UNIVERSIDADE D
COIMBRA

Júlia Freire Soares

FROM NEUROSCIENCE TO THE IDENTIFICATION OF STRUCTURAL MEDICAL IMAGING MARKERS IN MULTIPLE SCLEROSIS

VOLUME 1

Dissertation to obtain a Master Degree in Biomedical Engineering supervised by João Valente Duarte (PhD) and Teresa Sousa (PhD) and presented to the Physics Department of Faculty of Sciences and Technologies of the University of Coimbra

July, 2019

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D COIMBRA

From Neuroscience to the Identification of Structural Medical Imaging Markers in Multiple Sclerosis

Dissertation presented to the Faculty of Sciences and Technology of the University of Coimbra to obtain a Master Degree in Biomedical Engineering

Júlia Freire Soares

Supervisors:
João Valente Duarte (PhD)
Teresa Sousa (PhD)

Coimbra, 2019

This work was developed in collaboration with:

Institute of Nuclear Sciences Applied to Health

Coimbra Institute for Biomedical Imaging and Translational Research

Serviço de Neurologia dos Hospitais da Universidade de Coimbra

Agradecimentos

"It never hurts to keep looking for sunshine"
Alan Alexander Milne

Exprimo a minha gratidão ao Doutor João Duarte, pelo voto de confiança, pelo incentivo constante e pelas oportunidades que me proporcionou. A sua ajuda e dedicação foram cruciais para a concretização deste trabalho. Graças à partilha dos seus vastos conhecimentos, que contribuíram tanto para o meu enriquecimento académico como pessoal, termino esta etapa com a sensação de dever cumprido e cada vez mais fascinada pelas temáticas abordadas.

Agradeço à Doutora Teresa Sousa, pelo apoio inestimável, pela disponibilidade sempre manifestada e por todo o aconselhamento essencial para a construção e amadurecimento deste projeto.

Aos meus amigos do coração, um enorme obrigada por terem feito Coimbra uma segunda casa. Por todos os momentos jamais esquecidos e por terem contribuído cada um com um pouco de si para o desenvolvimento da minha pessoa ao longo desta jornada.

Agradeço ao meu João Pedro, pelo companheirismo e pela amizade ao longo destes 5 anos. Pela partilha de sonhos e por sempre acreditar desde o primeiro dia que eu seria capaz.

À minha família deixo o maior agradecimento. Obrigada mãe, pai, Filipe e Lucas por terem tomado este meu sonho como vosso e por celebrarem sempre as minhas conquistas. Obrigada pelo apoio desmedido e acima de tudo pelos valores que me transmitiram e que fazem de mim o que sou. Obrigada mãe e pai por me ensinarem

desde muito cedo o valor do trabalho e da dedicação e que não importa de onde vimos, mas sim onde chegamos.

Abstract

Multiple Sclerosis (MS) is a neurological disease of the central nervous system, thought to predominantly affect white matter (WM). As such, the detection of T2-hyperintense WM lesions on magnetic resonance imaging (MRI) scans of the brain has become a crucial criterion for diagnosis and predicting prognosis in early disease. Despite extensive criteria to perform a MS diagnosis and the available treatments to manage the disease, MS remains a complex disease with several questions to be solved. The cause of the disease depends on many factors and is hardly explained, and MS misdiagnosis is still a concern. These questions have motivated the study of imaging biomarkers.

Beyond identification of WM lesions, there are other structural changes in the brain that can characterize MS, such as grey matter (GM) lesions and regional WM and GM atrophy, which have been associated with cognitive deficit and disease severity. Understanding their cause, how early they appear, and which regions are primarily affected can provide important clues about the pathophysiology of the disease and lead to a more accurate diagnosis. The aim of this work is to identify structural neuroimaging biomarkers of interest for the discrimination of MS patients and different subtypes, and also for investigation of MS pathophysiology.

MRI measures of voxel-wise GM and WM volume, cortical thickness and gyrification index were used to investigate patterns of structural deficits in MS. These measures were extracted from T1-MPRAGE and T2-FLAIR images from 64 healthy controls and 59 MS patients. The images underwent univariate statistical analyses, such as voxel-based morphometry (VBM) and surface-based morphometry (SBM), to investigate regional morphometric differences. Multivariate pattern analysis (MVPA) using a Support Vector Machine (SVM) classifier was also applied to the same dataset, to obtain an automatic classification of MS patients and MS subtypes and to explore the potential of pattern recognition techniques in the analysis of neuroimaging biomarkers of disease.

Either VBM analysis, testing regional GM and WM volume differences, and SBM analysis, testing surface differences in cortical thickness and gyrification index, revealed several morphometric differences between patients and controls. Such

results demonstrated the possibility to track morphological MS-related changes with these measures. SVM classification based on GM and WM volumes yielded quite accurate discrimination of MS patients. The classification of MS subtypes, on the other hand, was much lower than the discrimination of MS patients from healthy controls. Nonetheless, the significant results obtained suggest that including a higher number of participants is expected to allow an accurate discrimination of MS subtypes.

In conclusion, univariate analysis of different morphometric features parameters can be used to highlight disease effects in MS. Multivariate pattern analysis can be used to distinguish between individuals diagnosed with MS and control participants on the basis of neuroanatomical differences. Furthermore, this data-driven analysis can reveal subtle distributed networks of abnormal structural patterns in MS and might be explored further to derive useful predictive tools with potential in assisting diagnosis and accurate discrimination of MS subtypes.

Keywords: Multiple Sclerosis, brain structure, univariate analysis, multivariate pattern classification

Resumo

A Esclerose Múltipla (EM) é uma doença neurológica do sistema nervoso central, que afeta predominantemente a substância branca. Como tal, a deteção de lesões na substância branca que aparecem hiperintensas no sinal T2 de ressonância magnética (RM) do cérebro tornou-se um critério crucial para o diagnóstico e previsão do prognóstico no início da doença. Apesar dos extensos critérios necessários para realizar um diagnóstico de EM e os tratamentos disponíveis no decorrer da doença, a EM continua a ser uma doença complexa, com várias questões a serem resolvidas. A causa da doença depende de muitos fatores sendo dificilmente explicada, e o diagnóstico incorreto da EM é também uma preocupação. O estudo de biomarcadores de imagem é motivado por estas questões.

Além da identificação de lesões na substância branca, existem outras alterações estruturais no cérebro que podem caracterizar a EM. Lesões na substância cinzenta e atrofia regional na substância branca e na substância cinzenta têm sido associadas ao défice cognitivo e ao agravamento da doença. Perceber as sua causas, quão cedo estas surgem e quais as regiões principalmente afetadas, pode fornecer pistas importantes sobre a fisiopatologia da doença conduzindo a um diagnóstico mais preciso. O objetivo deste trabalho é identificar biomarcadores estruturais de neuroimagem de interesse para a discriminação de doentes com EM e de diferentes subtipos e também para investigar acerca da fisiopatologia da EM

Medidas obtidas com as imagens de ressonância magnética como volume de substância cinzenta e substância branca, espessura cortical e índice de girificação foram usadas para investigar padrões que revelam alterações estruturais na EM. Estas medidas foram extraídas de imagens T1-MPRAGE e T2-FLAIR de 64 controlos saudáveis 59 doentes com EM. As imagens foram submetidas a análises estatísticas univariadas, como *voxel-based morphometry* (VBM) e *surface-based morphometry* (SBM), para investigar diferenças morfológicas regionais. A análise multivariada de padrões (MVPA), usando um classificador *Support Vector Machine* (SVM) também foi aplicada no mesmo conjunto de dados, para obter uma classificação automática de doentes com EM e subtipos de EM e explorar o potencial das técnicas de

reconhecimento de padrões na análise de biomarcadores de neuroimagem da doença.

Tanto a análise VBM, testando diferenças regionais de volume de matéria cinzenta e matéria branca, quanto a análise SBM, testando as diferenças de superfície na espessura cortical e índice de girificação, revelaram várias diferenças estruturais entre doentes e controlos. Tais resultados demonstraram a possibilidade de detetar alterações morfológicas na EM com estas medidas. A classificação SVM baseada no volume de substância cinzenta e substância branca resultou numa discriminação bastante precisa dos doentes com EM. A classificação dos subtipos de EM, por outro lado, foi muito menos precisa do que a discriminação entre doentes com EM e controlos saudáveis. No entanto, os resultados significativos obtidos sugerem que se espera que um maior número de participantes permita uma discriminação precisa dos subtipos de EM.

Concluindo, análises univariadas em diferentes parâmetros estruturais pode ser usada para destacar os efeitos da doença na EM. A análise multivariada pode ser usada para distinguir indivíduos diagnosticados com EM e participantes saudáveis com base em diferenças neuroanatómicas. Além disso, esta análise pode revelar padrões estruturais subtis anormais na EM, e pode ser mais explorada para investigar biomarcadores preditivos úteis com potencial para auxiliar o diagnóstico e para discriminar com precisão subtipos de EM.

Palavras-chave: Esclerose Múltipla, estrutura do cérebro, análises univariadas, classificação multivariada

List of Figures

List of Tables

List of Equations

Abbreviations and Symbols

Abbreviations

2D Two Dimensional

3D Three Dimensional

BBSI Brain Boundary Shift Integral

BET Brain Extraction Tool

BICCR Brain to Intracranial Capacity Ratio

BPF Brain Parenchymal Fraction

CAT12 Computational Anatomy Toolbox

CIS Clinically Isolated Syndrome

CNS Central Nervous System

CSF Cerebro-Spinal Fluid

CV Cross Validation

DARTEL Diffeomorphic Anatomic Registration Through Exponentiated Lie

DICOM Digital Imaging and Communications in Medicine

DTI Diffusion Tensor Imaging

FDR False Discovery Rate

FWE Family Wise Correction

FLAIR Fluid Attenuated Inversion Recovery

FN False Negative

FP True Positive

GBD Global Burden of Diseases

GLM General Linear Model

GM Grey Matter

GWAS Genome Wide Association Studies

IBA Index of Brain Atrophy

LGA Lesion Growth Algorithm

LOGO-CV Leave One Subject per Group Out- Cross Validation

LOO-CV Leave One Subject Out – Cross Validation

LPA Lesion Prediction Algorithm

LST Lesion Segmentation Toolbox

MAGNIMS Magnetic Ressonance Imaging in Multiple Sclerosis

MNI Montreal Neurological Institute

MOG Myelin Oligodendrocyte Glycoproteins

MPRAGE Magnetization-Prepared Rapid Gradient Echo

MRI Magnetic Ressonance Imaging

MS Multiple Sclerosis

MVPA Multivariate Pattern Analysis

NAWM Normal Appearing White Matter

NIfTI Neuroimaging Informatics Technology Initiative

PBT Projection-Based Thickness

PD Proton Density

PRoNTo Pattern Recognition for Neuroimaging Toolbox

RRMS Relapsing-Remitting Multiple Sclerosis

RTF Random Theory Field

SBM Surface-Based Morphometry

SD Standard Deviation

SIENA Structural Image Evaluation using Normalization of Atrophy

SNR Signal Noise Ratio

SPMS Secondary Progressive Multiple Sclerosis

SVM Support Vector Machine

TAL Tailarach Space

TDS Template-Driven Segmentation

TIV Total Intracranial Volume

TLV Total Lesion Volume

TN True Negative

TP True Positive

VBM Voxel-Based Morphometry

WBR Whole-Brain Ratio

WM White Matter

Symbols

β Parameters estimates matrix/vector

$\boldsymbol{X}$ Design Matrix

$\boldsymbol{Y}$ Observation matrix/vector

E Error

$\boldsymbol{c}$ Contrast vector

$\widehat{\boldsymbol{\beta}}$ Distributed parameters

$\boldsymbol{var}$ Variance

$\boldsymbol{p}$ *p*-value

$\boldsymbol{D}$ Dataset

$\boldsymbol{x_i}$ features

$\boldsymbol{y_i}$ labels

w classification weights

$\boldsymbol{b_0}$ bias term

k number of permutations

Contents

1 Introduction

1.1 Background and Motivation

1.1.1 Multiple Sclerosis

Multiple Sclerosis (MS) is a chronic, autoimmune, progressive, inflammatory neurological disease with no known cure that affects the central nervous system (CNS) [1]. The term *Multiple* refers to multiple areas of the brain and spinal cord that are affected, whereas the term *Sclerosis* is due to the presence of multifocal demyelinated regions of the CNS, essentially known as white matter (WM) lesions. *Sclerosis* is a Latin word which stands for scars and MS literally means *many scars*. WM lesions, the macro structural manifestation of demyelination, are characterized by myelin oligodendrocyte glycoproteins (MOG) loss and infiltrated inflammatory cells, namely lymphocytes and macrophages [1], [2]. The immune system recognizes MOG proteins as being unknown and harmful and the subsequent inflammatory response leads to myelin sheath degradation [2], [3]. Myelin sheaths around the axons have the function of protecting axons and conducting nerve impulses [2]–[4]. Demyelination is characterized by the degradation of myelin sheaths of the axons which leads to impairment or interruption of transmission of neuronal signals, and ultimately death of nerve cells.

MS is a progressive disease, its course is quite varied and unpredictable, and is characterized by motor, sensory and cognitive deterioration in 65 % of the patients [5], [6]. The presence of relatively preserved axons and neurons within the WM lesions in the early stages of the disease distinguishes MS from other destructive pathophysiological conditions that are followed by focal inflammation [1]. However, as disease progresses axonal damage becomes irreversible [3].

In 2016, the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) estimated 2221188 prevalent cases of Multiple Sclerosis worldwide, and 8367 in Portugal [7]. Women are twice more affected by MS than men and MS is mainly incident in Caucasian subjects (more specifically descendants from northern

Europe). The disease usually manifests in adults between the ages of 20-50 years old. Occasionally it is present during childhood or late middle age [1]–[3].

The cause of MS is not fully understood. However, it is believed that the emergence of the disease results from a genetic predisposition combined with non-genetic factors (*triggers*). These triggers can be a virus, such as Epstein–Barr virus, or metabolic and environmental factors, such as smoking, child obesity and low levels of Vitamin D [1]–[3]. Genome Wide Association Studies (GWAS) identified more than 100 variants associated with MS, conferring a minimal percentage to the risk of the disease [1]–[3].

The onset of the disease is marked by a first single attack (isolated in time) known by "clinically isolated syndrome" (CIS). Subjects who have experienced CIS manifest symptoms that are suggestive of MS. Clinical symptoms of MS vary depending on the location and degree of the lesions. Symptoms include muscle weakness, loss of vision, dizziness, fatigue, balance problems, sensory disturbances, bladder dysfunction, constipation, sexual problems and depression. In 85% of young adults, these may be a consequence of monofocal (isolated in space) or multifocal (dissemination in space) damages in CNS. These damages involve the optic nerve, brainstem, cerebellum, spinal cord, or cerebral hemispheres. To be termed as CIS, the episode should last for at least 24h and should occur in the absence of infections, inflammatory disorders, genetic diseases, neoplasms, vascular disease, or other autoimmune diseases [8].

Sensory disturbances constitute the initial clinical findings in MS patients. The most common are paraesthesia (tingling or "pins and needles"), dysesthesias (unpleasant or painful sense of touch, burning sensation), diplopia, ataxia, vertigo, trigeminal neuralgia, unilateral numbness in one leg that spreads to other parts of the body and bladder (urinary sphincter) disturbances. These disturbances are typically solved but sometimes they can evolve into neuropathic pain. Regarding muscle weakness, over 30 % of MS patients exhibit spasticity generally in the legs. Due to spasticity, more than 50 % of MS patients find problems in walking and need help to do common tasks. Bladder dysfunction is present in 90 % of the patients and is characterized by weekly or more frequent incontinence episodes. Optic neuritis is also common in MS and is characterized by complete or partial vision loss. Notably,

the principal cause of work-related disability is fatigue, which affects 90 % of the patients [1]–[3].

During life, MS patients have relapses, or flare-ups, where new symptoms occur, or old symptoms become worse. The symptoms can disappear completely after the relapse. However, the neurological damage is permanent.

Typically, 10 to 20 years after MS diagnosis, approximately 15 % of patients evolve to a progressive clinical course from onset, which may lead to impaired mobility and cognition [2]. Patients can be grouped into four major MS subtypes based on the duration, severity of the symptoms, and the number of relapses that occur on a time frame of a subtle progression [1], [3]:

- **Relapsing remitting MS (RMMS)**

  This is the most common MS form affecting about 85 % of MS patients. It is noticeable by acute attacks (relapses), where new symptoms occur or there is exacerbation of old MS symptoms due to higher inflammatory response. The relapses are followed by periods of remission, when symptoms improve or disappear [1], [3], [9]. In RRMS, the disease does not worsen between relapses.

- **Secondary progressive MS (SPMS)**

  Patients typically (up to 50 %) evolve to SPMS about 15 years after RRMS onset. SPMS is characterised by a gradual increase in disability level caused by ongoing relapses and remissions. Despite a few periods of remission or improvement of symptoms, the patients' condition gets continuously worse [1]–[3], [9].

- **Primary progressive MS (PPMS)**

  Approximately 5 % to 15 % of MS patients develop primary progressive MS. There is a regular patients' condition deterioration, reflected on a progressive incapacity such as spastic paraparesis, sensory ataxia, cerebellar ataxia and cognitive and visual failure. This form of MS is resistant to pharmacological treatments [1], [3], [9].

- **Relapsing-Progressive MS (RPMS)**

This subtype is a rare form of MS, affecting less than 5 % of MS patients. RPMS continuously progresses, with sporadic outbreaks of symptoms worsening. There are no periods of remission [1]–[3], [9].

1.1.2 MS Diagnosis

Magnetic resonance imaging (MRI) is the cornerstone technique to identify biomarkers of MS and to diagnose MS. The term "biomarker" comes from the combination of the terms "biological" and "marker". There are several definitions for the term biomarker but all of them agree on the same statement. Biomarkers are objective indications of the biological processes observed from outside the patient, which can be accurately and reproducibly measured. These can indicate normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention. Biomarkers can be seen as a subcategory of medical signs in which a medical state is designated [10].

MS diagnosis is based on the McDonald's criteria and on MAGNIMS guidelines The McDonald's criteria consist of a diagnostic scheme based on "rules" to support neurologists for a better and more reliable diagnosis of MS. The European collaborative research network for the study of MS with MRI (Magnetic Resonance Imaging in Multiple Sclerosis - MAGNIMS) reviews the findings of studies about MRI criteria for MS diagnosis and proposes new ones. According to the revised McDonald's criteria and MAGNIMS guidelines, propagation in time (temporal progression of the disease) and space (spatial patterns of disease-related changes in multiple regions of the CNS) are the main conditions to determine MS diagnosis [11], [12]. Demonstration of cerebrospinal fluid (CSF) oligoclonal bands in the absence of atypical CSF findings also allows a diagnosis of Multiple Sclerosis to be made. Dissemination in space can be demonstrated by one or more T2-hyperintense lesions in two or more of four areas of the CNS: periventricular, cortical or juxtacortical, and infratentorial brain regions, and the spinal cord. Dissemination in time can be demonstrated by the simultaneous presence of gadolinium-enhancing and non-enhancing lesions. For example, by a new T2-hyperintense or gadolinium-enhancing lesion on a follow-up MRI, or by two clinical attacks separated by at least three months [11]. These criteria are summarized in Table 1.1.

Table 1.1: The 2017 McDonald criteria for diagnosis of Multiple Sclerosis [11].

Number of clinical attacks	Number of lesions with clinical evidence	Additional data needed
≥ 2 clinical attacks	≥ 2	None
≥ 2 clinical attacks	1 (since there is evidence of a lesion in a different anatomical location from the previous attack)	None
≥ 2 clinical attacks	1	Dissemination in space demonstrated by an additional clinical attack involving a different CNS site or by MRI
1 clinical attack	≥ 2	Dissemination in time demonstrated by an additional clinical attack or by MRI OR demonstration of CSF-specific oligoclonal bands
1 clinical attack	1	Dissemination in space demonstrated by an additional clinical attack involving a different CNS site or by MRI AND Dissemination in time demonstrated by an additional clinical attack or by MRI OR demonstration of CSF-specific oligoclonal bands

1.1.3 Importance of supporting tools for MS diagnosis and treatment

MS is a complex disease with several remaining questions to be solved. The cause of the disease depends on many factors and is hardly explained. Treatments available to manage the disease course (including shortening the duration of acute exacerbations, decreasing their frequency, and providing symptomatic relief) are only partially effective and without curative purposes [1]. Moreover, there are several downsides related to the applied therapies: high costs, inconvenient administration (parenteral administration), frequency of adverse effects (especially 'flu-like' symptoms for several hours, diarrhoea, nausea, abdominal pain) and a relatively modest overall impact on disease course (for example, reductions in relapse rate of less than 35 %)[13], [14] .

MS misdiagnosis is also still a concern. There are several reports of this situation. Subjects diagnosed with MS, which did not fulfil the McDonald Criteria are likely candidates to an alternate diagnosis. MS misdiagnosis included small vessel cerebrovascular disease, migraine, fibromyalgia, and functional neurological disorders [15], [16] [17] . The underlying pathogenic heterogeneity of MS disease and the impracticality of directly sampling CNS tissue (as opposed to blood or CSF) have limited the investigation of biomarkers. Such non-invasive tools would not only help to perform a proper MS diagnosis, but also add to the understanding of its cause and progress [2]. In this context, imaging appears has a strong candidate to allow the identification of clinically and scientifically useful biomarkers of MS. In fact, beyond identification of WM lesions, there are other structural changes which can characterize MS, e.g. grey matter lesions (GM) and regional WM and GM atrophy are associated with cognitive deficit and disease severity in MS [9], [18].
The attempt to find a solution to the described current problems has motivated the study of these imaging biomarkers and the search for new ones. Understanding their cause, how early they appear, and which regions are primarily affected can provide cues about the pathophysiology of the disease and lead to a more accurate diagnosis [9], [18], [19].

This work aims to compare different imaging-derived biomarkers of MS via standard univariate and pattern recognition analysis of MRI data. The dataset used belongs to Doctor Sónia Batista, neurologist at Centro Hospitalar e Universitário de Coimbra (CHUC) and was previously described and analysed in other research studies [20]–[22].

1.2 Goals

The main goal of this work is to assess and compare different imaging biomarkers of MS based on T1-MPRAGE and T2-FLAIR MR images. Such MS biomarkers consist in structural changes, which are here studied in a dataset of MS patients and healthy controls by applying univariate and multivariate statistical analyses of MR-derived brain morphometric measures. It is hypothesized that MS structural signatures might help discriminate MS patients from healthy controls, and furthermore help

solving the challenging differentiation of MS subtypes. Notably, the analysis of structural biomarkers also has the advantage of guiding scientific research about the pathophysiology of MS and its cause. Thus, these tools might in the future facilitate the evaluation of predisposition to the disease and its early diagnosis, with potential benefits in the prevention and adequacy of clinical interventions.

The univariate statistical analyses to be applied are Voxel Based Morphometry (VBM) and Surface Based Morphometry (SBM). It is aimed to quantify structural changes such as volume changes (VBM) and surface changes (SBM). VBM will be applied to investigate changes in GM and WM volumes, and SBM to investigate changes in surface measures of cortical thickness and gyrification index.

The multivariate statistical analyses to be applied are based on pattern recognition algorithms. Multivariate pattern analyses (MVPA) are implemented by mean of classification models using a Support Vector Machine (SVM). These classification models aim to differentiate MS patients from controls and to discriminate MS subtypes based on GM and WM volume features.

The novelty of this work is the study of less common features like the gyrification index. Another innovative exploration is the differentiation of MS subtypes based on MRI brain scans, which was not investigated before, as to our knowledge. Moreover, an important objective is to extract different morphometric measures and analysis approaches (VBM, SBM and MVPA) in the same dataset. This allows for comparisons between the different MS biomarkers studied with both univariate and multivariate methods and to discuss about the most discriminative ones.

1.3 Research Contribution

The work developed in this dissertation resulted in the submission of the paper "*Investigating whole-brain MRI markers in Multiple Sclerosis – emerging dimensions in morphometric space*" to the *15th Mediterranean Conference on Medical and Biological Engineering and Computing* (MEDICON 2019 - www.medicon2019.org/) to be held in Coimbra from 26 to 28 of September 2019. The work which has been reviewed by peers and accepted as full paper, will be

presented and discussed in the special session *INT4DAT – Intelligent Systems and Technologies for Diagnostic, Assistance, and Therapeutics*.

1.4 Dissertation Outline

The upcoming contents addressed in this dissertation are organized into four major chapters. In Chapter 2, a literature review about the primary MS biomarkers currently in use is presented. Chapter 3 includes a detailed description of the applied methods during the course of this work. In Chapter 4, the results are presented and discussed taking into account the research hypotheses of this project. Finally, in Chapter 5, conclusions about the work developed in this dissertation and future work approaches are presented.

2 Literature Review

2.1 MRI in the detection of MS

The structure of the brain can be accessed in a non-invasive way and with no biological damage through MRI. Anatomical scans with strong imaging contrast between different soft tissues and with high resolution are granted, and allow identifying biomarkers that other imaging modalities find difficult to achieve [23]. MRI is the main tool to produce early and reliable diagnosis of MS and to monitor disease progression. Through MRI it is possible to track illness progression including differentiation of the types of MS, or to evaluate the effectiveness of the treatment in order to approve therapies or develop medication that may lead to any improvements [11], [24]. MRI is thus the ideal candidate to serve the investigation of imaging biomarkers in MS (for more details see methods section).

Different types of MRI images can be generated, depending on the acquisition parameters. These can be proton density (PD), two-dimensional (2D) or three dimensional (3D) T2-weighted/T2-fluid attenuated inversion recovery (FLAIR) spin echo or turbo spin echo, and 2D or 3D T1-weighted spin echo or turbo spin echo [12], [24] . According to the MAGNIMS guidelines, there are specific MRI sequences for the appropriate MS diagnosis and to assess morphological changes such as described below [12].

2.2 Brain Structural Changes in MS

2.2.1 Volume Changes

There is a lot of research dedicated to the study of the WM lesions in MS. These lesions are considered the hallmark of MS because it was believed this disease was defined pathologically as an inflammatory process confined to the WM. Nowadays, GM lesions and brain atrophy are also considered essential volumetric brain

biomarkers because they are associated with cognitive deficit and correlated with the severity of the disease [9], [18], [24], [25].

WM and GM Lesions

The main disease-related changes in the brain are focal WM lesions, which constitute the MRI criteria for MS diagnosis. The distribution, morphology, evolution, and signal abnormalities of these lesions are usually studied with conventional MRI sequences [11], [12], [24]. WM lesions are best visualized with T2 and T2-FLAIR sequences. However, T1-weighted imaging after contrast agent application is preferred for initial evaluation of MS [24]. In T2 and PD images, WM lesions are observable as hyperintense spots, looking brighter compared to the surrounding tissue [9], [18], [24], [26]. In T1-weighted MR images, lesions are hypointense, appearing as "black holes" compared to the surrounding WM tissue intensities.

Up to 40 % of the WM lesions might persist for around 6 months and are characterized by significant demyelination, axonal destruction and irreversible damage. Previous studies have demonstrated that the number, volume and localization of WM lesions, in both T2 and T1 images, are associated with neurodegeneration and increased MS patients disability [11], [18], [24], [26] [27], [28].

In the last years, a large effort has been devoted to the study of the GM lesions in order to improve the understanding of its clinical relevance [25], [29], [30]. Double inversion recovery (DIR) MR sequences are used to detect and characterize these lesions.

These sequences have allowed to identify four types of cortical lesions (CL) based on their location [18], [30]:

- **Type I**

  WM and GM lesion are combined (known by cortical/juxtacortical lesions).

- Type II

Lesions of type II are entirely located within the cerebral cortex, so they are not in direct contact with subcortical WM or pia mater. These are in general small and are identified as perivascular lesions.

- Type III

Lesions of type III are present in subpial areas, usually confined within the layers 3 and 4 of the cortex. These are called band-like subpial lesions.

- Type IV

Lesions of type IV are the intracortical lesions because they comprise the entire cortex's width and do not extend to the subcortical WM. These can spread to large areas of the gyri or entire lobes.

MRI has also demonstrated that CL are a frequent finding at the earliest clinical MS stages, being present in about 35–40 % of patients with clinically isolated syndrome [25], [29], [30]. Additionally, the number and volume of CL were found to correlate with cognitive impairment (deficits in attention, concentration, speed of information processing, and memory) and physical disability. The existence of CL has been hypostasized as one of the pathological factors leading to cortical atrophy of MS patients [25], [29], [30].

Whole Brain Atrophy

Whole brain atrophy provides an estimation of the global brain tissue loss, which is present in all stages of the MS disease. As such, its evaluation has been suggested as an important parameter in clinical studies. It is strongly correlated with clinical disease progression. Increased atrophy over time leads to deterioration of motor, sensory and cognitive functions [9], [18], [24], [31]. Importantly, the evaluation of whole brain atrophy also informs about the diffuse or aggregate nature of the tissue injury [18]. Considering the whole brain to evaluate cortical atrophy provides great precision and statistical power to the analysis, due to the huge quantity of voxels considered.

When studying brain atrophy two issues must be considered. Normal aging causes brain atrophy, with an annual loss of brain volume of approximately 0.2–0.5 %, whereas in MS the atrophy rate is estimated to be 0.5–1.3 % per year [24], [31]. Thus, it is important to account for age-related effects in the investigation of brain atrophy. In MS, it is also important to have in consideration the "pseudoatrophy" phenomenon. Initial treatments reduce CNS inflammation, which translates in a decrease of brain volume and regular stabilization. However, brain tissue loss may have not occurred [18], [24], [31].

Regional Atrophy

Besides the loss of whole-brain volume, regional atrophy (with involvement of WM and GM separately) has been deeply studied and constitutes one of the main advances in atrophy quantification. As an alternative to whole-brain examination of atrophy, segmentation of WM and GM, and quantification of their atrophy independently, has shown progresses in the prediction of disability over time [18], [24], [31]. When compared to whole brain measures, regional measures of atrophy are more attentive to detail. Regional measures are not diluted by areas where there is little or no change, therefore anatomical changes are detected with more precision [31].

Although atrophy evaluation involves both GM and WM, the study of GM atrophy has revealed clinical information more relevant. As GM atrophy occurs early in MS, it has a higher impact to the annual loss of brain volume and shows a better correlation with lesion load than WM atrophy [18], [24], [31]. There is evidence of multiple regions marked by GM loss that are associated to disability especially in verbal memory and cognitive impairment [18], [24], [32] [33]. Thalamus is the region with the strongest correlation with disease progression, affecting cognitive function and being linked to depression and fatigue [18], [24]. Recently, atrophy of the amygdala was shown to be the main predictor of impairment of social cognition in multiple sclerosis [20].

Regarding WM atrophy, previous studies have shown its negative impact in the performance of mental processing and working memory tasks [33]. Moreover, the volume decrease of corpus callosum has been related to the prediction of developing MS [24].

Quantification of brain atrophy

Although MRI is extremely useful in detection of structural changes, the access to lesion volume, as well as the delineation of regions of interest, is usually done manually and may be prone to errors. Automatic methods have emerged to automatically handle these difficulties [34].

The volume of any brain structure can be quantified if a good MRI image contrast is provided (clear differentiation of tissues brightness). Brain atrophy is mostly measured from T1-weighted images, where there is a good contrast between a bright parenchyma against a dark CSF background and good GM/WM contrast [31]. Several software tools are currently available in clinical research to quantify brain atrophy and to give information about other morphological changes. These tools may differ on parameters like levels of user intervention, the target anatomy being examined (e.g. cortex or deep GM), and the target feature to be quantified (e.g. volume or thickness). However, there are standard methods like normalization, registration and automation [31], [35], which are common concepts to the different analysis tools (later approached in this dissertation). Brain atrophy assessment tools can be divided into two groups: registration and segmentation based, according to their dependence mainly on registration and segmentation routines, respectively. Nevertheless, most software use a combination of both methods [31], [35].

At the simplest level, brain atrophy can be quantified by segmenting the brain and calculating the volume of WM, GM and CSF. However, if these methods are not fully automated, it will be difficult to achieve reliable results. In addition, segmentation methods are time-consuming and vulnerable to operator bias. Currently, there are already fully automated segmentation methods available to overcome these problems. Some examples of such methods are: brain parenchymal fraction (BPF), index of brain atrophy (IBA), whole-brain ratio (WBR), brain to intracranial capacity ratio (BICCR), fuzzy connectedness/Udupa's method, 3DVIEWNIX, the Alfano method, and SIENAX [31], [35].

As alternative to segmentation-based methods, brain atrophy might be estimated using registration-based methods. In these approaches, the brain volume is directly quantified through image subtraction. However, it requires images spatially matched to produce meaningful results [36]. There are several of these methods

currently available as the brain boundary shift integral (BBSI), the structural image evaluation using normalization of atrophy (SIENA), the statistical parametric mapping (SPM), the template-driven segmentation (TDS), and the Voxel Based Morphometry (VBM).

In this work, VBM was chosen to assess morphological changes between groups. VBM is an automated and relatively easy-to–use, time-efficient, whole-brain tool. This tool can detect focal microstructural differences in brain anatomy, *in vivo*, between groups of individuals, without requiring any a priori decision concerning which structure to evaluate [37], [38]. This method will be described in more detail in the methods section.

Limitations of brain atrophy quantification

The quantification of brain atrophy with MRI is affected by several factors that can be source of errors. This might lead to either overestimation or underestimation of the atrophy values, causing misinterpretation of the results. Brain volume might be affected by image acquisition and processing factors, lifestyle (e.g. alcohol, smoking, diet, and dehydration), genetics, and other diseases (e.g. diabetes, cardiovascular risks). Importantly, age is also a factor to consider, since there is a decrease of brain volume with aging as referred before. Fluctuations of brain volume throughout the day must also be taken into account when measuring brain atrophy, particularly in matched case-control and longitudinal studies, as the brain volume seems to decrease from morning to evening [39].

Despite current methods' precision, brain atrophy measures are not commonly used in clinical practice. There is still a lack of comparative studies, a high variability between different methods, uncertainty in the results when applied to a single patient over the scale of months or a few years, and difficulty in developing completely automatic methods [31]. In order to put these measures into clinical practice, results need to be confirmed on larger datasets. There is also the need of establish normative values for inter-subject brain volume variability (both for healthy individuals and patients with MS). Finally, the creation of a worldwide standardized protocol for image acquisition, not only for research studies, but also for individual patient management, is required [31].

2.2.2 Surface Changes

In the context of brain morphology, cortical thickness and cortical folding are less common but relevant features to study in MS. Cortical thickness comprises the distance between the white matter surface and the pial surface [40]. Measurement of cortical thickness is a way of assessing cortical (GM) atrophy, being accessible by brain atrophy quantification methods [18], [28]. Studies have proved that cortical thinning is related with lesion volume, clinical symptoms, cognitive deficits and clinical disability [42], [43].

Cortical folding involves sulci and gyri. The gyrification index and the sulcal scale are examples of measures used to assess folding pattern abnormalities [44]. Gyrification index is a metric that quantifies the amount of cortex buried within the sulcal folds, as compared with the amount of cortex on the outer visible cortex. The mathematical approximation to measure gyrification index is defined as the ratio of the total folded cortical surface over the perimeter of the brain. Gyrification index can be seen as a degree of cortical folding: a cortex with extensive folding has a large gyrification index, whereas a cortex with limited folding has a small gyrification index [45], [46]. It is known that there is an increase in gyrification index during early stages of brain development, after childhood there is stabilization. Therefore, gyrification index can reflect the fundamental principles of cortical development and brain organization [47].

Measurement of cortical thickness and gyrification index is illustrated in Figure 2.1.

There are several studies showing the correlation of gyrification index and cortical thickness with aging and neurodevelopment disorders [44], [46]–[48]. Studying how these measures vary in the context of MS can provide valuable information for diagnosis and add to the understanding of disease.

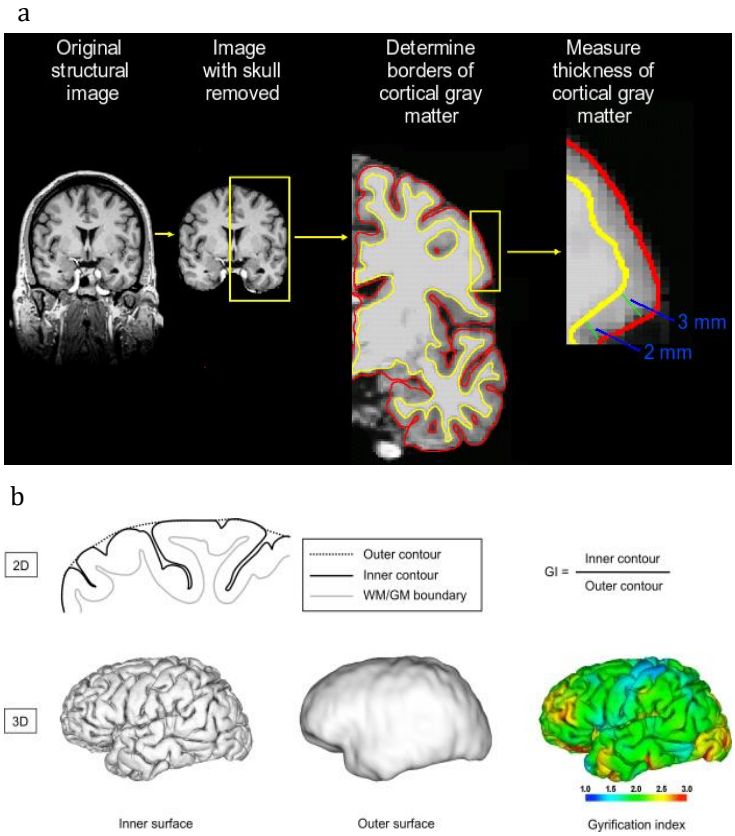


Figure 2.1: Measurements of surface changes. a) Measurement of cortical thickness. Yellow contours represent the frontier between WM and GM. Red contours represent pial surface. **b)** Measurement of gyrification index. Adapted from [45][46]

2.3 Statistical Analysis Methods

2.3.1 Univariate Statistical Analysis

Morphological alterations in MS are usually studied based in mass-univariate statistical analyses. This class of neuroimaging analysis is focused on the identification of signal changes at the level of individual image elements (i.e. voxels) at specific points in time. The most common approach consists in statistical parametric mapping, which performs individual statistical tests (in each voxel) to create whole-brain maps showing structural differences between groups. The voxels with significant differences highlight structural alterations in the presence of a disease, for example. Univariate analyses are advantageous when isolated features are able to support a diagnostic and are directly related to a certain pathology [49]–

[51]. However, this type of statistical evaluation accounts only for individual effects and does not provide information about the complex spatial organization of the brain structural alterations in MS.

Whole-brain, mass-univariate analyses (e.g. VBM) accounts for a considerable number of voxels in which each one undergoes a statistical test. The higher the number of tests, the higher is the probability of false positives occurrence (for example, tests that show a significant test statistic but no relevant structural differences between groups). A correction for multiple comparisons, which will be further elaborated, is then required to limit the occurrence of false positives. Correction for multiple comparisons moderates the statistical power of these approaches, which might turn them too conservative to detect subtle morphological differences in the brain [51]–[54]. Furthermore, a smoothing of the images, applied as a preprocessing step in these type of analysis (explained later in the methods section), leads to a decrease of spatial resolution to detect differences, which in turns leads to reduced ability to detect regional morphological changes.

2.3.2 Multivariate Statistical Analysis

An alternative to produce biomarkers, which accounts for the complex spatial correlations and multivariate relationships among image elements, are the multivariate pattern statistical analysis (MVPA)[51], [52]. This family of analyses methods focus on the examination of distributed imaging patterns to detect differences between groups in the absence of a *priori* hypothesis. It takes into account information ranging from meso to large scale, distributed across multiple brain systems from all available brain data. In other words, it considers information from multiple voxels rather than information from isolated brain regions (individual voxels).

Identifying subtle multivariate structural alterations in the brain leads to a better understanding of MS disease processes. Importantly, it may also allow the development of an objective and automated tool, with a strong potential for diagnostic and characterization of MS subtypes and progression. Therefore, these characteristics has led these techniques to be considered as biomarkers of great interest in the field of neuroimaging [49]–[51].

Random forest, logistic regression, neuronal networks, decision trees, or support vector machines (SVM) are examples of multivariate methods used in the field of neuroimaging data, especially when applied to the study of neurological disorders [55]–[57]. SVM is a common choice for estimating multivariate patterns in the brain. It is prepared to deal with different types of samples, from high to low dimensional spaces, with high and low sample sizes. As a supervised method, it has the ability to learn the classification of complex and high dimensional training data and to apply the learned classification rules to unseen data [50], [56], [58].

Several classification studies in the context of MS have applied SVM classifiers. SVM has proved to reliably identify patterns that can separate MS patients from healthy controls with high sensitivity and specificity [55], [56], [58], [59]. On the other hand, SVM classification of MS subtypes, to the best of our knowledge, has never been tested and might be a useful approach to help diagnose different subtypes of MS. This classifier was applied in this work and will be described in more detail in the methods section (for more details see section 3.8.1).

Multivariate analyses methods have been suggested as an important approach to the detection of structural brain patterns related to MS. However, a systematic study of its performance as compared to the univariate analysis methods is still missing.

3 Methods

3.1 Experimental Design

MRI data from MS patients and controls were analysed with three different purposes: to quantify MS related volume and surface changes and automatically classify patients and controls and two MS subtypes. VBM was applied to investigate GM and WM volume changes, and surface-based morphometry (SBM) was used to study between-group differences in measures of cortical thickness and gyrification index. A MVPA was implemented using a SVM to differentiate MS patients from controls and to discriminate MS subtypes based on GM and WM volume features.

As MS is characterized by white matter lesions, which can deteriorate tissue segmentation, leading to suboptimal results of these kind of analyses. We also tested the same approaches (VBM, SBM and MVPA) after lesion filling to analyse how structural measures are influenced by this procedure

T2-weighted imaging may, in theory, be more sensitive to microscopic neurodegenerative processes than T1-weighted imaging. Furthermore, brain atrophy should also be detectable in T2-weighted scans due to the strong signal intensity contrast between cerebrospinal fluid and brain tissue.

Thus, another aim of this study was therefore to test the hypothesis that VBM using T2-weighted scans (T2-VBM) could provide additional information of GM alterations that are not due to atrophy than standard T1-VBM using MPRAGE, at identical spatial resolution.

T1-MPRAGE images were used as input to VBM, SBM and MVPA analyses, whereas T2-FLAIR images were used as input to VBM analysis and MVPA analyses.

T1-MPRAGE images were preprocessed using SPM12 [60] CAT12 [61] and Lesion Segmentation Toolbox (LST) [62] software tools in MATLAB [63] (version R2019a) environment. LST was used to segment T2 hyperintense lesions in FLAIR images and to calculate lesion probability maps to fill lesions in T1-weighted MR images.

VBM and SBM analyses were performed in CAT12. MVPA analyses were carried with Pattern Recognition for Neuroimaging Toolbox (PRoNTo)[64], using the same

T1-weigthed preprocessed images as for VBM analysis, so that we could qualitatively compare the results obtained with different methods.

T2-FLAIR images were preprocessed using MRIcron software [65], BET: Brain Extraction Tool [66] function from MATLAB for skull-striping and SPM12 for remaining preprocessing and segmentation. CAT12 is an extension of the segmentation in SPM12 but uses a completely different segmentation approach. The only common process is the initial removal of non-brain tissue from the image to get a starting estimate for the segmentation [67]. Because CAT12 is not optimized for the contrast of T2-weighted images, it was therefore unable to perform a precise tissue segmentation and we resolved to use SPM12 in the case of T2-weighted images. VBM statistical analysis was also performed in SPM12. Since SBM analysis is only available in CAT12, it was not possible to perform this analysis in T2 images.

Further details of preprocessing steps and data analyses are described in the next subsections. Scripts of the methods implementation are available as supplementary material (Annex A).

3.2 Participants

This study is based on a dataset including 59 MS patients, 51 of which were diagnosed with the relapsing-remitting MS subtype and 8 with the secondary progressive MS subtype. All patients were recruited by Doctor Sónia Batista at the Neurology Department of the Centro Hospitalar e Universitário de Coimbra and met the criteria for MS diagnosis according to the McDonald Criteria [11]. A group of 64 healthy volunteers, matched on age, sex, and educational level with MS patients, was recruited from the community and served as healthy controls. These cohorts were previously described in research studies for which approval for MRI experiments was granted by local ethics committee and all participants gave written informed consent [20]–[22]. Demographic information about the participants is summarized in Table 3.1.

Table 3.1: Demographic Data on the Study Cohort

Groups	RRMS	SPMS	Controls
n	51	8	64
Age (mean ± SD) (years)	36.35 ± 6.56	40.88 ± 9,47	36.84 ± 9.79
Gender (F/M)	32/19	7/1	42/22
Disease Duration (years)	9.90 ± 6.45	15.25 ± 5.52	—

RRMS referrers to the group of MS patients characterized with relapsing-remitting subtype and SPMS is the group of MS patients characterized with secondary progressive subtype.

3.3 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique that produces detailed anatomical (and functional) images [68], [69]. Its non-invasive nature and versatility make MRI a powerful method for obtaining detailed information about living tissue without the use of damaging radiation.

The principle of MRI is based on the magnetic properties of the atoms and the molecular interactions that constitute the tissues under observation. As the human body is constituted mostly by water, the majority of MRI techniques rely on abundant and particularly favourable magnetic properties of hydrogen unpaired protons. Protons are positive electrical charged particles which are present in the nuclei of hydrogen atoms in water. These particles possess a spin and they rotate about its axis at a constant velocity. A moving electrical charge is in fact an electrical current. Consequently, spinning protons create their own small magnetic fields.

When applying a strong magnetic field in the MRI scanner, together with radiofrequency (RF) pulses, we can disrupt the protons natural alignment and force them into a non-natural realignment with the static magnetic field of the MR scanner, which is known as magnetization. By absorbing energy, this process has an effect in protons own small magnetic fields. When the RF pulse is turned off, the excited protons relax back to their lower energy state, which is known as relaxation, generating radio frequency waves. Ultimately this is the signal measured (received) in the receiver coil (antenna) of the MR scanner. Through advanced computer techniques, the signal is processed, and the MR image is formed.

MRI is capable to distinguish many different normal and pathologic tissues. There are several ways to contrast different tissues. The two main factors to create different MRI contrasts are the imaging sequences of RF pulses being used [68], [69] and the existence of two simultaneous types of relaxation effects: the longitudinal relaxation and the transversal relaxation, characterized by time constants T1 and T2, respectively [70] [71]. There are three types of MRI images, as exemplified in Figure 3.1: proton density images (PD), which have also the influence of the protons present in the tissues being examined, T1-weighted images and T2-weighted images [71].

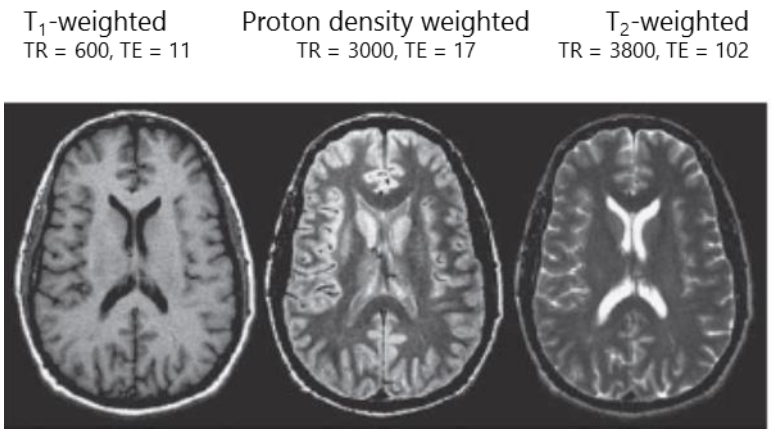


Figure 3.1: Magnetic resonance images of the same anatomical section showing a range of tissue contrasts. TR stands for repetition time and TE stands for echo time. In the first image, cerebrospinal fluid is black, whereas in the last image it is bright. Contrast is manipulated during image acquisition by adjusting several parameters, such as the TR and the TE (times given in milliseconds), which control the sensitivity of the signal to the local tissue relaxation times T1 and T2 and the local proton density. In the proton density weighted image GM is brighter than CSF, despite CSF has the highest proton density. The reason is that the T1 of CSF is too long, and for TR = 3 seconds the CSF signal is still substantially T1-weighted. Adapted from [122] .

For quantitative structural brain analysis normally only T1-weighted and T2-weighted images are used. T1-weighted imaging offers the greatest segmentation clarity between GM, WM and CSF, and is therefore most frequently used for quantitative MRI studies of brain morphology. On the other hand, T2-weighted imaging may be preferably used for quantification of intracranial volume. These images have increased signal intensity of CSF that allows easier quantification of CSF and brain parenchyma together. T2 images are conceived to study conditions with,

for example, abnormal inflammation or lesions associated to a disease [71]. In general, tissue signal intensities visualized on PD images are very similar to the ones pictured on T2-weighted images. Yet, some white matter lesions, such as periventricular MS plaques, and small superficial lesions are more hyperintense relative to CSF on PD images. Therefore, these images are useful to detect such lesions, which are difficult to identify on T2-weighted images, especially when lesions are adjacent to CSF [71].

In order to obtain more information about a disease, to produce reliable diagnosis and to monitor treatment, integration of various types of images is a common and sensible approach.

3.4 MRI Acquisition

MRI data were collected in a Siemens Magnetom TIM Trio 3 Tesla scanner (Siemens, Munich, Germany) with a phased array 12-channel birdcage head coil. For each participant two structural sequences were acquired. First, it was acquired an anatomical T1-weighted 3D magnetization prepared rapid gradient-echo (MPRAGE) sequence: TR 2530 ms; TE 3.42 ms; TI = 1100 ms; flip angle 7°; 176 slices with no gap, with isotropic voxel size 1 x 1 x 1 mm^3; field of view (FOV) 256 x 256 mm^2. Then, a sagittal 3D fluid-attenuated inversion recovery (FLAIR) sequence was also acquired: TR 5000 ms; TE 388 ms; TI 1800 ms; 160 slices with 1 x 1 x 1 mm^3 voxel size; FOV 250 x 250 mm^2.

3.5 Preprocessing

Firstly, MRI data were converted into a standard format, aligned to match the SPM template orientation, T1 images were coregistered, and the WM lesions were filled. Then, VBM and SPM processing were followed separately. Common preprocessing steps for both univariate analyses are described below.

Conversion into NIfTI format

Many neuroimaging quantitative software tools, including SPM12, CAT12 and PRoNTo, deal with MR images in NIfTI (Neuroimaging Informatics Technology Initiative) format. NIfTI format offers a standardized way to represent images, it is the most commonly used type of analytic file and it was developed to facilitate cross-platform and cross software interpretability [72]. During MRI acquisition, images were saved in DICOM (Digital Imaging and Communications in Medicine) format. Once SPM does not accept DICOM formats as inputs it was necessary to convert images to NIfTI format.

Correcting images orientation

Orientation of each structural scan was manually corrected by aligning each image onto the axis of the anterior and posterior commissures and setting the centre of the image in the anterior commissure. This step is needed to match the SPM template and is an important step to further processing less prone to errors, and to give the spatial normalization a good place to start from [73].

Coregistration of within-section anatomical scans and mean image

For each subject T1-MPRAGE images and one T2-FLAIR image were acquired in the same session. In order to increase the signal to noise ratio (SNR) of the data, the two T1 images were coregistered. A better-quality image was obtained with calculation of the mean of the two T1-MPRAGE scans from each participant. Coregistration is a fundamental step that also serves to align the scans that may be in different orientations, modalities, or acquired in different scanning sessions. The coregistration method is based on work by Collignon *et al*. using voxel based (VB) registration algorithms [74].

Lesion Filling

It has been shown that the heterogeneity of the tissue with WM lesions can distort registration, normalization and tissue segmentation. Consequently, if lesions are processed within the images, it might lead to an underestimation or overestimation of different brain metrics or analyses [75]–[78].

In T1-weighted images, WM lesion are hypointense, so they are displayed with MR signal intensities close to GM and cerebrospinal fluid (CSF). After brain segmentation and subsequent tissue classification, these lesions are categorized as GM or, in some cases, as CSF. It can also occur the inverse case: if WM lesion voxels are classified as WM (which is hyperintense in T1-weighted images), lesion voxels with hypointense signal intensities are added into the WM tissue distribution, increasing the probability of GM voxels with similar intensity to be misclassified as WM [75]–[77].

Several authors have proposed different methods to overcome the issues caused by the effects of WM lesions in MS patients. The commonly proposed solution consists in filling WM lesions with intensities like those of WM before performing tissue segmentation and image registration. By segmenting and filling the lesions, it is admitted that lesions are not present and therefore that their effect is cancelled. These methods can be divided into two groups: methods which use local intensities from the surrounding neighbouring voxels of lesions and methods which use global WM intensities from the whole brain [75].

Here, lesion filling was performed using the LST version 2.0.15 for SPM [62] T1 MPRAGE images from all participants were filled with WM lesion maps extracted automatically from T2 FLAIR images from the same corresponding participants. The following steps were required:

- **Lesion segmentation**

  LST offers two algorithms to obtain lesion probability maps, Lesion Prediction Algorithm (LPA) and Lesion Growth Algorithm (LGA). LGA demands an initial threshold defined by the user, while LPA, despite being a beta version in LST, requires no user-defined parameters. Moreover, LPA is faster and more sensible in some regions when compared to LGA [62]. For this reason, lesions were segmented using the LPA [79], as implemented in the LST toolbox.

  This algorithm consists of a binary classifier, in the form of a logistic regression model, previously trained on a dataset of 53 MS patients with severe lesion patterns. As covariates for this model, a similar lesion belief map as for the lesion growth algorithm [80] was used, as well as a spatial covariate that takes into account voxel specific changes in lesion probability. Parameters of this model fit

are used to segment lesions in new T2-weighted images by providing an estimate of the lesion probability for each voxel. The main output of this segmentation was the lesion probability map of each subject. Segmentation of lesions is a critical aspect and it is recommended to be done by experts (or at least reviewed by experts in case it is done by automatic algorithms). Since this is an exploratory study, lesion probability maps were used without a revision by an expert. Yet, it is important to have the notion that manual segmentation by experts is one factor that could improve the performance of analyses of this type of images.

- **Lesion Filling**

  Lesion probability maps were then coregistered with corresponding T1 MPRAGE images, and the filling algorithm was applied, resulting in filled T1 MPRAGE images. The LST filling method uses global WM intensities from the whole brain. As such, the T1 images were filled using a method inspired by a technique proposed on the work of Chard *et al.* [81], where lesion voxels are replaced by random intensities from a Gaussian distribution. This distribution is generated from the normal-appearing WM (NAWM) intensities and then filtered to re-establish the original spatial variation in WM [82].

A reliable tissue classification is crucial not only for quantify volume, but also for assessing more complex features of brain structure, for example surface measures, cortical thickness and gyrification index [76]. Filling the lesions with intensities corresponding to the surrounding NAWM granted an accurate tissue classification and trusty measurements of brain volumes [78]–[82].

3.5.1 VBM Preprocessing

VBM preprocessing includes steps such as normalization to a reference brain (template), modulation, tissue segmentation (GM, WM and CSF) and spatial smoothing of segmented images. VBM preprocessing of T2-weighted images required an additional prior step: skull stripping.

T2-Skull stripping

Skull stripping is a method to exclude brain areas that are not relevant to the study of structural differences between groups. These areas are non-brain tissues such as skin, fat, muscle, neck, and eyeballs which constitute a major obstacle for automatic brain image segmentation and analysis techniques. Therefore, quantitative morphometric brain studies often require a preliminary processing to separate the effective brain region accountable for the study from extra-cranial or non-brain tissues [83]. Performance of further preprocessing algorithms such as warping algorithms (normalization) and consequently results from statistical analyses, VBM and SBM, can be improved [84], [85].

In CAT12, skull stripping is part of the tissue segmentation step, by segmenting the anatomical scan and using a version of the sum of GM and WM probability maps restricted based on a threshold to mask out the bias corrected structural scan [73]. Due to the poor contrast of T2-weighted images in differentiating brain areas, skull tripping was performed manually with MRIcron software and BET function.

Spatial Normalization

Spatial normalization is the method to match all the brain images in the same stereotactic space in order to ensure a voxel wise correspondence across different images, which is crucial for group data analyses. All scans were normalized into a reference brain (template), which in SPM is in the standard Montreal Neurological Institute (MNI) brain space [86], [87]. By removing variability between brain sizes and shapes, a considerable amount of errors in further analysis is avoided [88]–[91].

Normalization can be divided into two steps: the linear (affine) registration and the nonlinear registration. The linear registration consists in linear transformations applied to the brain images (rotations, translations, scaling). This type of normalization considers the brain image as a rigid body, and only apply transformation steps that do not affect the brain shape. The goal is to remove global differences between the subject and the template by matching the MR images with the template in overall shape. The nonlinear registration accounts for local nonlinear shape differences by expanding or contracting brain regions, and thus changing local volume of specific structures [89], [91].

Both SPM12 and CAT12, during the segmentation step, provide a template for the normalization algorithm (the high-dimensional Diffeomorphic Anatomic Registration Through Exponentiated Lie - DARTEL) to spatially align each subject's image with the corresponding template [92]. The idea behind DARTEL is to increase the accuracy of intersubject alignment by modelling the shape of each brain using millions of parameters (three parameters for each voxel) [93]. The normalization step in CAT12 includes both linear (affine) and nonlinear options. We applied the affine normalization. Although nonlinear normalization was defined by default in previous VBM-toolbox versions, this method is no longer recommended. The use of standard modulation (described shortly later) in combination with total intracranial volume (TIV) as covariate provides more reliable results [67].

Importantly, normalization was not meant to make changes on a fine scale, i.e. to match every brain feature and transform one brain into another, exactly. Normalization was necessary to detect regional differences between groups disregarding global shape differences. Indeed, if the spatial normalization was made at such fine scale, all the segmented images from all participants would be identical and no significant differences would be detected at a local scale [90]. To correct for the effects of spatial normalization in our study, we applied modulation (described below) to our data and used TIV as covariate to correct different brain sizes.

Tissue Segmentation

Brain segmentation consists of classifying brain tissues into GM, WM and CSF. A brain image with a clear differentiation between tissues is required to an accurate segmentation. As such, T1-weighted images are preferred to perform brain tissue segmentation.

There are several ways to perform segmentation, usually prior probability maps, as well as voxel intensities, are the base of segmentation algorithms. Tissue probability maps from normal subjects provide prior spatial information, i.e. contain information about each voxel intensities' distribution for the various tissue types. This way, each voxel is assigned the corresponding (most likely) tissue type. Segmentation steps also consider smooth intensity variations (non-uniformity of image intensity), for example magnetic field variations or variations caused by

different positions of cranial structures within the MRI scanner and corrects for them [88]–[90].

Modulation

As explained before, in the case of an extremely precise spatial normalization, all MRI segments would look identical and no intersubject structural differences would be detected. Modulation serves to compensate for these possible effects and involves resizing brain structures that were expanded or contracted. The total amount of GM/WM in the modulated images remains the same as it would be in the original images [88]–[90]. This scaling is achieved through multiplication of the segmented images by the relative voxel volumes before and after warping. These relative volumes are called the Jacobian determinants of the deformation field [88].

Without this adjustment, VBM can be thought as an analysis that compares concentration of GM or WM tissues between groups. This is because brain structures are decreased or increased and therefore have less or more voxels (variations in concentration). With the adjustment, VBM compares the absolute volume of GM or WM structures between groups. The two approaches are known as "non-modulated" and "modulated" VBM, respectively [89].

At this point warped, modulated and segmented images were obtained from all participants.

Extraction of additional volume measures

TIV was computed trough "Estimate TIV" function in CAT12 to be used as covariate in the statistical models, with both no filled and filled data, and correct for different brain sizes. A good reason to use TIV as covariate is that at a later age, a reduction in brain volume is normal. However, the aim is to test the volume difference between controls and patients regardless of their age or gender.

Two other additional variables were extracted through "Extract Values of Interest" function of LST to be used as covariates in the statistical model with the filled data: the total lesion volume (TLV) and the number of lesions. Large lesions or multiple lesions with small volume might bias the volume measures. These two covariates were used to eliminate or at least account for these effects in the statistical analysis.

Smoothing (and bias correction)

Normalized, segmented and modulated images were smoothed with an isotropic Gaussian kernel. Smoothing can be thought of as a filter for bias correction of intensity non-uniformities, required to guarantee normality of the data for subsequent statistical parametric analysis [88]–[91]. After smoothing, each voxel intensity represents the weighted average of the surrounding voxels intensities, which results in a blurred image [90]. The size of the smoothing kernel determines the number of surrounding voxels that accounts for the average intensity, which typically varies between 4-16mm [88]. The size depends on the image resolution, i.e. it must be of the same order of magnitude of the expected regional differences between the groups [89], [91].

Smoothing of the images makes the regions around the voxels encompassed by the smoothing kernel contain a more homogeneous GM or WM intensities, leading to noise suppression and to improve the quality of the analysis. Smoothing also removes fine scale structure from the data, i.e. brain structures that are not constant (same position, same shape, same size) from subject to subject. One example is the primary sulci, quite deep brain structures, which are formed at early age and are conserved over subjects, however, sulci developing later generate more variability. Therefore, some structures can precisely match, whereas others cannot [88]. Smoothing increases the sensitivity of statistical analysis to detect variations by reducing the inter-subject variability, although excessive smoothing may diminish the accuracy to localize such differences in the brain [90].

During normalization, registration errors may occur and smoothing also compensates for the inexact nature of this step by blurring anatomical differences [89], [91]. These beneficial features of smoothing render the data more normally distributed, increasing the validity of parametric statistical tests and reduces the effective number of statistical comparisons, making the correction for multiple comparisons less severe [88]–[91].

Normalized and modulated images of both GM and WM from T1 filled/no filled and T2 images were smoothed in SPM12 using an 8 mm full-width-at-half-maximum (FWHM) isotropic Gaussian kernel. After these steps, normalized, segmented, modulated and smoothed images were obtained, which will be used as inputs to the statistical analyses. VBM preprocessing steps are illustrated in Figure 3.2.

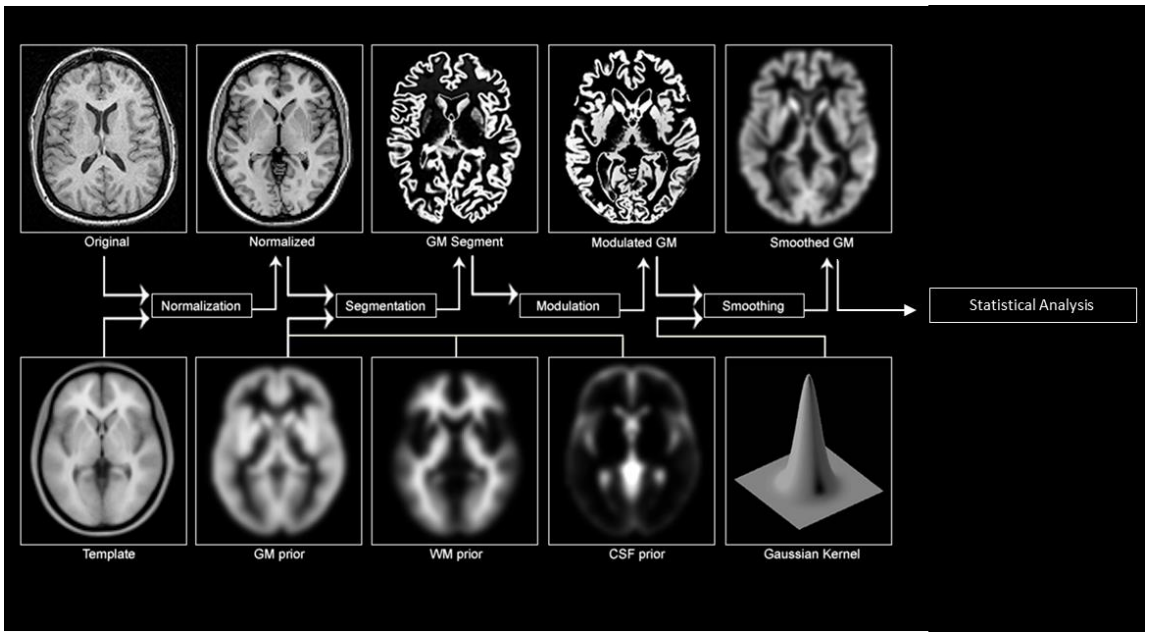


Figure 3.2: Main steps of voxel-based morphometry preprocessing. In this case it is shown GM VBM as an example. Adapted from [123].

3.5.2 SBM Preprocessing

SBM analysis of T1 images deals with surface measures, which requires further preprocessing steps. Similarly to VBM, images have to be normalized, segmented and modulated, and then the same images that undergo smoothing and VBM statistical analysis, undergo in the case of SBM: surface reconstruction, topological correction, spherical mapping, spherical normalization, extraction of surface measures of cortical thickness and gyrification index, spatial smoothing of extracted metrics and SBM statistical analysis. Cortical thickness is obtained through delimitation of the outer and inner boundaries of GM. The determination of the distance between the two borders in each voxel give us a cortical thickness metric. On the other hand, gyrification index is obtained by delimitating the outer and inner counters of the cortex. The mathematical approximation to measure gyrification index is defined as the ratio of the total folded cortical surface over the perimeter of the brain. The extraction of these surface parameters was obtained with the function "Extract Additional Surface Parameters" in CAT12. Then, the extracted metrics were corrected for inhomogeneities using the CAT12 "Resample and Smooth Surface Data" function.

Surface Reconstruction

After segmentation step in CAT12, a surface mesh is generated (surface reconstruction) from volume images, to allow the estimation of cortical thickness and gyrification index. This surface mesh is obtained through a fully automated method called projection-based thickness (PBT). Through brain segmentation, WM distance to outer GM is estimated and local maxima (which corresponds to cortical thickness) is projected onto GM voxels using a neighbouring relationship described by the WM distance [94].

Topological Correction

Surface mesh contain topological defects which can interfere with estimation of surface measures. Errors during segmentation, noise, partial volume effects, and artifacts during the MRI data acquisition are examples of possible causes of topological defects. Topological defects usually are in the form of handles and holes that prevent the deformation of the flat surface mesh into a sphere. A method called topological correction is needed to deal with these defects. During the topology correction process, handles should be cut, and the holes should be filled. Besides this correction, smoothing is also necessary because cortical surface contains sharp edges, caused by artifacts such as noise, that need to be minimized [95].

Spherical Mapping

Brain mapping analyses require the deformation the flattened surface mesh into a suitable inflated brain coordinated system. Since the cortical surface is roughly homeomorphic to a sphere, the most commonly used coordinate system is a spherical one. This process is called spherical mapping. The spherical mapping method provides an isometric map that conserves all angle and area information in the original cortical mesh, which facilitates an inter-subject analysis. Furthermore, an inflated brain makes regions with differences in surface measures more easily visible [96]. In addition, reparameterization is performed to reduce distortion inherent to the initial spherical maps. The reparameterization algorithm used by CAT12 is described in Yotter *et al.* work [96].

Spherical Normalization

Finally, spherical normalization is performed by adapting the volume-based diffeomorphic DARTEL algorithm, used by CAT12, to work with spherical maps and achieve spatial normalization of surfaces [97].

Smoothing

For the same reasons as described for VBM preprocessing, smoothing is applied to surface measures prior to statistical analysis. Cortical thickness and gyrification index images were smoothed using 15 mm and 20 mm FWHM kernel respectively through "Resample and Smooth Surface Data" function in CAT12.

Steps of SBM preprocessing are illustrated in Figure 3.3.

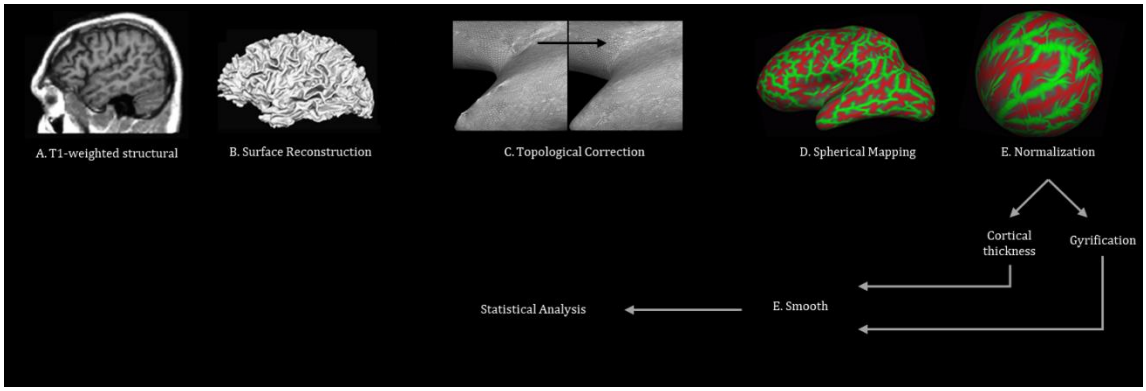


Figure 3.3: Main steps of surface-based morphometry preprocessing. Green areas represent gyri and red areas represent sulci. Adapted from [124].

3.6 Sample Homogeneity

Sample homogeneity of all images was performed to identify images with poor quality, possibly containing artifacts, and to check if enough data quality was assured to proceed to further analyses. This quality check step would help to identify and optimize problematic processing steps and to exclude problematic images.

The method behind this process is based in the standard deviation (SD) of each image across the sample. In order to obtain standard deviation values, firstly it was

needed to calculate sample mean homogeneity, then the squared distance of each image to the sample mean, and finally the sum of the squared distance of each image from the sample mean, which resulted in SD. The squared distance of one image from the sample mean is the measure that represents how much that image homogeneity deviates from the sample mean homogeneity. Images that presented a mean correlation below two standard deviations were carefully checked [98].

All the volume segments of GM and WM from T1 and T2 images, as well as surface images from T1 images were checked for sample homogeneity.

3.7 Univariate Statistical Analyses

In the framework of neuroimaging studies, statistics are applied in order to find patterns in the data. These patterns can be achieved by studying the variables of interest in a sample. A variable is a characteristic that can be measured, or it is observed on a subject. This variable can be dependent, if it is on the influence of other variables, or independent. In clinical studies, a sample is collected, a number of variables of interest are considered, and a statistical test is performed to compare variables. When the goal is to study only one variable, this is called univariate statistical analysis [99]. VBM and SBM analyses are univariate approaches, in the sense that only one dependent variable is tested for differences between independent groups at a time (GM, WM, cortical thickness or gyrification index). In fact, a mass univariate approach is behind VBM and SBM analyses, as the same independent statistical test is performed on every single voxel or vertex, respectively.

3.7.1 Univariate Statistical Models

General Linear Model

VBM and SBM analyses are based on the application of General Linear Models (GLMs). A GLM is statistical method to relate continuous or categorical predictors (independent variables and confounds) to a single outcome variable (dependent variable). It describes a response (Y) in terms of all its contributing factors (X) in a

linear combination (with weights β), while also accounting for the contribution of error (ε). This model is built on a simple equation:

$$Y = X\beta + \varepsilon \qquad \text{(Equation 3.1)}$$

Y is the observed data, the dependent variable in which we are interested in study (e.g. GM volume), which will be modelled/predicted with as much as different factors that can account for its variation between images. X is the experimental design matrix, which embodies as predictors all experimentally controlled factors and potential confounds. β represents the regression coefficients, also known as beta weights, quantifying how much each predictor X independently influences the dependent variable Y. Finally, ε is the error, the variance in the data (Y) which is not explained by the linear combination of predictors (X) [100], [101].

Confounds are variables (covariates) that may produce effects with no interest for the study and bias the results. They may also capture interesting effects correlated with data. For example, if we are interested in observing the GM atrophy in patients when compared to controls, we want to look at differences that appear exclusively due to the disease and not due to effects related to age or gender. However, we can be interested in study how the brain atrophy develops with aging, by correlating the effects of age with brain atrophy.

We want to see whether our design matrix has an effect on brain structure (the dependent variable Y). In other words, we want to see how GM or WM volume in each voxel varies across different scans. The null hypothesis is that there are no differences between controls and patients, i.e. our design matrix has no effects on brain structure.

In this work, as there are many samples (multiple scans of the same voxel) of a response Y, it resulted in a mass-univariate analysis. Equation 3.1 was applied for each voxel of the dataset resulting in the following equation:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & \cdots & X_{1k} \\ X_{21} & \cdots & X_{2k} \\ X_{31} & \cdots & X_{3k} \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{nk} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad \text{(Equation 3.2)}$$

Each row of the design matrix is one scan and each column is a predictor or confound. At each voxel, the optimal linear combination of the effects that may

influence the outcome is computed. At this point, all the β parameters were estimated for each voxel and ready to proceed to statistic tests.

- **Design matrix of VBM**

  For VBM analysis performed with each image type (no filled T1-MPRAGE, filled T1-MPRAGE, T2-FLAIR), X corresponds to scans containing brain segments of WM or GM from the control and MS groups; Age, gender and TIV are the covariates for all models, and TLV and number of lesions are additionally included as covariates in the model with T1 scans after lesion filling.

- **Design matrix of SBM**

  For SBM analysis performed with T1-MPRAGE no-filled or T1-MPRAGE filled, X corresponds to scans containing cortical thickness or gyrification index from the control and MS groups. Age and gender are included as covariates in both models.

Contrasts

Contrasts are linear combinations of predictors coefficients corresponding to the null hypothesis, represented by a row vector in the form $c'\beta$. As here it was intended to test only a single hypothesis, volume/surface differences between two groups, disregarding the effects of the covariates, t-contrasts were used for this purpose.

By defining a contrast, relationships between data are generated disregarding others, for example ignoring the effects of the confounds initially defined in the model that have no interest for the study. The linear combination established by the contrast supresses the effects of these variables. The null hypothesis is that the linear combination $c'B$ is zero. To model the analysis towards the goal of testing groups' differences, the predictors weights must be specified in the contrast vector. The construction of the design matrix will determine the order of the weights to be specified in the contrast vector c. The order of the predictors in the β column must have a correspondence with the order of the weights in the contrast vector c [102] .

Example:

The hypothesis to be tested is if there are GM volume differences between groups (controls and patients). GM volume is the dependent variable (Y) whereas T1 GM

segments of controls, T1 GM segments of patients, age, gender and TIV are the independent variables. These variables may or not produce an effect in the observed data (*Y*), depending on the defined contrast. In this context, a *t*-contrast specified by [1 -1 0 0 0] tests whether there are brain regions where GM volume in the control group is higher than the one of disease group, excluding from the analysis the effects caused by age, gender and TIV.

In this work, it was only tested one hypothesis in each model – to test for differences between the patients and controls data specified in each VBM and SBM analysis. The groups are all the same in each analysis, controls and patients. The inputs of the analyses vary depending on the type of image and what was intended to study. In each analysis there was only one type of image, thus one dependent variable being studied.

- **VBM *t*-contrasts**

  Differences in GM/WM volume are inspected by an increase or decrease of the GM/WM volume in one group relative to the other. Therefore, the following *t*-contrasts in each of the 6 VBM analyses were applied (3 VBM analyses of GM and 3 VBM analyses of WM):

  T1-MPRAGE no-filled scans: *t*-contrast $= [1 \quad -1 \quad 0 \quad 0 \quad 0]$ to find brain regions where there is an increase of GM/WM volume in the control group relative to MS group, ignoring the effects of age, gender and TIV.

  T1-MPRAGE filled scans: *t*-contrast $= [1 \quad -1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$ to find brain regions where there is an increase of GM/WM volume in the control group relative to MS group, ignoring the effects of age, gender TIV, TLV and number of lesions.

  T2-FLAIR scans: *t*-contrast $= [1 \quad -1 \quad 0 \quad 0 \quad 0]$ to find brain regions where there is an increase of GM/WM volume in control group relative to MS group, ignoring the effects of age, gender and TIV.

  The contrasts destined to find increases of GM/WM volume in MS group were also tested, by changing the contrast weights to their symmetrical.

- **SBM *t*-contrasts**

  Similarly, the following *t*-contrasts were applied in SBM analyses. The *t*-contrast $= [1 \quad -1 \quad 0 \quad 0]$ was applied to cortical thickness and gyrification index maps from both T1-MPRAGE no-filled scans or T1-MPRAGE filled scans. It allowed to find brain regions where there is an increase of cortical thickness/gyrification index in the control group relative to the MS group, ignoring the effects of age and gender.

  A contrast to find an increase of cortical thickness/gyrification index in MS group was also applied, by changing the contrast weights to their symmetrical: *t*-contrast $= [-1 \quad 1 \quad 0 \quad 0]$.

_t_-test

A *t*-test is a statistical test to make inferences about parameters of populations. Usually, it is used to compare the mean of a variable in one group against a given mean value or to compare the means between two groups. *t*-tests belong to class of parametric methods, in which the data are assumed to be normally distributed.

A brain scan contains millions of voxels (in volume space) and vertexes (in surface space). Each voxel contains a certain volume of GM/WM and each vertex has cortical thickness and gyrification index values associated. Therefore, each group containing all the subject has a mean value of GM/WM volume and a mean value of cortical thickness and gyrification index in each voxel, all assumed to be normally distributed due also to the smoothing preprocessing step. The *t*-test will compare if the two gaussian samples have the same mean in each voxel. In this case, it is a two-sample *t*-test where the two sample are the two independent groups.

In this study, the null hypothesis assumes that there are no differences between the groups and so no linear combination between the regressors of the design matrix exists, $c^T\beta = 0$. As the linear combinations of regressors are normal distributed, $c^T\hat{\beta}$, the mean value of the combination of regressors was tested against zero in order to find regional differences.

The mathematical approximation to perform these tests is given by:

$$T = \frac{c^T\hat{\beta} - c^T\beta}{\sqrt{var(c^T\hat{\beta})}} = \frac{c^T\hat{\beta}}{\sqrt{var(c^T\hat{\beta})}} \qquad \text{(Equation 3.3)}$$

Where, $var\left(c^T\hat{\beta}\right)$ is the standard deviation of the linear combinations.

From each *t*-test performed in each voxel results a value of t which comes from a *t*-distribution. A small *t*-value indicates that the null hypothesis has a low probability of being true and a large *t*-value that the null hypothesis is likely false. However, to evaluate if these values provide sufficient evidence to reject the null hypothesis, it is necessary to compare the computed *t*-values to a given value, a specified threshold. This threshold is a quantity, a percentile from the *t*-distribution. The selected percentile has a significance level α associated, specified by the user. If the *p*-value associated to a given *t*-value is below the significance level α, the null hypothesis should be rejected in favour of the alternative one. [102] [103].

Correction for Multiple Comparisons

In statistics, the multiple comparisons problem occurs when a statistical analysis involves multiple simultaneous statistical tests. A comparison of the p-value with the chosen confidence level is generally applied to each test individually. However, it is desirable to have a confidence level for the whole family of simultaneous tests.

In VBM and SBM analyses, the results of voxel-wise statistical tests require a correction for multiple comparisons to determine the significance of an effect in any given voxel, accounting for the simultaneous tests. The correction for multiple comparisons decreases the false positives probability, because the more tests are carried, the higher the probability that some reject the null hypothesis by chance. Without any correction, the number of false positives would be proportional to the number of the independent *t*-tests. At the brain scale, there is a considerable number of false positives due to the very high number of voxels or vertexes tested with VBM or SBM.

There are several methods to solve this problem. In the context of VBM, the most used is called the family wise error (FWE) correction. Within these class of methods, the Bonferroni correction is often used, although its conservative nature. This method controls the false positives adjusting the statistical threshold: the *p*-value threshold is divided by the number of performed comparisons. Here, the number of comparisons is equal to the number of voxels or vertexes being study. Bonferroni correction assumes that all the tests are independent from each other, but at the brain scale this is not entirely true. The structure of the spatial correlations of the

brain data is very complex and difficult to understand, in such a way that one voxel/vertex is not completely independent from the surrounding ones.

An alternative approach is the random field theory (RFT). RTF considers the spatial variance of the data, and therefore assumes the dependent relationships between the voxels/vertexes. To control the FWE rate, RTF calculates the smoothness of a given statistical map, analyses voxels/vertexes (or clusters or patterns) with particular statistical levels and estimates how these voxels/vertexes would appear by chance in an image with that local smoothness. This way, RTF controls the number of false positives at a regional level (e.g. within clusters) instead at the voxel level [104], [105]. Instead of RTF, there are other FWE correction methods that can be put into practice, such as the false discovery rate (FDR), permutation testing and small-volume correction. Nonetheless, RTF was here applied in both SPM12 and CAT12 analyses.

Statistical Parametric Maps

Statistical Parametric Maps are the resultant outputs of the VBM and SBM analyses. These maps are images that show the outcome of the parametric *t*-test performed in each voxel. At a specified threshold, with correction for multiple comparisons, SPM highlight the brain regions where there are significant differences between the groups.

3.7.2 Implementation

VBM Model

Based on VBM analysis the hypothesis of regional GM and WM differences, between the MS patients and controls, was tested.

The statistical models based on no filled T1-MPRAGE data and T2-FLAIR data was built with 2 independent samples, i.e. the two groups, and with 3 covariates (gender, age and TIV). For this reason, it was chosen a two-sample *t*-test. It was applied a 0.1 absolute masking threshold to compute the statistics only in areas that are above this intensity threshold. Areas below this threshold will be excluded. This allows to eliminate voxels with very low intensity values (e.g. voxels outside the brain or voxels with noise) and helps preventing the analysis of voxels in WM when

performing the analysis in GM and vice versa [106]. A statistical threshold of $p < 0.05$ with FWE correction for multiple comparisons was also applied.

The statistical models based on the filled T1-MPRAGE data was similarly constructed, except for the covariates. In this case, 5 covariates were used: gender, age, TIV, TLV and number of lesions.

To summarize, the following models were built:

- One statistical model based on GM volume extracted from no filled T1-MPRAGE images;
- One statistical model based on WM volume extracted from no filled T1-MPRAGE images;
- One statistical model based on GM volume extracted from filled T1-MPRAGE images
- One statistical model based on WM volume extracted from filled T1-MPRAGE images;
- One statistical model based on GM volume extracted from T2-FLAIR images;
- One statistical model on WM volume extracted from T2-FLAIR images;

These statistical tests allowed a qualitative comparison between the results of filled and no filled data, as well as between T1 and T2 images.

SBM model

Based on SBM analysis the hypothesis of regional cortical thickness and gyrification index differences, between MS patients and controls, was tested.

The statistical models of both filled and no filled T1-MPRAGE data were built similarly to VBM analysis, with 2 samples i.e. two groups, 64 controls and 59 patients and with 2 covariates, gender and age. Surface measures do not depend on volume measures, so TIV is not needed as covariate in the model [107]–[109]. No masking threshold was applied to the data. Cortical thickness maps were corrected with FWE correction ($p < 0.05$) while gyrification index maps were obtained with no correction ($p < 0.001$) for both analyses.

The following models were built:

- One statistical model based on cortical thickness extracted from no filled T1-MPRAGE images;
- One statistical model based on gyrification index extracted from no filled T1-MPRAGE images;
- One statistical model based on cortical thickness extracted from filled T1-MPRAGE images;
- One statistical model based on gyrification index extracted from filled T1-MPRAGE images.

3.7.3 Descriptive Univariate Plots

Descriptive univariate scatter graphics are a useful and practical way to show results derived from a comparative analysis. These graphics were obtained through MATLAB function "UnivarScatter" [110] with the purpose of plot the features values and to see how they differ between groups.

Graphics of GM and WM volume, as well as cortical thickness and gyrification index values in the regions that show differences between groups were obtained. To access GM and WM volume values, a mask of each cluster (region) that carry significant differences was saved in SPM. These masks were applied to GM and WM volumes of each subject and the volume values were extracted through "get_totals" function from MATLAB. To obtain surface values, cortical thickness and gyrification index, in the regions that show evidence of surface differences, the "Display Surface Results" and "Plot mean inside cluster" functions from CAT12 were used.

3.8 Multivariate Pattern Statistical Analysis

Statistical pattern recognition is a field within the area of machine learning, which is concerned with automatic discovery of regularities (patterns) in data using computer algorithms. Computer algorithms take advantage of these regularities to take actions for example to classify the data into different categories. In the field of neuroimaging studies, brain scans contain numerous spatial patterns which are

used by learning models to discriminate between experimental conditions or groups of subjects [64].

3.8.1 SVM Classification Models

In this work, SVM classification was applied to investigate if MS patients and controls can be discriminated through morphometric brain measures. The automatic classification of different groups of patients, SPMS vs RRMS patients, was also tested.

SVM is a discriminative classifier formally defined by a separating hyperplane. Given labelled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two-dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side (Figure 3.4). The idea behind SVM algorithm consists in defining a decision boundary that separates the study objects in different classes, based on their features, as generally as possible. Then SVM determines the maximal distance between the decision boundary and the closest objects of each class, to define the support vectors. The largest margin between support vectors ensures the maximal generalization capacity of the classifier regarding new data (unknown subject)[111]–[113] .

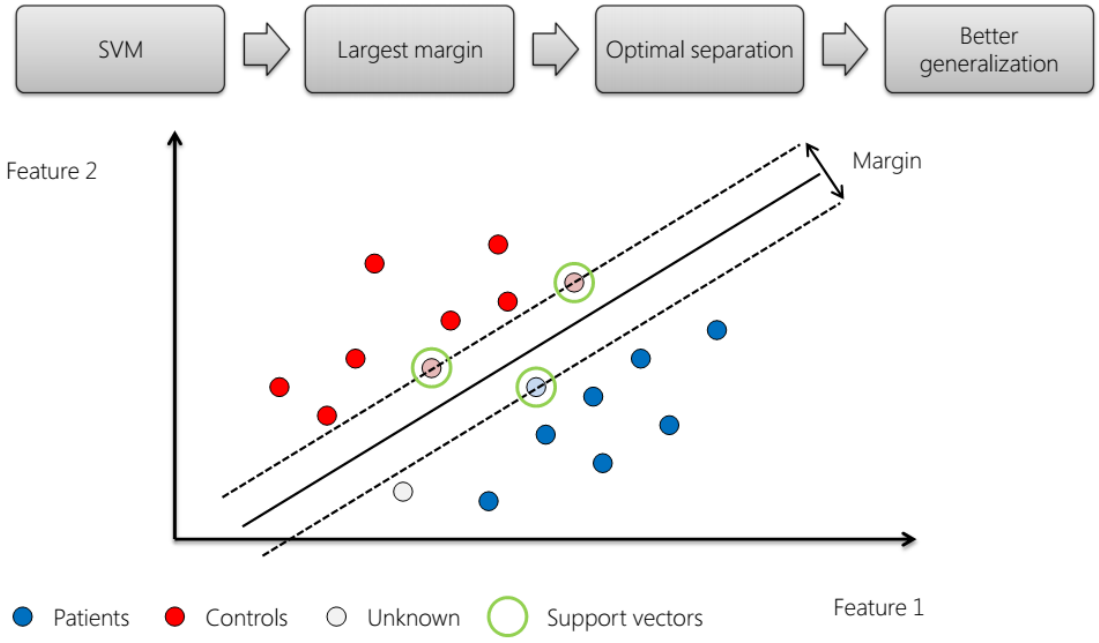


Figure 3.4: Schematic representation of a binary SVM classifier. In this example two groups of subjects (two classes, patients and controls) can be separated based in two distinct features of each group (feature 1 and feature 2). The support vectors mark the largest margin with the optimal separation between groups. The largest margin ensures the maximal generalization capacity of the classifier regarding new data (unknown subject). If more features are available a higher dimensional space is used. Adapted from [113]

Decision function

Given a dataset D constituted by the brain scans from the groups we want to discriminate and considering the classification between MS and controls as an example,

$$D = \{x_i, y_i\}, i = 1 \dots N$$

where x_i represents the features being evaluated (voxels) and y_i the labels (controls and MS patients), a feature is categorized as part of a class by the decision function or classifier model:

$$f(x_i) = y_i \qquad \text{(Equation 3.4)}$$

For all sample data x, SVM should find weights (w) such that the data points will be separated according to a decision rule:

$$y_i = w^T x_i + b_0 \qquad \text{(Equation 3.5)}$$

In our case, y_i represents our decision value, i.e. data to be classified (T1-MPRAGE/T2-FLAIR images to be labelled as controls or MS patients), x represents the features we want to evaluate, i.e. GM or WM volume in each voxel and w represent the parameters to be learned, the weights attributed to each voxel. In this analysis each brain image is assigned to MS or control group according to the decision value [112]–[114]

High Dimensional spaces

Real life classification problems are usually complex and cannot be demonstrated by a linear relation. A great advantage of the SVM is to overcome this issue by mapping the original feature space into a higher dimensional space and by finding a linear boundary through decision hyperplanes. Then, it returns to the original input space and draws the obtained solution of the classification problem that was learned. This method is called "kernel trick"[113]. Schematic representation of the "kernel trick" is illustrated in Figure 3.5.

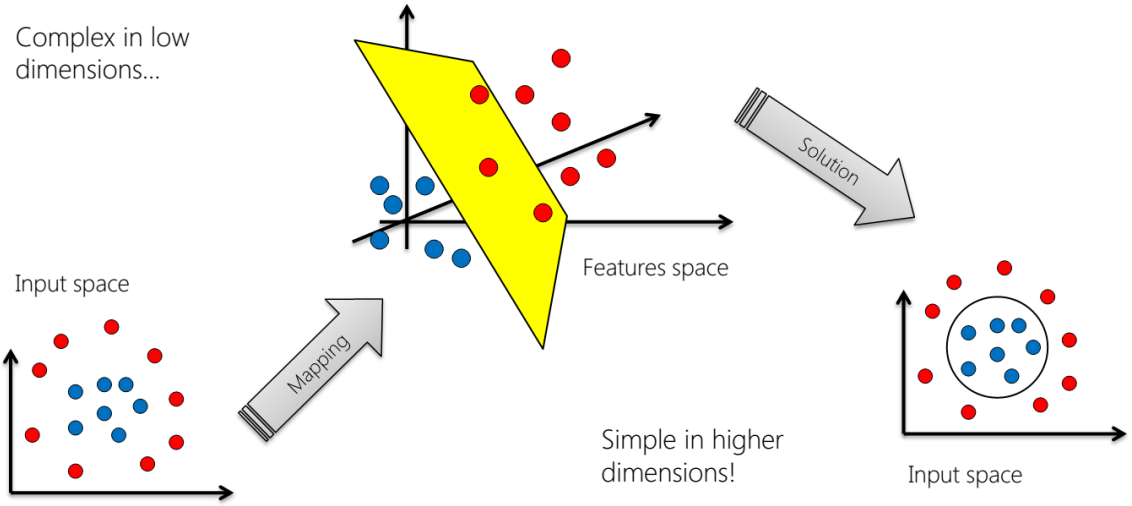


Figure 3.5: Schematic representation of the kernel trick. SVM algorithm maps the input feature space into a higher-dimensional space, in which the decision boundary is linear Then, it returns to the original input space and draws the solution of the classification problem that was learned. Adapted from [113]

In this higher dimensional space, two parallel hyperplanes are constructed on each side of a hyperplane that optimally separates the data. The separating hyperplane is

the hyperplane that represents the distance between the two parallel hyperplanes, built upon the support vectors. This distance, in its turn, represents how well the SVM could distinguish the groups. Subjects who are relatively close to this frontier are subjects that were more difficult to be assigned to a class. Therefore, it can be assumed the larger the distance between these parallel hyperplanes the better is the performance (the higher is the accuracy of the classifier).
For a better understanding of the mathematical functioning of the SVM classifier, assuming the two classes are represented by 1 and -1,

$$y = 1 \; or \; y = -1 \quad \text{(Equation 3.6 and Equation 3.7)}$$

and if

$$w^T x_i + b_0 = 0 \quad \text{(Equation 3.8)}$$

it means no discrimination was achieved for this example. In other words, the classifier could not place the example on any side of the separating hyperplane. This equation (3.8) represents the separating hyperplane, where no difference between the subjects exists. If the classifier can place a given unseen example at a distance from the separating hyperplane, it means the example will be assigned the class corresponding to the side of the hyperplane in which it falls. If the distance from the example is negative (whatever value), the example will be classified as belonging to class -1, whereas if the distance is positive (whatever value) the example will be classified as belonging to class 1. The higher the distance from the hyperplane, the easier it is to classify the example.

If all data examples, brain images of controls and MS patients in this case, would fall on the correct side of the hyperplane, we would have an ideal classifier, with an accuracy performance of 100 %. Accuracy is a metric for the ability of the classifier to predict the real labels of unseen examples. What happens in reality is that not all examples are correctly classified, there are controls being misclassified as MS patients and MS patients being misclassified as controls. From here, measures of performance such as sensitivity and specificity can be estimated [115], [116].

Cross-Validation

To be able to perform this assignment, SVM as a supervised method, needs to learn this function from data for which the labels are known – training data. Then the

function is applied to unseen brain images – test data - without knowing its label, so that it can test the learned function for its accuracy. To accomplish this, part of the dataset is given to the SVM algorithm along with the labels' assignment (classes to which every example belongs to). This is the training set. From the training set the SVM will learn the decision boundary between the classes (labels). This is called the training phase. Then, after the training phase the remaining examples, which constitute the test set, are given to the SVM without the labels' to be assigned to one of the classes.

The ideal would be to train the classifier with as much data as possible but the size of the available data is typically limited. The goal is to obtain a classifier model as general as possible, thus able to label any future image belonging to test set correctly. Using just a few examples for training does not seem a good approach, as few brain scans do not provide a representative amount of all possible patterns to discriminate groups. On the other hand, training with all the data available causes the problem of overfitting, which means the classifiers learns a very good relation between the features and the labels for that particular set of data but it will not be able to generalize it and accurately classify new unseen (potentially different) data. To optimize the data exploration, to increase the size of the test set and to avoid overfitting (loss of generalization ability of the model), a cross-validation scheme is commonly used. The idea of cross-validation is to determine a scheme of partition of the whole data set into training and test sets that tells the algorithm what data are going to be the training set (and consequently the remaining data is the test set), in which the classifier will learn the decision function**.** There can be several partition schemes, and cross-validation means that every split set of data is both used as training and test set in different iteration. The simplest one would be to split the dataset in two, so that half the data is used as the training set and the other half is used as the test set in one iteration, and in a second iteration the sets are used the other way around. However, in order to optimize the use of often limited amount of data, a cross-validation scheme is used in which larger parts of data are used for training and fewer for testing in each iteration. This could be e.g. splitting the data set in 5 and use 4 parts for training and 1 part for testing, with different parts being used for training and testing in a number of iterations that allow every example f data to be used at least once for testing, without ever being used for training and

testing in the same iteration. This is called *k*-fold cross-validation, where *k* is the number of parts into which the dataset is divided. Usually it is used *k*=10 or *k*=5, corresponding to leaving out 10% or 20% of the data on each testing fold [64], [115]. The training phase will be performed with *k*-1 subsets and the test phase test with the remaining unseen set. This is repeated *k* times and the classification results are averaged to obtain performance measures of the classifier as accuracy, sensitivity and specificity.

In neuroimaging, a context in which the number of subjects is usually low, a more extreme cross-validation is often used, a leave one out cross-validation (LOO-CV) or, better, a leave one per group out cross-validation (LOGO-CV). This means that the dataset containing *N* subjects of each class is split in *N* folds, such that each fold contains only one subject of each class. The training and testing will be performed *N* times, each time leaving a different fold out for testing, while training on all the remaining folders. This scheme of LOGO-CV follows these steps:

- leave one subject per group out, train on the remaining ones (training set), make a prediction for these subjects (test)
- repeat the previous step for each pair of subjects (in case it is a binary classification of two groups), leaving in each iteration one different subject per group out.
- compute the accuracy, sensitivity and specificity of the predictions made for all the iterations.

The cross-validation scheme we used is the LOGO-CV because it allows to maximize the amount of data used for training while using all the data for testing as well, while maintaining the balance between classes in training/test sets.

Performance

The most commonly used measures to evaluate the performance of a classifier are sensitivity, specificity, and accuracy. When classifying images of the test set, any given image can be classified correctly or incorrectly. Considering two classes with labels 1 and -1, we can refer to the classes as positive and negative. Typically, in clinical context, the group of patients is considered as the positive class and the controls as the negative one. The following cases might occur during automatic classification:

TP = true positives = subjects who were correctly labelled as MS patients;
TN = true negatives = subjects who were correctly labelled as controls;
FN = false negatives = subjects who were mislabelled as controls;
FP = false positives = subjects who were mislabelled as MS patients.

Sensitivity tell us the proportion of patients, among all the patients' samples, correctly classified.

$$Sensitivity = \frac{TP}{TP+FN} \quad \text{(Equation 3.9)}$$

Specificity tell us the proportion of controls, among all the controls' samples, correctly classified.

$$Specificity = \frac{TN}{TN+FP} \quad \text{(Equation 3.10)}$$

Accuracy is an estimative of the overall classifier performance. It translates the proportion of controls and patients correctly classified.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{(Equation 3.11)}$$

It is considered that the classification algorithm has learned properly the relation between the features that describe the data and the labels when: the performance measures significantly exceed a level considered as reasonable (it depends on the context of the problem); and if it is significantly higher than chance level (50 % in the case of a binary classifier). In such case, the null hypothesis that there is no discriminative information in the data (no difference between the classes) is rejected [64], [115].

Statistical significance of the accuracy value

In order to evaluate if the classifier was sufficient to reject the null hypothesis, the classification results were validated by a permutation test. A permutation test is used to obtain a level of statistical significance, the *p*-value. The *p*-value gives the probability of classification performance occurs given that the null hypothesis is true. As such, the smaller the *p*-value value, the more relevant are the accuracy, sensitivity and specificity classification values[104], [111], [115] .

A permutation test repeats the classification many times, each time shuffling the labels assigned to the samples. This way any true relation between the data and the labels is destroyed. The classification performance *p*-value was then computed as the proportion of times the classification performance of randomly assigned labels was equal or higher than the one obtained with true labels. Essentially, the permutation test procedure measures how likely the observed accuracy would be obtained by chance.

The same test was repeated 1000 times (cross validation LOGO-CV) with differently randomized labels each time. At this point, the *p*-value can be computed as low as $1/1000 = 0.0001$.

3.8.2 Implementation

PRoNTo software has five main analysis modules: dataset specification, feature set selection, model specification, model estimation and weights computation. It receives as input any NIfTI images [117].

In dataset specification, the brain scans were introduced with the features to be analysed (e.g. GM segments). Each image was assigned to the corresponding group (e.g. controls or MS patients). In this module a first-level mask was also entered, which was used to optimise the feature set preparation, by discarding all uninteresting features, such as voxels outside the brain [117] .

In feature set preparation, the entered images were selected, and the program computed both the feature matrix to be written into an appropriate format and the linear kernel (dot product of the images, representing a similarity measure). The result was used as input to the classification algorithm. Any operations on the features were performed using the original features matrix to compute a new kernel matrix.

In model specification, the classes to be discriminated, the algorithm to be used (in this case a SVM), and the cross-validation scheme (in this case a LOGO-CV) were defined.

Finally, the model was estimated, and the performance values were obtained. From this step, results outputs as the histogram of the decision function values, the ROC/area under the curve plot, the model predictions and the confusion matrix,

could be extracted. These outputs help to understand the behaviour of SVM and to interpret the performance of the classifier.

In addition, PRoNTo provides the displaying of the weights in the original features (voxels) space. Weights are displayed into maps that contain at each voxel the corresponding weight of the linear model, which relates to how much this particular voxel contributed to the classification.

MS patients vs Controls Classification models

Depending on the features and type of images being analysed, 6 SVM classifiers to discriminate MS patients and controls were built:

- One SVM classifier with GM volumes from no filled T1-MPRAGE images;
- One SVM classifier with WM volumes from no filled T1-MPRAGE images;
- One SVM classifier with GM volumes from filled T1-MPRAGE images;
- One SVM classifier with WM volumes from filled T1-MPRAGE images;
- One SVM classifier with GM volumes from T2-FLAIR images;
- One SVM classifier with WM volumes from T2-FLAIR images.

MS Subtypes Classification models

The available dataset includes only eight patients diagnosed with secondary progressive MS (SPMS patients), which is a very small sample to train the classifier. In order to achieve a reasonable classification performance, 100 different SVM models were built with the eight SPMS patients and with different sets of 8 participants diagnosed with relapsing-remitting MS (RRMS patients). For each model, the eight RRMS patients were randomly selected (and different in each time).

However, SVM classification of MS subtypes with T2 scans accounted with even less subjects, since images of certain participants had to be excluded due to excessive heterogeneities.

To discriminate SPMS patients and RRMS patients, four SVM classifiers were built:

- One SVM classifier with GM volumes from filled T1-MPRAGE images;

- One SVM classifier with WM volumes from no filled T1-MPRAGE images;
- One SVM classifier with GM volumes from T2-FLAIR images;
- One SVM classifier with WM volumes from T2-FLAIR image.

4 Results and Discussion

4.1 Global Measures

This section presents how global measures such as total intracranial volume (TIV), total grey matter (GM) volume and total white matter (WM) volume are distributed in controls and MS patients. Number of lesions and total lesion volume (TLV) in both groups are also illustrated. Graphics of these measures are illustrated in Figures 4.1 and 4.2.

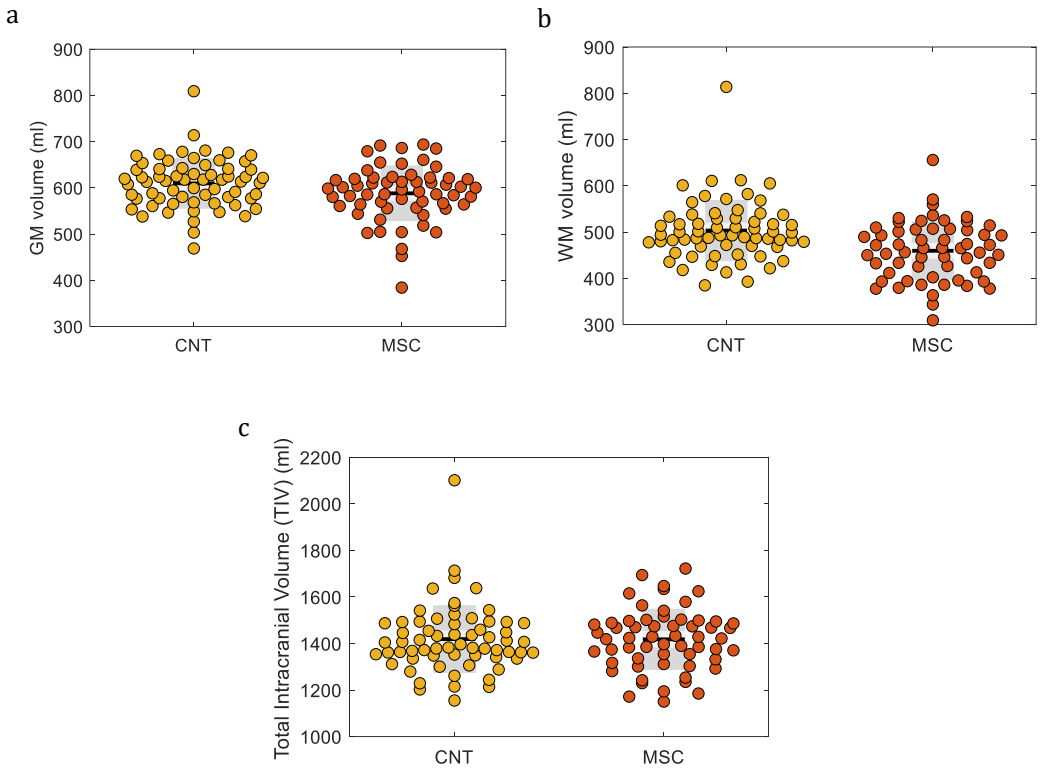


Figure 4.1: Global Volume Measures from filled T1-MPRAGE of controls and MS patients. CNT stands for controls and MSC stands for MS patients. Yellow represents the global measures values of control group and red represents to the global measures values of MS patients group **a)** GM distribution values. **b)** WM distribution values **c)** Total Intracranial Volume distribution values.

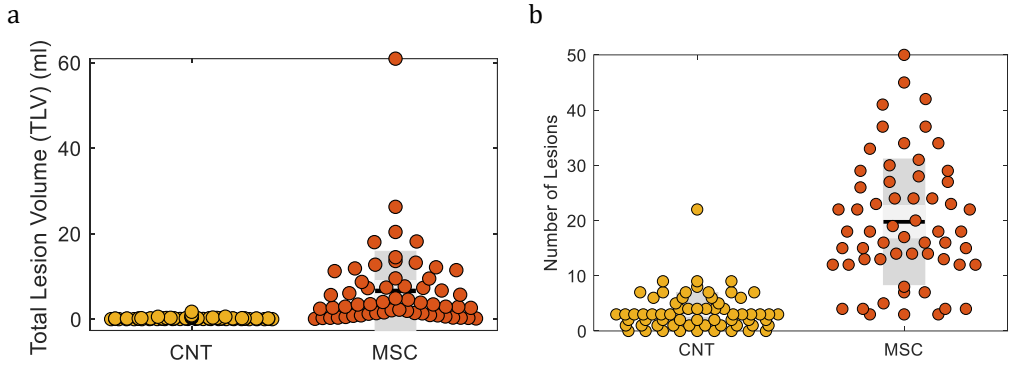


Figure 4.2: Lesion measures from filled T1-MPRAGE of controls and MS patients. CNT stands for control group and MSC stands for MS patients group. Yellow represents the global measures values of control group and red represents to the global measures values of MS patients group **a)** Total Lesion Volume distribution values. **b)** Number of Lesions distribution values.

Discussion

The overlap in the distribution of global volume measures suggests that globally there are no significant differences between the brains of MS patients and controls. From these results it is assumed that these measures should not be considered as structural differences between groups. However, we can only guarantee that there are no significant differences if the *t*-value and *p*-value, resulting from a t-test performed between these measures of each group, indicate that there are in fact no such differences. The distribution of the number of lesions and their total volume illustrate how these features can indeed be clinically important biomarkers of MS. As expected, images from controls revealed a very low number of lesions and volume when compared to MS patients, as this is a well-known characteristic of the disease.

In what concern global surface measures of cortical thickness and gyrification index, CAT12 does not provide an option to quantify these measures globally. It is debatable if a global value of cortical thickness or gyrification is meaningful, as these can greatly vary across subjects. Therefore, the distribution of these measures among groups is not shown here.

4.2 Voxel-Based Morphometry Results

4.2.1 Sample homogeneity

CAT12 manual recommends using the unsmoothed segmentations to check sample homogeneity, which provide more anatomic detail. Boxplots showing homogeneity values of normalized and modulated segments of GM and WM from no filled T1-MPRAGE, filled T1-MPRAGE and T2-FLAIR images for both groups can be found as supplementary material to this dissertation (annex B).

From all the boxplots resulting from the tests performed in the three types of images (no filled/filled T1 images and T2 images), in annex, it is evident which are the images that have a mean correlation below two standard deviations. Yet, in the case of T1-weighted images, these have high correlation values, so there was no need to exclude any subjects from the subsequent analyses. T1-weighted images showed correlation values close to one, meaning that these images are globally like each other in terms of intensity homogeneity. However, there are T2-weighted images that showed very low correlation values. When visually inspected, these images seemed abnormal when compared to the remaining images. Their heterogeneities were probably due to errors during the segmentation step. As such, images with excessive deviation from the mean were excluded from analyses (VBM and classification), as follows:

- 3 normalized and modulated GM segments from MS patients group;
- 5 normalized and modulated GM segments from controls group;
- 6 normalized and modulated WM segments from MS patients group;
- 5 normalized and modulated WM segments from controls group.

Thus, the dataset with T1-weighted images remained complete, with 59 MS patients and 64 controls, while the dataset with T2-weighted images was reduced:

- 56 normalized and modulated GM segments from MS patients' group;
- 59 normalized and modulated GM segments from controls group;
- 53 normalized and modulated WM segments from MS patients' group;
- 59 normalized and modulated WM segments from controls group.

4.2.2 Statistical Parametric Maps

The statistical parametric maps for VBM analyses of T1-weighted and T2-weighted GM and WM images were obtained considering differences to be significant at $p < 0.05$ (FWE corrected). Clusters of contiguous voxels with significant differences were overlaid onto brain templates and their coordinates identified the corresponding brain regions. These coordinates correspond to Montreal Neurological Institute (MNI) space and to Talairach (TAL) space [118]. Tables with coordinates and corresponding name of regions are available as supplementary material to this dissertation (annex B, Tables B.1, B.2, B.3, B.4, B.5, B.6). In all VBM analyses with GM and WM we tested the contrast controls > MS patients, i.e. the regions where there was more GM/WM volume in controls. The defined contrast would reveal the opposite effect (higher GM/WM volume in MS patients) by significant negative *t*-values. In any case, the opposite effect (higher GM/WM volume in MS patients) would be revealed by significant negative t-values with the defined contrast

Grey Matter results

Statistical parametric maps from GM VBM analysis yielded several significant clusters ($p < 0.05$, FWE corrected) showing GM atrophy in MS patients. These are illustrated in Figure 4.3.

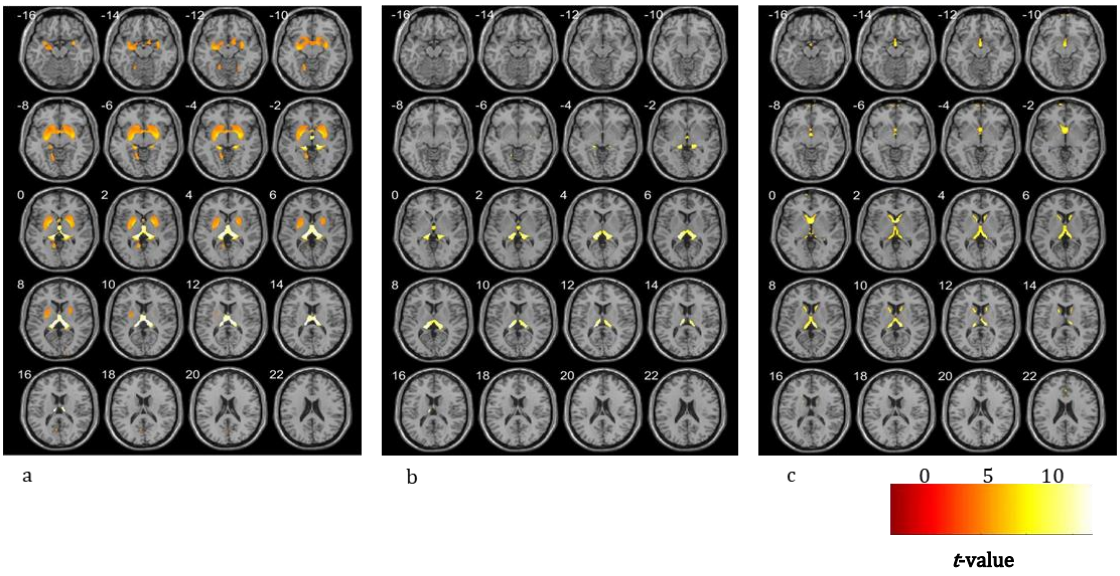


Figure 4.3: Results of VBM analysis of GM volume (controls > MS). Results are presented at a voxel-level *p*-value < 0.05, FWE corrected. Color bar scale represents *t*-values. a) Results of VBM analysis of GM segments obtained from no filled T1-MPRAGE images. **b)** Results of VBM analysis of GM segments obtained from filled T1-MPRAGE images. **c)** Results of VBM analysis of GM segments obtained from T2-FLAIR images. All maps are overlaid onto a representative MR image in standard space in SPM.

Volume differences in GM obtained from no filled T1-MPRAGE images showed evidence of GM atrophy in MS patients. This was found for clusters in right and left pulvinar of the thalamus, right and left lentiform nucleus, right cingulate gyrus, left medial frontal gyrus, left lingual gyrus, right middle frontal gyrus, right rectal gyrus, and in right fusiform gyrus.

Volume differences in GM obtained from filled T1-MPRAGE images showed evidence of GM atrophy in MS patients in clusters corresponding to regions in right and left pulvinar of the thalamus, right thalamus, right and left culmen (cerebellum), left lingual gyrus, and right lentiform nucleus.

Maps of volume differences in GM obtained from T2-FLAIR images showed evidence of GM atrophy in MS patients in clusters in the right pulvinar of the thalamus, right caudate body, left medial frontal gyrus, right superior frontal gyrus, left paracentral

lobule, right middle frontal gyrus, right inferior temporal gyrus, and right fusiform gyrus.

Discussion

These results show that GM atrophy is indeed a biomarker of MS, being present in a considerable number of brain regions. The results regarding the location of GM atrophy are in accordance with previous literature of VBM studies in MS [32][77]. The only common region resultant from each of the three analyses (different types of brain images) is the thalamus, which is the most significantly affected region. It is highly associated with cognitive deficits, which is commonly a disease effect [119].

Most of the differences that appear in no filled T1 images are present in T2 images, which were not filled either. It is known that the lesion filling in T1-MPRAGE images provides more reliable VBM results [75]–[77], [82]. As such, differences between groups in no filled T1 images and T2 images, which do not appear in maps resulting from the analysis of GM segments of T1 images after lesion filling, might be less reliable. On the other hand, in theory T2-weighted imaging might be more sensitive to microscopic neurodegenerative processes than T1-weighted imaging. This might also explain the differences visible in T2 images that did not appear in T1 images. These differences correspond to superior frontal gyrus and right inferior temporal gyrus. Therefore, these results should be considered to further investigation of brain regions revealing atrophy.

White Matter results

Statistical parametric maps from WM VBM analysis yielded several significant clusters ($p < 0.05$, FWE corrected) showing WM atrophy in MS patients. These are illustrated in Figure 4.4.

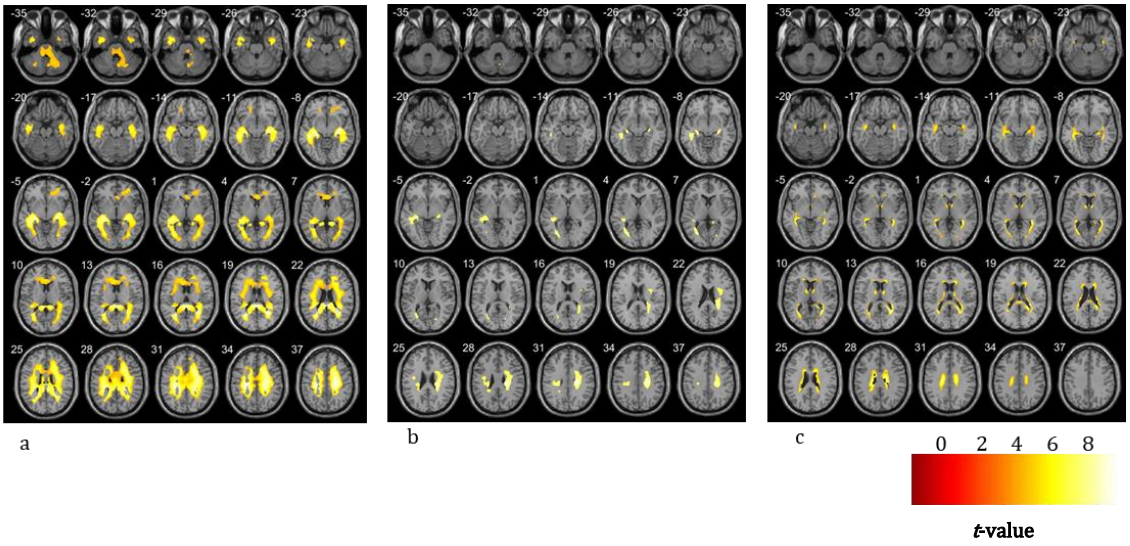


Figure 4.4: Results of VBM analysis of WM volume (controls > MS). Results are presented at a voxel-level *p*-value < 0.05, FWE corrected. Color bar scale represents *t*-values. a) Results of VBM analysis of WM segments obtained from no filled T1-MPRAGE images. **b)** Results of VBM analysis of WM segments obtained from filled T1-MPRAGE images. **c)** Results of VBM analysis of WM segments obtained from T2-FLAIR images. All maps are overlaid onto a representative MR image in standard space in SPM.

Volume differences in WM obtained from no filled T1-MPRAGE images showed evidence of WM atrophy in MS patients in clusters corresponding to regions in right and left parahippocampal gyrus and regions in frontal and temporal lobe.

Volume differences in WM obtained from filled T1-MPRAGE images showed evidence of WM atrophy in MS patients in clusters corresponding to regions in temporal and frontal lobes, in left hippocampus, in left precentral gyrus, in right cuneus.

Volume differences in WM obtained from T2-FLAIR images showed evidence of WM atrophy in clusters corresponding to regions in temporal lobe and brainstem.

Discussion

Despite the lower sensitivity for WM analysis, were found significant volume differences between patients and controls in the 3 VBM analyses performed. The VBM results show significant differences in hippocampus and parahippocampal gyrus, which are regions associated to memory function. These results are in accordance with a previous study regarding atrophy in MS, which has correlated WM atrophy with a negative impact in working memory tasks [120].

Even though WM atrophy is not commonly studied, the results here presented add to the MS understanding by demonstrating that WM atrophy is in fact a feature in this disease. However, there are methods more appropriate to quantify WM atrophy, as for example diffusion tensor imaging [21].

4.2.3 Descriptive Univariate Plots Results

The distribution of GM and WM volume in the brain regions that have shown the most significant differences between patients and controls in VBM analyses are illustrated below. GM volumes distribution from T1 images and T2 images are illustrated in Figures 4.5 and 4.6, respectively. Whereas, WM volumes distribution from T1 images and T2 images are illustrated in Figures 4.7 and 4.8, respectively.

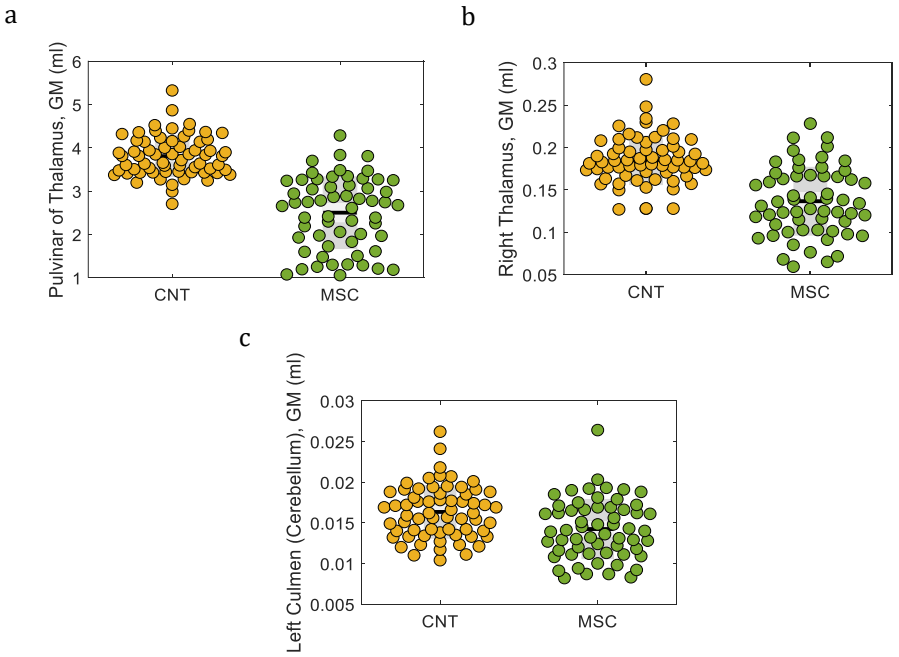


Figure 4.5: GM volumes extracted from filled T1-MPRAGE images in the regions that contain the most significant differences between controls and MS patients. CNT stands for control group and MSC stands for MS patients' group. Yellow represents the GM volume values of control group and green represents the GM volume values of MS patients group **a)** GM volume of the pulvinar of the thalamus. **b)** GM volume of the right thalamus. **c)** GM volume of the left culmen (cerebellum).

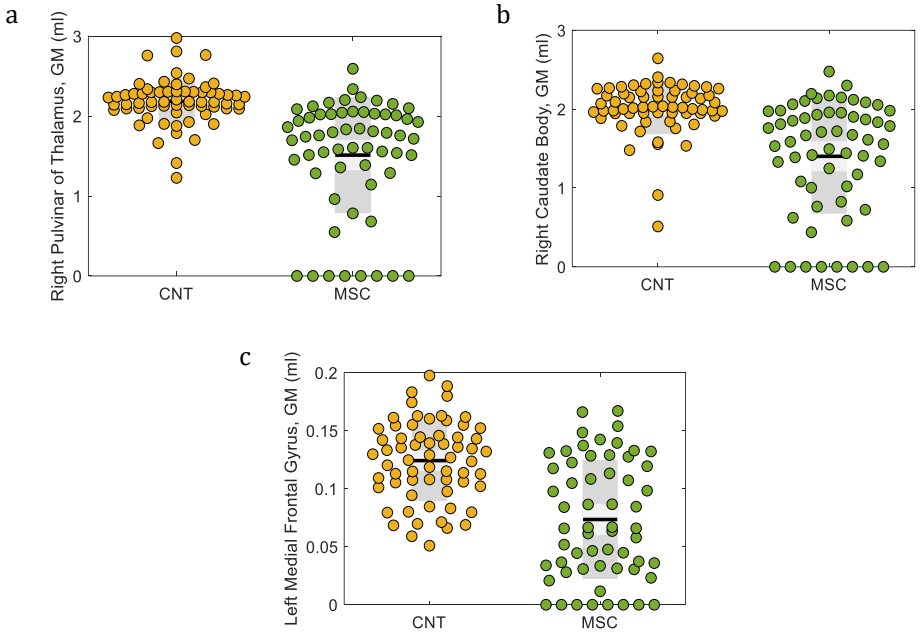


Figure 4.6: GM volumes extracted from T2-FLAIR images in the regions that contain the most significant differences between controls and MS patients. CNT stands for control group and MSC stands for MS patients' group. Yellow represents the GM volume values of control group and green represents the GM volume values of MS patients group **a)** GM volume of the right pulvinar of the thalamus. **b)** GM volume of the right caudate body. **c)** GM volume of the left medial frontal gyrus.

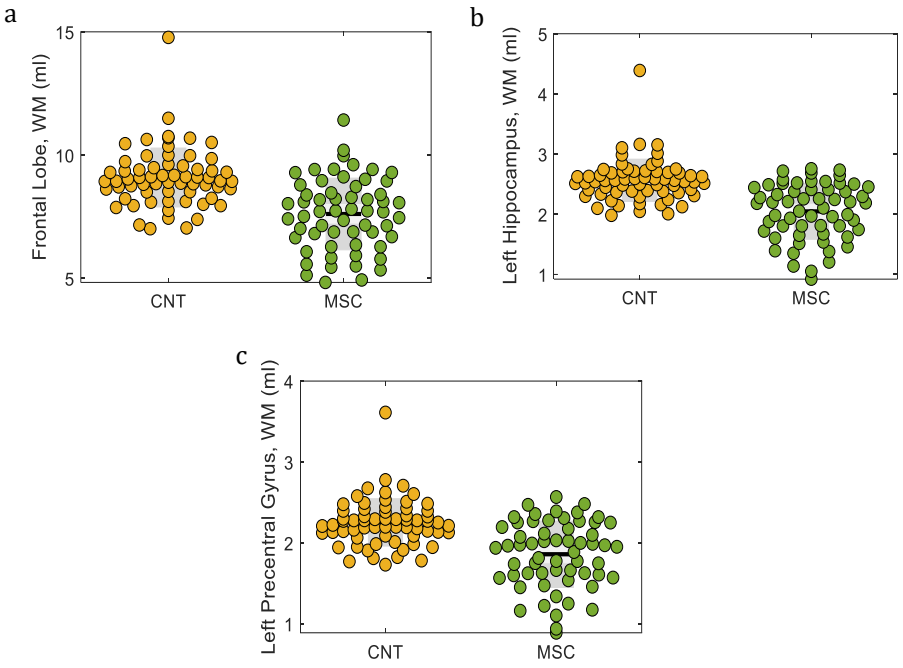


Figure 4.7: WM volumes extracted from filled T1-MPRAGE images in the regions that contain the most significant differences between controls and MS patients. CNT stands for control group and MSC stands for MS patients' group. Yellow represents the WM volume values of control group and green represents the WM volume values of MS patients group **a)** WM volume of frontal lobe. **b)** WM volume of the left hippocampus. **c)** WM volume of the left precentral gyrus.

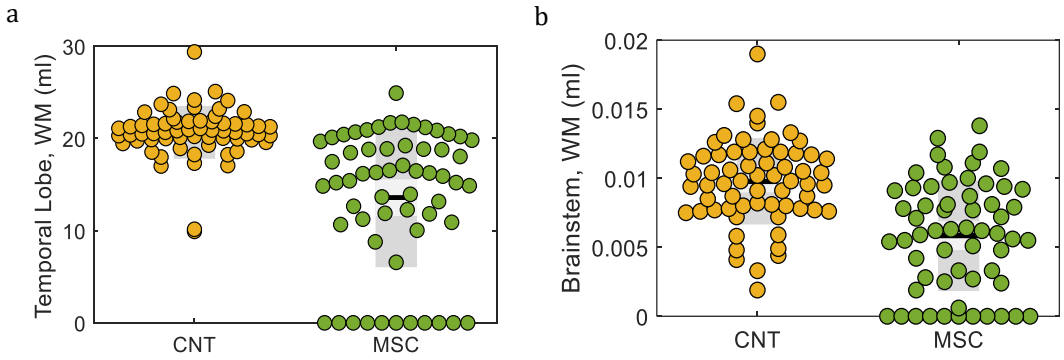


Figure 4.8: WM values from T2-FLAIR images of the regions of that contain the most significant differences between controls and MS patients. CNT stands for control group and MSC stands for MS patients' group. Yellow represents the GM volume values of control group and green represents the GM volume values of MS patients' group. **a)** WM volume of temporal lobe. **b)** WM volume of brainstem.

Discussion

These results illustrate how both GM and WM atrophy is present in MS patients. However, not all patients exhibited brain atrophy. There are MS patients with regional GM/WM volume similar to controls. This suggests that atrophy measures in addition to be investigated as potential MS biomarkers, might also be explored to personalized patient characterization.

General VBM Discussion

In all VBM analyses the contrast used was controls > MS patients. The inverse contrast, MS patients > controls, did not reveal any region where patients have significantly higher tissue volume than controls, as expected. As there is atrophy in MS patients, an increase of volume relatively to controls is not expected in any brain region. However, without correction for multiple comparisons, a few regions survive to this contrast. This was probably due to physiognomy differences in brain or hyperintensity of images (data not shown).

4.3 Surface-Based Morphometry Results

4.3.1 Sample homogeneity

Despite the better performance of this function when using unsmoothed images, in the case of surface data this analysis was performed in smoothed images. This was because the extraction of cortical thickness and gyrification index and preprocessing were entirely done in one single step. Therefore, it was not possible to obtain unsmoothed cortical thickness and gyrification index data.

Homogeneity values of smoothed images of cortical thickness and gyrification index of no filled T1-MPRAGE and filled T1-MPRAGE images for both groups can be found as supplementary material of this dissertation (annex B). No subjects were excluded from the analyses as they all have shown high values of correlation and all close to one.

4.3.2 Statistical Parametric Maps

Clusters of contiguous vertexes with significant cortical thickness and gyrification index differences were overlaid onto brain templates for visualization. Then, their coordinates in standard space identified the corresponding brain regions. Tables with coordinates and corresponding names of the regions are available as supplementary material to this dissertation (annex C, tables C.1, C.2, C.3, C.4, C.5, C.6).

Cortical thickness results

SBM analysis of cortical thickness values extracted from no filled T1-MPRAGE and filled T1-MPRAGE images yielded several significant clusters ($p < 0.05$, FWE corrected). Differences were tested with the contrast controls > MS patients and the results are shown in Figure 4.9.

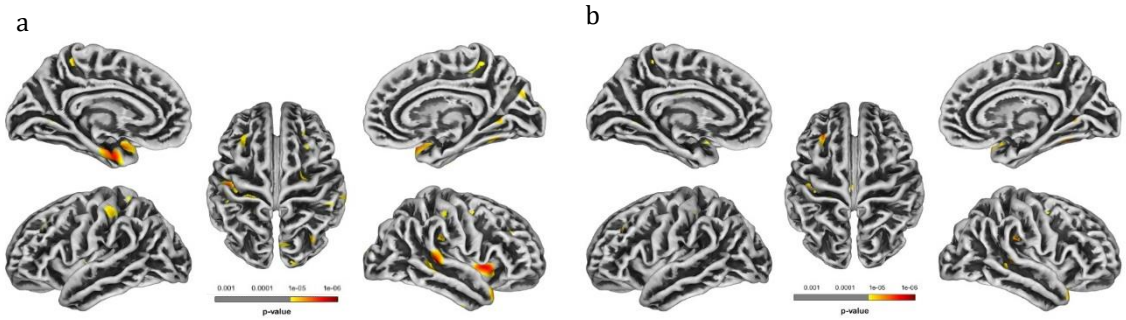


Figure 4.9: Results of SBM analysis of cortical thickness (controls > MS) showing a decrease of cortical thickness in MS patients. Results are presented at a voxel-level p-value < 0.05, FWE corrected. Color bar scale represents log p-values. **a)** Results of SBM analysis of cortical thickness extracted from no filled T1-MPRAGE images. **b)** Results of SBM analysis of cortical thickness extracted from filled T1-MPRAGE images. Maps are overlaid onto a representative inflated brain in standard space in CAT12.

Maps of differences in **cortical thickness obtained from no filled T1-MPRAGE** images show a decrease of cortical thickness in MS patients, as compared to controls, in the following regions: right and left superior temporal gyrus, right and left inferior parietal lobule, right and left middle temporal gyrus, left parahippocampal gyrus, left postcentral gyrus, right and left precentral gyrus, right posterior cingulate, right posterior lobe, declive (cerebellum), right precuneus, right cuneus, left paracentral

lobule, right and left middle frontal gyrus, right superior frontal gyrus, right middle occipital gyrus, and left insula.

Maps of differences in **cortical thickness obtained from filled T1-MPRAGE** images show a decrease of cortical thickness in MS patients as compared to controls, in the following regions: right and left superior temporal gyrus, right posterior lobe, declive (cerebellum), left middle temporal gyrus, left middle frontal gyrus, right and left inferior parietal lobule, right posterior cingulate, left postcentral gyrus, left cingulate gyrus, left lingual gyrus, right and left paracentral lobule.

Discussion

SBM analysis revealed a substantial decrease of cortical thickness in MS patients, which firstly confirms that cortical thinning may be a promising neural marker for clinical trials and secondly stresses the need of investigate its relationship with the disease effects.

Most of the differences observed in measures of cortical thickness extracted from no filled T1 images are also found when using filled T1 images. The investigation of the mechanisms underlying the decrease of cortical thickness in these regions may help to understand the pathophysiology of the disease. However, there were also surface differences that did not survive after lesion filling and others that only appeared after lesion filling.

The symmetric contrast (MS > controls) did not reveal regions with significant differences (data not shown).

Gyrification index results

SBM analysis of gyrification index in no filled T1-MPRAGE and filled T1-MPRAGE yielded several significant clusters ($p < 0.001$, uncorrected). Differences were tested with both contrasts, controls > MS patients and MS patients > controls. Gyrification index maps showing these differences for both contrasts for no filled T1-MPRAGE and filled T1-MPRAGE are represented in Figure 4.10.

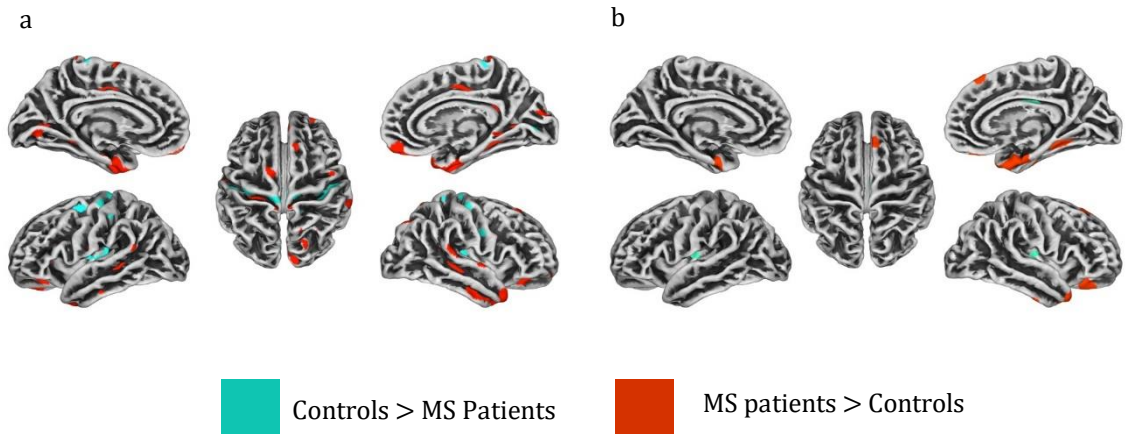


Figure 4.10: Results of SBM analysis of gyrification index. Results are presented at a voxel-level *p*-value < 0.001 (uncorrected). Blue and red indicates decreased (controls > MS patients) and increased gyrification index (MS patients > controls), respectively. a) Results of SBM analysis of gyrification index measures extracted from no filled T1-MPRAGE images. **b)** Results of SBM analysis of gyrification index measures extracted from filled T1-MPRAGE images. Maps are overlaid onto a

Maps of differences in gyrification index obtained from no filled T1-MPRAGE images show decreased gyrification index in MS patients in: right and left insula, right and left precentral gyrus, right and left postcentral gyrus, right and left superior temporal gyrus, right and left cuneus, left superior occipital gyrus, right medial frontal gyrus, and right cingulate gyrus.

Increased gyrification index in MS patients was found in: right and left inferior frontal gyrus, right and left precentral gyrus, right and left parahippocampal gyrus, right and left superior temporal gyrus, left inferior temporal gyrus, right subcallosal gyrus, right and left cuneus, left lingual gyrus, right and left cingulate gyrus, left rectal gyrus, left superior frontal gyrus, right paracentral lobule, right postcentral gyrus, right posterior cingulate, right middle frontal gyrus, and right precuneus.

Maps of differences in gyrification index obtained from filled T1-MPRAGE images show decreased gyrification index in MS patients in: right cingulate gyrus and in right and left insula cortex.

Increased gyrification index was found in: right parahippocampal gyrus, right orbital gyrus, right superior frontal gyrus, right uncus, and right inferior temporal gyrus.

Discussion

Since this is an exploratory study of a less commonly explored morphometric measure, we opted to use a more liberal threshold, $p < 0.001$ (uncorrected). This option does not allow a strong control of false positives and reports suggestive rather than significant results.

Gyrification maps showing differences between groups, in each contrast at a time, with the respective *p*-values are available as supplementary material of this dissertation (annex C, Figures C.5, C.6). These maps allow to inspect which regions are the most/less different between groups. The gyrification maps illustrated above are more informative in the sense that they allow to see the results of both contrasts in the same brain image. They allow seeing what regions have a higher/lower gyrification index in MS patients compared to healthy controls. However, they do not provide information about the *p*-values, and therefore does not allow to infer which are the most significant differences among all the differences.

Both decreased and increased gyrification index in MS patients was observed in different brain regions. Nevertheless, only increases of gyrification index cases survived the statistical correction for multiple comparisons, $p < 0.05$ (FWE corrected). Gyrification maps with FWE correction as well brain coordinates and corresponding name of the regions showing differences are available in annex C, Figure C.7.

Studies addressing correlations between surface measures and aging have suggested that increased gyrification index is linked to brain plasticity mechanisms to facilitate brain connectivity and functional development [48]. On the other hand, it is believed that MS patients have their brain connectivity altered [121]. As such, increased gyrification index in MS patients might be related with reflecting compensatory mechanisms.

4.3.3 Descriptive Univariate Plots Results

Descriptive univariate plots of the regions that show the most significance differences between groups resulting from the previously analyses in filled T1 images are represented in Figures 4.11, 4.12 and 4.13.

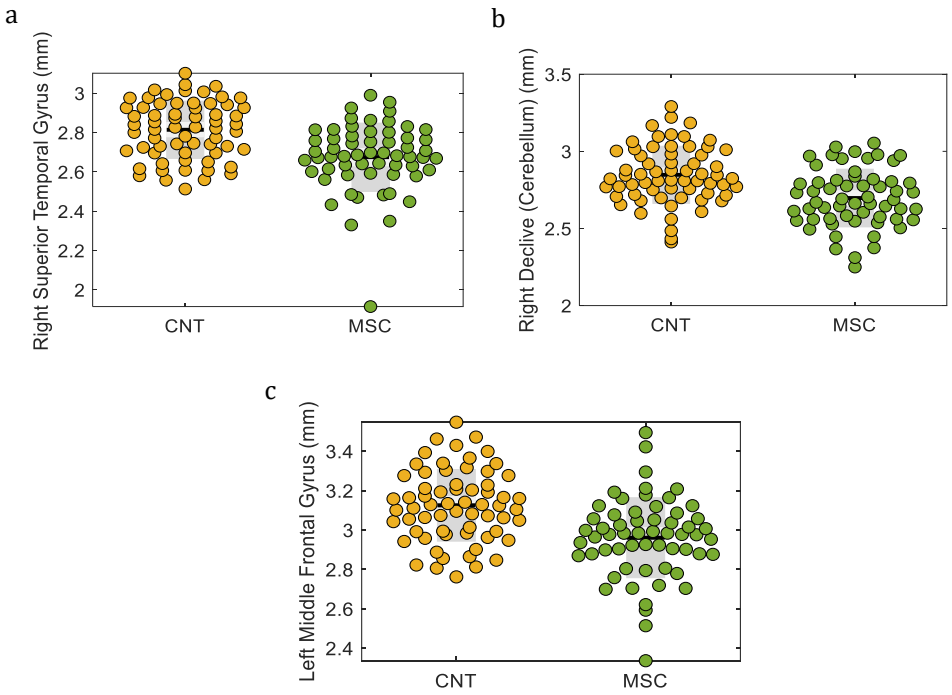


Figure 4.11: Distribution of cortical thickness values extracted from filled T1-MPRAGE images in the regions that contain the most significant decrease in MS patients. CNT stands for control group and MSC stands for MS patients' group. Yellow represents the cortical thickness values of control group and green represents the cortical thickness values of MS patients' group**. a)** cortical thickness of the right superior temporal gyrus. **b)** cortical thickness of the right declive (cerebellum) **c)** cortical thickness of the left middle frontal gyrus.

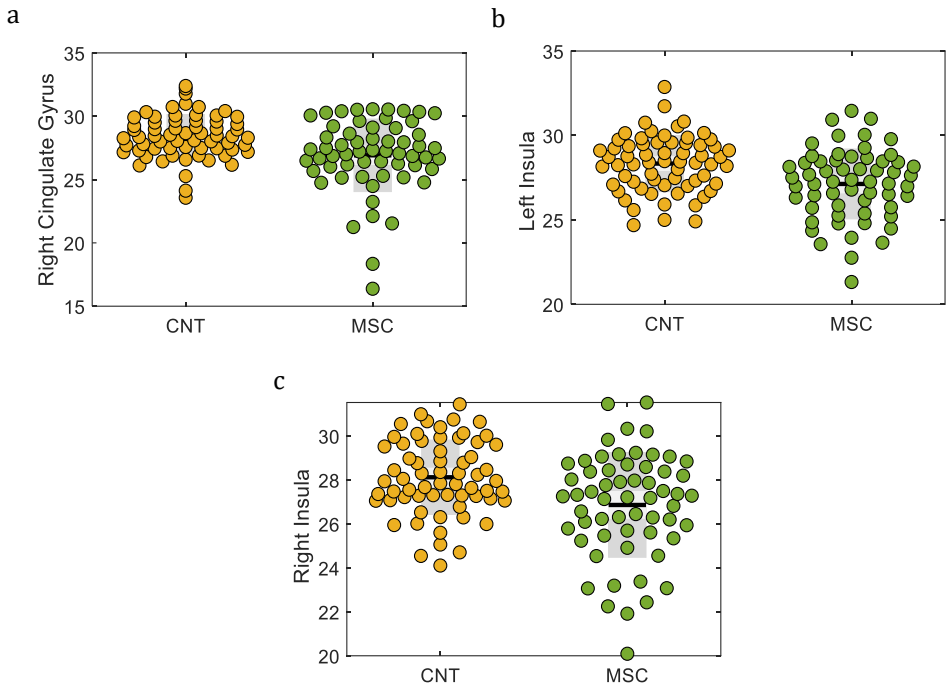


Figure 4.12: Distribution of gyrification index extracted from filled T1-MPRAGE images in regions that contain the most significant decrease in MS patients. CNT stands for control group and MSC stands for MS patients' group. Yellow represents the gyrification index values of control group and green represents the gyrification index values of MS patients' group. **a)** Gyrification index of the right cingulate gyrus. **b)** Gyrification index of the left insula **c)** Gyrification index of the right insula.

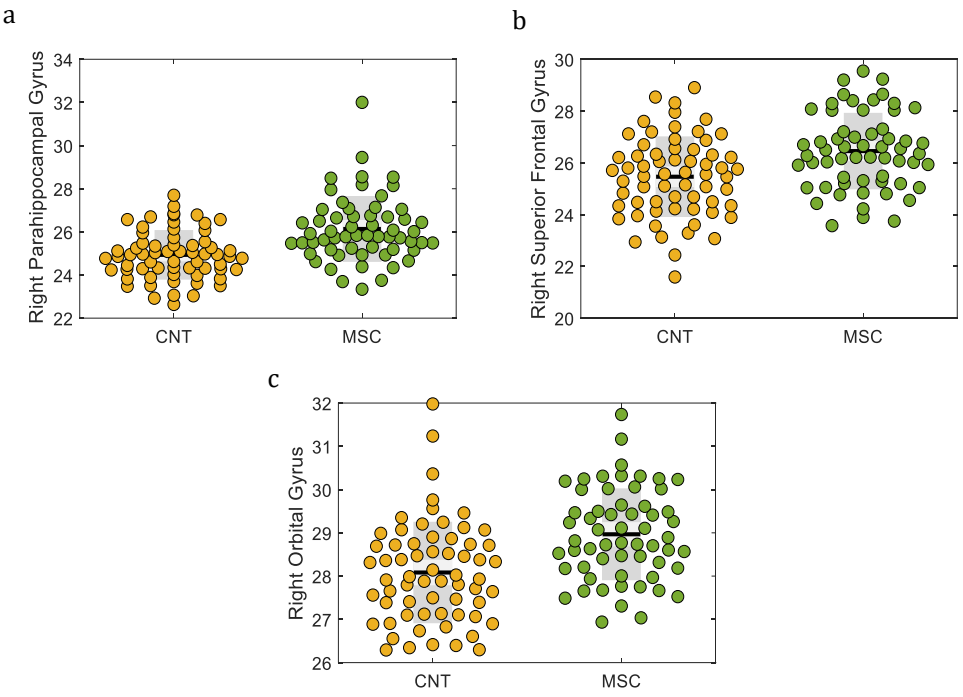


Figure 4.13: **Distribution of Gyrification Index extracted from filled T1-MPRAGE images in regions that contain the most significant increase in MS patients.** CNT stands for control group and MSC stands for MS patients' group. Yellow represents the gyrification index values of control group and green represents the gyrification index values of MS patients' group **a)** Gyrification index of right parahippocampal gyrus. **b)** Gyrification index of right superior frontal gyrus. **c)**. Gyrification index orbital gyrus

These results illustrate how cortical thickness and gyrification index are altered in MS and support the hypothesis that beyond regional volumetric differences of GM and WM volumes, there are surface regional differences between groups.

4.4 Support Vector Machine Classification Results

4.4.1 MS patients vs Controls Classification

Classification performance based on no filled T1-MPRAGE

SVM classification applied to GM segments extracted from no filled T1-MPRAGE images correctly classified 74.58 % (sensitivity) of patients and 98.44 % (specificity) of controls, achieving an overall classification balanced accuracy of 86.51 % ($p \leq 0.001$). In what concerns, WM segments, SVM correctly classified 67.80 % (sensitivity) of patients and 96.88 % (specificity) of controls achieving an overall classification balanced accuracy of 82.34 % ($p \leq 0.001$).

The confusion matrixes showing the true positives (TP), false negatives (FN), true negatives (TN) and false positives (FP) values for the classification based on GM and WM segments are illustrated in Figure 4.14. Plots showing the decision boundary and the prediction of each test example are illustrated in Figure 4.15.

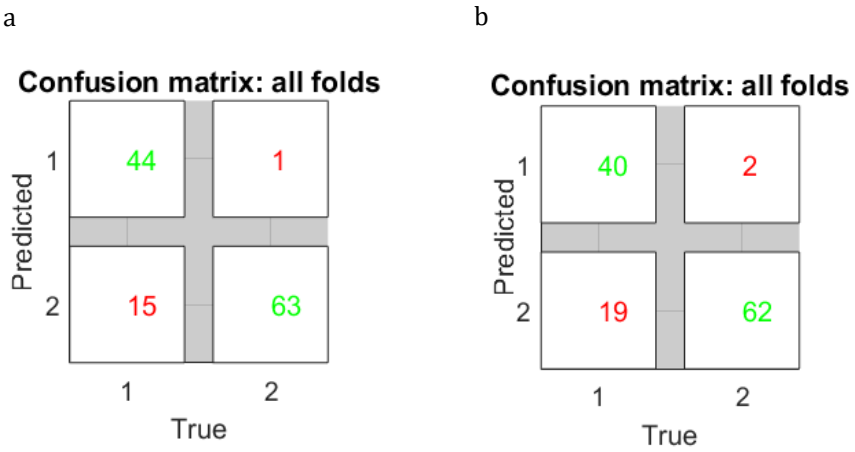


Figure 4.14: Confusion matrixes resulting from SVM classification of GM (a) and WM (b) segments extracted from no filled T1-MPRAGE images. True are the true labels of each example and predicted are the labels assigned by the SVM classifier. In these plots, 1 refers to MS patients' class and 2 refers to controls class. With the information of the first column we calculate sensitivity and with information of second column we calculate specificity.

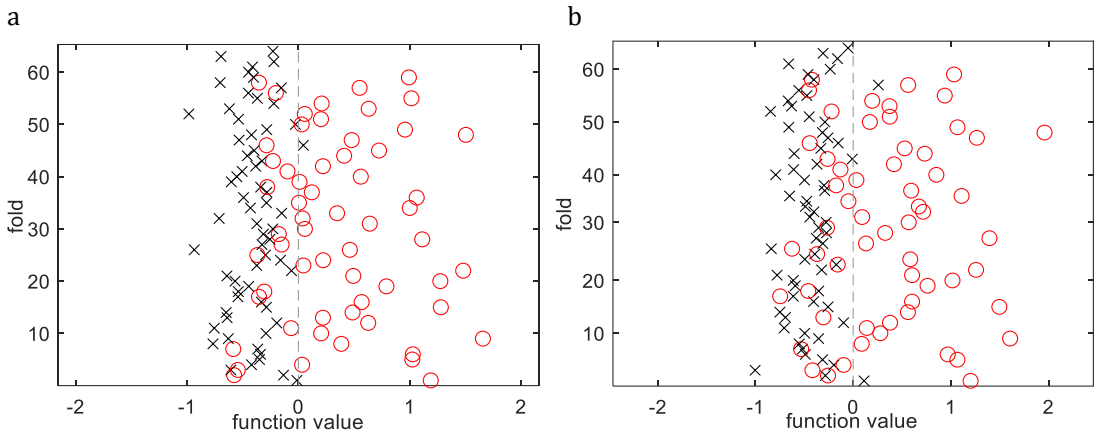


Figure 4.15: Predictions resulting from SVM classification of GM (a) and WM (b) segments extracted from no filled T1-MPRAGE images. Crosses represent the controls and circles represent the MS patients. The dashed line represents the decision boundary determined by the classifier during the training phase. Examples falling on the positive part of the decision space (function value) are classified as class 1 = MS patients, while examples falling on the negative half of the decision space are classified as class -1 = controls. We can observe that there are examples easily classified (very far from the decision boundary) while other are much closer to the decision boundary and a few are on the wrong side, ended up misclassified.

Classification performance based on Filled T1-MPRAGE

SVM classification applied to GM from filled T1-MPRAGE correctly classified 71.19 % (sensitivity) of patients and 98.44 % (specificity) of controls. The overall classification was performed with 84.81 % of accuracy ($p \leq 0.001$). In what concerns WM, SVM correctly classified 69.49 % of patients and 95.31 % of controls, achieving an overall classification of 82.40 % ($p \leq 0.001$).

Confusion matrix showing the TP, FN, TN and FP is illustrated in Figure 4.16. Predictions showing boundary decision and hyperplane dimension is illustrated in Figure 4.17.

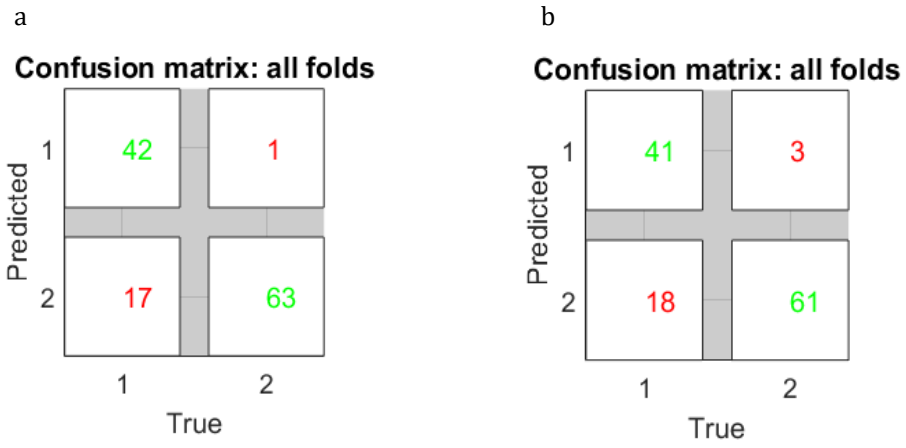


Figure 4.16: Confusion matrixes resulting from SVM classification of GM (a) and WM (b) segments extracted from filled T1-MPRAGE images. True are the true labels of each example and predicted are the labels assigned by the SVM classifier. In these plots, 1 refers to MS patients' class and 2 refers to controls class. With the information of the first column we calculate sensitivity and with information of second column we calculate specificity.

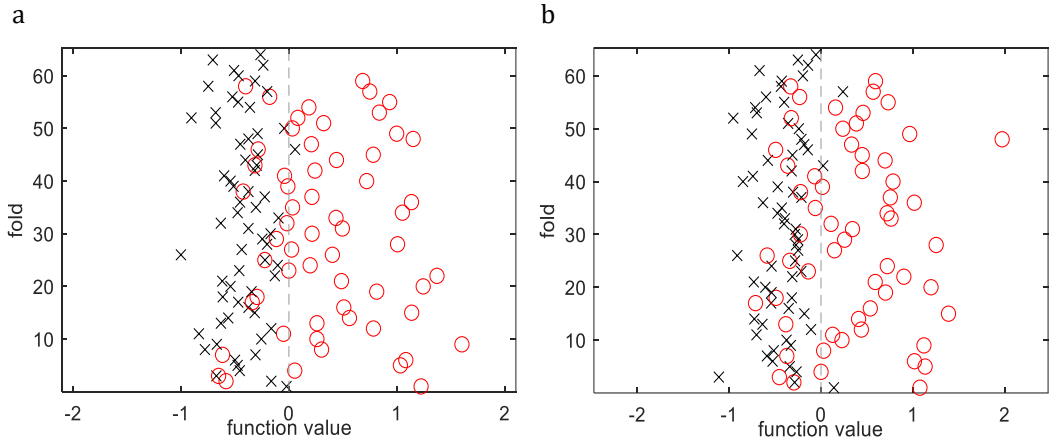


Figure 4.17: Predictions resulting from SVM classification of GM (a) and WM (b) segments extracted from filled T1-MPRAGE images. Crosses represent the controls and circles represent the MS patients. The dashed line represents the decision boundary determined by the classifier during the training phase. Examples falling on the positive part of the decision space (function value) are classified as class 1 = MS patients, while examples falling on the negative half of the decision space are classified as class -1 = controls. We can observe that there are examples easily classified (very far from the decision boundary) while other are much closer to the decision boundary and a few are on the wrong side, ended up misclassified.

Classification performance based on T2-FLAIR

SVM classification applied to GM from T2-FLAIR correctly classified 71.43 % (sensitivity) of patients and 91.53 % (specificity) of controls, achieving an overall classification of 81.48 % ($p = 0.001$). In what concerns WM, SVM correctly classified 71.40 % of patients and 96.61 % of controls, achieving an overall classification of 84.15 % ($p = 0.001$).

Confusion matrix showing the TP, FN, TN and FP is illustrated in Figure 4.18. Predictions showing boundary decision and hyperplane dimension is illustrated in Figure 4.19.

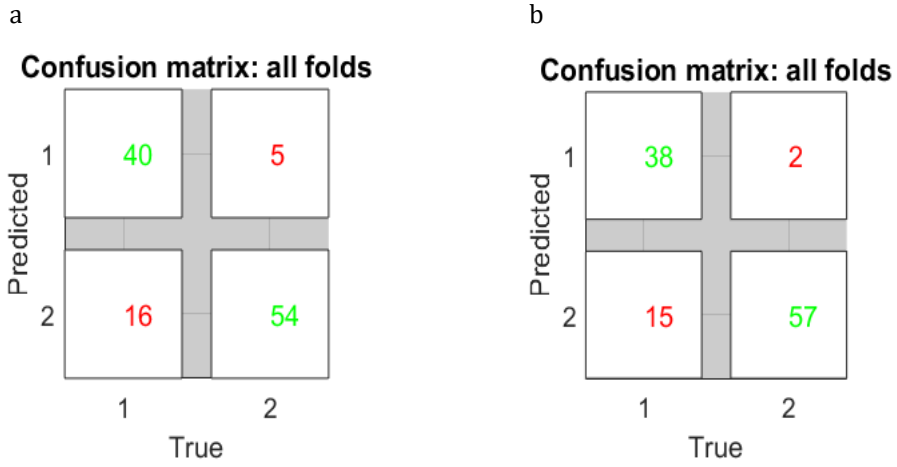


Figure 4.18: Confusion matrixes resulting from SVM classification of GM (a) and WM (b) segments extracted from T2-FLAIR images. True are the true labels of each example and predicted are the labels assigned by the SVM classifier. In these plots, 1 refers to MS patients' class and 2 refers to controls class. With the information of the first column we calculate sensitivity and with information of second column we calculate specificity.

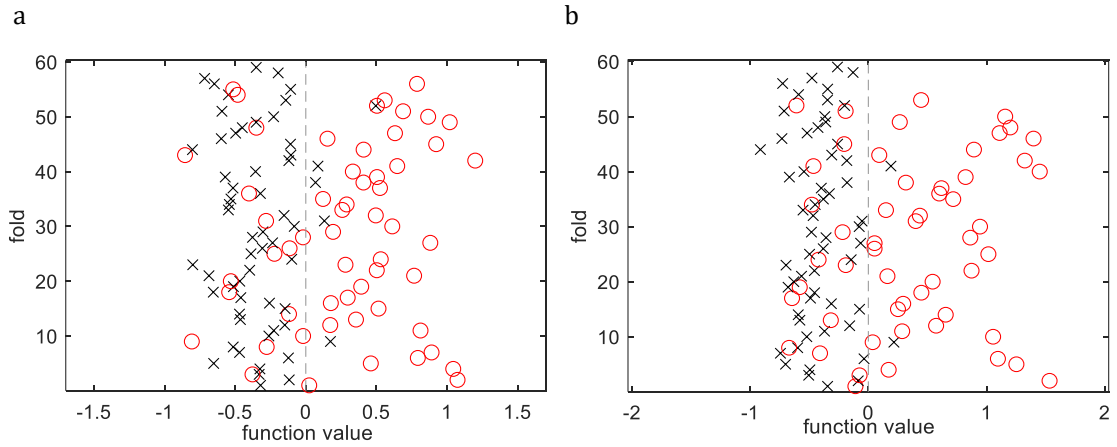


Figure 4.19: Predictions resulting from SVM classification of GM (a) and WM (b) segments extracted from filled T1-MPRAGE images. Crosses represent the controls and circles represent the MS patients. The dashed line represents the decision boundary determined by the classifier during the training phase. Examples falling on the positive part of the decision space (function value) are classified as class 1 = MS patients, while examples falling on the negative half of the decision space are classified as class -1 = controls. We can observe that there are examples easily classified (very far from the decision boundary) while other are much closer to the decision boundary and a few are on the wrong side, ended up misclassified.

The performance values of all SVM classifier models are summarized in Figure 4.20.

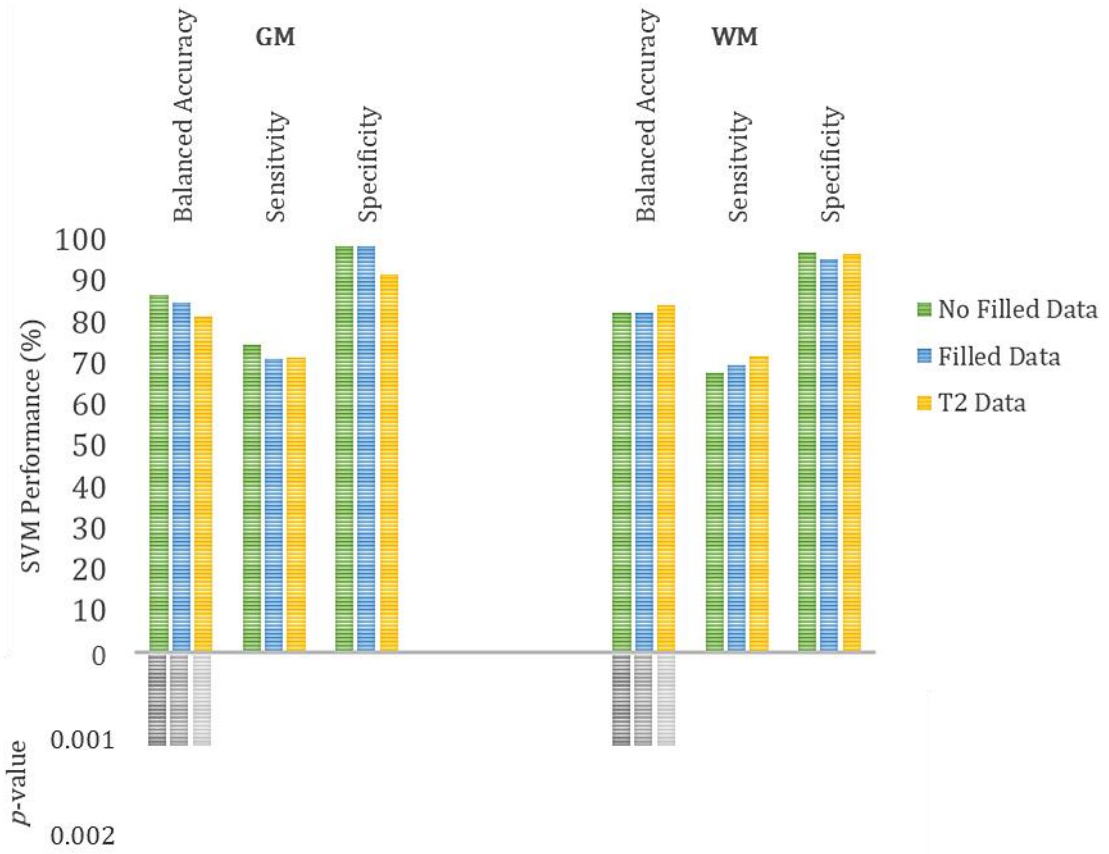


Figure 4.20: Sensitivity, specificity, balanced accuracy and *p*-values of MS patients vs controls classification of GM and WM segments extracted from no filled T1-MPRAGE images, filled T1-MPRAGE images and T2-FLAIR images

Discussion

Despite the three models based on GM segments present similar balanced accuracies, the one with the highest balanced accuracy corresponds to the model using no filled T1-MPRAGE images. However, no filled data contains the effects of lesions which can lead to an overestimation of the classification results. In what concerns WM, all the three models present similar values for balanced accuracy as well, nevertheless the highest balanced accuracy is observed in the model using T2 images. T2 data is not lesion filled either which can bias the classification results. However, when compared to T1 images, T2 images can provide additional information crucial to discriminate groups and to produce a more efficient classification. In both GM and WM models, it is denoted that specificity presents

higher values compared to sensitivity. This demonstrates a greater facility of the SVM to classify subjects as controls.

From the presented results, it is demonstrated a reliable detection of MS patients, which reveals that MVPA is a useful predictive tool that could be used as biomarker with potential for support of MS diagnosis.

4.4.2 MS patients and controls classification weight maps

The weight maps representing the contribution of each voxel for a classification procedure were overlaid onto a template brain image. Figures 4.21 and 4.22 present, on left panels, the slices of the brain containing the most contributing voxels of GM and WM classification, respectively. On the right panels of the figures the same slices are presented but with the statistical parametric maps of VBM analyses of GM and WM, respectively. Comprehensive weights maps of the whole brain for each classification are available in annex D, Figures D.1 and D.2 of this dissertation. In these weight maps, also called discriminative maps, negative weights represent the contribution of a feature to classify an example as a control, whereas positive weights represent the contribution of that particular feature (voxel) to classify the image as an MS patient.

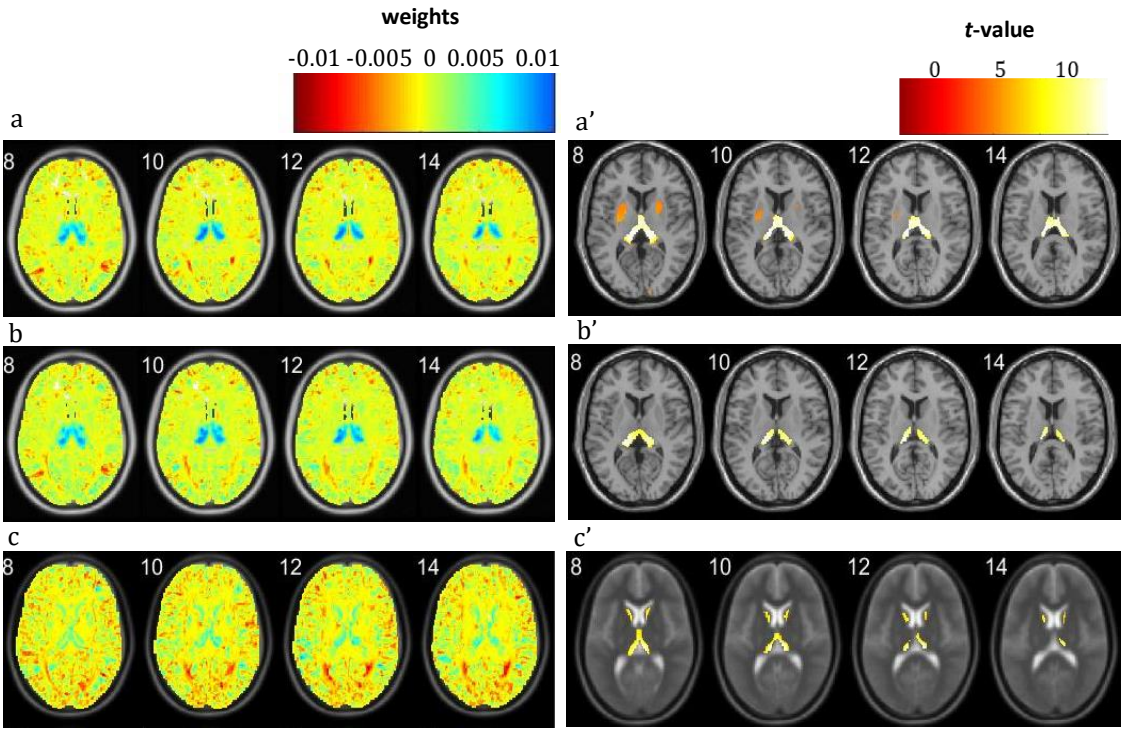


Figure 4.21: Comparison between GM SVM classification and GM VBM results. The left panel presents the brain slices with the highest weights of SVM classification of GM images. In color bar scale, positively weighted voxels are displayed in blue/green while negatively weighted voxels are displayed in orange/red. The right panel presents the same slices with the results of VBM analysis of GM. The contrast here represented is controls > MS patients and the color scale represents t-values. Results are presented at a voxel-level *p*-value < 0.05, FWE corrected. a) Weights of GM classification from no filled T1-MPRAGE images. b) Weights of GM classification from filled T1-MPRAGE images. c) Weights of GM classification from T2-FLAIR images. a') VBM analysis of GM extracted from no filled T1-MPRAGE images. b') VBM analysis of GM extracted from filled T1-MPRAGE images. c') VBM analysis of GM extracted from T2-FLAIR images.

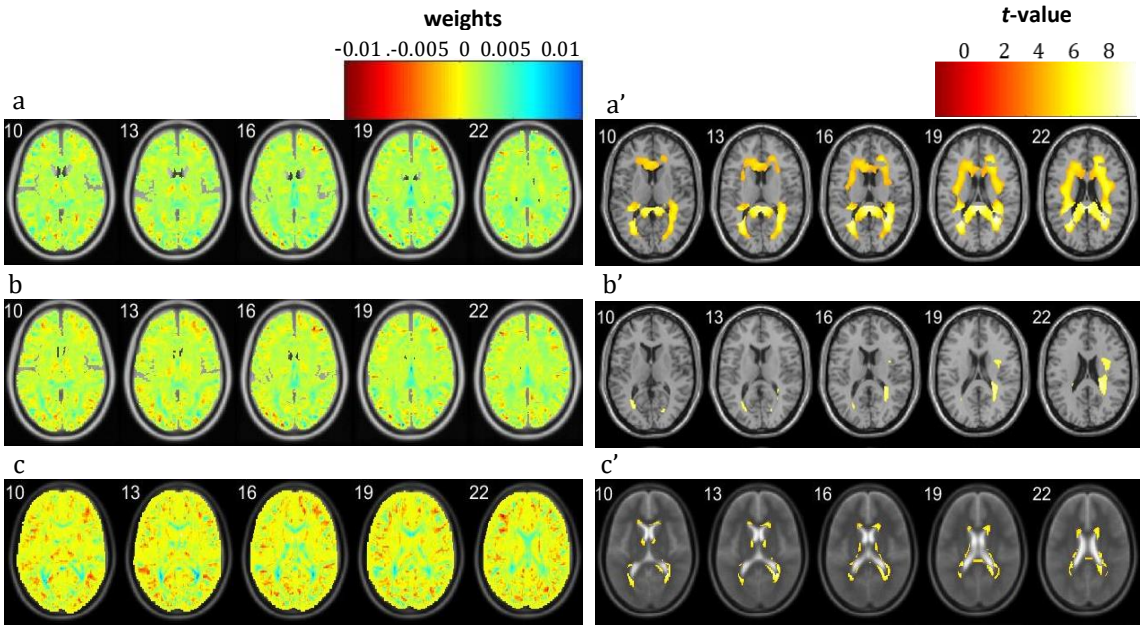


Figure 4.22: Comparison between WM SVM classification and WM VBM results. The left panel presents the brain slices with the highest weights of SVM classification of WM images. In color bar scale, positively weighted voxels are displayed in blue/green while negatively weighted voxels are displayed in orange/red. The right panel presents the same slices with the results of VBM analysis of WM. The contrast here represented is controls > MS patients and the color scale represents t-values. Results are presented at a voxel-level p-value < 0.05, FWE corrected. a) Weights of WM classification from no filled T1-MPRAGE images. b) Weights of WM classification from filled T1-MPRAGE images. c) Weights of WM classification from T2-FLAIR images. a') VBM analysis of WM extracted from no filled T1-MPRAGE images. b') VBM analysis of WM extracted from filled T1-MPRAGE images. c') VBM analysis of WM extracted from T2-FLAIR images.

Discussion

Weight maps from SVM classification highlighted similar regions to those found with VBM analysis. In GM maps, higher weights were assigned to thalamic regions. In WM maps, higher weights were assigned to temporal regions. These results suggest that differences in such regions are crucial for the discrimination between MS patients and control groups.

4.4.3 MS Subtypes Classification: SPMS vs. RRMS

Classification performance based on filled T1-MPRAGE

SVM classifier applied to GM volume segments extracted from filled T1-MPRAGE data correctly classified 46.50 % of SPMS patients (mean sensitivity) and 67.88 % of RRMS patients (mean specificity). The overall mean classification balanced accuracy of was 57.19 %. In what concerns WM, volume segments extracted from filled T1-MPRAGE images, SVM correctly classified 54.25 % of SPMS and 63.75 % of RRMS patients. The overall mean classification balanced accuracy was 59.00 %.

Classification performance based on T2-FLAIR

SVM classifier applied to GM volume segments extracted from T2-FLAIR data correctly classified 44.86 % of SPMS and 56.43 % of RRMS patients, achieving an overall mean classification balanced accuracy of 50.64 %. In what concerns WM volume, SVM correctly classified 35.17 % of SPMS patients and of 57.83 % RRMS patients, achieving an overall mean classification balanced accuracy of 46.50 %.

Validity of the performance values

Given these low balanced accuracy values, it was important to test whether these classifications resulted by chance or not. To accomplish this, a one sample *t*-test was performed with the mean accuracy values for each classification against a chance level of 50 %. 100 repetitions of each classification model were performed, randomly selecting the RRMS patients to compare with SPMS patients. From each individual *t*-test resulted a *p*-value which indicates the probability of these results occur by chance. The mean *p*-value of each classification is listed in the Table 4.1.

Table 4.1: mean *p*-values of MS subtypes classification

	Filled T1-MPRAGE	T2-FLAIR
Grey Matter	5.53e-10	0.689
White Matter	2.61e-14	0.017

Performance values of SVM discrimination between MS subtypes are summarized in Figure 4.23

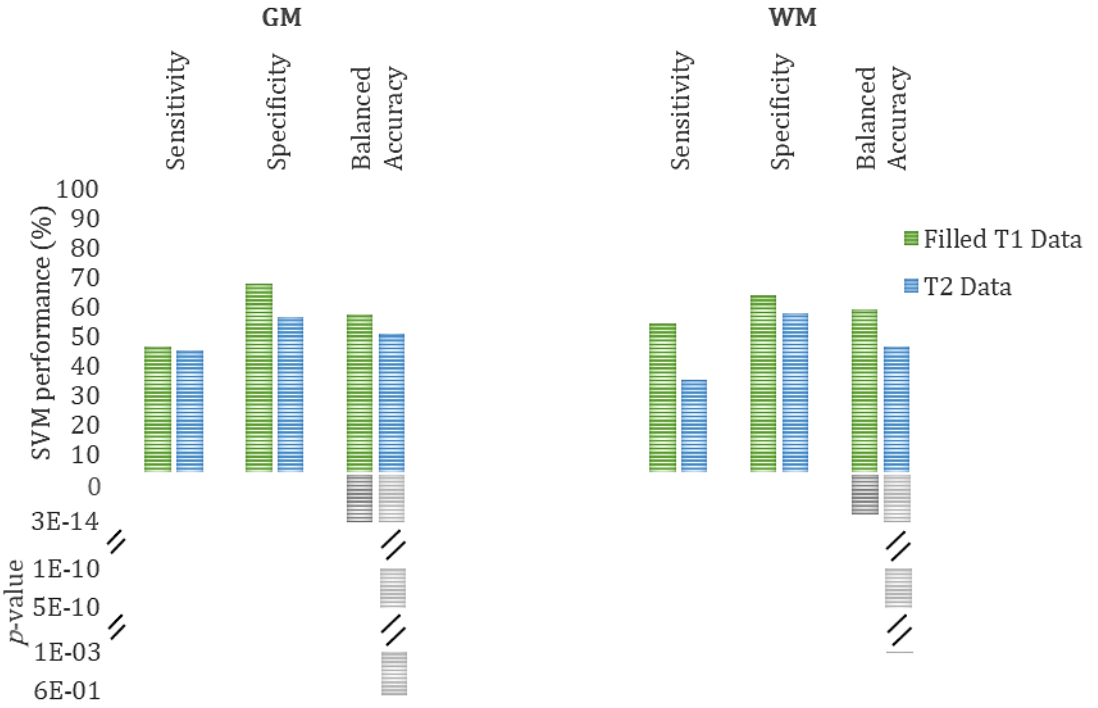


Figure 4.23: Sensitivity, specificity, accuracy and p-values of MS subtypes classification applied to GM and WM volume segments extracted from filled T1-MPRAGE images and T2-FLAIR images.

Discussion

At a first glance these are not the most optimal results, since the values of balanced accuracy are around 50 % - 60 %. For both GM and WM models, models using filled T1 images present the highest balanced accuracy. Despite the model being performed 100 times with different RRMS patients each time, the limited number of subjects of each class contributes to a less reliable classification and therefore to lower accuracy values. However, obtained *p*-values of classification using T1 images reveal that the probability of these results happening by chance is very low. Therefore, it is still possible to discriminate subjects with different MS subtypes based on T1 images.

During sample homogeneity testing, T2 images of certain subjects had to be excluded which limited even more the number of subjects per group (7 subjects per group in GM model and 6 subjects per group in WM model), which might explain the lower performance values obtained using T2 images. The *p*-value classification of GM extracted from T2 images indicates that the hypothesis that our classification results happen by chance cannot be rejected. On the other hand, the *p*-value of

classification of WM extracted from T2 images indicates that the hypothesis that our results happen by chance, i.e. with a mean accuracy of 50 %, can be rejected. However, WM classification revealed a mean accuracy below 50 % which means that it was not possible to perform a discrimination between MS subtypes.

Classification values were obtained through the mean of 100 classification models, each with a different subset of participants (RRMS). As the obtained discrimination was not successful, we do not show the weight map. In future work, when reaching a satisfactory discrimination, the discriminative map(s) shall be investigated in order to understand which features are behind the distinction of MS subtypes.

5 Conclusions and Future Work

Assessing different structural medical MR imaging-derived biomarkers, with different types of analyses, in the context of MS was the focus of this dissertation. It was hypothesized that structural signatures might help discriminate MS patients from healthy controls, and furthermore help solving the challenging differentiation of MS subtypes, ultimately leading to an earlier diagnosis of new patients or progression to a clinical subtype of MS. To assess structural changes in the brain of MS patients, univariate and data driven multivariate statistical analyses were implemented.

With univariate analyses, it was aimed to find brain regions with alterations in regional morphometric features (GM and WM volume, cortical thickness and gyrification index) in MS. On the other hand, with multivariate pattern analysis, it was aimed to discriminate between MS patients and controls and identify the most discriminative features (voxels). In addition, in the framework of multivariate methods, the discrimination between MS subtypes was tested. It was also aimed to perform a comparison between the performance of univariate and multivariate methods in the same dataset.

Voxel based morphometry (VBM) and surface-based morphometry (SBM) were the univariate methods applied. VBM results confirmed the presence of GM and WM atrophy in MS patients. Since VBM is more sensible to detect changes in GM, this atrophy was more evident than WM atrophy in MS. To assess WM atrophy, more appropriate methods should be applied, for example Diffusion Tensor Imaging (DTI). Surface based morphometry (SBM) confirmed that in MS the cortical thickness is decreased and revealed alterations of the gyrification index as well. Both decreased and increased gyrification index was found in MS, although the increase of gyrification index is more prominent.

In all analyses, clusters with between-group differences in GM/WM/cortical thickness/gyrification index are assigned to a brain location, which is identified by their anatomical correspondence with coordinates in a standard space. These allow a qualitative description of disease discrimination features based on macrostructural measures from MRI. Quantitative analysis of the statistical maps,

e.g. number of voxels per cluster and peak localization, is required for further investigation of disease pathophysiology.

Support Vector Machine (SVM) classification was the multivariate method used to implement pattern recognition analysis. SVM classification of MS patients or MS subtypes was based on GM and WM volume features. Regarding the discrimination of MS patients and controls, performance values demonstrated this is a useful predictive tool that could be used as biomarker with potential for diagnosis. Weight maps of this classification revealed structural differences between groups. The most contributive weights revealed patterns of differences very similar to the results obtained with VBM analysis. Besides these, weight maps of SVM classification revealed additional differences between groups, demonstrating its ability to detect subtle differences that univariate methods cannot achieve. Such differences may explain effects in MS and should be further investigated, both qualitatively and quantitatively, to provide precise clues about the disease.

Classification of MS subtypes, relapsing-remitting MS (RRMS) patients and secondary progressive MS (SPMS) patients, yielded low performance values. The limited number of subjects in each group probably explains these results. With a higher number of subjects and considering performance values of MS patients and controls classification, accurate discrimination of subtypes within MS should be possible to achieve.

Another aim of this study was therefore to test whether these types of analysis using T2-weighted scans could provide information of structural alterations additional to more commonly used T1 data, at identical spatial resolution, as T2-weighted MRI signal can be more sensitive to microscopic neurodegenerative processes. Indeed, we were able to detect morphometric differences that were not visible with analysis of T1 images. Furthermore, the highest accuracy of discrimination between MS patients and healthy controls based on WM content was achieved using T2 images.

Univariate and multivariate findings were compared in the three types of images used. All types of images revealed very similar affected regions with univariate analyses, and similar performance and weight maps with SVM classification. However, it should be noted that the results obtained with the analysis of images after lesion filling are those that should be interpreted as the most reliable. Even if

other models yield better classification results or morphometric regional differences in very different areas, these should be interpreted with caution, as (at least part) of these can arise from artifacts due to lesions.

Both methods have advantages and disadvantages. On the one hand, results of univariate analysis are derived from *a priori* hypothesis. An *a priori* hypothesis requires the definition of the groups and what differences between them, defined by contrasts, are intended to test. In univariate analysis, it was intended to find in which voxels individually MS patients have higher or lower values of measured morphometric features when compared to controls. The assessment of other regional variances between groups is not possible to achieve unless an indication of what is intended to test is given. One the other hand, results from multivariate analysis do not derive from an indication of what is supposed to find. Through pattern recognition of the relationship between given features and labels, differences between subjects are algorithmically found and used to discriminate groups. To sum things up, univariate analyses provide knowledge about what is being tested and about what represents the differences between groups. Multivariate analyses allow finding subtle differences that are not detected by other methods and which indeed allow discriminating groups, but with little knowledge of what these represent. To conclude, this data-driven analysis indicated spatially distributed networks of brain regions with abnormal structure in individuals with MS. It might provide important clues for the pathophysiology of this disorder, with potential benefits in the prevention and adequacy of clinical interventions.

Notably, we applied different analysis approaches (VBM, SBM and MVPA) to investigate a set morphometric measures in the same dataset. This allowed for comparisons between the different MS biomarkers studied with both univariate and multivariate methods and to discuss about the most discriminative ones. Notably, the analysis of structural biomarkers also has the advantage of guiding scientific research about the pathophysiology of MS and its cause.

Considering the results obtained in this work, future approaches should be considered. A multimodal imaging classification could provide a more accurate discrimination between groups. Since surface measures were proved to be altered in MS, classification based on surface features together with volumetric features could also be a promising approach in the field of MS diagnosis.

Funding

This research was supported by grants from Biogen, Foundation for Science and Technology (FCT), COMPETE and FEDER [grants CENTRO-01-0145-FEDER-000016(reference SAICT 000016 BIGDATIMAGE), POCI-01-0145-FEDER-007440 (reference UID/NEU/04539/2013), and CENTRO-01-0145-FEDER-031973]. The sponsors did not participate in any aspect of the design or performance of the study, including the data collection, management, analysis, and interpretation or preparation of the manuscript. The authors declare no competing financial or other conflict of interest.

Bibliography

[1] M. M. Goldenberg, "Multiple sclerosis review.," *P T*, vol. 37, no. 3, pp. 175–84, 2012.

[2] D. S. Reich, C. F. Lucchinetti, and P. A. Calabresi, "Multiple Sclerosis," *N. Engl. J. Med.*, vol. 378, no. 2, pp. 169–180, 2018.

[3] R. Dobson and G. Giovannoni, "Multiple sclerosis – a review," *Eur. J. Neurol.*, vol. 26, no. 1, pp. 27–40, 2019.

[4] J. E. Joy and R. B. Johnston, "Clinical and Biological Features," in *Multiple Sclerosis: Current Status and Strategies for the Future*, J. Editors, Ed. Institute of Medicine, 2001, pp. 29–114.

[5] K. Rahn, B. Slusher, and A. Kaplin, "Cognitive impairment in multiple sclerosis: a forgotten disability remembered.," *Cerebrum*, vol. 2012, p. 14, 2012.

[6] N. P. Staff, C. F. Lucchinetti, and B. M. Keegan, "Multiple sclerosis with predominant, severe cognitive impairment," *Arch. Neurol.*, vol. 66, no. 9, pp. 1139–1143, 2009.

[7] M. T. Wallin *et al.*, "Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016," *Lancet Neurol.*, vol. 18, no. 3, pp. 269–285, 2019.

[8] D. H. Miller, D. T. Chard, and O. Ciccarelli, "Clinically isolated syndromes," *Lancet Neurol.*, vol. 11, no. 2, pp. 157–169, 2012.

[9] D. M. Sima, D. Loeckx, D. Smeets, S. Jain, P. M. Parizel, and W. V Hecke, "Use Case I: Imaging Biomarkers in Neurological Disease. Focus on Multiple Sclerosis," in *Imaging Biomarkers*, L. Martí-Bonmatí and A. Alberich-Bayarr, Eds. Springer, 2016, pp. 169–180.

[10] K. Strimbu and J. A. Travel, "What are Biomarkers?," *Curr. Opin. HIV AIDS*, vol. 5, no. 6, pp. 463–466, 2010.

[11] A. J. Thompson, B. L. Banwell, F. Barkhof, and et.al., "Diagnosis of multiple sclerosis: revision of the McDonald criteria 2017," *Lancet Neurol.*, vol. 17, pp. 162–173, 2018.

[12] M. Filippi *et al.*, "MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines," *Lancet Neurol.*, vol. S1474-4422, no. 15, 2016.

[13] H. Wiendl and R. Hohlfeld, "Multiple sclerosis therapeutics Unexpected outcomes clouding undisputed successes," *Neurology*, vol. 72, no. 11, pp. 1008–1015, 2009.

[14] J. H. Noseworthy, "Progress in determining the causes and treatment of multiple sclerosis," *Nature*, vol. 399, pp. A40–A47, 1999.

[15] M. Kaisey, A. J. Solomon, and et.al., "Incidence of multiple sclerosis misdiagnosis in referrals to two academic centers," *Mult. Scler. Relat. Disord.*, vol. 30, pp. 51–56, 2019.

[16] W. J. Brownlee, T. A. Hardy, F. Fazeka, and D. H. Miller, "Diagnosis of multiple sclerosis: progress and challenges," *Lancet*, vol. 389, no. 10076, pp. 1336–1346, 2017.

[17] B. I. Yamout *et al.*, "Alternative diagnoses in patients referred to specialized centers for suspected MS," *Mult. Scler. Relat. Disord.*, vol. 18, no. June, pp. 85–89, 2017.

[18] D. Ontaneda and R. J. Fox, "Imaging as an Outcome Measure in Multiple Sclerosis," *Neurotherapeutics*, vol. 14, no. 1, pp. 24–34, 2017.

[19] M. Absinta, P. Sati, and D. S. Reich, "Advanced MRI and staging of multiple sclerosis lesions.," *Nat. Rev. Neurol.*, vol. 12, no. 6, pp. 358–368, 2016.

[20] S. Batista *et al.*, "Impairment of social cognition in multiple sclerosis: Amygdala atrophy is the main predictor," *Mult. Scler.*, vol. 23, no. 10, pp. 1358–1366, 2016.

[21] S. Batista *et al.*, "Disconnection as a mechanism for social cognition impairment in multiple sclerosis," *Neurology*, vol. 89, no. 1, pp. 38–45, 2017.

[22] C. Alves *et al.*, "The retinal ganglion cell layer predicts normal-appearing white matter tract integrity in multiple sclerosis: A combined diffusion tensor imaging and optical coherence tomography approach," *Hum. Brain Mapp.*, vol. 39, no. 4, pp. 1712–1720, 2018.

[23] I. Blockx, R. D. Palacios, and A. V. der Linden, "MRI with applications in neurological disorders," in *MRI with applications in neurological disorders*, F. Gazeau, Ed. Future Science, 2013, pp. 50–64.

[24] U. W. Kaunzner and S. A. Gauthier, "MRI in the assessment and monitoring of multiple sclerosis: an update on best practice," *Ther. Adv. Neurol. Disord.*, vol. 10, no. 6, pp. 247–261, 2017.

[25] M. Calabrese *et al.*, "Cortical Lesions and Atrophy Associated with Cognitive Impairment in Relapsing-Remitting Multiple Sclerosis," *Arch. Neurol.*, vol. 66, no. 9, pp. 1144–1150, 2009.

[26] À. Rovira, C. Auger, and J. Alonso, "Magnetic resonance monitoring of lesion evolution in multiple sclerosis," *Ther. Adv. Neurol. Disord.*, vol. 6, no. 5, pp. 298–310, 2013.

[27] R. A. Rudick, J.-C. Lee, J. Simon, and E. Fisher, "Significance of T2 Lesions in Multiple Sclerosis: A 13-Year Longitudinal Study," *Ann. Neurol.*, vol. 60, no. 2, pp. 236–242, 2006.
[28] R. A. Bermel *et al.*, "Predictors of Long-Term Outcome in Multiple Sclerosis Patients Treated with Interferon Beta," *Ann. Neurol.*, vol. 73, no. 1, pp. 95–103, 2013.
[29] M. Calabrese, M. Filippi, and P. Gallo, "Cortical lesions in multiple sclerosis," *Nat. Rev. Neurol.*, vol. 6, no. 8, pp. 438–444, 2010.
[30] M. Calabrese, A. Favaretto, V. Martini, and P. Gallo, "Grey matter lesions in MS from histology to clinical implications," *Prion*, vol. 7, no. 1, pp. 20–27, 2013.
[31] M. A. Rocca *et al.*, "Brain MRI atrophy quantification in MS from methods to clinical application," *Neurology*, vol. 88, no. 4, pp. 403–413, 2017.
[32] A. Eshaghi *et al.*, "Progression of regional grey matter atrophy in multiple sclerosis," *Brain*, vol. 141, no. 6, pp. 1665–1677, 2018.
[33] M. P. Sanfilipo, R. H. B. Benedict, B. Weinstock-Guttman, and R. Bakshi, "Gray and white matter brain atrophy and neuropsychological impairment in multiple sclerosis," *Neurology*, vol. 66, no. 5, pp. 685–692, 2006.
[34] D. García-Lorenzo, S. Francis, S. Narayanan, D. L. Arnold, and D. L. Collins, "Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging," *Med. Image Anal.*, vol. 17, no. 1, pp. 1–18, 2013.
[35] D. Pelletier, K. Garrison, and R. Henry, "Measurement of Whole-Brain Atrophy in Multiple Sclerosis," *J. Neuroimaging*, vol. 14, pp. 115–119, 2004.
[36] V. M. Anderson, N. C. Fox, and D. H. Miller, "Magnetic resonance imaging measures of brain atrophy in multiple sclerosis," *J. Magn. Reson. Imaging*, vol. 23, no. 5, pp. 605–618, 2006.
[37] F. Kurth, E. Luders, and L. Angeles, "Voxel-Based Morphometry," *Brain Mapp. An Encycl. Ref.*, vol. 1, pp. 345–349, 2015.
[38] C. Scarpazza and M. S. De Simone, "Voxel-based morphometry: current perspectives," *Neurosci. Neuroeconomics*, vol. 5, pp. 19–35, 2016.
[39] K. Nakamura, R. A. Brown, S. Narayanan, D. L. Collins, and D. L. Arnol, "Diurnal fluctuations in brain volume: Statistical analyses of MRI from large population," *Neuroimage*, vol. 118, pp. 126–132, 2015.
[40] B. Fischl and A. M. Dale, "Measuring the thickness of the human cerebral cortex from magnetic resonance images.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 20, pp. 11050–11055, 2000.
[41] R. Righart *et al.*, "Volume versus surface-based cortical thickness measurements: A comparative study with healthy controls and multiple sclerosis patients," *PLoS One*, vol. 12, no. 7, pp. 1–10, 2017.
[42] M. Calabrese *et al.*, "Cortical atrophy is relevant in multiple sclerosis at clinical onset," *J. Neurol.*, vol. 254, no. 9, pp. 1212–1220, 2007.
[43] M. Sailer *et al.*, "Focal thinning of the cerebral cortex in multiple sclerosis," *Brain*, vol. 126, no. 8, pp. 1734–1744, 2003.
[44] Y. Yang, S. Chefer, X. Geng, H. Gu, X. Chen, and E. A. Stein, "Structural and Functional Neuroimaging Methods: Applications to Substance Abuse and Addiction," in *Neuroimaging in Addiction*, B. Adinoff and E. A. Stein, Eds. John Wiley & Sons, Ltd, 2011, pp. 37–82.
[45] M. Schaer, M. B. Cuadra, N. Schmansky, B. Fischl, J.-P. Thiran, and S. Eliez, "How to Measure Cortical Folding from MR Images: a Step-by-Step Tutorial to Compute Local Gyrification Index," *J. Vis. Exp.*, no. 59, pp. 1–8, 2012.
[46] M. Schaer, M. Bach Cuadra, L. Tamarit, F. Lazeyras, S. Eliez, and J. P. Thiran, "A Surface-based approach to quantify local cortical gyrification," *IEEE Trans. Med. Imaging*, vol. 27, no. 2, pp. 161–170, 2008.
[47] R. Spalthoff, C. Gaser, and I. Nenadić, "Altered gyrification in schizophrenia and its relation to other morphometric markers," *Schizophr. Res.*, vol. 202, pp. 195–202, 2018.
[48] L. J. Hogstrom, L. T. Westlye, K. B. Walhovd, and A. M. Fjell, "The structure of the cerebral cortex across adult life: Age-related patterns of surface area, thickness, and gyrification," *Cereb. Cortex*, vol. 23, no. 11, pp. 2521–2530, 2013.
[49] A. R. McIntosha and N. J. Lobaughb, "Partial least squares analysis of neuroimaging data: applications and advances," *Neuroimage*, vol. 23, pp. 250–263, 2004.
[50] K. A. Linn, B. Gaonkar, T. D. Satterthwaite, J. Doshi, C. Davatzikos, and R. T. Shinohara, "Control-group feature normalization for multivariate pattern analysis of structural MRI data using the support vector machine," *Neuroimage*, vol. 132, pp. 157–166, 2016.
[51] C.-W. Woo, L. J. Chang, M. A. Lindquist, and T. D. Wager, "Building better biomarkers: brain models in translational neuroimaging," *Nat. Neurosci.*, vol. 20, no. 3, pp. 365–377, 2017.

[52] J. V. Duarte, M. J. Ribeiro, I. R. Violante, G. Cunha, E. Silva, and M. Castelo-Branco, "Multivariate Pattern Analysis Reveals Subtle Brain Anomalies Relevant to the Cognitive Phenotype in Neurofibromatosis Type 1," *Hum. Brain Mapp.*, vol. 35, no. 1, pp. 89–106, 2012.
[53] C. Ecker *et al.*, "Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach," *Neuroimage*, vol. 49, no. 1, pp. 44–56, 2010.
[54] M. J. Marzelli, F. Hoeft, D. S. Hong, and A. L. Reiss, "Neuroanatomical Spatial Patterns in Turner Syndrome," *Neuroimage*, vol. 55, no. 2, pp. 439–447, 2011.
[55] J. M. Mateos-Pérez, M. Dadar, M. Lacalle-Aurioles, Y. Iturria-Medina, Y. Zeighami, and A. C. Evans, "Structural neuroimaging as clinical predictor: A review of machine learning applications," *NeuroImage Clin.*, vol. 20, pp. 506–522, 2018.
[56] E. E. Tulay, B. Metin, N. Tarhan, and M. K. Arıkan, "Multimodal Neuroimaging: Basic Concepts and Classification of Neuropsychiatric Diseases," *Clin. EEG Neurosci.*, vol. 50, no. 1, pp. 20–33, 2018.
[57] Y. Zhang *et al.*, "Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision tree, k-nearest neighbors, and support vector machine," *Simulation*, vol. 92, no. 2, pp. 861–871, 2016.
[58] K. Bendfeldt *et al.*, "Multivariate pattern classification of gray matter pathology in multiple sclerosis," *Neuroimage*, vol. 60, pp. 400–408, 2012.
[59] M. Zurita *et al.*, "Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data," *NeuroImage Clin.*, vol. 20, pp. 724–730, 2018.
[60] U. Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, London, "SPM." [Online]. Available: https://www.fil.ion.ucl.ac.uk/spm/software/spm12/.
[61] C. Gaser and R. Dahnke, "CAT12, A computational Anatomy Toolbox for SPM." [Online]. Available: http://www.neuro.uni-jena.de/cat/.
[62] G. Morphometry Group, Department of Neurology, Technische Universität München (TUM), Munich, Germany; Department of Statistics, Ludwig-Maximilians-University, Munich, Germany; and Structural Brain Mapping Group, Departments of Neurology and Psychiatry, Fried, "LST, A Lesion Segmentation Tool for SPM." [Online]. Available: www.applied-statistics.de/lst.html.
[63] "Mathworks." [Online]. Available: https://www.mathworks.com/.
[64] J. Schrouff *et al.*, "PRoNTo: Pattern recognition for neuroimaging toolbox," *Neuroinformatics*, vol. 11, no. 3, pp. 319–337, 2013.
[65] "MRIcron." [Online]. Available: https://www.nitrc.org/projects/mricron.
[66] S. M. Smith, "BET:Brain Extraction Tool." .
[67] "Manual Computational Anatomy Toolbox - CAT12." [Online]. Available: http://www.neuro.uni-jena.de/cat12/CAT12-Manual.pdf.
[68] E.-W. Radue, M. Weigel, R. Wiest, and H. Urbach, "Introduction to Magnetic Resonance Imaging for Neurologists," *Contin. Lifelong Learn. Neurol.*, vol. 22, no. 5, pp. 1379–1398, 2016.
[69] "Basics of Magnetic Resonance Imaging.," *Semin. Vasc. Surg.*, vol. 17, no. 2, pp. 66–82, 2004.
[70] H. H. Schild, *MRI made easy*. Berlin, Germany: Schering AG, 1990.
[71] E. F.Jackson, L. E.Ginsberg, D. F.Schomer, and N. E.Leeds, "A Review of MRI Pulse Sequences and Techniques in Neuroimaging," *Surg. Neurol.*, vol. 47, no. 2, pp. 185–199, 1997.
[72] "NIfTI: — Neuroimaging Informatics Technology Initiative." [Online]. Available: https://nifti.nimh.nih.gov/.
[73] "SPM12 Manual." [Online]. Available: https://www.fil.ion.ucl.ac.uk/spm/doc/spm12_manual.pdf.
[74] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information," *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 187–198, 1997.
[75] S. Valverde, A. Oliver, and X. Lladó, "A white matter lesion-filling approach to improve brain tissue volume measurements," *NeuroImage Clin.*, vol. 6, pp. 86–92, 2014.
[76] S. Magon *et al.*, "White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: A longitudinal study," *BMC Neurosci.*, vol. 15, no. 106, pp. 1–10, 2014.
[77] V. Popescu, N. C. G. Ran, F. Barkhof, D. T. Chard, C. A. Wheeler-Kingshott, and H. Vrenken, "Accurate GM atrophy quantification in MS using lesion-filling with co-registered 2D lesion masks," *NeuroImage Clin.*, vol. 4, pp. 366–373, 2014.
[78] M. Battaglini, M. Jenkinson, and N. De Stefano, "Evaluating and Reducing the Impact of White

Matter Lesions on Brain Volume Measurements," *Hum. Brain Mapp.*, vol. 33, no. 9, pp. 2062–2071, 2012.
[79] P. Schmidt, "Bayesian inference for structured additive regression models for large-scale problems with applications to medical imaging," LMU München, 2016.
[80] P. Schmidt *et al.*, "An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis," *Neuroimage*, vol. 59, no. 4, pp. 3774–3783, 2012.
[81] D. T. Chard, J. S. Jackson, D. H. Miller, and C. Wheeler-Kingshott, "Reducing the Impact of White Matter Lesions on Automated Measures of Brain Gray and White Matter Volumes," *J. Magn. Reson. Imaging*, vol. 32, no. 1, pp. 223–228, 2010.
[82] S. Valverde *et al.*, "Quantifying brain tissue volume in multiple sclerosis with automated lesion segmentation and filling," *NeuroImage Clin.*, vol. 9, pp. 640–647, 2015.
[83] P. Kalavathi and V. B. S. Prasath, "Methods on Skull Stripping of MRI Head Scan Images—a Review," *J. Digit. Imaging*, vol. 29, no. 3, pp. 365–379, 2016.
[84] J. M. S. Pereira, L. Xiong, J. Acosta-Cabronero, G. Pengas, G. B. Williams, and P. J. Nestor, "Registration accuracy for VBM studies varies according to region and degenerative disease grouping," *Neuroimage*, vol. 49, no. 3, pp. 2205–2215, 2010.
[85] L. Z. Diaz-de-Grenu, J. Acosta-Cabronero, J. M. S. Pereira, G. Pengas, G. B. Williams, and P. J. Nestor, "MRI detection of tissue pathology beyond atrophy in Alzheimer's disease: Introducing T2-VBM," *Neuroimage*, vol. 56, no. 4, pp. 1946–1953, 2011.
[86] A. . Evans, D. L. Collins, S. R. Mills, E. D. Brown, R. L. Kelly, and T. M. Peters, "3D statistical neuroanatomical models from 305 MRI volumes"," in *IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, 1993.
[87] W. Chau and A. R. McIntosh, "The Talairach coordinate of a point in the MNI space: How to interpret it," *Neuroimage*, vol. 25, no. 2, pp. 408–416, 2005.
[88] J. Ashburner and K. J. Friston, "Voxel-Based Morphometry," in *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, W. Penny, K. Friston, J. Ashburner, S. Kiebel, and T. Nichols, Eds. Elsevier, 2011, pp. 92–98.
[89] A. Mechelli, C. Price, K. Friston, and J. Ashburner, "Voxel-Based Morphometry of the Human Brain: Methods and Applications," *Curr. Med. Imaging Rev.*, vol. 1, no. 2, pp. 105–113, 2005.
[90] J. L. Whitwell, "Voxel-Based Morphometry: An Automated Technique for Assessing Structural Changes in the Brain," *J. Neurosci.*, vol. 29, no. 31, pp. 9661–9664, 2009.
[91] J. Ashburner and K. J. Friston, "Voxel-based morphometry - The methods," *Neuroimage*, vol. 11, no. 6 I, pp. 805–821, 2000.
[92] J. Ashburner, "A fast diffeomorphic image registration algorithm," *Neuroimage*, vol. 38, no. 1, pp. 95–113, 2007.
[93] J. Ashburner, "VBM tutorial," 2010. [Online]. Available: https://www.fil.ion.ucl.ac.uk/~john/misc/VBMclass10.pdf%0A%0A.
[94] R. Dahnke, R. A. Yotter, and C. Gaser, "Cortical thickness and central surface estimation," *Neuroimage*, vol. 65, pp. 336–348, 2013.
[95] R. A. Yotter, R. Dahnke, P. M. Thompson, and C. Gaser, "Topological Correction of Brain Surface Meshes Using Spherical Harmonics," *Hum. Brain Mapp.*, vol. 32, no. 7, pp. 1109–1124, 2011.
[96] R. A. Yotter, P. M. Thompson, and C. Gaser, "Algorithms to Improve the Reparameterization of Spherical Mappings of Brain Surface Meshes," *J. Neuroimaging*, vol. 21, no. 2, pp. e134–e147, 2011.
[97] R. A. Yotter, G. Ziegler, P. M. Thompson, and C. Gaser, "Diffeometric Anatomical Registration on the Surface," *HBM*, 2011.
[98] C. Gaser, "Check Sample Homogeneity at Structural Brain Mapping Group." [Online]. Available: http://www.neuro.uni-jena.de/vbm/check-sample-homogeneity/.
[99] D. (Dallie) Sandilands, "Univariate Analysis," in *Encyclopedia of Quality of Life and Well-Being Research*, A. C. Michalos, Ed. Dordrecht: Springer Netherlands, 2014, pp. 6815–6817.
[100] T. R. Oakes, A. S. Fox, T. Johnstone, M. K. Chung, N. Kalin, and R. J. Davidson, "Integrating VBM into the General Linear Model with voxelwise anatomical covariates," *Neuroimage*, vol. 34, no. 2, pp. 500–508, 2007.
[101] S. J. Kiebel and A. P. Holmes, "The General Linear Model," in *Human Brain Function: Second Edition*, 2003, pp. 730–760.
[102] J. B. Poline, F. Kherif, and W. Penny, "Contrasts and Classical Inference," in *Human Brain Function: Second Edition*, 2003, pp. 761–779.
[103] D. De Canditiis, "Statistical Inference Techniques," *Encycl. Bioinforma. Comput. Biol.*, pp. 698–705, 2018.

[104] "P threshold FAQ." [Online]. Available: http://mindhive.mit.edu/book/export/html/90.
[105] M. Brett, W. Penny, and S. Kiebel, "Introduction to Random Field Theory," in *Human Brain Function: Second Edition*, 2003, pp. 867–879.
[106] C. Gaser, "Statistical Analysis at Structural Brain Mapping Group." [Online]. Available: http://www.neuro.uni-jena.de/vbm/vbm2-for-spm2/statistical-analysis/.
[107] A. M. Winkler *et al.*, "Cortical thickness or Grey Matter volume," *Neuroimage*, vol. 53, no. 3, pp. 1135–1146, 2011.
[108] K. Im, J. M. Lee, O. Lyttelton, S. H. Kim, A. C. Evans, and S. I. Kim, "Brain size and cortical structure in the adult human brain," *Cereb. Cortex*, vol. 18, no. 9, pp. 2181–2191, 2008.
[109] K. Zilles, E. Armstrong, A. Schleicher, and H. Kretschmann, "Anatomy and Embryology The human pattern of gyrification in the cerebral cortex," *Anat Embryol*, vol. 179, pp. 173–179, 1988.
[110] M. L. Ramírez, "UnivarScatter." .
[111] M. Ojala and G. C. Garriga, "Permutation Tests for Studying Classifier Performance - ojala10a.pdf," *J. Mach. Learn. Res.*, vol. 11, pp. 1833–1863, 2010.
[112] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: a tutorial overview," *Neuroimage*, vol. 45, Mar. 2009.
[113] J. V. Duarte, "The role of long-range neural oscillatory synchrony as a mechanism underlying in perceptual coherence," Coimbra, 2016.
[114] D. Srivastava, "Data Classification Using Support Vector," *J. Theor. Appl. Inf. Technol.*, vol. 12, no. 1, pp. 1–7, 2010.
[115] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fMRI: a tutorial overview.," *Neuroimage*, vol. 45, no. 1 Suppl, pp. S199–S209, 2009.
[116] R. Gholami and N. Fakhari, "Support Vector Machine: Principles, Parameters, and Applications," in *Handbook of Neural Computation*, 1st ed., Elsevier Inc., 2017, pp. 515–535.
[117] J. Ashburner *et al.*, "PRoNTo Manual." [Online]. Available: http://www.mlnl.cs.ucl.ac.uk/pronto/prt_manual.pdf.
[118] "MniTalairach-MCR CBU Imaging Wiki." [Online]. Available: http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach.
[119] R. Fama and E. V. Sullivan, "Thalamic structures and associated cognitive functions: Relations with age and aging," *Neurosci. Biobehav. Rev.*, vol. 54, pp. 29–37, 2015.
[120] M. P. Sanfilipo, R. H. Benedict, B. Weinstock-Guttman, and R. Bakshi, "Gray and white matter brain atrophy and neuropsychological impairment in multiple sclerosis," *Neurology*, vol. 66, no. 5, pp. 685–692, 2006.
[121] M. Tahedl, S. M. Levine, M. W. Greenlee, R. Weissert, and J. V. Schwarzbach, "Functional connectivity in multiple sclerosis: Recent findings and future directions," *Front. Neurol.*, vol. 9, pp. 1–18, 2018.
[122] R. B. Buxton, "An Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques," Cambridge University Press, 2002.
[123] N. Hoobs and M. Novak, "Methods for Dummies," *Voxel-Based Morphometry*, 2008. [Online]. Available: https://www.fil.ion.ucl.ac.uk/mfd_archive/2007/powerpoint presentations/VBM_MfD_full.ppt.
[124] A. U. Turken *et al.*, "Multimodal surface-based morphometry reveals diffuse cortical atrophy in traumatic brain injury," *BMC Med. Imaging*, vol. 9, no. 20, pp. 1–15, 2009.

Annexes

A. Methods Scripts

DICOM to NIfTI format: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/dicomtonifti_job.m

Corregistration of T1-MPRAGE images : https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/coregistration_job.m

Mean of T1-MPRAGE images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/meanimgs_job.m

Segmentation of T1-MPRAGE images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/segimgs_job.m

Segmentation of T2-MPRAGE images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Segment_Data_T2_job.m

Lesion Probability Maps of T2-FLAIR images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Lesion_Probability_Maps_job.m

Lesion Filling: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Lesion_Filling_job.m

Extraction of Total Lesion Volume and Number of Lesions: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Extract_Values_of_Interest_job.m

Extraction of Global Measures: Total Intracranial Volume, GM, WM: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Estimate_TIV_GM_WM_job.m

Extraction of Surface Parameters from no filled T1-MPRAGE images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Extract_Additional_Surface_Parameters_job.m

Extraction of Surface Parameters from filled T1-MPRAGE images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Extract_Additional_Surface_Parameters_Filled_Images_job.m

Resample and Smooth of surface parameters from no filled T1-MPRAGE images*: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Resample_and_Smooth_Surface_no_filled_job.m

Resample and Smooth of surface parameters from filled T1-MPRAGE images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Resample_and_Smooth_Surface_Data_Filled_Images_job.m

Check Sample Homogeneity of VBM data: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Check_Sample_Homogeneity_VBM_Data_job.m

Check Sample Homogeneity of SBM Data: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Check_Sample_Homogeneity_Surface_Data_job.m

Data and Design of GM classification from no filled T1-MPRAGE images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Data_and_Design_GM_from_no_filled_T1.m

Data and Design of WM classification from no filled T1-MPRAGE images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Data_and_Design_WM_from_no_filled_T1.m

Data and Design of GM classification from filled T1-MPRAGE images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Data_and_Design_GM_from_filled_T1.m

Data and Design of WM classification from filled T1-MPRAGE images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Data_and_Design_WM_from_filled_T1.m

Data and Design of GM classification from T2-FLAIR images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Data_and_Design_GM_from_T2.m

Data and Design of WM classification from T2-FLAIR images: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Data_and_Design_WM_from_T2.m

Data and Design of subtypes classification: https://github.com/juufsoares/Structural-Imaging-Biomarkers-MS/blob/master/Data_and_Designx100.m

***Smooth of VBM Data was performed directly on SPM interface**

B. Voxel-Based Morphometry Results

Sample Homogeneity Values

a

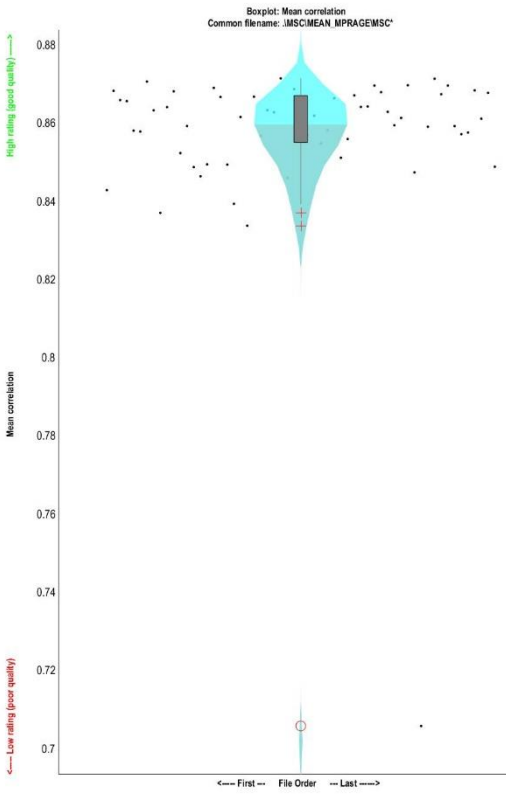


Figure B.1: Boxplots of homogeneity of normalized and modulated GM segments from no filled T1-MPRAGE images. a) Correlation of normalized modulated GM segments in MS patients' group with sample mean. **b)** Correlation of normalized modulated GM segments in controls' group with sample mean. Images obtained in CAT12.

a

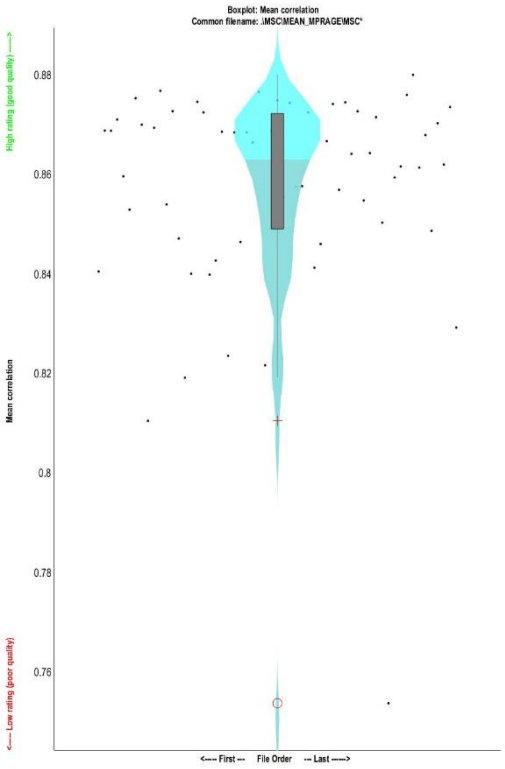
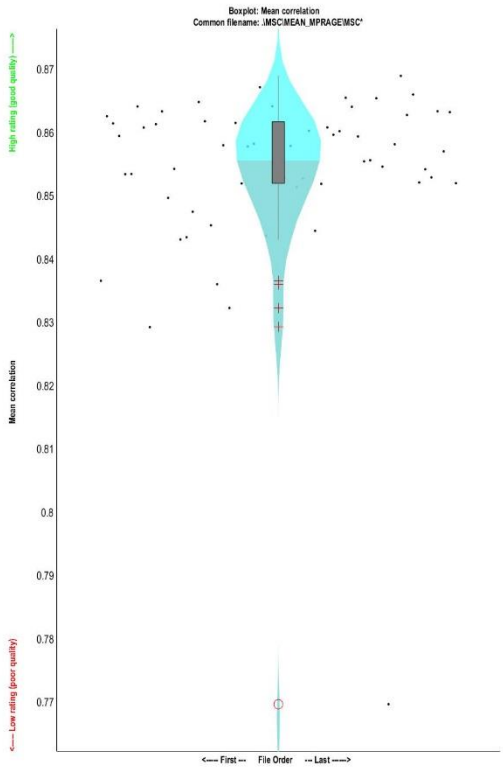


Figure B.2: Boxplots of homogeneity of normalized and modulated WM segments from no filled T1-MPRAGE images. a) Correlation of normalized modulated WM segments in MS patients' group with sample mean. **b)** Correlation of normalized modulated WM segments in controls' group with sample mean. Images obtained in CAT12.

a



b

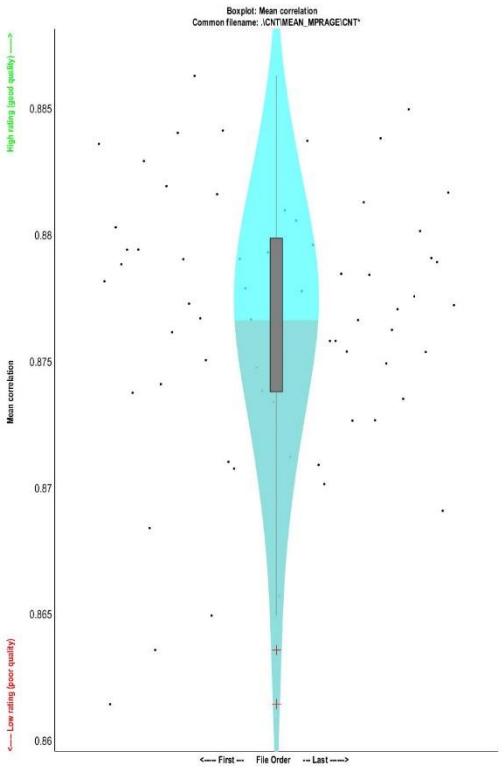


Figure B.3: Boxplots of homogeneity of normalized and modulated GM segments from filled T1-MPRAGE images. a) Correlation of normalized modulated GM segments in MS patients' group with sample mean. **b)** Correlation of normalized modulated GM segments in controls' group with sample mean. Images obtained in CAT12.

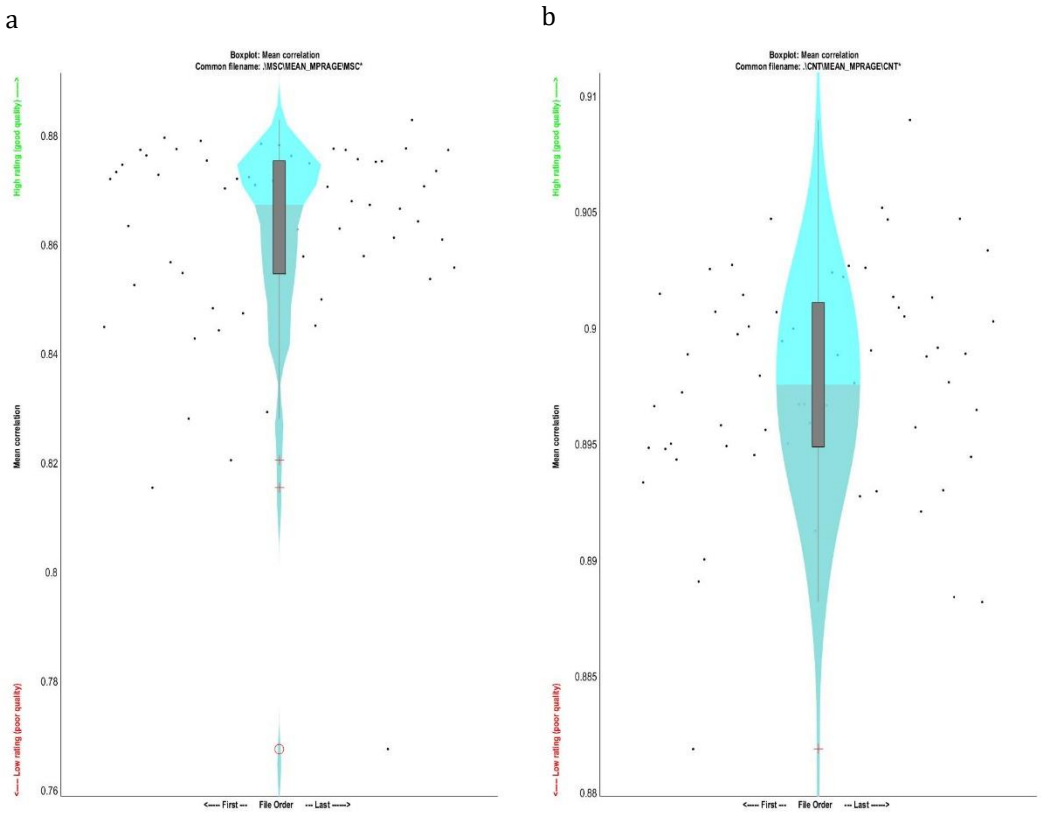


Figure B.4: Boxplots of homogeneity of normalized and modulated WM segments from filled T1-MPRAGE images. a) Correlation of normalized modulated WM segments in MS patients' group with sample mean. **b)** Correlation of normalized modulated WM segments in controls' group with sample mean. Images obtained in CAT12.

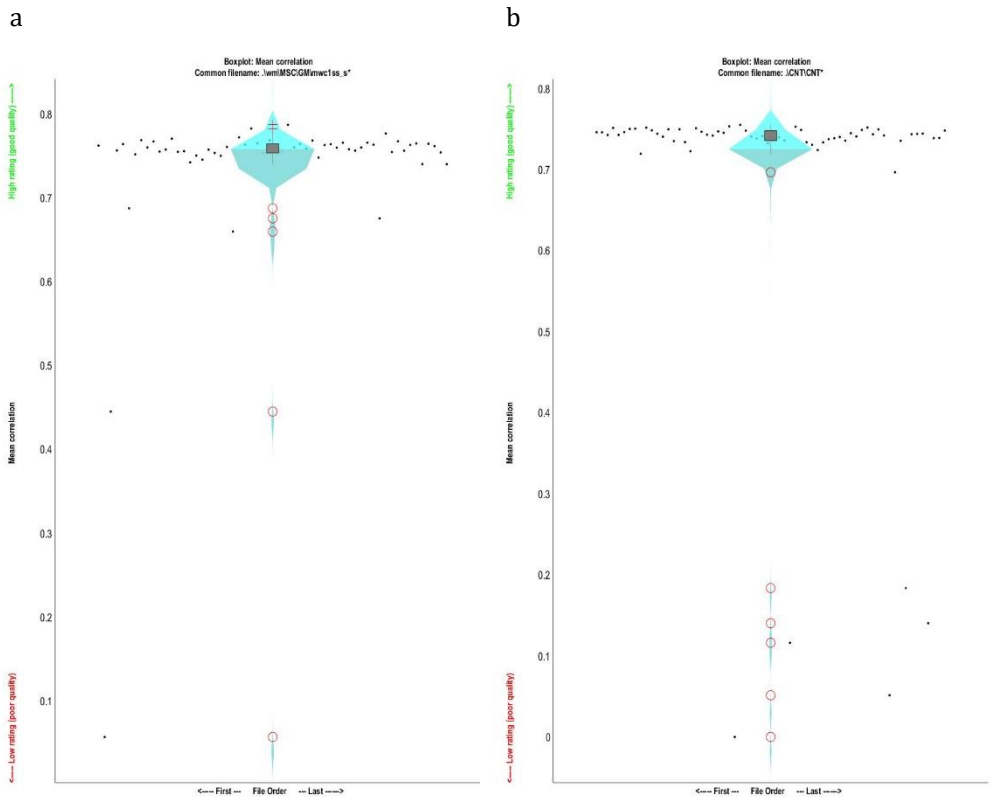


Figure B.5: Boxplots of homogeneity of normalized and modulated GM segments from T2-FLAIR images. a) Correlation of normalized modulated GM segments in MS patients' group with sample mean. **b)** Correlation of normalized modulated GM segments in controls' group with sample mean. Images obtained in CAT12.

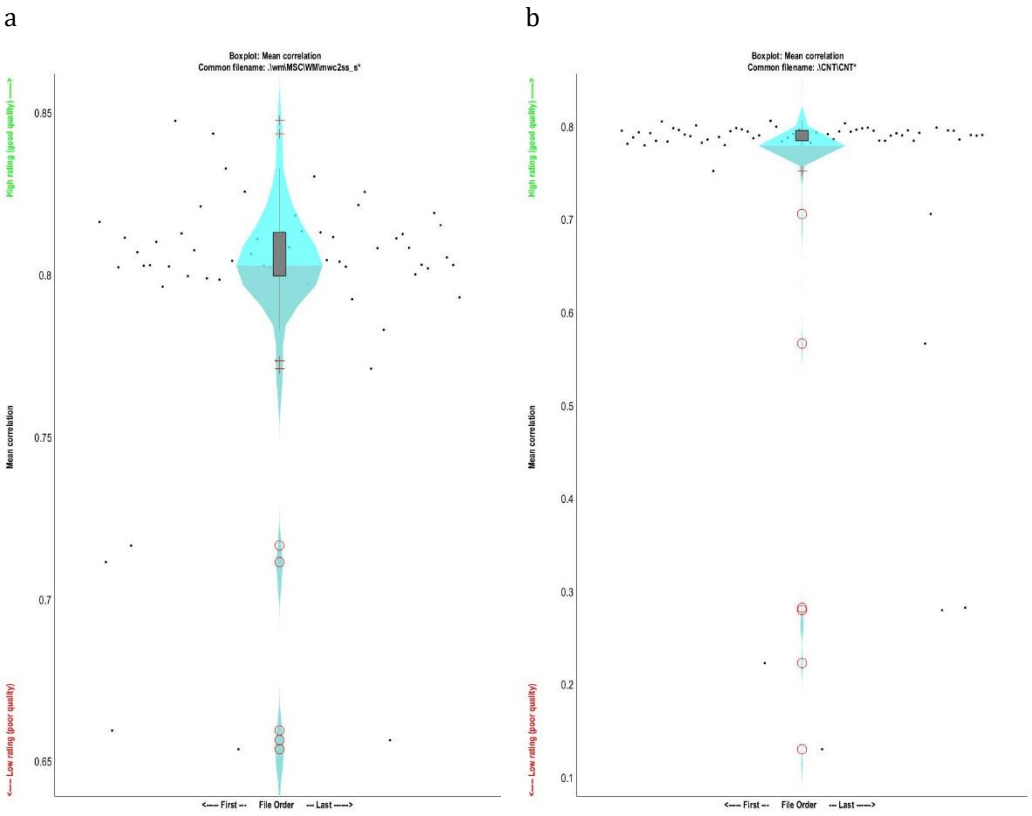


Figure B.6: Boxplots of homogeneity of normalized and modulated WM segments from T2-FLAIR images. a) Correlation of normalized modulated WM segments in MS patients' group with sample mean. **b)** Correlation of normalized modulated WM segments in controls' group with sample mean. Images obtained in CAT12.

Brain coordinates and corresponding regions

Table B.1: Standard space coordinates and corresponding regions with clusters yielding differences of GM volume extracted from no filled T1-MPRAGE images

	Coordinates (MNI) X, Y, Z	Coordinates (TAL) X, Y, Z	Brain Regions
Grey Matter	-15, -30, 5	-14, -30, 8	Left Thalamus, Pulvinar
	17, -30, 5	16, -30, 8	Right Thalamus, Pulvinar
	24, -6, -3	23, -8, 1	Right Lentiform Nucleus, Lateral Globus Pallidus
	-21, -3, -2	-20, -5, 1	Left Lentiform Nucleus, Lateral Globus Pallidus
	2, -11, 48	2, -8, 44	Right Cingulate Gyrus
	-2, -15, 56	-1, -12, 51	Left Medial Frontal Gyrus,
	-12, -57, -3	-11, -56, 1	Left Lingual Gyrus, Brodmann area 19
	35, 24, 42	35, 26, 40	Right Middle Frontal Gyrus, Brodmann area 8
	38, 30, 32	37, 31, 31	Right Middle Frontal Gyrus, Brodmann area 9
	6, 35, -24	5, 30, -22	Right Rectal Gyrus, Brodmann area 11
	38, -38, -24	36, -39, -16	Right Fusiform Gyrus, Brodmann area 20

Table B.2: Standard space coordinates and corresponding regions with clusters yielding differences of GM volume extracted from filled T1-MPRAGE images

	Coordinates (MNI) X, Y, Z	Coordinates (TAL) X, Y, Z	Brain Regions
Grey Matter	-18,-30, 8	-17, -30, 10	Left Thalamus, Pulvinar
	20, -29, 5	19, -29, 8	Right Thalamus, Pulvinar
	5, -8, 2	4, -9, 5	Right Cerebrum, Thalamus
	-8, -69, -9	-7, -67, -3	Left Cerebellum, Anterior Lobe, Culmen
	-15, -51, -5	-14, -51, 0	Left Lingual Gyrus
	29, -6, -8	28, -8, -3	Right Lentiform Nucleus, Putamen
	14, -47, -11	14, -47, -5	Right Cerebellum, Anterior Lobe, Culmen,

Table B.3: Standard space coordinates and corresponding regions with clusters yielding differences of GM volume extracted from T2-FLAIR images

Coordinates (MNI) X, Y, Z	Coordinates (TAL) X, Y, Z	Brain Regions

	9, -26, 11	9, -26, 13	Right Thalamus, Pulvinar
	-11, -27, 12	-11, -27, 14	Right Thalamus, Pulvinar
	11, 11, 11	10, 9, 12	Right Caudate, Caudate Body
	-9, 71, -3	-9, 65, -4	Left, Medial Frontal Gyrus, Brodmann area 10
	-6, 66, -14	-6, 60, -15	Left Medial Frontal Gyrus, Brodmann area 11
Grey Matter	8, 69, -8	7, 63, -9	Right, Superior Frontal Gyrus, Brodmann area 10
	0, -21, 48	0, -18, 44	Left Paracentral Lobule
	29, 32, 51	29, 35, 47	Right Middle Frontal Gyrus, Brodmann area 8
	54, 0, -41	52, -5, -32	Right Inferior Temporal Gyrus Brodmann area 20
	14, 62, 32	13, 60, 28	Right Superior Frontal Gyrus, Brodmann area 9
	42, -75, -21	43, -75, -13	Right Fusiform Gyrus, Brodmann area 19
	-2, 63, -17	-2, 56, -18	Left, Medial Frontal Gyrus, Brodmann area 11
	12, 59, 36	11, 58, 32	Right Superior Frontal Gyrus, Brodmann area 9

Table B.4: Standard space coordinates and corresponding regions with clusters yielding differences of WM volume extracted from no filled T1-MPRAGE images

	Coordinates (MNI) X, Y, Z	Coordinates (TAL) X, Y, Z	Brain Regions
	-29, -29, -8	-27, -30, -3	Left Parahippocampal Gyrus, White Matter
	30, -53, 20	30, -51, 21	Right Cerebrum, Temporal Lobe, Sub-Gyral, White Matter
	29, -27, -6	28, -28, 0	Right Cerebrum, Sub-lobar, Extra-Nuclear, White Matter
White Matter	-9, -36, -39	-8, -37, -29	Left Brainstem, Pons
	32, -21, -21	30, -23, -14	Right Parahippocampal Gyrus
	-14, 38, -12	-14, 33, -11	Left Cerebrum, Frontal Lobe, Sub-Gyral, White Matter
	11, -3, -15	10, -5, -9	Right Cerebrum, Sub-lobar, Extra-Nuclear, White Matter
	5, -27, -21	4, -28, -14	Right Brainstem, Midbrain
	-27, -3, -12	-26, -5, -7	Left Cerebrum, Sub-lobar, Extra-Nuclear, White Matter

Table B.5: Standard space coordinates and corresponding regions with clusters yielding differences of WM volume extracted from filled T1-MPRAGE images

	Coordinates (MNI) X, Y, Z	Coordinates (TAL) X, Y, Z	Brain Regions
White Matter	23, -24, 30	23, -23, 29	Right Cerebrum, Frontal Lobe, Sub-Gyral, White Matter

26, -36, 23	26, -35, 23	Right Cerebrum, Sub-lobar, Extra-Nuclear
33, -9, 26	32, -9, 26	Right Cerebrum, Frontal Lobe, Sub-Gyral
-29, -27, -9	-27, -28, -3	Left Hippocampus
-44, -33, -8	-41, -34, -2	Left Hippocampus
-39, -33, 2	-37, -33, 5	Left Hippocampus
-26, -26, 32	-25, -25, 31	Left Cerebrum, Frontal Lobe, Sub-Gyral
-24, -44, 29	-23, -42, 28	Left Cerebrum, Frontal Lobe, Sub-Gyral
-36, -11, 27	-35, -11, 26	Left Precentral Gyrus
27, -26, -8	26, -27, -2	Right Cerebrum, Sub-lobar, Extra-Nuclear
21, -75, 11	21, -72, 13	Right Cuneus
14, -41, 8	14, -40, 11	Right Cerebrum, Sub-lobar, Extra-Nuclear, Corpus Callosum
-14, -42, 6	13, -41, 9	Right Cerebrum, Sub-lobar, Extra-Nuclear, Corpus Callosum
-20, -9, 35	-19, -8, 33	Left Cerebrum, Frontal Lobe, Sub-Gyral
-27, -57, 18	-26, -55, 19	Left Cerebrum, Temporal Lobe, Sub-Gyral

Table B.6: Standard space coordinates and corresponding regions with clusters yielding differences of WM volume extracted from T2-FLAIR images

	Coordinates (MNI)	Coordinates (TAL)	Regions
White Matter	27, -50, 15	25, -48, 17	Right Cerebrum, Temporal Lobe, Sub-Gyral
	30, -42, 17	29, -41, 18	Right Cerebrum, Sub-lobar, Extra-Nuclear
	-14, -5, -15	13, -7, -9	Right Brainstem, Midbrain

C. Surface-Based Morphometry Results

Sample Homogeneity Values

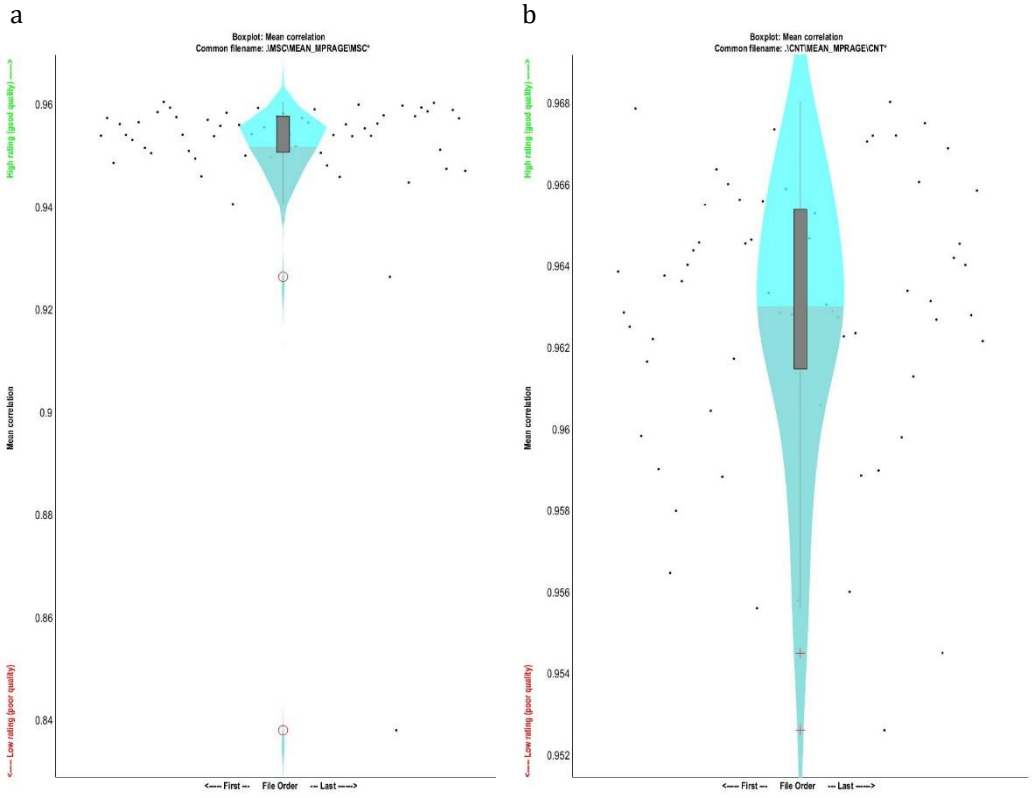


Figure C.1: Boxplots of homogeneity of smoothed cortical thickness extracted from no filled T1-MPRAGE images. a) Correlation of smoothed cortical thickness in MS patients' group with sample mean. **b)** Correlation of smoothed cortical thickness in controls' group with sample mean. Images obtained in CAT12.

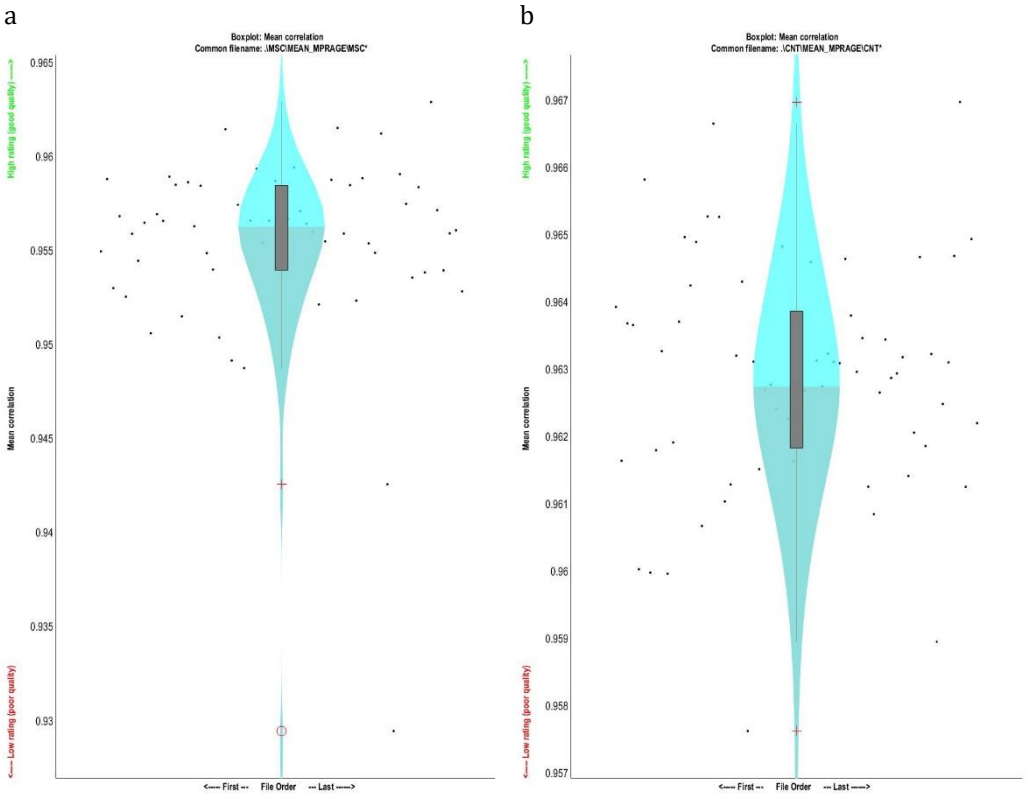


Figure C.2: Boxplots of homogeneity of smoothed gyrification index extracted from no filled T1-MPRAGE images. a) Correlation of smoothed gyrification index in MS patients' group with sample mean. **b)** Correlation of smoothed gyrification index in controls' group with sample mean. Images obtained in CAT12.

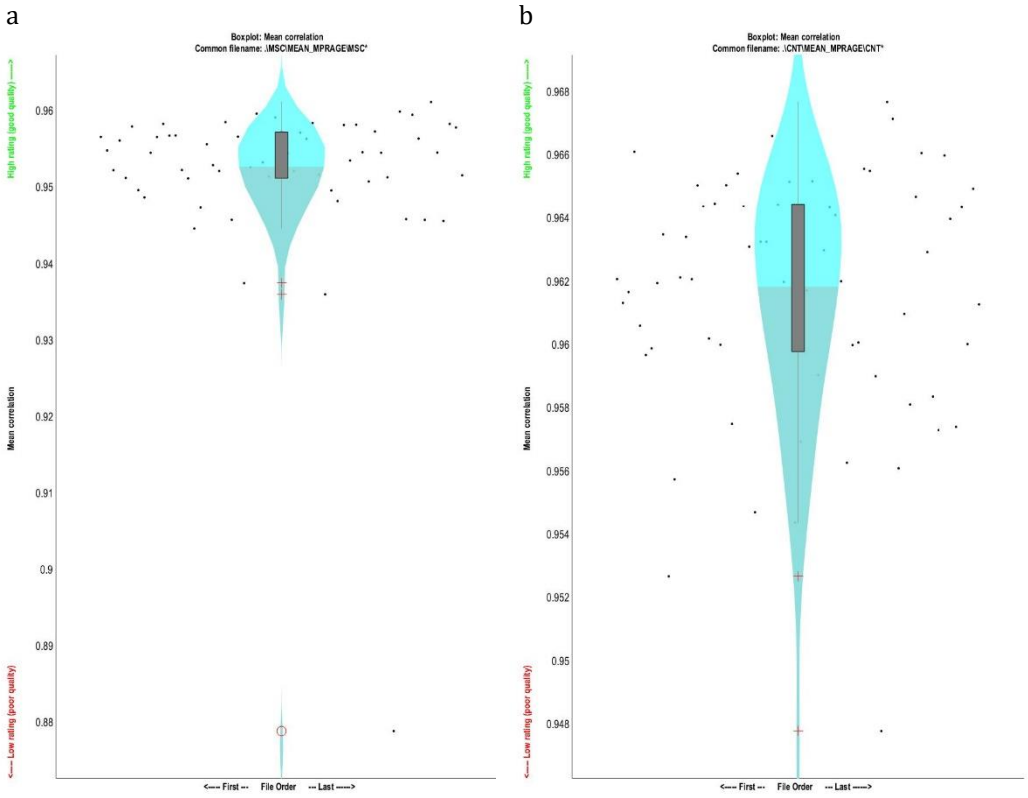


Figure C.3: Boxplots of homogeneity of smoothed cortical thickness extracted from filled T1-MPRAGE images. a) Correlation of smoothed cortical thickness in MS patients' group with sample mean. **b)** Correlation of smoothed cortical thickness in controls' group with sample mean. Images obtained in CAT12.

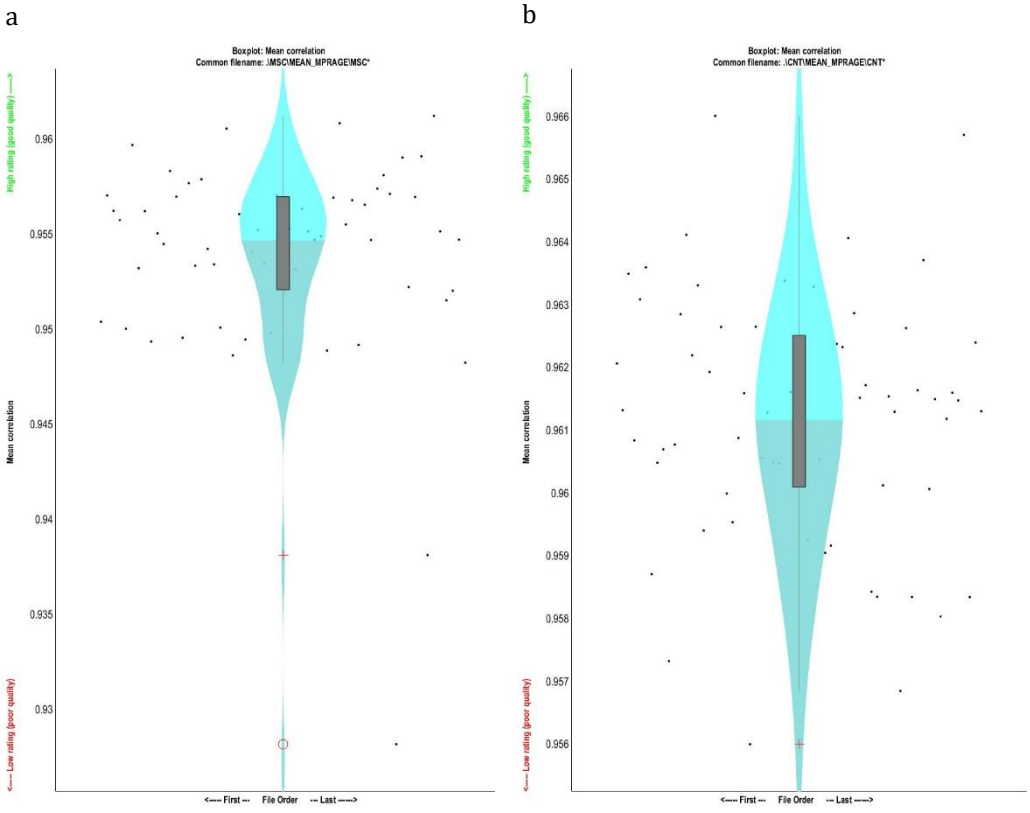


Figure C.4: Boxplots of homogeneity of smoothed gyrification index extracted from filled T1-MPRAGE images. a) Correlation of smoothed gyrification index in MS patients' group with sample mean. **b)** Correlation of smoothed gyrification index in controls' group with sample mean. Images obtained in CAT12.

Brain Coordinates and Corresponding Regions

Table C.1: Standard space coordinates and corresponding regions with clusters yielding differences of cortical thickness extracted from no filled T1-MPRAGE images

	Coordinates (MNI) X, Y, Z	Coordinates (TAL) X, Y, Z	Brain Regions
Cortical Thickness	60, -34, 4	57, -34, 8	Right Superior Temporal Gyrus
	49, -30, 27	48, -28, 27	Right Inferior Parietal Lobule
	50, -41, -2	48, -41, 3	Right Cerebrum, Temporal Lobe, Sub-Gyral
	61, -35, -6	58, -36, 0	Right Middle Temporal Gyrus
	39, 12, -3	37, 8, 1	Right Cerebrum, Sub-lobar, Insula
	44, 18, -26	42, 12, -18	Right Superior Temporal Gyrus, Brodmann area 38
	-26, -11, -27	-25, -13, -19	Left Parahippocampal Gyrus
	-21, -5, -33	-21, -8, -25	Left Parahippocampal Gyrus
	-46, -18, 35	-45, -17, 33	Left Postcentral Gyrus
	-42, 3, -23	-41, -1, -17	Left Temporal Lobe, Sub-Gyral
	-46, -18, 35	-45, -17, 33	Left Frontal Lobe, Postcentral Gyrus
	-36, -21, 43	-35, -19, 40	Left Frontal Lobe, Sub-Gyral
	-27, -29, 53	-27, -26, 49	Left Precentral Gyrus, Brodmann area 4
	16, -60, 3	16, -59, 7	Right Posterior Cingulate
	31, -52, -18	30, -52, -11	Right Cerebellum, Posterior Lobe, Declive
	34, -74, 41	35, -70, 38	Right Precuneus
	-10, -65, 3	10, -63, 7	Right Cuneus
	-15, -42, 59	-15, -38, 53	Left Paracentral Lobule, Brodmann area 3
	-21, -47, 63	-21, -43, 57	Left Parietal Lobe, Sub-Gyral
	-32, 24, 44	-31, 25, 41	Left Middle Frontal Gyrus, Brodmann area 8
	28, -13, 53	28, -9, 49	Right Precentral Gyrus, Brodmann area 6
	23, -2, 50	23, 0, 46	Right Cerebrum, Frontal Lobe, Sub-Gyral
	28, 35, 32	27, 35, 30	Right Superior Frontal Gyrus
	4, -81, 26	4, -77, 25	Right Cuneus, Brodmann area 18
	18, -98, 15	18, -94, 17	Right Middle Occipital Gyrus
	19, -69, 32	19, -66, 30	Right Precuneus
	-55, -17, -17	-51, -19, -11	Left Middle Temporal Gyrus

	-47, -29, 51	-47, -26, 47	Left Postcentral Gyrus, Brodmann area 2
	28, 35, 32	27, 35, 30	Right Superior Frontal Gyrus
	29, 21, 46	29, 23, 43	Right Middle Frontal Gyrus, Brodmann area 8
	45, -28, 50	45, -25, 46	Right Postcentral Gyrus, Brodmann area 2
	52, -37, 48	52, -34, 45	Right Inferior Parietal Lobule, White Matter
	15, -44, 57	16, -40, 51	Right Paracentral Lobule, Brodmann area 7
	11, -33, 49	11, -30, 45	Right Frontal Lobe, Sub-Gyral
	-47, -35, 38	-46, -33, 36	Left Inferior Parietal Lobule, Brodmann area 40
	-41, -24, 3	-39, -25, 6	Left Superior Temporal Gyrus
	-39, -7, 2	-37, -8, 5	Left Insula

Table C.2: Standard space coordinates and corresponding regions with clusters yielding differences of cortical thickness extracted from filled T1-MPRAGE images

	Coordinates (MNI) X, Y, Z	**Coordinates (TAL) X, Y, Z**	**Brain Regions**
Cortical Thickness	57,-34,4	55, -34, 8	Right Superior Temporal Gyrus
	31, -52, -18	30, -52, -11	Right Cerebellum Posterior Lobe, Declive
	-56, -19, -15	-52, -21, -9	Left Middle Temporal Gyrus
	-29, 32, 39	-28, 32, 36	Left Middle Frontal Gyrus, Brodmann area 9
	-31, 25, 42	-30, 26, 39	Left Middle Frontal Gyrus, Brodmann area 8
	48, -30, 26	47, -29, 26	Right Inferior Parietal Lobule
	44, 18, -26	42, 12, -18	Right Superior Temporal Gyrus, Brodmann area 38
	8, -59, 2	8, -58, 6	Right Posterior Cingulate
	-35, -22, 43	-34, -20, 40	Left Frontal Lobe, Sub-Gyral
	48, -41, 1	46, -41, 3	Right Temporal Lobe, Sub-Gyral
	-43, -19, 34	-42, -18, 32	Left Postcentral Gyrus
	-21, -32, 57	-21, -29, 52	Left Postcentral Gyrus
	-1, -19, 28	0, -18, 27	Left Cingulate Gyrus
	-10, -57, 0	-9, -56, 4	Left Lingual Gyrus, Brodmann area 18
	23, 0, 50	23, 2, 46	Right Superior Frontal Gyrus
	-16, -42, 58	-16, -39, 53	Left Paracentral Lobule
	-46, 7, -19	-44, 2, -13	Left Superior Temporal Gyrus
	-43, -33, 50	-43, -30, 46	Left Inferior Parietal Lobule, Brodmann area 40
	15, -44, 57	16, -40, 51	Right Paracentral Lobule, Brodmann area 7

Table C.3: Standard space coordinates and corresponding regions with clusters yielding differences of decreased (controls > MS patients) gyrification index extracted from no filled T1-MPRAGE images

	Coordinates (MNI) X, Y, Z	Coordinates (TAL) X, Y, Z	Brain Regions
Decreased Gyrification index	-46, -14, -8	-43, -16, -3	Left Insula
	48, -15, 36	47, -13, 35	Right Precentral Gyrus
	45, -17, 5	42, -18, 0	Right Insula, Brodmann area 13
	-34, 2, 9	-33, 0, 10	Left Insula
	-32, -26, 14	-31, -26, 15	Left Insula, Brodmann area 13
	41, -24, 53	41, -20, 49	Right Postcentral Gyrus, Brodmann area 3
	-49, -18, 38	-48, -17, 36	Left Precentral Gyrus
	38, -20, 59	39, -16, 54	Right Precentral Gyrus, Brodmann area 4
	-9, -29, 74	-9, -24, 66	Left Precentral Gyrus
	-20, 0, 58	-20, 3, 53	Left Frontal Lobe, Sub-Gyral, Brodmann area 6
	-51, -34, 9	-48, -34, 11	Left Superior Temporal Gyrus
	54, -5, 33	52, -4, 32	Right Precentral Gyrus
	-36, -22, 63	-36, -18, 57	Left Precentral Gyrus
	14, -66, 4	14, -64, 7	Right Cuneus
	6, -78, 4	6, -75, 7	Right Cuneus
	57, -36, 5	55, -36, 9	Right Superior Temporal Gyrus
	13, -25, 71	13, -20, 64	Right Precentral Gyrus
	-32, -79, 26	-32, -76, 26	Left Superior Occipital Gyrus
	-12, -70, 3	-12, -68, 6	Left Cuneus
	2, -31, 69	2, -27, 62	Right Medial Frontal Gyrus
	18, -43, 69	19, -38, 62	Right Postcentral Gyrus
	3, -23, 27	3, -22, 26	Right Cingulate Gyrus, Brodmann area 23
	-47, -29, 49	-47, -26, 45	Left Postcentral Gyrus, Brodmann area 2
	14, -37, 41	14, -34, 38	Right Cingulate Gyrus
	45, -36, 17	44, -35, 18	Right Insula, Brodmann area 13

Table C.4: Standard space coordinates and corresponding regions with clusters yielding increased (MS patients > controls) gyrification index extracted from no filled T1-MPRAGE images

	Coordinates (MNI) X, Y, Z	Coordinates (TAL) X, Y, Z	Brain Regions

	24, 34, -12	23, 29, -9	Right Inferior Frontal Gyrus, Brodmann area 47
	-34, -19, 40	-33, -17, 37	Left Precentral Gyrus
	-27, -26, 51	-27, -23, 47	Left Precentral Gyrus
	34, -18, 41	34, -16, 39	Right Precentral Gyrus
	-24, 36, -12	-24, 31, -9	Left Inferior Frontal Gyrus
	35, -46, -10	34, -46, -4	Right Parahippocampal Gyrus, Brodmann area 19
	37, -24, -25	35, -26, -17	Right Parahippocampal Gyrus
	65, -33, 15	62, -32, 17	Right Superior Temporal Gyrus
	30, 8, -41	29, 2, -32	Right Temporal Lobe, Sub-Gyral
	30, -4, -35	29, -7, -26	Right Parahippocampal Gyrus
	39, 18, -36	37, 11, -28	Right Superior Temporal Gyrus
	-31, 1, -43	-31, -3, -35	Left Inferior Temporal Gyrus
	-25, -13, -27	-24, -15, -19	Left Parahippocampal Gyrus
	10, 45, -22	9, 39, -22	Right Frontal Lobe, Sub-Gyral
	12, 13, -16	11, 9, -11	Right Frontal Lobe, Subcallosal Gyrus
	-23, -71, 6	-22, -69, 9	Left Occipital Lobe, Cuneus
	-6, -77, -2	-5, -75, 2	Left Lingual Gyrus, Brodmann area 18
	45, -30, -4	43, -31, 1	Right Temporal Lobe, Sub-Gyral
	11, -7, 45	11, -5, 41	Right Cingulate Gyrus
	11, -93, 13	11, -86, 16	Right Cuneus
	-6, 41, -24	-6, 35, -23	Left Rectal Gyrus
	-5, 61, -18	-5, 54, -19	Left Superior Frontal Gyrus, Brodmann area 11
Increased Gyrification index	-29, 28, 5	-28, 25, 6	Left Frontal Lobe, Sub-Gyral
	-62, -51, 18	-60, -50, 20	Left Superior Temporal Gyrus
	6, -37, 75	6, -32, 67	Right Paracentral Lobule
	19, -45, -4	19, -45, 1	Right Parahippocampal Gyrus
	-48, -38, -2	-45, -38, 2	Left Temporal Lobe, Sub-Gyral
	60, -9, 20	57, -9, 21	Right Parietal Lobe, Postcentral Gyrus
	7, -43, 24	7, -41, 24	Right Posterior Cingulate
	24, 59, -8	23, 54, -8	Right Middle Frontal Gyrus, Brodmann area 10
	-13, -16, 42	-12, -14, 39	Left Cingulate Gyrus, Brodmann area 24
	-11, -5, 43	-10, -3, 40	Left Cingulate Gyrus
	20, -75, 45	21, -71, 41	Right Precuneus
	-35, -25, -24	-33, -26, -17	Left Parahippocampal Gyrus
	-34, -45, -10	-32, -45, -4	Left Parahippocampal Gyrus, Brodmann area 19
	63, -6, 0	59, -8, 4	Right Superior Temporal Gyrus
	-12, -1, 70	-12, 3, 63	Left Superior Frontal Gyrus
	47, -1, 47	46, 1, 45	Right Precentral Gyrus
	-59, -9, 18	-56, -10, 19	Left Postcentral Gyrus

	-10, 9, -17	-10, 5, -12	Left Frontal Lobe, Subcallosal Gyrus
	-6, -40, 74	-5, -35, 66	Left Paracentral Lobule, Brodmann area 4
	-54, -20, -28	-51, -22, -21	Left Inferior Temporal Gyrus
	11, 29, 57	11, 32, 52	Right Superior Frontal Gyrus
	34, 8, 29	33, 8, 28	Right Frontal Lobe, Sub-Gyral
	14, -90, -8	14, -88, -2	Right Lingual Gyrus
	33, 19, 11	31, 17, 13	Right Insula,
	32, -64, 45	33, -60, 41	Right Inferior Parietal lobule
	-38, 22, 22	-37, 21, 22	Left Frontal Lobe, Sub-Gyral
	17, -63, 63	18, -59, 56	Right Precuneus, Brodmann area 7
	45, -46, 15	44, -45, 17	Right Superior Temporal Gyrus
	21, -27, 54	21, -23, 49	Right Precentral Gyrus
	-35, 55, -7	-35, 51, -5	Left Middle Frontal Gyrus
	22, -61, 22	22, -59, 22	Right Precuneus

Table C.5: Standard space coordinates and corresponding regions with clusters yielding decreased (controls > MS patients) gyrification index extracted from filled T1-MPRAGE images

	Coordinates (MNI) X, Y, Z	Coordinates (TAL) X, Y, Z	Brain Regions
Decreased Gyrification index	2, -13, 29	2, -12, 28	Right Cingulate Gyrus
	-35, -10, 11	-33, -11, 12	Left Insula
	38, -18, 10	36, -18, 12	Right Insula

Table C.6: Standard space coordinates and corresponding regions with clusters yielding increased (MS patients > controls) gyrification index extracted from filled T1-MPRAGE images

	Coordinates (MNI) X, Y, Z	Coordinates (TAL) X, Y, Z	Brain Regions
Increased Gyrification index	23,-3,-34	22, -6, -26	Right Parahippocampal Gyrus, Brodmann area 35
	38, 15, -39	37, 8, -30	Right Superior Temporal Gyrus
	29, -12, -33	28, -15, -24	Right Parahippocampal Gyrus, Brodmann area 35
	14, 33, -27	13, 27, -24	Right Orbital Gyrus, Brodmann area 47
	5, 36, -26	4, 30, -24	Right Rectal Gyrus, Brodmann area 11
	33, -43, -14	32, -44, -7	Right, Parahippocampal Gyrus, Brodmann area 37
	7,33,56	7, 36, 50	Right Superior Frontal Gyrus
	-22, -1, -34	-22, -4, -26	Left Uncus

41, -24, -23	39, -26, -15	Right Parahippocampal Gyrus, Brodmann area 36
52, -18, -37	49, -21, -28	Right Inferior Temporal Gyrus, Brodmann area 20

Table C.7: Standard space coordinates and corresponding regions with clusters yielding increased (MS patients > controls) gyrification index in filled T1-MPRAGE images, corrected for multiple comparisons ($p<0.05$, FWE corrected).

	Coordinates (MNI) X, Y, Z	Coordinates TAL X, Y, Z	Brain Regions
Increased Gyrification index	24, 34, -12	23, 29, -9	Right Inferior Frontal Gyrus, Brodmann area 47
	-34, -19, 40	-33, -17, 37	Left Frontal Lobe, Sub-Gyral
	34, -18, 41	34, -16, 39	Right Frontal Lobe, Sub-Gyral
	-24, 36, -12	-24, 31, -9	Left Inferior Frontal Gyrus
	35, -46, -10	34, -46, -4	Right Parahippocampal Gyrus, Brodmann area 19
	65, -33, 15	62, -32, 17	Right Superior Temporal Gyrus
	30, 8, -41	29, 2, -32	Right Temporal Lobe, Sub-Gyral
	30, -4, -35	29, -7, -26	Right Parahippocampal Gyrus
	39, 18, -36	37, 11, -28	Right Superior Temporal Gyrus
	-31, 1, -43	-31, -3, -35	Left Inferior Temporal Gyrus
	10, 45, -22	10, 45, -22	Right Frontal Lobe, Sub-Gyral
	-25, -13, -27	-24, -15, -19	Left Parahippocampal Gyrus
	-23, -71, 6	-22, -69, 9	Left Cuneus
	-6, -77, -2	-5, -75, 2	Left Lingual Gyrus, Brodmann area 18
	45, -30, -4	43, -30, 7	Right Superior Temporal Gyrus Brodmann area 41
	11, -7, 45	11, -5, 41	Right Cingulate Gyrus
	11, -93, 15	11, -89, 16	Right Middle Occipital Gyrus, Brodmann area 18
	24, -8, -29	23, -11, -21	Right Parahippocampal Gyrus
	54, -17, -28	51, -20, -20	Right Inferior Temporal Gyrus

Table C.8: Standard space coordinates and corresponding regions with clusters yielding increased (MS patients > controls) gyrification index extracted from filled T1-MPRAGE images, corrected for multiple comparisons ($p<0.05$, FWE corrected).

	Coordinates (MNI) X, Y, Z	Coordinates (TAL) X, Y, Z	Brain Regions
Increased Gyrification	23, -3, -34	22, -6, -26	Right Parahippocampal Gyrus, Brodmann area 35

Gyrification Maps

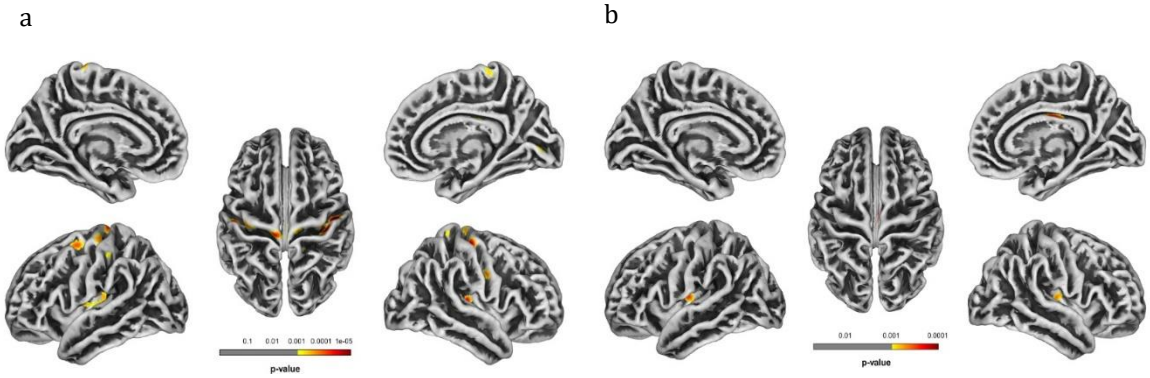


Figure C.5: Results of SBM analysis of decreased (controls > MS patients) gyrification index in MS, extracted from no filled T1-MPRAGE images **(a)** and from filled T1-MPRAGE images **(b)**. Results are presented at a voxel-level *p*-value < 0.001 (uncorrected). Color bar scale represents log *p*-values. Images obtained in CAT12.

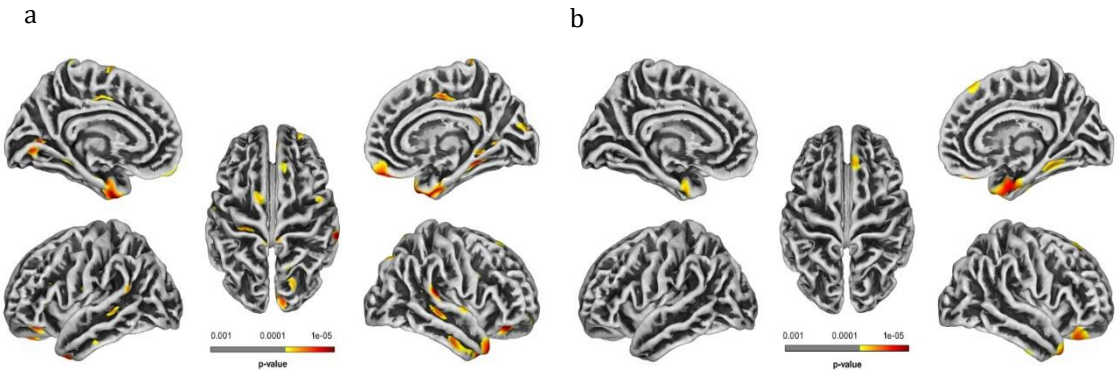


Figure C.6: Results of SBM analysis of increased (MS patients > controls) gyrification index in MS, extracted from no filled T1-MPRAGE images **(a)** and from filled T1-MPRAGE images **(b)** Results are presented at a voxel-level *p*-value < 0.001 (uncorrected). Color bar scale represents log *p*-values. Images obtained in CAT12.

a b

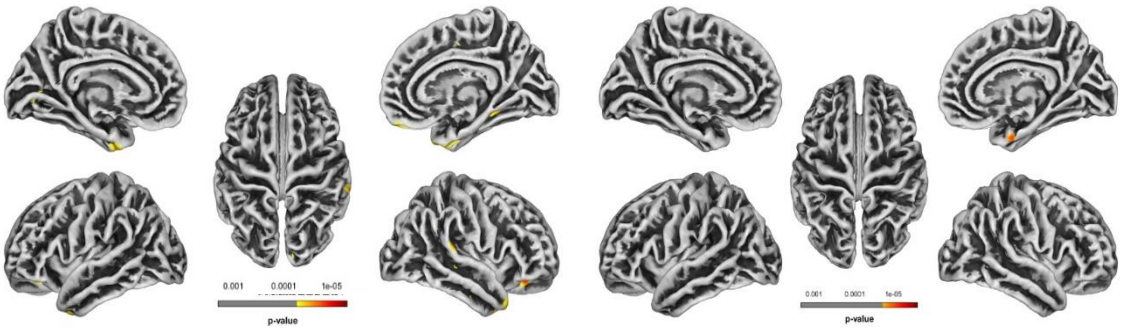


Figure C.7: Results of SBM analysis of increased (MS patients > controls) gyrification index in MS, extracted from no filled T1-MPRAGE images **(a)** and from filled T1-MPRAGE images **(b)** Results are presented at a voxel-level *p*-value < 0.005 (FWE corrected). Color bar scale represents log *p*-values. Images obtained in CAT12.

D. Support Vector Machine Classification Results

Weights maps

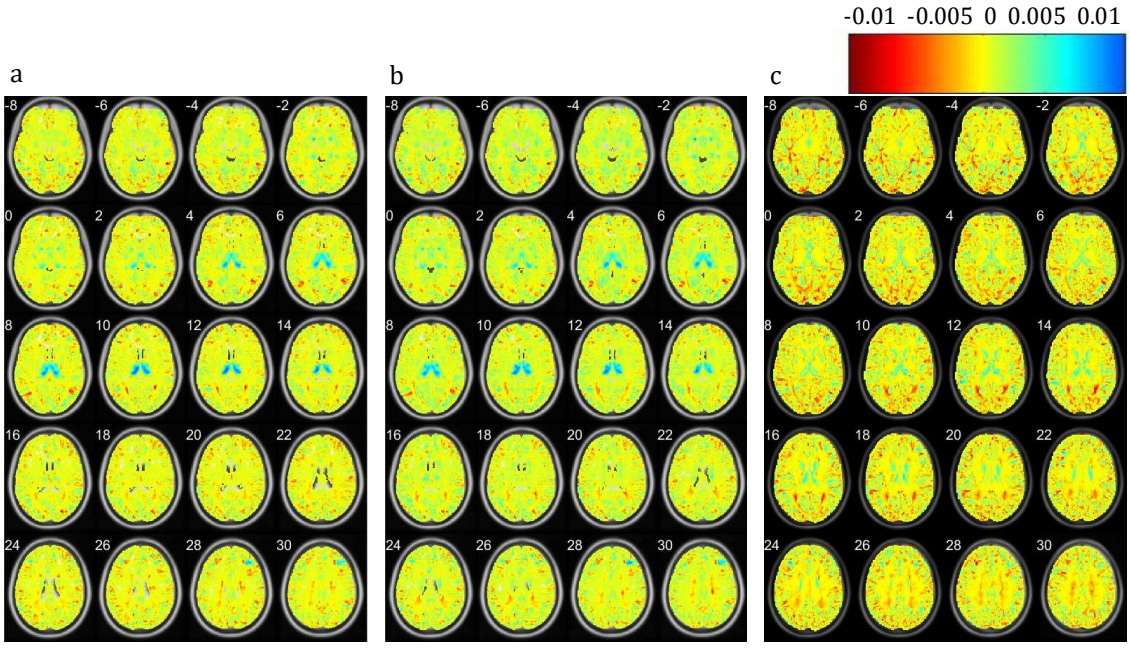


Figure D.1: Whole-brain representation of the discriminative map obtained from SVM classification of GM volume. In color bar scale, positively weighted voxels are displayed in blue/green and negatively weighted voxels are displayed in red/orange**. a)** Weight maps obtained with classification of GM segments extracted from no filled T1-MPRAGE images. **b)** Weight maps obtained with classification of GM segments extracted from filled T1-MPRAGE images. **c)** Weight maps obtained with classification of GM segments extracted from T2-FLAIR images.

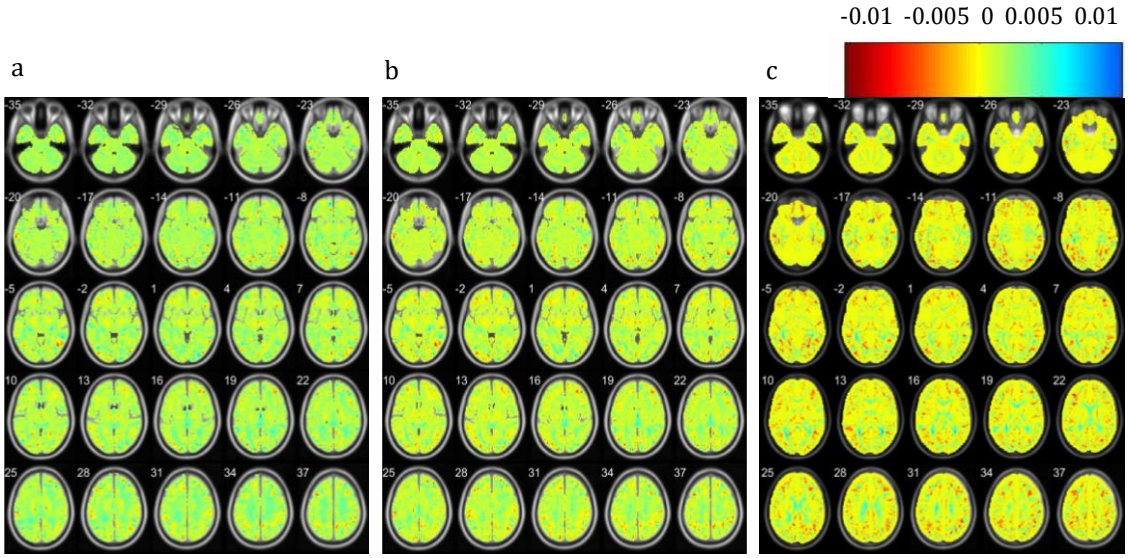


Figure D.2: Whole-brain representation of the discriminative map obtained from SVM classification of WM volume. In color bar scale, positively weighted voxels are displayed in blue/green and negatively weighted voxels are displayed in red/orange**. a)** Weight maps obtained with classification of WM segments extracted from no filled T1-MPRAGE images. **b)** Weight maps obtained with classification of WM segments extracted from filled T1-MPRAGE images. **c)** Weight maps obtained with classification of WM segments extracted from T2-FLAIR images.