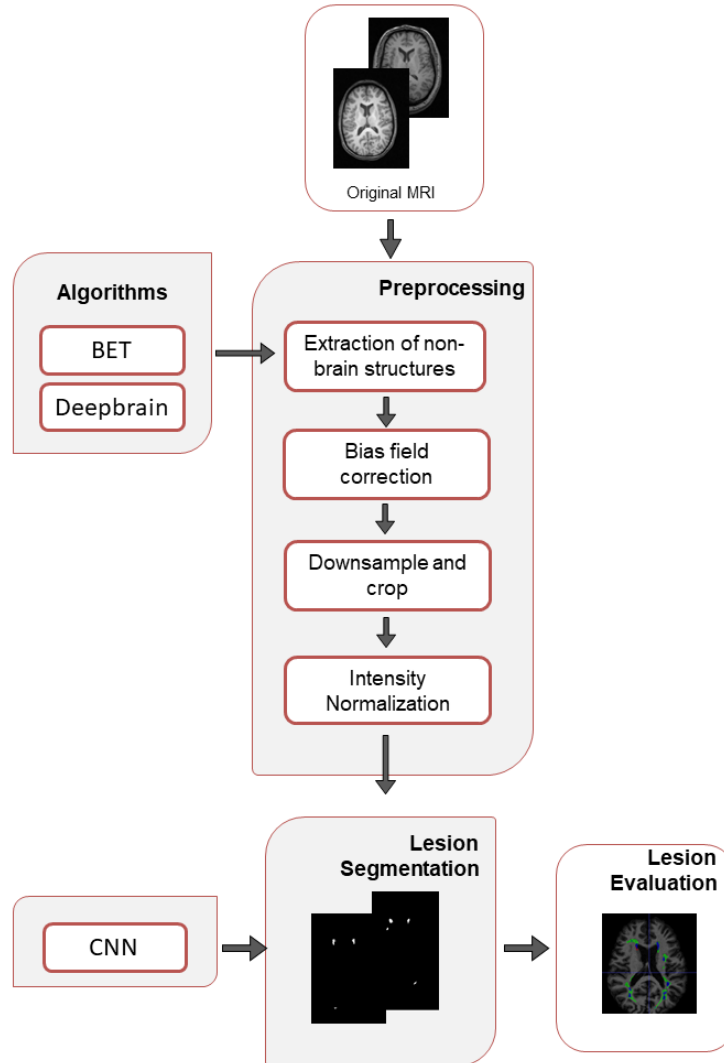


Figure 5.1: Pipeline of the proposed methodology.

5.2 Dataset

The proposed model was evaluated on the most known public dataset related to MS lesions, the Medical Image Computing and Computer Assisted Intervention (MICCAI) 2008 MS Lesion Segmentation Challenge. The data was acquired at two different sites, the Children's Hospital Boston

(CHB, Siemens) and the University of North Carolina (UNC, 3T Siemens Allegra). The dataset is composed of 45 scans that were grouped as followed:

- 20 training cases: 10 CHB and 10 UNC with manual segmentation from experts
- 25 testing cases: 15 CHB and 10 UNC cases with no segmentation annotations provided.

The lesion annotations present in the training set were primarily available with segmentation performed by the CHB rater. Nevertheless, the present dataset offers masks for the UNC MR images executed by two in-house experts [105].

From each subject was obtained a T1-W, T2-W, and FLAIR MRI modalities with a slice thickness of 1mm. T2-W and FLAIR images were rigidly registered to the T1-w images and resliced to the isotropic voxel spacing applying the b-spline-based interpolation. Concerning the shape of the provided images, both UNC and CHB centers have images with the exact dimensions of 512x512x512.

5.3 Image Preprocessing

As previously introduced, medical images require an initial enhancement step, the most common procedures performed were skull-stripping, bias field correction, noise reduction, intensity normalization, and registration.

The following subsections describe the preprocessing techniques which were only applied to the MICCAI dataset.

5.3.1 Registration

As referred to in the earlier chapter, registration is a prerequisite step while processing medical images since its objective is to align one image with another geometrically [106]. Considering the attempt to correctly compares images from different imaging equipment acquire at specific times and angles results in "different" images. Exist various types of algorithms to perform the registration, such as rigid-body registration, which corresponds to a transformation composed of rotations and translations of the space.

Even though this step was not implemented in this project, considering that the dataset authors performed it, it is worth referencing, where the MICCAI 2008 dataset was used rigid registration.

5.3.2 Extraction of Non cerebral Structures - Skull stripping

This step is revealed as an essential procedure for various experiments like cerebral volume estimation, extraction of the cortical surface, and tissue segmentation. Skull-stripping is very important since the skull, fat, skin, and other structure could introduce classification errors due to the intensity similarity with other cerebral systems, like MS lesions.

On the other hand, this step is also pertinent given that it reduces the number of structures to be analyzed, so the computation time is reduced. Therefore, there were two automatic algorithms,

Brain Extraction Tool (BET) [107], a standard tool used by numerous authors available in the FMRIB Software Library (FSL), and Deepbrain [108].

A. BET

BET is an automatic method that allows the removal of non-brain tissue from an image. Additionally, it can evaluate the inner and outer skull surfaces. It is an iterative algorithm that bases the estimation to differentiate the brain and non-brain tissue in an intensity threshold. The center of gravity of the head is discovered, accompanied by its approximate dimensions. A sphere surface is initialized, inside the brain, with a triangular tessellation, permitting a slow deformation that follows forces keeping the surface well-spaced and even meanwhile seeking to shift towards the edge of the brain. This process is repeated until it reaches a fitting solution.

BET was applied through the FSL software [109], which requires the input image and the name for the output file. Moreover, BET allows the user to set some parameters, such as fraction intensity threshold. This variable establishes the stripping severity of the extraction, varies from 0 to 1, and with smaller values reflects a more conservative extraction, yielding significant brain outline estimations, while higher values perform smaller brain outline estimations. By default, the fraction value is set for 0.5. Additionally, some arbitrary options that "customize" the extraction could be joined, like the head radius or center-of-gravity, both in millimeters.

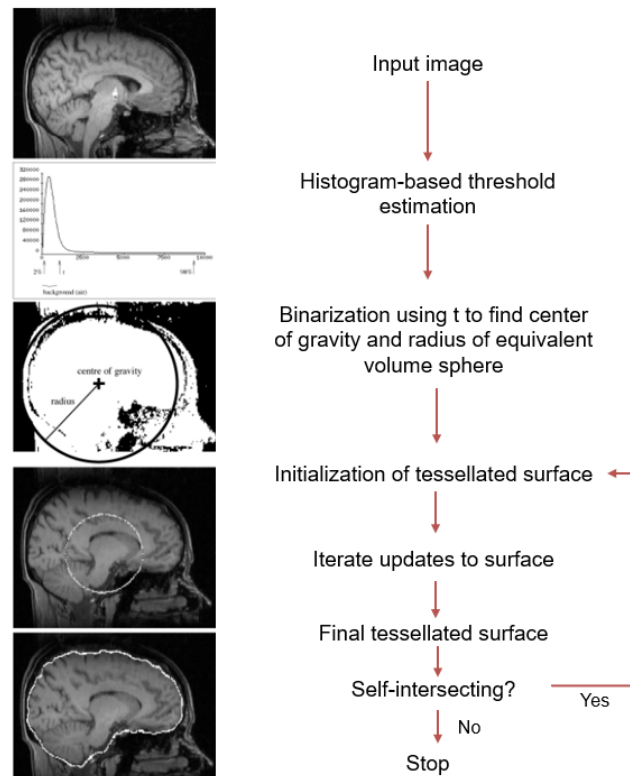


Figure 5.2: Flowchart of the BET algorithm (Adapted from [107]).

B. Deepbrain

Deepbrain is a Python algorithm based on U-Net, which was designed to produce a low amount of parameters, and unlike other similar software, does not require specific information such as brain templates or prior probability masks.

The U-Net is composed of two main branches, convolutional and deconvolutional. It is initialized with two convolutional layers with eight filters applied to the input images. In between the convolutional layers, max-polling and dropout layers were used to reduce the input size and prevent overfitting, respectively. The quantity of filters increases, in a factor of 2, between layers until it reaches 32 filters. Afterward, the deconvolutional part of the net, the ascending branch, mirrors the layers setting implemented in the previous branch, as presented in the Figure 5.3. Convolutional layers were executed with a kernel size of 5^3 , stride, and padding of 1^3 . ReLU was selected as the activation function. Max-polling layer settings were also equal throughout the network, with a stride and pooling size of 2^3 . Similarly, dropout layers were set with a drop percentage of 0.3. A sigmoid function was used to obtain the probabilistic map, which was formerly thresholded to obtain the final brain mask.

The CNN was trained on large datasets, like the Alzheimer's Disease Neuroimaging Initiative (ADNI), that own thousands of cases, using the data augmentation technique, with diverse rotations and orientations of the brain.

The mask used to extract brain tissue is estimated using a probability value which the user can assign, being the suggested value 0.5. The model performs admirably in T1-w images, considering the high contrast of the tissue in this MRI modality. Consequently, those images were directly applied to the algorithm, and the output masks of the extracted tissues were later multiplied by the corresponding remaining modalities, preserving uniformity between sequences.

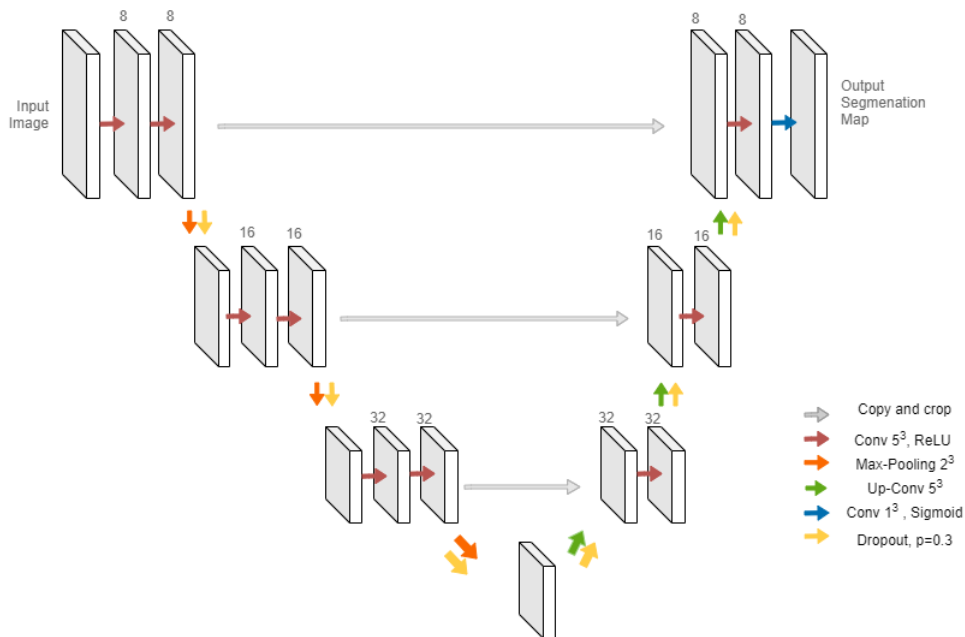


Figure 5.3: Deepbrain architecture schematic.

5.3.3 Bias Field Correction

Bias field signal is a low-frequency and smooth signal that corrupts MRI images, and the MR machine produces it. So the correction of these not uniformities is necessary to reduce misclassifications. Therefore it was implemented the N4 algorithm from the Simple ITK library. N4 is a variant of the N3 algorithm (Non-parametric Non-uniformity Normalization), an iterative method, working in log-transform space of image intensities and bias field. It tries the smooth multiplicative field that maximizes the high-frequency content of the distribution of tissue intensity [110]. The method is entirely automatic and does not require prior information.

The algorithm begins with the creation of a background mask. The iterative block follows this step in the space of the log-transform space of the image intensities and bias field. The logarithmic transformation can move into difficulties, especially in the areas where the signal intensity approximates zero, the same as the air-filled background regions. Then, masking the background supports in minimizing these concerns. In the iterative segment, each iteration has three steps:

- The bias field histogram is estimated so that the corrected image is sharpened, which is accomplished by utilizing a deconvolution filter to the image intensity histogram. The filter uses a Gaussian kernel with an FWHM selected by the user.
- The bias field estimation is smoothed by fitting it with a 3D b-spline field with a specified grid. The size of the grid controls the degrees of smoothing, which is typically around 200mm for images acquired with a body coil, and less for those obtain with localized coils.
- The termination criterion is tested on a stipulated threshold. In practice, the algorithm concludes following an established number of iterations.

Once completed, the bias fields are extrapolated over the entire field of view. Finally, the original image is divided by the bias field to obtain the corrected image. The process is summarized in the Figure 5.4.

5.3.4 Downsample, Crop and Intensity Normalization

The original images were large (512x512x512) and possessed some slices, particularly in the edges of the image, that did not detain relevant information since most voxels correspond to background voxels non-brain tissues. Accordingly, the images were cropped, producing a shape of 352x352x312. Subsequently, to decrease the computational power and time required to complete the process, the images were downsampled with a fraction of 0.5, resulting in a final image shape of 176x176x156.

Regarding the images intensity normalization, it was obtained by subtracting the the images to their mean intensity value and divided by their variance.

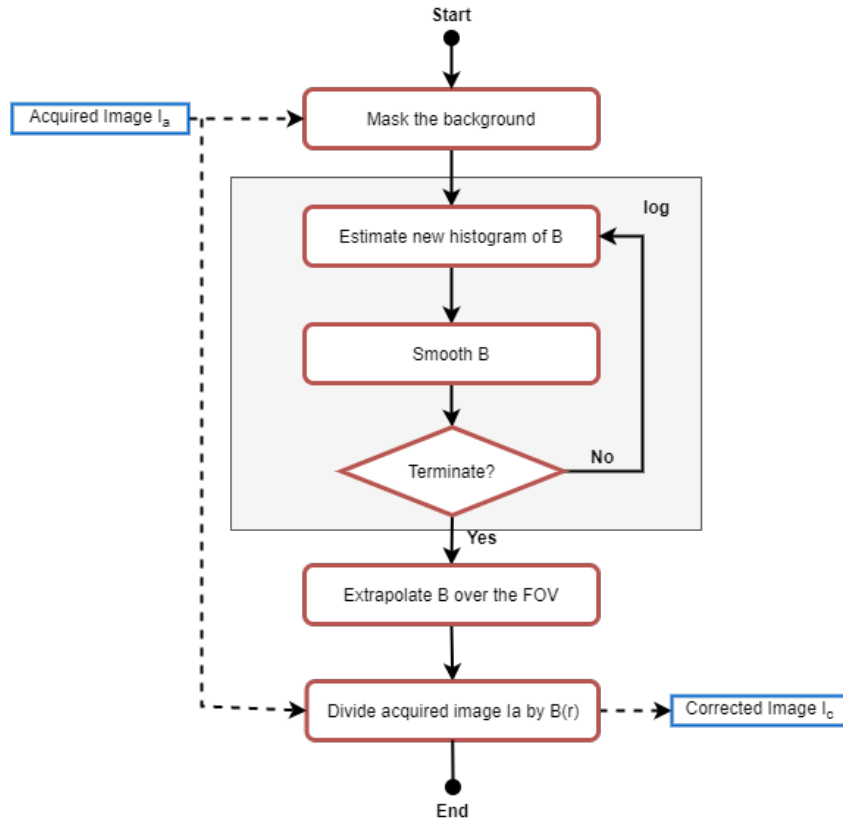


Figure 5.4: Flowchart of the N4 procedure, where B represents the bias field.

5.4 Lesion Segmentation

Based on the research conducted in Chapter 4 Deep Learning models application are increasing in various areas. Of the DL algorithms most used in segmentation tasks, CNN models are one of the most adopted. Although the increase of the depth of their architecture has been recently proposed, it could lead to overfitting, especially when using a small dataset. The presented model proposes a patch-wise cascade model with two CNN models based on the model proposed by Valverde et al.[74]. The original methodology was implemented using the Lasagne and Theano libraries. However, in this study, we used the Tensorflow[111] and Keras [112] which are open-source high-level APIs that allow an efficient design and training of the machine learning models.

5.4.1 Model Inputs and the Cascade Configuration

The input of the model are patches, which are stored in a vector of computed multi-channel 3D patches obtained from the intensity normalized images. The 3D axial patches with a dimension p centered on a voxel of interest are computed from each training image. The voxels selection is executed by the value $t_{\min}$, for voxels, in the FLAIR channel, whose intensity values are smaller than $t_{\min}$ are discarded. With the list of all voxels of interest, the patches are produced for all the inputted sequences. The set of all the computed patches is stacked into an array $P=[n \times p \times p \times p \times c]$,

where n denotes the number of patches and c the number of sequences used in training, such as T1-w FLAIR. Furthermore, a list of labels L_n is also estimated from the lesion annotations of the experts, where $L_n=1$ if the current voxel n is a lesion and $L_n=0$ otherwise. To avoid error due to class imbalance, the negative class was randomly under-sample, meaning that the training vector F_1 has the same sample of both classes. After training the CNN_1 , the patches array is used to obtain the probability Y_n^1 of each voxel n to belong to the positive class. According to Y_n^1 , a new balance training feature set is obtained using all positive voxels of P , and the same number randomly selected negative was miss predicted. Therefore $Y_n^1 > 0.5$ with $L_n=0$. Lastly, the following network, CNN_2 , is trained using the F_2 set, in which the output is the probability Y_n^2 for each voxel of being part of a WM lesion, Figure 5.5.

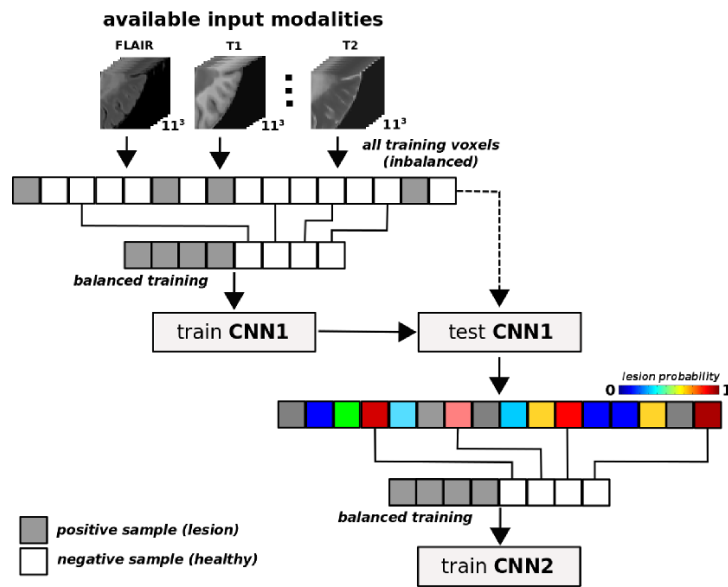


Figure 5.5: Training Pipeline. Input MRI sequences are processed to create an array of patches, which is included a balanced number of patches from each of the classes. This array is inputted to CNN_1 to be trained. The originally unbalanced array is then tested to this network, and its results are threshold, the classes are balanced, and used as input to CNN_2 , concluding the training process (From[74]).

5.4.2 Model Architecture

Therefore, as earlier referred, the model is composed of two seven layers CNNs (CNN_1 and CNN_2). Each network is comprised of two convolutional layers accompanied by max-polling layers with 32 and 64 filters, respectively. The network culminates with a fully connected (FC) layer of size 256, a softmax, and an FC layer of size two, which outputs the probability maps of each voxel to belong to either the positive or negative class. The architecture is resumed in Figure 5.6.

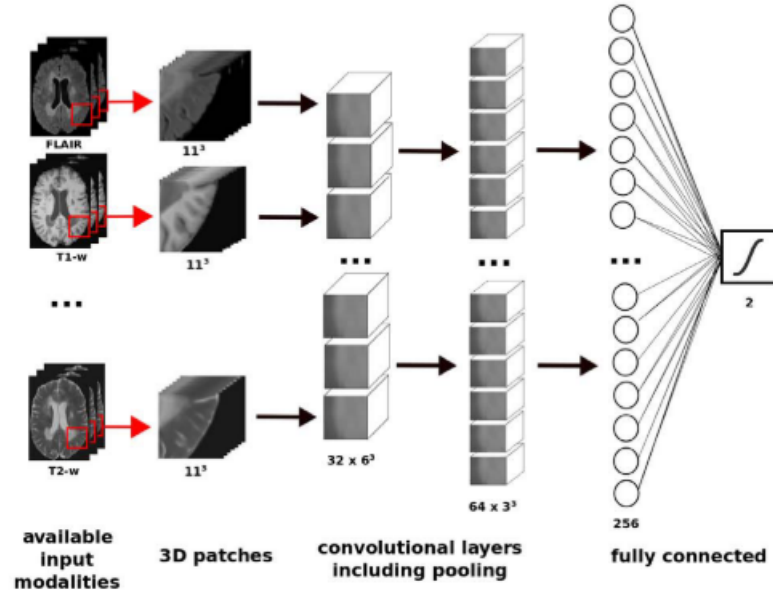


Figure 5.6: CNN Architecture (From [74]).

5.4.3 CNN Training

In order to optimize CNN weights, training data was split into training and validation sets. Network parameters are learned by adopting the adaptive learning rate method *Adadelta* and *Adam* optimizer, with a batch size of 128 and sparse categorical cross-entropy as the lost cost. Overfitting is a constant concern in this process, therefore to reduce it, batch-normalization and dropout were employed before the first fully connected layer. The drop percentage was set to $p=0.5$, which means that 50% of the neurons are set to zero, not influence the final results of the network. Moreover, early-stopping was implemented, terminating the training of the model after a certain number of epochs with no improvement, particularly the decrease of the validation loss.

5.4.4 Data Augmentation

Data augmentation, as previously mentioned in Section 4, is a technique that, similarly to dropout or batch normalization, aids in reducing the overfitting and increase overall accuracy since it consists of increasing the data used for training carried on transformations applied to the original dataset. These transformations operate from rotation and translation to noise addition or even changes in brightness. This technique is common in medical applications, considering the limited data available to train the algorithms. Accordingly, to the original images were applied a rotation of 180 degrees in the axial plane. Other rotations were withdrawn in order to preserve the symmetry and evade artificial rotations of brain structures, as suggested by the author [74].

5.4.5 CNN Testing

Achieved the training step, new images are processed by the CNN cascade model, such as illustrated in Figure 5.7. For each new patient used for testing, the input patches are obtained with a similar approach as previously described. CNN₁ discards all voxels with a low probability of being classified as a lesion. The remaining voxels are re-evaluated with CNN₂, resulting in the probabilistic lesion mask. The final binary mask is calculated with a linear threshold t_{bin} of CNN₂ output. Subsequently, a post-processing step is executed to reduce the false positives by rejecting the lesion with a dimension below l_{min} .

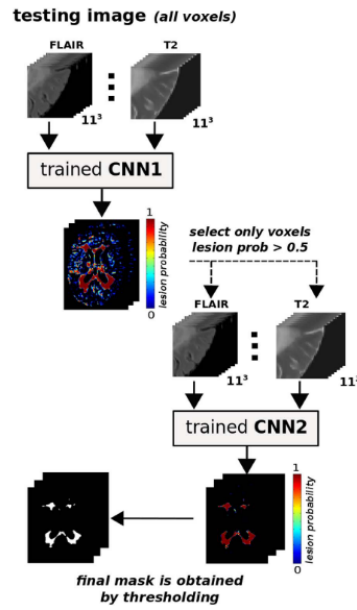


Figure 5.7: Flowchart of the testing procedure of the proposed cascade model. Similar to training, an array of patches is formed and applied sequentially to each network to be evaluated (From [74]).

5.4.6 Evaluation

The evaluation of the lesion segmentation results, using the test set, would have to be performed blind since it would require submitting the results to the challenge website to obtained feedback. So, to overcome this obstacle, tests step were conducted on images from the training set but excluded from the training process. The evaluation metrics to those images are based on the following scores:

- The percentage error of the absolute difference in lesion volume (VD) between manual masks and the segmentation output:

$$VD = \frac{|TP_s - TP_{gt}|}{TP_{gt}} * 100 \quad (5.1)$$

where TP_s and TP_{gt} represent the number of segmented voxels in the output and manual mask, respectively.

- Sensitivity of the method in terms of the True Positive Rate (TPR) between the experts segmentation and the model output, expressed in percentage:

$$TPR = \frac{TP_d}{TP_d + FN_d} * 100 \quad (5.2)$$

where TP_d and FN_d denote the number of correctly and missed lesion region candidates, respectively.

- False discovery rate of the method in terms of the False Positive Rate (FPR) between manual lesion and output segmentation mask, presented in percentage:

$$FPR = \frac{FP_d}{FP_d + TP_d} * 100 \quad (5.3)$$

where FP_d denotes the number of lesion region candidates wrongly classified as a lesion.

5.4.7 Comparison with other methods

The performance of the proposed approach is compared with other available state-of-the-art MS lesion segmentation algorithms such as Lesion Segmentation Toolbox [113]. In this toolbox, the author presents two algorithms, the Lesion Growth Algorithm (LGA) [114], and the Lesion Prediction Algorithm (LPA) [115].

LGA performs the segmentation based on a lesion growth method, unsupervised algorithm. FLAIR intensity distributions are estimated for each of the three tissue classes to detect FLAIR-hyperintense outliers weighted according to their spatial probability of being WM, resulting in the lesion belief maps for each tissue type. They are then summed, and the binary version of the lesion belief maps for the GM is thresholded with a set initial value, being $k=0.3$ by default, which is then used as the seed for the lesion growth. The growth model expands towards the summed belief map, which occurs until no voxel is assigned to lesions. Besides the FLAIR images, this algorithm requires T1 images to estimate the segmentation of the different tissues. The initial threshold value is ranged from 0 to 1.

LPA is a voxel-wise binary logistic regression classification model used to segment lesion that appears hyperintense in FLAIR images. The model was trained on 53 MS patients from a cohort of 312 subjects, whose total lesion volume was superior to 10 ml. Additionally, the trained involved the T1-w and FLAIR. Previous to training, two features were extracted from the data, the position from each brain voxel in a common space and the lesion belief map. The positions are obtained based on T1-w coregister FLAIR images normalized to the MNI space, generating an inverse deformation field utilized to map the MNI coordinates into the particular native space of the patient in the process. The lesion belief map is obtained with a tissue segmentation of the

FLAIR image, and intensities are standardized, dividing each voxel by the mean of segmented GM, subtracted from the FLAIR intensities. From the previous operation are kept the positions with positive results. Moreover, the remaining voxels are multiplied by the tissue probability map for WM, collected from the inverse deformation field from the former step. The two features are joined, during training, adopting a voxel-wise logistic regression model. Unlike the previous methods, LPA only requires FLAIR images, being optional to input the T1 image.

Nonetheless, in both methods, the inputs need to be the original images with no processing applied onto them, i.e., the MRI scans with not skull stripping or bias field correction applications. Before applying either one of the methods, the images were downsampled for half of the initial size, enabling the program to run without any issues.

5.5 Lesion Evaluation

After the segmentation stage, some steps were applied to analyze the obtained results quantitatively. The volume brain was estimated using the outputted binary mask used to extract non-brain structures based on the skull-stripping step. The estimated lesion segmentation was combined with the annotations from the experts. The combination was accomplished in two steps:

- The estimated mask is altered, i.e., the voxels whose intensity values were one got switch to value two.
- The masks were summed and overlayed in the ITK-Snap with its corresponding T1-w, which gives the volume of each label.

The ITK-Snap software permits collecting metrics, like the number of voxels and volume, based on the segmentation labels. These labels correspond to the joined mask values, intending value one corresponded to initial lesion segmentation executed by the experts, two express the estimated mask, and three for the intersection of the previous two. Moreover, having both the brain volume and the lesion volume, the percentage of the lesion is obtained by the quotient of lesion volume and brain volume.

The volume computation was executed by multiplying the number of voxels from masks, N_{voxels} , and the voxels dimension of the image, $1 \times 1 \times 1 \text{ mm}^3$. The resultant volume, in mm^3 , is then converted into ml.

5.6 Adopted Resources

The present section details the various tools, either hardware or software, used throughout the development of this project.

5.6.1 ITK-SNAP

ITK-SNAP is a software application that allows the navigation of 3D medical images. Besides that, it also facilitates the manual and semi-automatic segmentation of those images. It is an open-source platform distributed by the General Public License commonly adopted to segmentation CT, and MR images [116]. This application supports numerous file formats, including NHDR and Nifti, and supports multi-variate imaging, like DTI.

The MS lesion annotations of the MICCAI dataset were performed with a resource of this software. As visible in the Figure 5.8, the top panel and the bottom right allow the visualization and navigation in the three anatomical planes (axial, coronal, and sagittal). The fourth panel allows observing a 3D model of the segmentation annotations, not illustrated in this Figure 5.8.

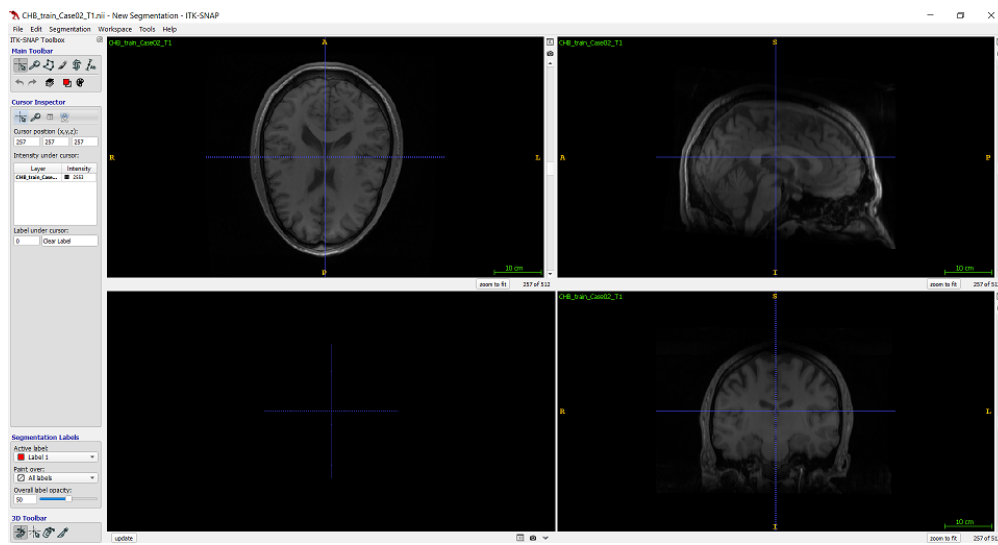


Figure 5.8: ITK-SNAP interface with an example from the MICCAI 2008 dataset. The four black panels allow the visualization of the three anatomical planes of the medical image, and the fourth, the bottom left panel, permits a 3D visualization of the segmentation information.

5.6.2 Anaconda - Jupyter Notebook

Anaconda is an open-source package and environment management system which runs in the major operating systems. It is a distribution of Python and R programming languages, especially for the data science and machine learning areas. It is available and maintained by an organization with the same name, which was founded in 2012 [117, 118].

Jupyter Notebook, similarly to Anaconda, is an open-source web application that permits the production and sharing of documents, which include live code, visualizations, such as images or charts, and explanatory text [119]. Besides, the web application Jupyter notebook combines a second part, the notebook documentation, rendering all content visible in the previous part, including inputs and outputs of the computations, images, and the narrative text [120].

These two applications operated collectively and were adopted to run the code and implement the developed methodology, mainly to execute the code on the server end.

5.6.3 Tensorflow and Keras

As referred in the section 5.4, the model was implemented employing Tensorflow [111] and Keras[112]. They are packages available for Python programming language that aim to facilitate machine learning development. Initially, they were two distinct packages, but Keras was recently merged into the TensorFlow API, strengthening this tool. The Tensorflow was developed by Google Brain, a team inside Google corporation. The most recent version available is Tensorflow 2.5. The various tasks that can be performed using this software, due to its adjustable architecture, can easily facilitate the use of different hardware to run the code, meaning that it permits the execution of the code in various platforms, like CPUs, GPUs, and TPUs, which are the Central, Graphical and Tensor processing units, respectively.

5.6.4 Matlab, SPM and LST

Matlab is a well know interactive software and a programming language developed to solve mathematical problems. Unlike the remaining tools, this one is not open-source. Still it is available free of charge for students. It is available and maintained by MathWorks, a company founded by the authors of Matlab, Jack Little, and Cleve Moler [121].

Even though its primary intent is numeric computing, the numerous toolbox and software programs available permit the user to perform the most variety of operations. One of those programs is the Statistical Parametric Mapping software toolkit to analyze and process medical images. Its original release was in October of 2014, and the currently available version is the SPM12 [122].

The Lesion Segmentation Toolbox (LST) [113] is an open-source toolbox for SPM, which can segment T2 hyperintense lesion in FLAIR images, was primarily designed for MS lesion segmentation, but it has been used in other diseases like Alzheimer's. This toolbox is composed of two algorithms, the Lesion Growth Algorithm (LGA) [114] and Lesion Prediction Algorithm (LPA) [115], which are employed to compared with the results obtained for the proposed methodology. Furthermore, the toolbox permits creating a report with results for the algorithm, where among other information, presents the estimated lesion volume and the number of lesions. An example of this report is shown in Figure 5.9.

5.7 Summary

This chapter detailed the main steps and methods used in the development of this project. The workflow was divided into three main steps: preprocessing, lesion segmentation, lesion evaluation.

The preprocessing image phase involved the application of techniques for extraction of non-cerebral structures, skull-stripping, bias field correction, downsample, and intensity normalization with the resource of Python libraries like Simple ITK and Scipy implemented within the Jupyter Notebook. For the skull-stripping different automatic methods were tested, BET inserted in the FSL toolbox, and deepbrain an open-source python algorithm. The results were analyzed and compared.



Figure 5.9: The report retrieves from the LST software. The left part presents information related to the images used and algorithm name and parameters, like threshold and the maximum number of iterations if the selected algorithm is the LGA. Furthermore, it presents an illustration of the lesion location in the left bottom section. The right section exhibits the outcomes and a slide frame to visualize the overlay obtained lesion segmentation annotations.

For the lesion segmentation step, a cascade model composed of two patch-wise CNN was adopted. The architecture is equal for both networks, summing a total of seven layers. Subsequently, a post-processing algorithm was executed to reduce the false positives. The results obtained were evaluated by multiple metrics and corresponded to the manual annotation given by experts. Moreover, the models were compared to existing models, LGA and LPA from the Lesion Segmentation Toolbox.

Lastly, some of the resources used during the elaboration of this methodology are presented, like ITK-SNAP, Anaconda, and Matlab. The outcomes achieved by exercising the approach presented in this chapter and its examination are presented in the succeeding chapter.

Chapter 6

Results and Discussion

This chapter exhibits the results obtained after implementing the adopted methodology on the publicly available dataset and the respective analysis. Moreover, these outputs are compared to those of existing algorithms, ending with a summary.

6.1 Implementation

6.1.1 Computational Infrastructure

The development of this experiment was in a Linux server with NVIDIA DGX with 4 Teslas V100 (32GB) with Intel Xeon (40 cores) and 256GB of memory. The operating system was Ubuntu 16.04. The access to the server was performed with a Microsoft Windows 10 Home 64-bit PC, with 16GB of RAM which was compatible with all the additional tools used (Anaconda, MATLAB, ITK-Snap).

This methodology was implemented using Python 3.8 on a Jupyter Notebook. Furthermore, the Lesion Segmentation Toolbox(LST) version 3.0.0 contains the LGA and LPA algorithms used to compare the results, is executed on top of the Statistic Parametric Mapping toolbox, and the version install was the SPM12, all of those were performed with MATLAB R2021a. ITK-SNAP version utilizes the 3.8.0, released in 2019.

6.1.2 Image Formats

The MICCAI 2008 original dataset was available in an NHDR format, Nearly High Dynamic Range, characterized by having two files to save the information from one scan, one with the header information and the other a raw file with the image for each modality.

Considering the number of files produced with this format, the images were converted into a Nifti format, a well-known medical format. Nifti, Neuroimaging Informatics Technology Initiative, was an operation group in early 2000 at the US National Institute of Health (NIH) that developed this format to replace an existing file type presented with many issues, ANALYZE shortage of spatial orientation information [123]. This new format, nifti, provides information on

the image volume space in two different manners: one comprising a rotation and translation to map the voxel coordinates to the scanner frame of reference using a quaternion encoding, and the second save 12 parameters conferring a more general linear change, representing the alignment of the image to a standard or a template-based coordinate method [123]. Besides all the information stored inside a Nifti file, the primary difference to the NHDR is that it allows the storage of information in a single file.

The conversion from NHDR to Nifti format achieved with the `pynrrd` library [124], a Python library dedicated to manipulated nhdr files, combined with the `nibabel` [125], a package that handles Nifti, and other medical files formats. So, in a brief description, the original files were read by the `pynrrd`, the respective header, and the image data were merged and later saved with `nibabel` in a Nifti file.

6.2 Image Preprocessing

A. Skull-stripping

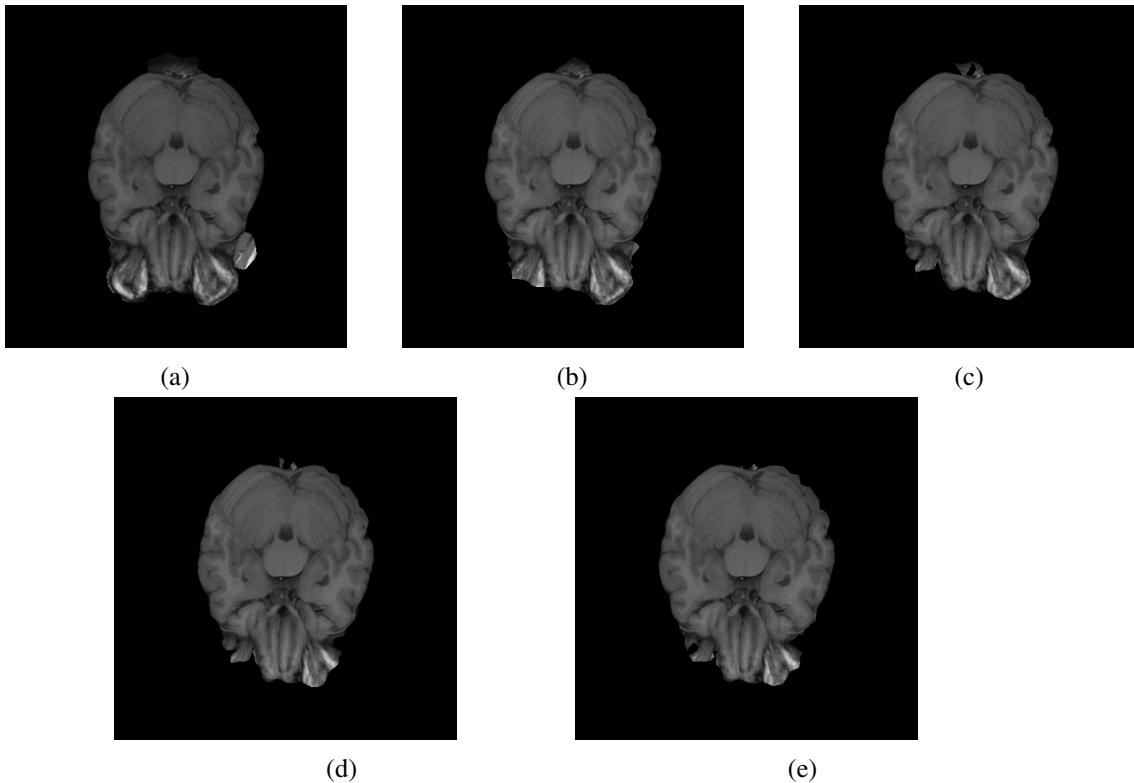


Figure 6.1: Results from the application of the BET algorithm, the images were obtained from the same axial slice and patient. The fractional intensity threshold used were: 6.1a with 0.4, 6.1b with 0.45, 6.1c 0.5, 6.1d with 0.55, 6.1e with 0.6. It is visible the eyes orbits structure being further extracted as the fractional threshold increases.

As mentioned before, this step is significant to reduce error in the segmentation task. BET and Deepbrain were tested to examine their extraction results. BET is integrated into the FSL toolbox

and allows the user to set some parameters, such as fraction intensity threshold. The experimental work of the algorithm was obtained using the option -R (robust), which performs more robust brain center estimation, in each iteration utilizing the same input image and main options. The iterations terminate when the center-of-gravity ceases moving, approximately around 10 iterations. The tests were conducted for different values of fractional intensity threshold within a range of 0.4 to 0.6, considering that the algorithm uses as a default value the 0.5. The Figure 6.1 shows the collected results applying the T1-w image.

Deepbrain requires the use of the T1 images to obtain a more accurate extraction. Similar to the BET algorithm, in the Deepbrain, the probability values to obtain the mask that segments the tissues are altered in a range of 0.45 to 0.55. The results are exhibited in the Figure 6.2.

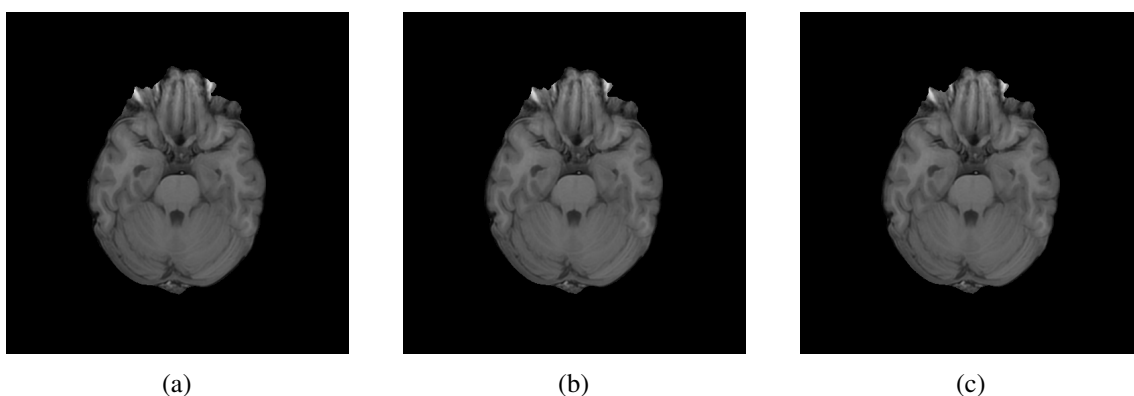


Figure 6.2: Results from the application of the Deepbrain algorithm, the images were obtained from the same axial slice and patient. The fractional intensity threshold used were: 6.2a with 0.45, 6.2b with 0.50 , 6.2c with 0.55.

Even though there is no expert reference to the brain surface, visually, it is clear that Deepbrain performs more precise results, leaving only the brain tissues. At the same time, in the BET algorithm, some non-cerebral structures remain after executing the algorithm, for instance, the eye orbits.

B. Bias Field Correction

The N4 algorithm was implemented using the Simple ITK library, as disclosed in the former chapter. The Figure 6.3 is visible the correction performed, the image on the left is the input image, and the image on the right was the output of the algorithm, where it is denoted a better distribution of the intensities, with darker areas being slightly attenuated. On the other hand, brighter areas with higher intensity values have been highlighted. The presented example was obtained for the T2-w modality of the subject designates as *CHB Case03*.

C. Intensity Normalization

The final pre-processing operation performed was the intensity normalization. This step is crucial since the intensity is used to estimate the lesion tissues. The normalization is performed by subtracting the images to their mean and divided from their variance [74].

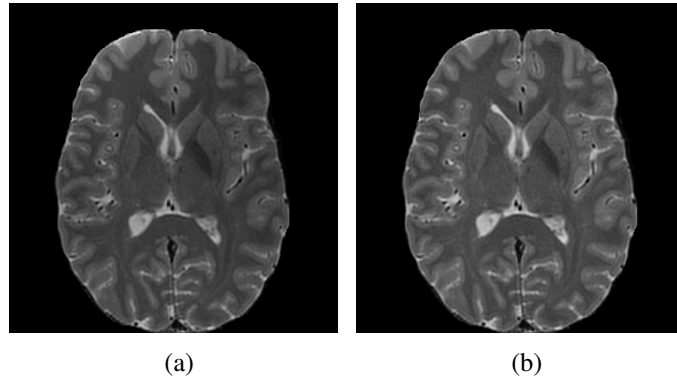


Figure 6.3: Results from the application of the N4 ITK algorithm. In 6.3a is the image inputted to the algorithm and in 6.3b the respective output, obtained for the training set T2-w sequence of subject *CHB Case03*.

6.3 Lesion Segmentation

The MICCAI dataset, as stated early, provides data from two different sites with distinct subjects, so to further explore this diversity of data were evaluated, based on the cascade approach, three models. Two independent models were trained on the data from the two different sites separately, i.e., one model was trained on data from the CHB center and the other with UNC images, and the third with a mixture of data from the two sites. Furthermore, since the challenge provided lesion annotation of the UNC center performed by the CHB expert, an additional model was trained based on those annotations, ultimately resulting in 4 distinct models.

Before examining these four models and understanding the influence of learning rate and the optimization methods adopted, some experiments were conducted to access the reasonable balance between the optimization functions and the network metrics, loss function, and accuracy. Therefore were selected two optimization methods, *Adam* and *Adadelta*, and the learning rate was varied in a range of 0.001 and 0.01. These experiments were performed on the MICCAI model trained on data from both sites. With *Adadelta*, the best balance was obtained with a learning rate of 0.01, whereas for the *Adam* optimizer, that balance was achieved with a value of 0.001. Since, acknowledging that with a low rate, the train time increase, *Adadelta* was selected as the optimizer for the reaming experiments.

Furthermore, considering the dataset provided three distinct MR sequences, FLAIR, T1-W, and T2-W, to investigate the impact of those modalities in the output results, different models were trained using distinct combinations of modalities. The FLAIR modality is an essential component of the model, so it was kept constant in all the combinations. The results for the different combinations are presented in Table 6.1, similarly to the previous assessment analysis was conducted using both centers model.

As expected FLAIR sequence has a significant influence on the obtained results. Compared with the other modalities combinations, the variations are minor for DSC, being the higher difference for the model trained T1-w and FLAIR. The more negligible volume difference was obtained

Table 6.1: Results obtained for different input modalities combination inputted on the model trained with both center data. Mean DSC, VD, TPF, and FPF are presented for each modality combination. The best results are presented in bold.

Input Modalities	DSC	VD	TPR	FPR
FLAIR	28.08	80.96	55.25	71.85
FLAIR, T1-w	35.1	57.34	46.56	62.90
FLAIR, T2-w	32.79	45.01	46.08	71.52
FLAIR, T1-w, T2-w	33.01	51.79	53.30	72.63

with the combination of FLAIR and T2-w. Overall, joining FLAIR with T1-w images presented plausible results with a lower FPR than the others combinations. These experiments corroborate with similar experiments found in the literature [67], with FLAIR expressing a significant weight in the lesion detection.

Once the input modalities and optimization functions and their respective learning rates were set, the initial min threshold was studied. This parameter is used on the FLAIR images to capture the voxels of interest and create the input array. The threshold values were altered from 0.5 to 0.6 using the same model with similar settings, and results are presented in Table 6.2. From the exper-

Table 6.2: Initial min threshold variations. The results were obtained working with the same model trained with the same setting, modifying the threshold values, and testing the same subjects. Mean DSC, VD, TPF, and FPF are presented for each threshold value.

Threshold	DSC	VD	TPR	FPR
0.50	23.73	83.69	84.6	86.2
0.55	23.46	84.32	86.53	86.43
0.60	19.17	87.89	88.73	89.25

iment and based on the content of Table 6.2 it is noteworthy that the best results were obtained for a threshold value of 0.50. In contrast, the least favorable results were obtained for a threshold of 0.6, with a significant difference presented in the DSC value. From the results, it is reasonable to assume that for higher threshold values, lower are the obtained DSC values. Even though the increase of threshold reduces the number of voxels of interest, it also reduces the generalization and increases the number of possible discarded lesion voxels candidates. Moreover, while comparing the values between threshold values 0.50 and 0.55, there are no significant variations for all the presented metrics, yet and based on the previously reported reasons, the min threshold value was set to 0.5 in further experiments.

Based on the former results, and considering the fact that there is no segmentation reference for the test sets provided by the challenges, the models were trained leave-one-out cross-validation, i.e., for each patient, an independent pipeline was trained from the rest of the patient images of the datasets. Resulting in 4 models, one trained with CHB data, one with a mixture of both sites, and two with UNC data, one with labels from UNC experts and the other from the CHB, the obtained

results are presented in Table 6.3.

The best results were obtained for the CHB, and the model trained with data from the two centers, with the CHB model exhibiting higher values. The UNC models performed more poorly compared with the previous. These results could be related to the presented lower lesion volume compared to those of the CHB subjects, which is even more intensified because of the downsample done in the preprocessing step. When examining the two UNC models, the models trained with in-house expert exhibits higher results, with a variation of 7% for DSC. In opposition, for volume difference, the model trained with CHB labels presents a significantly lower value, in contrast to the FPR, suggesting that the estimated lesion exhibit a larger volume than the annotations from the experts.

Table 6.3: Results obtained for the different models, trained on T1 and FLAIR modalities. The two models from the UNC center were trained one with the annotations from the in-house expert and the other with label from the CHB center expert. Mean DSC, VD, TPF, and FPF are presented for each model.

Model	DSC	VD	TPR	FPR
CHB	47.7	34.18	41.62	44.15
UNC	33.4	84.72	36.3	57.91
UNC (w/ labels from CHB)	26.42	53.38	41.14	73.88
CHB and UNC	46.89	37.03	51.28	50.7

Considering that the output from the second CNN is further processed, the minimum size (l_{min}) and the threshold used to obtained the segmentation mask (t_{bin}) are the parameters applied and demand to be optimized. So to understand their impact on the final results, these parameters were tested with different values ranging from 10 to 25 and 0 to 1, for l_{min} and t_{bin} , respectively. This examination was conducted on the UNC model, and the results are illustrated in Figure 6.4.

As observed, the various curves follow a similar pattern, with DSC values not exhibiting significant alterations. The curve from the l_{min} equal to 20 presents the most acceptable balance between higher DSC value and smaller minimum lesion size.

Data augmentation has been used to overcome the small training set, allowing the model to be trained with more data, even though the images are transformation applied to the original data. In the attempt to assess how much this technique is improving the output results, this test was performed in the models with CHB and UNC, using in-house expert labels and the two-center data model trained merely on the provided MR images. The Table 6.4 exposes the retrieved results.

As anticipated, the results were slightly worse compared to those using data augmentation presented in Table 6.3. The models trained with the CHB subset presented a less than 1% (0,67%) for the DSC. At the same time, UNC models exhibit a difference of 12.25%. Ultimately, model using the two center data, the variation was equal to 11,27% for the DSC and more than 18 % less for the FPR. Considering these variations, it is noteworthy that the UNC model had the most profitable and significant improvement with the application of data augmentation, being that from the three models in the study was the one with the higher difference percentage.

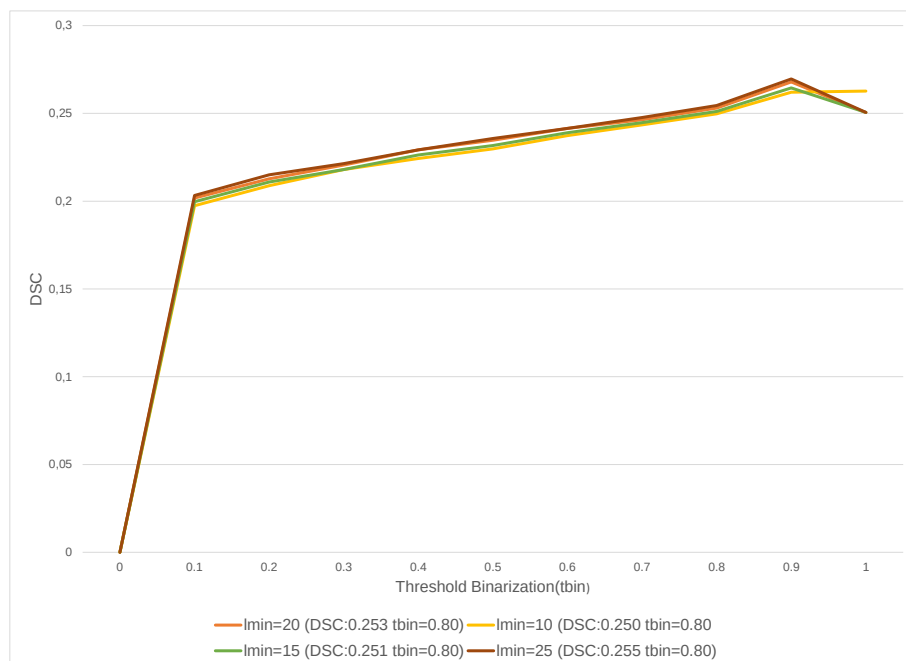


Figure 6.4: Parameter optimization for the cascaded CNN, tbin and lmin, versus DSC coefficient, obtained for the UNC model, using T1-w and FLAIR images as input sequences.

Table 6.4: Results obtained for the different models without applying data augmentation. The two models from the UNC center were trained one with the annotations from the in-house expert and the other with label from the CHB center expert. Mean DSC, VD, TPF, and FPF are presented for each threshold value.

Model	DSC	VD	TPR	FPR
CHB	47.03	46.57	39.73	42.3
UNC	21.15	52.23	32.64	83.37
CHB and UNC	35.62	48.0	56.17	68.98

• Comparing with other methods

The performance of the proposed approach is compared with other available state-of-the-art MS lesion segmentation algorithms such as Lesion Segmentation Toolbox [113]. In this toolbox, the author presents two algorithms, the Lesion Growth Algorithm (LGA) and Lesion Prediction Algorithm (LPA). The LGA required a T1-w and a FLAIR image and the value for the initial threshold. This last parameter was experimented with using a range of values from 0.1 to 0.25 with steps of 0.05. The results obtained with the toolbox function that determines the optimal threshold show that the balance between the higher DSC score and the sensitivity is obtained for an initial threshold value of 0.15.

The acquired results were then compared against the models trained with CHB and UNC images separately and are presented in Table 6.5. The CHB data shows better results than the UNC, as occur in the previous experiments, due to the higher lesion volume presented in the CHB subjects than for the UNC, which for the LPA algorithm may be similar to the images using in its training. LGA presented better results in both cases, with a slight difference for LPA with CHB data, whereas for UNC, the DSC values present a difference near the 50%. Besides the lower results obtained with the UNC data, the proposed methodology outperformed the LST algorithms, except for the volume difference where the LGA showed better results in the two studied situations.

Table 6.5: Segmentation results for each of the evaluated methods on the two MS datasets. Results are shown for LST algorithms, LGA and LPA and the proposed approach trained with the two center MICCAI dataset. Mean DSC, VD, TPF, and FPF are presented for each method and database.

Model	DSC	VD	TPR	FPR
CHB center data				
LGA	54.05	24.51	62.82	52.58
LPA	51.72	56.47	85.25	62.89
Proposed model	47.7	34.18	41.62	44.15
UNC center data				
LGA	21.30	56.54	35.17	84.72
LPA	13.83	85.62	55.02	92.09
Proposed model	33.4	84.72	36.3	57.91

6.4 Lesion Evaluation

After the segmentation, some examples from the models were evaluated. Based on these different labels, the ITK-Snap computes volume and other statistics presented in Table 6.6.

The masks were joined and visualized in the ITK-Snap, Figure 6.5, where the red color represents the lesion obtained from the annotations of the experts, the green illustrates the estimated lesions segmentation resultant from the cascade models, and the blue is the interception of the previous two lesion masks.

The Table 6.6 demonstrates the higher lesion volumes of the CHB subject compared to the UNC, which is a small representation of the entire subsets. The CHB trained model presents, even if a significant difference, the closest to the manual label, as visible in the Figure 6.5b. Moreover, in both situations, the models presented higher lesion volumes for UNC participants than the ones from the annotation of the experts, which once again corroborates the previous results. This situation can be explained by the presence of artifacts in the image, such as movements, which led to a misinterpretation of the lesion in the preprocessing step, and is transfer to the segmentation task.

Table 6.6: Characterization of the segmentation models for the different models. The same subjects are compared. The lesion percentage column presents the ratio of lesions with brain volume.

Subjects	Lesion Volume Expert, ml	Estimated Lesion Volume, ml	Estimated Lesion Volume, ml	Brain Volume, ml	Lesion Percentage
CHB Model					
CHB Case06	15.38	7.11	1342.51		0.53%
UNC Model					
UNC Case10	1.13	2.02	1446.46		0.14%
Mix Model					
CHB Case06	15.38	11.38	1342.51		0.85%
UNC Case10	1.13	4.79	1446.46		0.33%

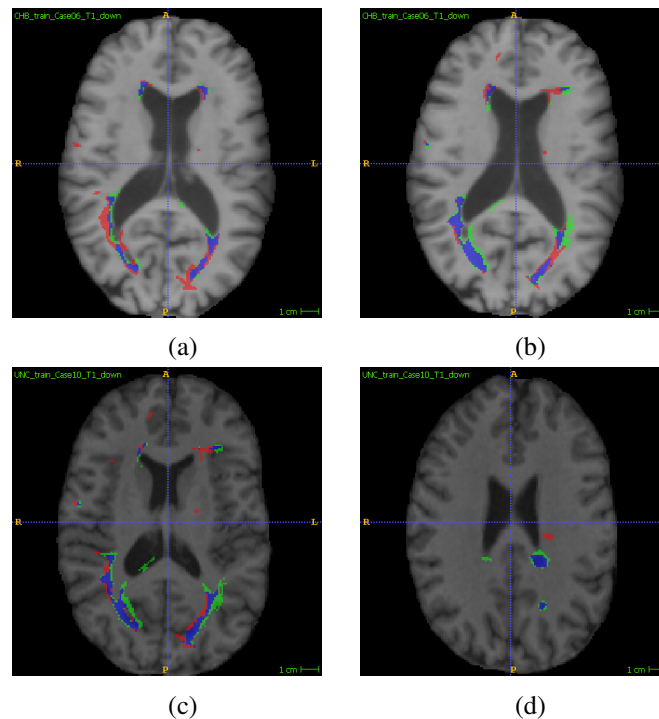


Figure 6.5: Obtained results for the joined masks visualized in ITK-Snap. The red color represents the lesion segmented by the expert, the green estimated lesion, and the blue intersection between the previous two. This example was obtained with the same subjects reported in the previous Table. Subject *CHB Case06* resultant masks for two center data model, 6.5a, and the CHB data trained model, 6.5b. Subject *UNC Case10* retrieved for the model using both center 6.5c, and the UNC data 6.5d. The presented slices are not the same in both subjects for the different models.

6.5 Summary

This chapter illustrated the outcomes of the processing and analysis techniques used to study MR neurological images of MS subjects. Of the preprocessing steps, four tasks were executed, the

skull stripping, with a comparative analysis of the two selected algorithms, BET and Deepbrain. Followed by the application of the N4 algorithms performing a bias correction. Ultimately, the downsample, cropped, and intensity normalization.

The models were trained on the MICCAI dataset, and given the multi-center data CHB and UNC included, was divided, permitting the trained of four similar models. Two with data from the separate centers, one of those was further split considering the different label sources, and one with combinations of both centers. The model trained with CHB data presented the best results associated with higher lesion volume subjects than the UNC data. The FLAIR modality shows a more decisive influence on the output, as stated in the literature, followed by combining FLAIR and T1-W for the MICCAI dataset used for all of the experiments.

Compared with existing methods, LPA and LGA, LGA presented higher results than the LPA algorithm. The best results were obtained for the models using the CHB data, yet the proposed models still had a lower result. Nonetheless, for the UNC data, the proposed model outperformed both LST algorithms.

Lastly, the estimated segmentation mask and the manual annotation were joined to access the volume of the lesion using the ITK-Snap software.

Chapter 7

Conclusion and Future Work

The last chapter of the Dissertation draws global appreciation and conclusions from the overall work, accompanied by a future work viewpoint.

7.1 Conclusion

The focus of this Dissertation was processing and analysis of the medical image, particularly the study of preprocessing techniques, image segmentation, and classification to identify multiple sclerosis lesions on MR images.

The developed work provided an understanding of the disease, afflicting millions of people worldwide, how it is characterized, and the diagnosis procedure used. Furthermore, concepts alongside the pathology were also acquired, such as magnetic resonance imaging technique and the main steps in the image analysis and processing methods.

Throughout the initial state-of-the-art search, it was evident that MS image analysis is a pertinent and current theme, based on the number of papers published each year. Moreover, the examination resulted in two main focuses related to the study of MS and computer image analysis- lesion segmentation and disease classification. Lesion segmentation is essential in MS characterization, recognizing that its amount (quantity and volume) helps determine disease stages and, most of all, assess progression. Another relevant factor about the evolution of the condition is the appearance of new lesions.

The experimental work was executed in the publicly available dataset, the MICCAI 2008. The pipeline is composed of three main stages preprocessing, lesion segmentation, and evaluation. The preprocessing steps reviewed and discussed to MR images reveal that such enhancement is essential for improving the subsequent steps. Registration, the first step, is crucial for adequately comparing images either from the same or distinct subjects, even though not performed, since the provided images have been registered before being available. Followed by the skull-stripping, which removes no-brain tissue, bias field correction to attenuate some inhomogeneities cause during the acquisition process, image size reduction (cropping and downsample), and finally, intensity normalization.

The lesion segmentation models proposed in this work rely on a cascade structure, which included two convolutional neural networks. The first is trained to be more sensitive, revealing possible candidate lesion voxels, whereas the second is trained to decrease wrongly assigned voxels originating from the former network. Both networks were designed with equal architectures, a seven-layer CNN.

As for the lesion segmentation, voxel and patch approaches are the more predominant strategy used based on intensity. Concerning the MRI sequence, T1-w images are more regularly utilized for the segmentation of brain tissue. Whereas the T2-w or PD are common in detecting lesions, and CSF is challenging to distinguish them. Therefore, FLAIR sequences have risen to overcome this issue, considering they allow the suppression of the CSF so present a better separation of lesion and brain tissues.

The experimental work was executed in the publicly available dataset, the MICCAI 2008. Based on the aforementioned methods, the preprocessing steps focus on extracting non-brain tissues, correcting inhomogeneities and the intensity normalization since the provided images were previously registered.

The lesion segmentation models proposed in this work rely on a cascade structure, which included two convolutional neural networks. The first is trained to be more sensitive, revealing possible candidate lesion voxels, whereas the second is trained to decrease wrongly assigned voxels originating from the former network.

As for the lesion evaluation step was performed a volumetric analysis using the estimated lesion segmentation and the annotations from the experts. This task was mainly executed with ITK-Snap software.

Some trials were conducted during the experimental phase to test the cascade model versatility and explored the dataset. In the literature, T1-w images are more commonly utilized for the segmentation of brain tissue. Whereas the T2-w or PD are common in detecting lesions, and CSF is challenging to distinguish them. Therefore, FLAIR sequences have risen to overcome this issue, considering they allow the suppression of the CSF so present a better separation of lesion and brain tissues. Consequently, and based on the three modalities available in the dataset, the experiment with different image sequences, corroborated founding in the literature, with overall models trained on FLAIR images, present higher results. However, the best results were collected for the models using FLAIR and T1 sequences.

Furthermore, the dataset was divided into four subsets, resultant from the different sites of acquisition, which origin fourth different models. So, the developed models used as input either data from one of the two centers, CHB or UNC, the latter being divided into two subgroups based on the source of the labels. The results have shown differences between the CHB and UNC data results, which is acceptable because, even though they were acquired with a similar protocol, the subjects in the CHB present higher lesion volumes.

Additionally, the trials conducted with the parameters used to adjust the output estimations, I_{min} and T_{bin} , resulted in a similar pattern for the selected values, with very slight differences

between them. However, the best balance was obtained for l_{min} and t_{bin} equal to 20 and 0.8, respectively.

In an attempt to overcome the data limitation, very common in studies using medical images, it was adopted the data augmentation technique. In the present study, the original images were rotated 180° to create more data. Those images were then applied to the models jointly with the original. From the analysis of trained models with and without the application of this technique, the outcomes show a general increase, being the higher obtained for the UNC models, with a DSC increase value of 12.25%. Contrastingly, the CHB models presented the smallest difference percentage with an increase near 1%.

Confronted with other existing methodologies, the UNC performed poorly, as occurred in the presented approach, yet outperformed the LST algorithms. Of the two LST algorithms, LGA presented better results in both situations, with higher results when experimented on CHB data. As for the LPA, as anticipated, the higher outcomes were obtained with CHB since this subset presents a higher lesion volume similar to those used for the training of the model. Nonetheless, for the UNC data, the proposed model outperformed both LST algorithms.

Concluding, the analysis and preprocessing techniques are essential for the optimization of the subsequent steps. The application of CNN for MS lesion segmentation is viable if the dataset combines a good representation of the fundamental characteristics of this kind of image. This combination is most significant to build a training set with a set of generalized data. Besides, CNN was one of the most reported algorithms within the deep learning branch reviewed in the researched literature. The overall obtained results were promising and demonstrate that computed approaches can assist the medical professional. Confidently, they can have a broader spread in clinical facilities in the near future.

7.2 Future Work

The developed work corroborates computed image processing analysis approaches applied for assisting medical professions to achieve an early diagnosis. The diagnosis of MS is very particular due to its complexity and lack of concrete causes. As for lesion segmentation is challenging because the lesions do not follow a definite pattern of distribution.

The medical image datasets, particularly with public access, are limited and under strict policies to ensure access. Even with techniques like data augmentation, that obstacle can not be entirely overcome. Since the model was tested with a small sample dataset, one possible future task would be to experiment with a large dataset. Furthermore, the dataset used in the experimental tasks focuses solely on the WM lesions, limiting the generalization of the methodology given that lesions also appear in the grey matter, even if in a small ratio compared to WM. As for lesion size, the proposed method is limited to lesions with a minimum size. So, it would be helpful to enhance the model to enable the minor detect lesions since they are a cause for concern, especially in the study to assess new lesion appearance or even for early diagnosis.

Moreover, in a clinical environment, images sometimes possess defects, artifacts that could interfere with the evaluation. Considering that, to estimate the robustness of the models, test them with those particular images, and analyze if the model could overlook these defects and still properly detect the lesion, this way mimicking the clinical context.

References

- [1] Number of people with ms | atlas of ms. <https://www.atlasofms.org/map/global/epidemiology/number-of-people-with-ms>. (Accessed on 01/04/2021).
- [2] J. M.S. Pearce. Historical descriptions of multiple sclerosis: The stories of Augustus d’Este and The Journal of a Disappointed Man. *European Neurology*, 54(1):49–53, 2005. doi:10.1159/000087387.
- [3] Mike P. Wattjes, Àlex Rovira, David Miller, Tarek A. Yousry, Maria P. Sormani, Nicola De Stefano, Mar Tintoré, Cristina Auger, Carmen Tur, Massimo Filippi, Maria A. Rocca, Franz Fazekas, Ludwig Kappos, Chris Polman, Frederik Barkhof, and Xavier Montalban. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis - Establishing disease prognosis and monitoring patients. *Nature Reviews Neurology*, 11(10):597–606, 2015. doi:10.1038/nrneurol.2015.157.
- [4] Vijay P.B. Grover, Joshua M. Tognarelli, Mary M.E. Crossey, I. Jane Cox, Simon D. Taylor-Robinson, and Mark J.W. McPhail. Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians. *Journal of Clinical and Experimental Hepatology*, 5(3):246–255, 2015. URL: <http://dx.doi.org/10.1016/j.jceh.2015.08.001>, doi:10.1016/j.jceh.2015.08.001.
- [5] Tayyabah Yousaf, George Dervenoulas, and Marios Politis. *Advances in MRI Methodology*, volume 141. Elsevier Inc., 1 edition, 2018. URL: <http://dx.doi.org/10.1016/bs.irn.2018.08.008>, doi:10.1016/bs.irn.2018.08.008.
- [6] I.N. Bankman. *Handbook of Medical Image Processing and Analysis*. 2009. cited By 132. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85013986437&doi=10.1016%2fB978-0-12-373904-9.X0001-4&partnerID=40&md5=620362eda2d9f70404d1e3ca7da8c01c>, doi:10.1016/B978-0-12-373904-9.X0001-4.
- [7] Alvaro Alonso and Miguel A. Hernán. Temporal trends in the incidence of multiple sclerosis: A systematic review. *Neurology*, 71(2):129–135, 2008. doi:10.1212/01.wnl.0000316802.35974.34.
- [8] N. Grigoriadis and V. van Pesch. A basic overview of multiple sclerosis immunopathology. *European Journal of Neurology*, 22:3–13, 2015. doi:10.1111/ene.12798.
- [9] Sylvia Klineova and Fred D. Lublin. Clinical course of multiple sclerosis. *Cold Spring Harbor Perspectives in Medicine*, 8(9):1–12, 2018. doi:10.1101/cshperspect.a028928.

- [10] Types of ms | national multiple sclerosis society. <https://www.nationalmssociety.org/What-is-MS/Types-of-MS>. (Accessed on 01/20/2021).
- [11] Hatim Omar and Joshua Chalkley. Multiple Sclerosis in Children and Adolescents. *Multiple Sclerosis in Children and Adolescents*, 7(2):1–202, 2018.
- [12] Clare Walton, Rachel King, Lindsay Rechtman, Wendy Kaye, Emmanuelle Leray, Ruth Ann Marrie, Neil Robertson, Nicholas La Rocca, Bernard Uitdehaag, Ingrid van der Mei, Mitchell Wallin, Anne Helme, Ceri Angood Napier, Nick Rijke, and Peer Baneke. Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Multiple Sclerosis Journal*, pages 1816–1821, 2020. doi:10.1177/1352458520970841.
- [13] Massimo Filippi, Amit Bar-Or, Fredrik Piehl, Paolo Preziosa, Alessandra Solari, Sandra Vukusic, and Maria A. Rocca. Multiple sclerosis. *Nature Reviews Disease Primers*, 4(1):1–27, 2018. URL: <http://dx.doi.org/10.1038/s41572-018-0041-4>, doi:10.1038/s41572-018-0041-4.
- [14] Charles M Poser and Vesna V Brinar. Diagnostic criteria for multiple sclerosis : an historical review. *Clinical Neurology and Neurosurgery*, 106(3):147–158, 2004. doi:10.1016/j.clineuro.2004.02.004.
- [15] Alan J. Thompson, Brenda L. Banwell, Frederik Barkhof, William M. Carroll, Timothy Coetzee, Giancarlo Comi, Jorge Correale, Franz Fazekas, Massimo Filippi, Mark S. Freedman, Kazuo Fujihara, Steven L. Galetta, Hans Peter Hartung, Ludwig Kappos, Fred D. Lublin, Ruth Ann Marrie, Aaron E. Miller, David H. Miller, Xavier Montalban, Ellen M. Mowry, Per Soelberg Sorensen, Mar Tintoré, Anthony L. Traboulsee, Maria Trojano, Bernard M.J. Uitdehaag, Sandra Vukusic, Emmanuelle Waubant, Brian G. Weinshenker, Stephen C. Reingold, and Jeffrey A. Cohen. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, 17(2):162–173, 2018. doi:10.1016/S1474-4422(17)30470-2.
- [16] R. Dobson and G. Giovannoni. Multiple sclerosis – a review. *European Journal of Neurology*, 26(1):27–40, 2019. doi:10.1111/ene.13819.
- [17] National Multiple Sclerosis Society. Rehabilitation | national multiple sclerosis society. <https://www.nationalmssociety.org/Treating-MS/Rehabilitation>, 2020. (Accessed on 01/22/2021).
- [18] Complementary & alternative medicines | national multiple sclerosis society. <https://www.nationalmssociety.org/Treating-MS/Complementary-Alternative-Medicines>, 2020. (Accessed on 01/22/2021).
- [19] Vijayshree Yadav, Lynne Shinto, and Dennis Bourdette. Complementary and alternative medicine for the treatment of multiple sclerosis. *Expert Review of Clinical Immunology*, 6(3):381–395, 2010. URL: <https://dx.doi.org/10.1586/eci.10.12>, doi:10.1586/eci.10.12.
- [20] Bushberg, Jerold, J. Anthony Seibert, Edwin M. Leidholdt, and John M. Boone. *X-ray Production, X-ray Tubes, and X-ray Generators*. Wolters Kluwer Health | Lippincott Williams Wilkins, 2012.

- [21] Andrew Murphy and Dr J.Ray Ballinger. Relaxation | radiology reference article | radiopaedia.org. <https://radiopaedia.org/articles/relaxation?lang=us>. (Accessed on 01/05/2021).
- [22] Dr Usman Bashir Andrew Murphy. T2 relaxation | radiology reference article | radiopaedia.org. <https://radiopaedia.org/articles/t2-relaxation>. (Accessed on 01/25/2021).
- [23] Andrew Murphy and Dr J.Ray Ballinger. Mri pulse sequences | radiology reference article | radiopaedia.org. <https://radiopaedia.org/articles/mri-pulse-sequences-1?lang=us>. (Accessed on 01/07/2021).
- [24] Anthony L. Traboulsee and David K. B. Li. The role of mri in the diagnosis of multiple sclerosis. *Advances in Neurology*, 98:125–46, 02 2006. (Accessed on 01/08/2021).
- [25] Massimo Filippi, Paolo Preziosa, Brenda L. Banwell, Frederik Barkhof, Olga Ciccarelli, Nicola De Stefano, Jeroen J.G. Geurts, Friedemann Paul, Daniel S. Reich, Ahmed T. Toosy, Anthony Traboulsee, Mike P. Wattjes, Tarek A. Yousry, Achim Gass, Catherine Lubetzki, Brian G. Weinshenker, and Maria A. Rocca. Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. *Brain*, 142(7):1858–1875, 2019. doi: [10.1093/brain/awz144](https://doi.org/10.1093/brain/awz144).
- [26] Katarzyna Krupa and Monika Bekiesińska-Figatowska. Artifacts in magnetic resonance imaging. *Polish Journal of Radiology*, 80(1):93–106, 2015. doi: [10.12659/PJR.892628](https://doi.org/10.12659/PJR.892628).
- [27] H. Kim. *Design and Optimization for 5G Wireless Communications - Machine Learning*, chapter 5, pages 151–193. Jonh Wiley Sons, Ltd, 2020. doi: [10.1002/9781119494492.ch5](https://doi.org/10.1002/9781119494492.ch5).
- [28] Abdulhamit Subasi. Practical machine learning for data analysis using python - machine learning techniques. chapter 3, pages 91 – 202. Academic Press, 2020. URL: <http://www.sciencedirect.com/science/article/pii/B9780128213797000035>, doi: <https://doi.org/10.1016/B978-0-12-821379-7.00003-5>.
- [29] Altug Akay and Henry Hess. Deep Learning: Current and Emerging Applications in Medicine and Technology. *IEEE Journal of Biomedical and Health Informatics*, 23(3):906–920, may 2019. URL: <https://ieeexplore.ieee.org/document/8624419/>, doi: [10.1109/JBHI.2019.2894713](https://doi.org/10.1109/JBHI.2019.2894713).
- [30] Susmita Ray. A Quick Review of Machine Learning Algorithms. *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Perspectives and Prospects, COMITCon 2019*, pages 35–39, 2019. doi: [10.1109/COMITCon.2019.8862451](https://doi.org/10.1109/COMITCon.2019.8862451).
- [31] Mohammadreza Sheykhmousa, Masoud Mahdianpari, Hamid Ghanbari, Fariba Mohammadimanesh, Pedram Ghamisi, and Saeid Homayouni. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:6308–6325, 2020. doi: [10.1109/JSTARS.2020.3026724](https://doi.org/10.1109/JSTARS.2020.3026724).

- [32] Buettner, Ricardo and Kuri, Tarik and Feist, Andreas and Hudak, Jannik. Overview of Machine Learning Approaches Applied in Disease Profiling. *2020 IEEE Symposium on Industrial Electronics and Applications, ISIEA 2020*, 2020. doi:10.1109/ISIEA49364.2020.9188140.
- [33] Idongesit Zion, Simeon Ozuomba, and Philip Asuquo. An Overview of Neural Network Architectures for Healthcare. *2020 International Conference in Mathematics, Computer Engineering and Computer Science, ICMCECS 2020*, 2020. doi:10.1109/ICMCECS47690.2020.246980.
- [34] Li Deng and Dong Yu. Deep learning: Methods and applications. *Foundations and Trends in Signal Processing*, 7(3-4):197–387, 2013. doi:10.1561/20000000039.
- [35] Maribel Torres-Velazquez, Wei-Jie Chen, Xue Li, and Alan B. McMillan. Application and Construction of Deep Learning Networks in Medical Imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences*, pages 1–1, 2020. URL: <https://ieeexplore.ieee.org/document/9222094/>, doi:10.1109/TRPMS.2020.3030611.
- [36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *LNCIS*, 9351:234–241, 10 2015. doi:10.1007/978-3-319-24574-4_28.
- [37] Stanford University. Cs231n convolutional neural networks for visual recognition. <https://cs231n.github.io/convolutional-networks/>. (Accessed on 05/28/2021).
- [38] Afshine Amidi and Shervine Amidi. Cs 230 - convolutional neural networks cheatsheet. <https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks#hyperparameters>, 2018. (Accessed on 05/27/2021).
- [39] Phil Kim. MATLAB Deep Learning. *MATLAB Deep Learning*, (November 2013):121–147, 2017. doi:10.1007/978-1-4842-2845-6.
- [40] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR. URL: <http://proceedings.mlr.press/v37/ioffe15.html>.
- [41] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 06 2014. URL: <http://www.cs.toronto.edu/~rsalakhu/papers/srivastava14a.pdf>.
- [42] Nurshazlyn Mohd Aszemi and Dhanapal Durai Dominic Panneer Selvam. Hyperparameter optimization in convolutional neural network using genetic algorithms. *International Journal of Advanced Computer Science and Applications*, 10:269 – 278, 06 2019. doi:10.14569/IJACSA.2019.0100638.
- [43] Matthew D. Zeiler. Adadelta: An adaptive learning rate method, 2012. arXiv:1212.5701.

- [44] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [45] U. Raghavendra, U. Rajendra Acharya, and Hojjat Adeli. Artificial Intelligence Techniques for Automated Diagnosis of Neurological Disorders. *European Neurology*, 82(1-3):41–64, 2020. [doi:10.1159/000504292](https://doi.org/10.1159/000504292).
- [46] Hidetaka Arimura, Taiki Magome, Yasuo Yamashita, and Daisuke Yamamoto. Computer-aided diagnosis systems for brain diseases in magnetic resonance images. *Algorithms*, 2(3):925–952, 2009. [doi:10.3390/a2030925](https://doi.org/10.3390/a2030925).
- [47] José María Mateos-Pérez, Mahsa Dadar, María Lacalle-Aurioles, Yasser Iturria-Medina, Yashar Zeighami, and Alan C. Evans. Structural neuroimaging as clinical predictor: A review of machine learning applications. *NeuroImage: Clinical*, 20(January):506–522, 2018. URL: <https://doi.org/10.1016/j.nicl.2018.08.019>, [arXiv:1806.03512](https://arxiv.org/abs/1806.03512), [doi:10.1016/j.nicl.2018.08.019](https://doi.org/10.1016/j.nicl.2018.08.019).
- [48] Begum Irmak On Seker, Kelly Reeve, Joachim Havla, Jacob Burns, Martina A. Gosteli, Andreas Lutterotti, Sven Schippling, Ulrich Mansmann, and Ulrike Held. Prognostic models for predicting clinical disease progression, worsening and activity in people with multiple sclerosis. *Cochrane Database of Systematic Reviews*, 2020, 5 2020. [doi:10.1002/14651858.CD013606](https://doi.org/10.1002/14651858.CD013606).
- [49] H. M.Rehan Afzal, Suhuai Luo, Saadallah Ramadan, and Jeannette Lechner-Scott. The emerging role of artificial intelligence in multiple sclerosis imaging. *Multiple Sclerosis Journal*, pages 1–10, 2020. [doi:10.1177/1352458520966298](https://doi.org/10.1177/1352458520966298).
- [50] Monika A. Myszczyńska, Poojitha N. Ojamies, Alix M.B. Lacoste, Daniel Neil, Amir Saffari, Richard Mead, Guillaume M. Hautbergue, Joanna D. Holbrook, and Laura Ferraiuolo. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nature Reviews Neurology*, 16(8):440–456, 2020. URL: <http://dx.doi.org/10.1038/s41582-020-0377-8>, [doi:10.1038/s41582-020-0377-8](https://doi.org/10.1038/s41582-020-0377-8).
- [51] Paul M. Matthews, Frederico Roncaroli, Adam Waldman, Maria Pia Sormani, Nicola De Stefano, Gavin Giovannoni, and Richard Reynolds. A practical review of the neuropathology and neuroimaging of multiple sclerosis. *Practical Neurology*, 16(4):279–287, 2016. [doi:10.1136/practneurol-2016-001381](https://doi.org/10.1136/practneurol-2016-001381).
- [52] Kelley M. Swanberg, Karl Landheer, David Pitt, and Christoph Juchem. Quantifying the Metabolic Signature of Multiple Sclerosis by in vivo Proton Magnetic Resonance Spectroscopy: Current Challenges and Future Outlook in the Translation From Proton Signal to Diagnostic Biomarker. *Frontiers in Neurology*, 10(November), 2019. [doi:10.3389/fneur.2019.01173](https://doi.org/10.3389/fneur.2019.01173).
- [53] Isbi 2017. <https://biomedicalimaging.org/2015/program/isbi-challenges/>. (Accessed on 01/20/2021).
- [54] NeuroImaging Tool Resources Collaboratory. Nitrc: 2008 miccai ms lesion segmentation challenge: Tool/resource info. <https://www.nitrc.org/projects/msseg>. (Accessed on 01/20/2021).

- [55] Fred D. Lublin, Stacey S. Cofield, Gary R. Cutter, Robin Conwit, Ponnada A. Narayana, Flavia Nelson, Amber R. Salter, Tarah Gustafson, and Jerry S. Wolinsky. Randomized study combining interferon and glatiramer acetate in multiple sclerosis. *Annals of Neurology*, 73(3):327–340, 2013. URL: <https://dx.doi.org/10.1002/ana.23863>, doi:10.1002/ana.23863.
- [56] Mostafa Salem, Mariano Cabezas, Sergi Valverde, Deborah Pareto, Arnau Oliver, Joaquim Salvi, Àlex Rovira, and Xavier Lladó. A supervised framework with intensity subtraction and deformation field features for the detection of new T2-w lesions in multiple sclerosis. *NeuroImage: Clinical*, 17(June 2017):607–615, 2018. URL: <https://doi.org/10.1016/j.nicl.2017.11.015>, doi:10.1016/j.nicl.2017.11.015.
- [57] Ehab Essa, Doaa Aldesouky, Sherif E. Hussein, and M. Z. Rashad. Neuro-fuzzy patch-wise R-CNN for multiple sclerosis segmentation. *Medical and Biological Engineering and Computing*, 58(9):2161–2175, 2020. doi:10.1007/s11517-020-02225-6.
- [58] 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, and Meritxell Bach Cuadra. RimNet: A deep 3D multimodal MRI architecture for paramagnetic rim lesion assessment in multiple sclerosis. *NeuroImage: Clinical*, 28(September):102412, 2020. URL: <https://doi.org/10.1016/j.nicl.2020.102412>, doi:10.1016/j.nicl.2020.102412.
- [59] Ariel Birenbaum and Hayit Greenspan. Multi-view longitudinal CNN for multiple sclerosis lesion segmentation. *Engineering Applications of Artificial Intelligence*, 65(January):111–118, 2017. URL: <http://dx.doi.org/10.1016/j.engappai.2017.06.006>, doi:10.1016/j.engappai.2017.06.006.
- [60] Annegreet Van Opbroek, M Arfan Ikram, Meike W Vernooij, and Marleen De Bruijne. Transfer Learning Improves Supervised Image Segmentation Across Imaging Protocols. *IEEE Transactions on Medical Imaging*, 34(5):1018–1030, 2015.
- [61] Mário João Fartaria, Guillaume Bonnier, Alexis Roche, Tobias Kober, Reto Meuli, David Rotzinger, Richard Frackowiak, Myriam Schluep, Renaud Du Pasquier, Jean Philippe Thiran, Gunnar Krueger, Meritxell Bach Cuadra, and Cristina Granziera. Automated detection of white matter and cortical lesions in early stages of multiple sclerosis. *Journal of Magnetic Resonance Imaging*, 43(6):1445–1454, 2016. doi:10.1002/jmri.25095.
- [62] B. Stasiak, P. Tarasiuk, I. Michalska, and A. Tomczyk. Application of convolutional neural networks with anatomical knowledge for brain MRI analysis in MS patients. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 66(6):857–868, 2018. doi:10.24425/bpas.2018.125933.
- [63] Tanya Nair, Doina Precup, Douglas L. Arnold, and Tal Arbel. Exploring uncertainty measures in deep networks for Multiple sclerosis lesion detection and segmentation. *Medical Image Analysis*, 59:101557, 2020. URL: <https://doi.org/10.1016/j.media.2019.101557>, doi:10.1016/j.media.2019.101557.

- [64] Erol Kazancli, Vesna Prechkovska, Paulo Rodrigues, Pablo Villoslada, and Laura Igual. Multiple sclerosis lesion segmentation using improved convolutional neural networks. *VISI-GRAPP 2018 - Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 4(Visigrapp):260–269, 2018. doi:[10.5220/0006540902600269](https://doi.org/10.5220/0006540902600269).
- [65] Tom Brosch, Lisa Y.W. Tang, Youngjin Yoo, David K.B. Li, Anthony Traboulsee, and Roger Tam. 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, 2016. doi:[10.1109/TMI.2016.2528821](https://doi.org/10.1109/TMI.2016.2528821).
- [66] Tom Brosch, Lisa Y.W. Tang, Youngjin Yoo, David K.B. Li, Anthony Traboulsee, and Roger Tam. 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, 2016. doi:[10.1109/TMI.2016.2528821](https://doi.org/10.1109/TMI.2016.2528821).
- [67] Ponnada A. Narayana, Ivan Coronado, Sheeba J. Sujit, Xiaojun Sun, Jerry S. Wolinsky, and Refaat E. Gabr. Are multi-contrast magnetic resonance images necessary for segmenting multiple sclerosis brains? a large cohort study based on deep learning. *Magnetic Resonance Imaging*, 65:8–14, 2020. URL: <https://doi.org/10.1016/j.mri.2019.10.003>, doi:[10.1016/j.mri.2019.10.003](https://doi.org/10.1016/j.mri.2019.10.003).
- [68] Ponnada A. Narayana, Ivan Coronado, Melvin Robinson, Sheeba J. Sujit, Sushmita Datta, Xiaojun Sun, Fred D. Lublin, Jerry S. Wolinsky, and Refaat E. Gabr. Multimodal mri segmentation of brain tissue and t2-hyperintense white matter lesions in multiple sclerosis using deep convolutional neural networks and a large multi-center image database. *2018 9th Cairo International Biomedical Engineering Conference, CIBEC 2018 - Proceedings*, pages 13–16, 2019. doi:[10.1109/CIBEC.2018.8641800](https://doi.org/10.1109/CIBEC.2018.8641800).
- [69] Refaat E. Gabr, Ivan Coronado, Melvin Robinson, Sheeba J. Sujit, Sushmita Datta, Xiaojun Sun, William J. Allen, Fred D. Lublin, Jerry S. Wolinsky, and Ponnada A. Narayana. Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: A large-scale study. *Multiple Sclerosis Journal*, 26:1217–1226, 2020. doi:[10.1177/1352458519856843](https://doi.org/10.1177/1352458519856843).
- [70] Mostafa Salem, Sergi Valverde, Mariano Cabezas, Deborah Pareto, Arnau Oliver, Joaquim Salvi, Àlex Rovira, and Xavier Lladó. A fully convolutional neural network for new t2-w lesion detection in multiple sclerosis. *NeuroImage: Clinical*, 25:102149, 2020. URL: <https://doi.org/10.1016/j.nicl.2019.102149>, doi:[10.1016/j.nicl.2019.102149](https://doi.org/10.1016/j.nicl.2019.102149).
- [71] Chaoyi Zhang, Yang Song, Sidong Liu, Scott Lill, Chenyu Wang, Zihao Tang, Yuyi You, Yang Gao, Alexander Klistorner, Michael Barnett, and Weidong Cai. MS-GAN: GAN-Based Semantic Segmentation of Multiple Sclerosis Lesions in Brain Magnetic Resonance Imaging. *2018 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2018*, 2019. doi:[10.1109/DICTA.2018.8615771](https://doi.org/10.1109/DICTA.2018.8615771).
- [72] Youngjin Yoo, Tom Brosch, Anthony Traboulsee, David K.B. Li, and Roger Tam. Deep learning of image features from unlabeled data for multiple sclerosis lesion segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8679:117–124, 2014. doi:[10.1007/978-3-319-10581-9_15](https://doi.org/10.1007/978-3-319-10581-9_15).

- [73] Nils Gessert, Julia Krüger, Roland Opfer, Ann Christin Ostwaldt, Praveena Manogaran, Hagen H. Kitzler, Sven Schippling, and Alexander Schlaefer. Multiple sclerosis lesion activity segmentation with attention-guided two-path CNNs. *Computerized Medical Imaging and Graphics*, 84(May):101772, 2020. URL: <https://doi.org/10.1016/j.compmedimag.2020.101772>, doi:10.1016/j.compmedimag.2020.101772.
- [74] Sergi Valverde, Mariano Cabezas, Eloy Roura, Sandra González-Villà, Deborah Pareto, Joan C. Vilanova, Lluís Ramió-Torrentà, Àlex Rovira, Arnau Oliver, and Xavier Lladó. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *NeuroImage*, 155(December 2016):159–168, 2017. URL: <http://dx.doi.org/10.1016/j.neuroimage.2017.04.034>, arXiv:1702.04869, doi:10.1016/j.neuroimage.2017.04.034.
- [75] Hrishikesh Deshpande, Pierre Maurel, and Christian Barillot. Classification of multiple sclerosis lesions using adaptive dictionary learning. *Computerized Medical Imaging and Graphics*, 46:2–10, 2015. URL: <http://dx.doi.org/10.1016/j.compmedimag.2015.05.003>, doi:10.1016/j.compmedimag.2015.05.003.
- [76] Snehashis Roy, Qing He, Elizabeth Sweeney, Aaron Carass, Daniel S. Reich, Jerry L. Prince, and Dzung L. Pham. Subject-Specific Sparse Dictionary Learning for Atlas-Based Brain MRI Segmentation. *IEEE Journal of Biomedical and Health Informatics*, 19(5):1598–1609, 2015. doi:10.1109/JBHI.2015.2439242.
- [77] Zahra Karimaghloo, Douglas L. Arnold, and Tal Arbel. Adaptive multi-level conditional random fields for detection and segmentation of small enhanced pathology in medical images. *Medical Image Analysis*, 27:17–30, 2016. URL: <http://dx.doi.org/10.1016/j.media.2015.06.004>, doi:10.1016/j.media.2015.06.004.
- [78] Colm Elliott, Douglas L. Arnold, D. Louis Collins, and Tal Arbel. Temporally consistent probabilistic detection of new multiple sclerosis lesions in brain MRI. *IEEE Transactions on Medical Imaging*, 32(8):1490–1503, 2013. doi:10.1109/TMI.2013.2258403.
- [79] Ozdemir Cetin, Volkan Seymen, and Unal Sakoglu. Multiple sclerosis lesion detection in multimodal MRI using simple clustering-based segmentation and classification. *Informatics in Medicine Unlocked*, 20:100409, 2020. URL: <https://doi.org/10.1016/j.imu.2020.100409>, doi:10.1016/j.imu.2020.100409.
- [80] Ali Alijamaat, Alireza NikravanShalmani, and Peyman Bayat. Multiple sclerosis identification in brain MRI images using wavelet convolutional neural networks. *International Journal of Imaging Systems and Technology*, pages 1–8, 2020. doi:10.1002/ima.22492.
- [81] Viktor Wottschel, Declan T. Chard, Christian Enzinger, Massimo Filippi, Jette L. Frederiksen, Claudio Gasperini, Antonio Giorgio, Maria A. Rocca, Alex Rovira, Nicola De Stefano, Mar Tintoré, Daniel C. Alexander, Frederik Barkhof, and Olga Ciccarelli. SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of multiple sclerosis. *NeuroImage: Clinical*, 24(September):102011, 2019. URL: <https://doi.org/10.1016/j.nicl.2019.102011>, doi:10.1016/j.nicl.2019.102011.
- [82] V. Wottschel, D. C. Alexander, P. P. Kwok, D. T. Chard, M. L. Stromillo, N. De Stefano, A. J. Thompson, D. H. Miller, and O. Ciccarelli. Predicting outcome in clinically isolated syndrome using machine learning. *NeuroImage: Clinical*, 7:281–287, 2015.

- URL: <http://dx.doi.org/10.1016/j.nicl.2014.11.021>, doi:10.1016/j.nicl.2014.11.021.
- [83] Haike Zhang, Esther Alberts, Viola Pongratz, Mark Mühlau, Claus Zimmer, Benedikt Wiestler, and Paul Eichinger. Predicting conversion from clinically isolated syndrome to multiple sclerosis—An imaging-based machine learning approach. *NeuroImage: Clinical*, 21(November 2018):101593, 2019. URL: <https://doi.org/10.1016/j.nicl.2018.11.003>, doi:10.1016/j.nicl.2018.11.003.
- [84] Mariana Zurita, Cristian Montalba, Tomás Labbé, Juan Pablo Cruz, Josué Dalboni da Rocha, Cristián Tejos, Ethel Ciampi, Claudia Cárcamo, Ranganatha Sitaram, and Sergio Uribe. Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion mri data. *NeuroImage: Clinical*, 20:724–730, 2018. URL: <https://doi.org/10.1016/j.nicl.2018.09.002>, doi:10.1016/j.nicl.2018.09.002.
- [85] Kerstin Bendfeldt, Stefan Klöppel, Thomas E. Nichols, Renata Smieskova, Pascal Kuster, Stefan Traud, Nicole Mueller-Lenke, Yvonne Naegelin, Ludwig Kappos, Ernst Wilhelm Radue, and Stefan J. Borgwardt. Multivariate pattern classification of gray matter pathology in multiple sclerosis. *NeuroImage*, 60:400–408, 2012. URL: <http://dx.doi.org/10.1016/j.neuroimage.2011.12.070>, doi:10.1016/j.neuroimage.2011.12.070.
- [86] Sarah C. Reitz, Stephanie Michelle Hof, Vinzenz Fleischer, Alla Brodski, Adriane Gröger, René Maxime Gracien, Amgad Droby, Helmuth Steinmetz, Ulf Ziemann, Frauke Zipp, Ralf Deichmann, and Johannes C. Klein. Multi-parametric quantitative MRI of normal appearing white matter in multiple sclerosis, and the effect of disease activity on T2. *Brain Imaging and Behavior*, 11(3):744–753, 2017. doi:10.1007/s11682-016-9550-5.
- [87] Heiko Neeb and Jochen Schenk. Multivariate prediction of multiple sclerosis using robust quantitative mr-based image metrics. *Zeitschrift für Medizinische Physik*, 29:262–271, 2019. URL: <https://doi.org/10.1016/j.zemedi.2018.10.004>, doi:10.1016/j.zemedi.2018.10.004.
- [88] Gabriel Mangeat, Russell Ouellette, Maxime Wabartha, Benjamin De Leener, Michael Plattén, Virginija Danylaitė Karrenbauer, Marcel Warntjes, Nikola Stikov, Caterina Mainero, Julien Cohen-Adad, and Tobias Granberg. Machine learning and multiparametric brain mri to differentiate hereditary diffuse leukodystrophy with spheroids from multiple sclerosis. *Journal of Neuroimaging*, 30:674–682, 2020. doi:10.1111/jon.12725.
- [89] Youngjin Yoo, Lisa Y.W. Tang, Tom Brosch, David K.B. Li, Shannon Kolind, Irene Vavasour, Alexander Rauscher, Alex L. MacKay, Anthony Traboulsee, and Roger C. Tam. Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls. *NeuroImage: Clinical*, 17(October 2017):169–178, 2018. URL: <https://doi.org/10.1016/j.nicl.2017.10.015>, doi:10.1016/j.nicl.2017.10.015.
- [90] Glen Pridham, Olayinka Oladosu, and Yunyan Zhang. Evaluation of discrete orthogonal versus polar Stockwell Transform for local multi-resolution texture analysis using brain MRI of multiple sclerosis patients. *Magnetic Resonance Imaging*, 72(July):150–158, 2020. URL: <https://doi.org/10.1016/j.mri.2020.07.007>, doi:10.1016/j.mri.2020.07.007.

- [91] Fabian Eitel, Emily Soehler, Judith Bellmann-Strobl, Alexander U. Brandt, Klemens Ruprecht, René M. Giess, Joseph Kuchling, Susanna Asseyer, Martin Weygandt, John Dylan Haynes, Michael Scheel, Friedemann Paul, and Kerstin Ritter. Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. *NeuroImage: Clinical*, 24(September):102003, 2019. URL: <https://doi.org/10.1016/j.nicl.2019.102003>, [arXiv:1904.08771](https://arxiv.org/abs/1904.08771), doi:10.1016/j.nicl.2019.102003.
- [92] Shui Hua Wang, Chaosheng Tang, Junding Sun, Jingyuan Yang, Chenxi Huang, Preetha Phillips, and Yu Dong Zhang. Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling. *Frontiers in Neuroscience*, 12(NOV):1–11, 2018. doi:10.3389/fnins.2018.00818.
- [93] Aldo Marzullo, Gabriel Kocivar, Claudio Stamile, Françoise Durand-Dubief, Giorgio Terracina, Francesco Calimeri, and Dominique Sappey-Marini. Classification of multiple sclerosis clinical profiles via graph convolutional neural networks. *Frontiers in Neuroscience*, 13:1–15, 2019. doi:10.3389/fnins.2019.00594.
- [94] 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 Granziere, Jonas Richiardi, and Tobias Kober. CVSnet: A machine learning approach for automated central vein sign assessment in multiple sclerosis. *NMR in Biomedicine*, 33(5):1–11, 2020. doi:10.1002/nbm.4283.
- [95] Annapareddy V.N. Reddy, Ch Phani Krishna, and Pradeep Kumar Mallick. An image classification framework exploring the capabilities of extreme learning machines and artificial bee colony. *Neural Computing and Applications*, 32(8):3079–3099, 2020. URL: <https://doi.org/10.1007/s00521-019-04385-5>, doi:10.1007/s00521-019-04385-5.
- [96] Yudong Zhang, Siyuan Lu, Xingxing Zhou, Ming Yang, Lenan Wu, Bin Liu, Preetha Phillips, and Shuihua Wang. Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: Decision tree, k -nearest neighbors, and support vector machine. *Simulation*, 92(9):861–871, 2016. doi:10.1177/0037549716666962.
- [97] Bernd Taschler, Tian Ge, Kerstin Bendfeldt, Nicole Müller-Lenke, Timothy D. Johnson, and Thomas E. Nichols. Spatial modeling of multiple sclerosis for disease subtype prediction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8674 LNCS(PART 2):797–804, 2014. doi:10.1007/978-3-319-10470-6_99.
- [98] Marco Pota, Massimo Esposito, Rosario Megna, Giuseppe De Pietro, Mario Quarantelli, Vincenzo Brescia Morra, and Bruno Alfano. Multivariate fuzzy analysis of brain tissue volumes and relaxation rates for supporting the diagnosis of relapsing-remitting multiple sclerosis. *Biomedical Signal Processing and Control*, 53:101591, 2019. URL: <https://doi.org/10.1016/j.bspc.2019.101591>, doi:10.1016/j.bspc.2019.101591.
- [99] Zezhong Ye, Ajit George, Anthony T. Wu, Xuan Niu, Joshua Lin, Gautam Adusumilli, Robert T. Naismith, Anne H. Cross, Peng Sun, and Sheng Kwei Song. Deep learning

- with diffusion basis spectrum imaging for classification of multiple sclerosis lesions. *Annals of Clinical and Translational Neurology*, 7(5):695–706, 2020. doi:[10.1002/acn3.51037](https://doi.org/10.1002/acn3.51037).
- [100] Adrian Ion-Mărgineanu, Gabriel Kocevar, Claudio Stamile, Diana M. Sima, Françoise Durand-Dubief, Sabine Van Huffel, and Dominique Sappey-Marinier. A comparison of machine learning approaches for classifying multiple sclerosis courses using MRSI and brain segmentations. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10614 LNCS:643–651, 2017. doi:[10.1007/978-3-319-68612-7_73](https://doi.org/10.1007/978-3-319-68612-7_73).
- [101] Yijun Zhao, Brian C. Healy, Dalia Rotstein, Charles R.G. Guttmann, Rohit Bakshi, Howard L. Weiner, Carla E. Brodley, and Tanuja Chitnis. Exploration of machine learning techniques in predicting multiple sclerosis disease course. *PLoS ONE*, 12(4):1–13, 2017. doi:[10.1371/journal.pone.0174866](https://doi.org/10.1371/journal.pone.0174866).
- [102] Gabriel Kocevar, Claudio Stamile, Salem Hannoun, François Cotton, Sandra Vukusic, Françoise Durand-Dubief, and Dominique Sappey-Marinier. Graph theory-based brain connectivity for automatic classification of multiple sclerosis clinical courses. *Frontiers in Neuroscience*, 10(OCT):1–11, 2016. doi:[10.3389/fnins.2016.00478](https://doi.org/10.3389/fnins.2016.00478).
- [103] Adrian Ion-Mărgineanu, Gabriel Kocevar, Claudio Stamile, Diana M. Sima, Françoise Durand-Dubief, Sabine Van Huffel, and Dominique Sappey-Marinier. Machine learning approach for classifying multiple sclerosis courses by combining clinical data with lesion loads and magnetic resonance metabolic features. *Frontiers in Neuroscience*, 11:1–11, 2017. doi:[10.3389/fnins.2017.00398](https://doi.org/10.3389/fnins.2017.00398).
- [104] Ben Bellows, Aman Bhandari, Mahad Ibrahim, and Jaspal S. Sandhu. Peering into the Black Box. *Information Communication Technologies*, (January):38–43, 2011. doi:[10.4018/9781599049496.ch067](https://doi.org/10.4018/9781599049496.ch067).
- [105] Martin Styner, Joohwi Lee, Brian Chin, Matthew Chin, Olivier Commowick, Hoai-Huong Tran, Valerie Jewells, and Simon Warfield. 3d segmentation in the clinic: A grand challenge ii: Ms lesion segmentation. *MIDAS Journal*, 11 2007.
- [106] J. Tohka. *Rigid-Body Registration*, volume 1. Elsevier Inc., 2015. URL: <http://dx.doi.org/10.1016/B978-0-12-397025-1.00299-2>, doi:[10.1016/B978-0-12-397025-1.00299-2](https://doi.org/10.1016/B978-0-12-397025-1.00299-2).
- [107] Stephen Smith. Fast Robust Automated Brain Extraction. *Human Brain Mapping*, 17(3):143–155, 2002. doi:[10.1002/hbm.10062](https://doi.org/10.1002/hbm.10062).
- [108] Ivan Itzcovich. Deepbrain | pypi. <https://pypi.org/project/deepbrain/>. (Accessed on 05/19/2021).
- [109] Bet/userguide - fslwiki. <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET/UserGuide#bet>, January 2018. (Accessed on 06/01/2021).
- [110] Nicholas J. Tustison, Brian B. Avants, Philip A. Cook, Yuanjie Zheng, Alexander Egan, Paul A. Yushkevich, and James C. Gee. N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, 2010. doi:[10.1109/TMI.2010.2046908](https://doi.org/10.1109/TMI.2010.2046908).

- [111] Tensorflow. <https://www.tensorflow.org/>. (Accessed on 06/01/2021).
- [112] Keras: the python deep learning api. <https://keras.io/>. (Accessed on 06/01/2021).
- [113] Lst – lesion segmentation for spm | paul schmidt – freelance statistician. <https://www.applied-statistics.de/lst.html>. (Accessed on 06/08/2021).
- [114] Paul Schmidt, Christian Gaser, Milan Arsic, Dorothea Buck, Annette Förchler, Achim Berthele, Muna Hoshi, Rüdiger Ilg, Volker J. Schmid, Claus Zimmer, Bernhard Hemmer, and Mark Mühlau. An automated tool for detection of flair-hyperintense white-matter lesions in multiple sclerosis. *NeuroImage*, 59(4):3774–3783, 2012. URL: <https://www.sciencedirect.com/science/article/pii/S1053811911013139>, doi:<https://doi.org/10.1016/j.neuroimage.2011.11.032>.
- [115] Paul Schmidt. Bayesian inference for structured additive regression models for large-scale problems with applications to medical imaging, January 2017. URL: <http://nbn-resolving.de/urn:nbn:de:bvb:19-203731>.
- [116] Paul A. Yushkevich, Artem Pashchinskiy, Ipek Oguz, Suyash Mohan, J. Eric Schmitt, Joel M. Stein, Dženan Zukić, Jared Vicory, Matthew McCormick, Natalie Yushkevich, Nadav Schwartz, Yang Gao, and Guido Gerig. User-guided segmentation of multi-modality medical imaging datasets with itk-snap. *Neuroinformatics*, 17(1):83–102, Jan 2019. 29946897[pmid]. URL: <https://pubmed.ncbi.nlm.nih.gov/29946897>, doi:[10.1007/s12021-018-9385-x](https://doi.org/10.1007/s12021-018-9385-x).
- [117] Anaconda | the world’s most popular data science platform. <https://www.anaconda.com/>, 2021. (Accessed on 05/23/2021).
- [118] Anaconda. Conda — conda documentation. <https://docs.conda.io/en/latest/>, 2017. (Accessed on 05/23/2021).
- [119] Project jupyter | home. <https://jupyter.org/>, May 2021. (Accessed on 05/23/2021).
- [120] The jupyter notebook — jupyter notebook 6.4.0 documentation. <https://jupyter-notebook.readthedocs.io/en/stable/notebook.html>, 2015. (Accessed on 05/23/2021).
- [121] Mathworks - founders - matlab & simulink. <https://www.mathworks.com/company/aboutus/founders.html>, 2021. (Accessed on 06/02/2021).
- [122] The Wellcome Centre for Human Neuroimaging. Spm12 software - statistical parametric mapping. <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>. (Accessed on 06/07/2021).
- [123] Michele Larobina and Loredana Murino. Medical image file formats. *Journal of Digital Imaging*, 27(2):200–206, Apr 2014. URL: <https://doi.org/10.1007/s10278-013-9657-9>, doi:[10.1007/s10278-013-9657-9](https://doi.org/10.1007/s10278-013-9657-9).
- [124] pynrrd · pypi. <https://pypi.org/project/pynrrd/>. (Accessed on 06/02/2021).
- [125] Neuroimaging in python — nibabel 3.2.1+100.g9fdf5e3e documentation. <https://nipy.org/nibabel/>. (Accessed on 06/02/2021).