frontiers

Exploratory radiomic analysis of conventional versus quantitative brain MRI: Towards automatic diagnosis of early multiple sclerosis

Elizaveta Lavrova[1, 2], Emilie Lommers[1, 3], Henry C. Woodruff[2, 4], Pierre Maquet[1, 3], Eric Salmon[1], Philippe Lambin[2, 4*], Christophe Phillips[1, 5]

[1]GIGA Cyclotron research center - In vivo imaging, University of Liège, Belgium, [2]The D-Lab, Department of Precision Medicine, GROW – School for Oncology, Maastricht University, Netherlands, [3]Clinical Neuroimmunology Unit, Neurology Department, University Hospital Center of Liège, Belgium, [4]Department of Radiology and Nuclear Medicine, Maastricht University Medical Centre, Netherlands, [5]GIGA-In silico Medicine, University of Liège, Belgium

Submitted to Journal:
Frontiers in Neuroscience

Specialty Section:
Brain Imaging Methods

Article type:
Original Research Article

Manuscript ID:
679941

Received on:
12 Mar 2021

Revised on:
23 May 2021

Journal website link:
www.frontiersin.org

Conflict of interest statement

The authors declare a potential conflict of interest and state it below

PL reports — within and outside the submitted work—grants or sponsored research agreements from Varian Medical, Oncoradiomics, ptTheragnostic/DNAmito, and Health Innovation Ventures. He received an advisor/presenter fee and/or reimbursements of travel costs/external grant writing fee and/or in-kind manpower contribution from Oncoradiomics, BHV, Merck, Varian, Elekta, ptTheragnostic, and Convert Pharmaceuticals. P.L. has shares in the company Oncoradiomics SA, Convert Pharmaceuticals, and The Medical Cloud Company SPRL, and is co-inventor of two issues patents with royalties on radiomics (PCT/NL2014/050248, PCT/NL2014/050728) licensed to Oncoradiomics, one issue patent on mtDNA (PCT/EP2014/059089) licensed to ptTheragnostic/DNAmito, three non-patented inventions (software) licensed to ptTheragnostic/DNAmito and Oncoradiomics and Health Innovation Ventures, and three non-issues, non-licensed patents on Deep Learning-Radiomics and LSRT (N2024482, N2024889, N2024889). HW has (minority) shares in the company Oncoradiomics.

Author contribution statement

ELa, HW, CP, ES, and PL contributed to study conceptualization. PL and ES acquired the funding. ELo, PM, and CP performed data acquisition and curation. ELa, HW, PL, CP, and ES developed methodology. ELa, HW, and CP performed analysis. ELa, HW, and CP wrote the original draft. All the authors contributed to manuscript revision, read, and approved the submitted version. The supervision was performed by HW, CP, PL, ES.

Keywords

Multiple Sclerosis, brain MRI, Radiomic feature, quantitative MRI (qMRI), Histological MRI

Abstract

Word count: 339

Conventional magnetic resonance imaging (cMRI) is poorly sensitive to pathological changes related to multiple sclerosis (MS) in normal-appearing white matter (NAWM) and grey matter (GM), with the added difficulty of not being very reproducible. Quantitative MRI (qMRI) on the other hand attempts to represent physical properties of tissues, making it an ideal candidate for quantitative medical image analysis, or radiomics. We therefore hypothesized that qMRI-based radiomic features have added diagnostic value in MS compared to cMRI. This study investigated the ability of cMRI (T1w) and qMRI features extracted from WM, NAWM, and GM to distinguish between MS patients (MSP) and healthy control subjects (HCS). We developed exploratory radiomic classification models on a dataset comprising 36 MSP and 36 HCS recruited in CHU Liege, Belgium, acquired with cMRI and qMRI. For each image type and region of interest, qMRI radiomic models for MS diagnosis were developed on a training subset and validated on a testing subset. Radiomic models based on cMRI were developed on the entire training dataset and externally validated on open-source datasets with 167 HCS and 10 MSP. Ranked by region of interest, the best diagnostic performance was achieved in the whole WM. Here the model based on magnetization transfer imaging (a type of qMRI) features yielded a median area under the receiver operating characteristic curve (AUC) of 1.00 in the testing sub-cohort. Ranked by image type, the best performance was achieved by the magnetization transfer models, with median AUCs of 0.79 (0.69-0.90 90% CI) in NAWM and 0.81 (0.71-0.90) in GM. External validation of the T1w models yielded an AUC of 0.78 (0.47-1.00) in whole WM, demonstrating a large 95% CI and low sensitivity of 0.30 (0.10-0.70). This exploratory study indicates that qMRI Radiomics could provide efficient diagnostic information using NAWM and GM analysis in MSP. T1w radiomics could be useful for a fast and automated check of conventional MRI for WM abnormalities once acquisition and reconstruction heterogeneities have been overcome. Further prospective validation is needed involving more data for better interpretation and generalization of the results.

Contribution to the field

Multiple sclerosis is a neurodegenerative disorder of the central nervous system, leading to physical and mental disability. It is essential to diagnose it at the early stages of the demyelination process. The reliable biomarkers are still under development. The demyelination leads to focal white matter lesions, originating the symptoms and being detectable with magnetic resonance imaging. However, clinical magnetic resonance images are expressed in arbitrary units, depending on many factors, which disturbs data comparison in multi-center studies. Additionally, they represent only the visual contrast, which compromises the objective analysis. Moreover, recent studies showed that diffuse damages appear at the early stages of the disease. However, clinical magnetic resonance imaging is not sensitive to these changes. Thus, we combined quantitative magnetic resonance imaging and radiomics to develop a reproducible and objective approach for diagnosing multiple sclerosis. For this, we used a unique dataset containing both clinical and quantitative magnetic resonance imaging of both multiple sclerosis patients and healthy control subjects. We used radiomic features extracted from different brain tissues and different image types to train machine-learning

models for binary classification between multiple sclerosis patients and healthy control subjects. We compared the models based on clinical and quantitative magnetic resonance images. For models based on clinical magnetic resonance images, we performed an external validation. We could not validate models based on quantitative magnetic resonance images because of the data uniqueness at the moment. Therefore, the performed study is exploratory, demonstrating the potential of quantitative magnetic resonance imaging and radiomics in multiple sclerosis studies. We reported the current limitations that are indicating directions for further research.

Funding statement

ELa is in a Maastricht-Liege University Imaging Valley PhD program.

Ethics statements

Studies involving animal subjects
Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects
Generated Statement: The studies involving human participants were reviewed and approved by Liege University Hospital-Faculty Ethics Committee, University Hospital Center of Liege, Liege, Belgium. The patients/participants provided their written informed consent to participate in this study.

Inclusion of identifiable human data
Generated Statement: No potentially identifiable human images or data is presented in this study.

Data availability statement

Generated Statement: The data analyzed in this study is subject to the following licenses/restrictions: DS2 and DS3 are public datasets, the accession details can be found in (Li et al., 2019), and (Souza et al., 2018). DS1 MRI data cannot be shared publicly. The code to perform the analysis and radiomic features values are publically available from GitHub URL: https://github.com/lavrovaliz/brain-tissue-radiomics-on-clinical-and-quantitative-MRI-for-MS. The details on the packages, with the indication of versions and functions used, can be found in the appendix Table S18.. Requests to access these datasets should be directed to elommers@chuliege.be.

Exploratory radiomic analysis of conventional versus quantitative brain MRI: Towards automatic diagnosis of early multiple sclerosis

Elizaveta Lavrova[1,2] †, Emilie Lommers[2,4] †, Henry C. Woodruff[1,3] †, Avishek Chatterjee[1], Pierre Maquet[2,4], Eric Salmon[2], Philippe Lambin[1,3†*], Christophe Phillips[2,5†]

[1]The D-Lab, Department of Precision Medicine, GROW – School for Oncology, Maastricht University, Maastricht, The Netherlands

[2]GIGA Cyclotron Research Centre In vivo imaging, University of Liège, Liège, Belgium

[3]Department of Radiology and Nuclear Imaging, GROW- School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht, The Netherlands

[4]Clinical Neuroimmunology Unit, Neurology Department, CHU Liège, Belgium

[5]GIGA In silico Medicine, University of Liège, Liège, Belgium

***Correspondence:**
Corresponding Author
philippe.lambin@maastrichtuniversity.nl

† These authors contributed equally to this work

Keywords: multiple sclerosis, brain MRI, radiomic features, quantitative MRI, histological MRI.

Abstract

Conventional magnetic resonance imaging (cMRI) is poorly sensitive to pathological changes related to multiple sclerosis (MS) in normal-appearing white matter (NAWM) and grey matter (GM), with the added difficulty of not being very reproducible. Quantitative MRI (qMRI) on the other hand attempts to represent physical properties of tissues, making it an ideal candidate for quantitative medical image analysis, or radiomics. We therefore hypothesized that qMRI-based radiomic features have added diagnostic value in MS compared to cMRI. This study investigated the ability of cMRI (T1w) and qMRI features extracted from WM, NAWM, and GM to distinguish between MS patients (MSP) and healthy control subjects (HCS). We developed exploratory radiomic classification models on a dataset comprising 36 MSP and 36 HCS recruited in CHU Liege, Belgium, acquired with cMRI and qMRI. For each image type and region of interest, qMRI radiomic models for MS diagnosis were developed on a training set and validated on a testing set. Radiomic models based on cMRI were developed on the entire training dataset and externally validated on open-source datasets with 167 HCS and 10 MSP. Ranked by region of interest, the best diagnostic performance was achieved in the whole WM. Here the model based on magnetization transfer imaging (a type of qMRI) features yielded a median area under the receiver operating characteristic curve (AUC) of 1.00 in the testing subset. Ranked by image type, the best performance was achieved by the magnetization transfer models, with median AUCs of 0.81 (0.74-0.89 90% CI) in NAWM and 0.88 (0.82-0.94) in GM. External validation of the T1w models yielded an AUC of 0.65 (0.30-0.85) in whole WM, demonstrating a large 95% CI and low sensitivity of 0.30 (0.10-0.70). This exploratory study indicates that qMRI radiomics could provide efficient diagnostic information using NAWM and GM analysis in MSP. T1w radiomics could be useful for a

fast and automated check of conventional MRI for WM abnormalities once acquisition and reconstruction heterogeneities have been overcome. Further prospective validation is needed involving more data for better interpretation and generalization of the results.

1 Introduction

Multiple sclerosis (MS) is an inflammatory disorder of the central nervous system, responsible for focal and diffuse damages, including both demyelination and neurodegeneration, and often leading to physical and mental disability (Lassmann, 2018; Chen et al., 2019). In 2016, there were more than 2 million prevalent cases globally (Wallin et al., 2019). In Europe, the overall mean costs per patient were more than €50k (adjusted to 2015 purchasing power parity) in severe disease (Kobelt et al., 2017).

Early diagnosis in MS is challenging, because pathology mechanisms are not yet completely understood, and disease biomarker discovery is still ongoing. The McDonald criteria is currently used for diagnosis (Thompson et al., 2018). It assimilates information about clinical relapses, focal white matter (WM) lesions (plaques) visualized with conventional magnetic resonance imaging (cMRI), and cerebrospinal fluid (CSF) analysis (Thompson et al., 2018; Kaunzner & Gauthier, 2017; Oh, Vidal-Jordana, & Montalban, 2018; Trip & Miller, 2005). If the patient does not meet the diagnostic criteria, the diagnosis of MS is provisionally not retained. Although cMRI is playing a valuable role in the routine clinical practice, it merely captures a very small proportion of MS-related pathological processes (Filippi et al., 2019; Zivadinov & Leist, 2005). It is particularly not sensitive to detect and track diffuse pathological changes occurring both in the normal appearing white matter (NAWM) and grey matter (GM). These changes appear in the early stages of the disease and better correlate with clinical outcomes than the only WM focal lesion load (Griffin et al., 2002; Yoo et al., 2018; Bonnier et al., 2014; Treaba et al., 2019; Davda, Tallantyre, & Robertson, 2019). Additionally, routine cMRI voxel intensities are expressed in arbitrary units, which vary based on a large number of factors, including the patient being examined, equipment, and protocol being used. This makes MRI analysis strongly dependent on the medical specialist's expertise, and hinders data reproducibility and comparison in follow-up and cross-sectional studies. Therefore, there is an unmet clinical need for development and automated detection of quantitative and objective early MS biomarkers.

Quantitative MRI (qMRI) potentially overcomes these limitations by quantifying physical micro-structural properties of brain tissues in standardized units. Commonly, some of the following parameters are estimated: longitudinal and effective transverse relaxation rates (R1 and R2*, respectively) or times (T1 and T2*, respectively), proton density (PD), magnetization transfer (MT) saturation, and a number of diffusion MRI (dMRI) metrics. Values in qMRI maps are linked to biological tissues' physical properties, such as axonal myelination (MT, R1, R2*, T1, dMRI), iron accumulation (R2*, T2*), and free water proportion (PD) (Nikolaus Weiskopf et al., 2013; N. Weiskopf, Mohammadi, Lutti, & Callaghan, 2015; Tabelow et al., 2019). It has been shown that qMRI data is fairly reproducible between different scanners and attractive for multi-center studies (R.-M. Gracien et al., 2020). Current MS research compares the qMRI properties of brain between healthy control subjects (HCS) and MS patients (MSP) (Yoo et al., 2018; Reitz et al., 2017; Lommers et al., 2019; Hagiwara, Hori, Yokoyama, Nakazawa, et al., 2017; Andica et al., 2018; Saccenti et al., 2019). It has been shown that with specific qMRI sequences, more MS-related damages can be detected compared with cMRI using similar acquisition times (Hagiwara, Hori, Yokoyama, Takemura, et al., 2017). Furthermore, it has been shown that qMRI reveals pathological GM alterations (Lommers et al., 2021) and early MS-related GM changes (R. M. Gracien et al., 2016).

Quantitative imaging biomarkers' discovery is currently experiencing a large increase in research interest, and radiomics is rapidly emerging as a major tool in radiology. Radiomics is a high-throughput imaging data quantification approach, aimed to calculate quantitative descriptors of medical images to characterize the underlying biology and establish correlation with clinical endpoints (Lambin et al., 2012; Philippe Lambin et al., 2017; Rogers et al., 2020). Radiomics has shown promise in personalized medicine for cancer treatment (Prasanna, Patel, Partovi, Madabhushi, & Tiwari, 2017; Coroller et al., 2015; P. Lambin et al., 2017; van Timmeren et al., 2017) and is already applied in neurology to predict epilepsy in patients with low-grade gliomas (Z. Liu et al., 2018), to distinguish between MS and neuromyelitis optica spectrum disorders on the spine MRI (Y. Liu et al., 2019; Ma et al., 2019), and to differentiate Alzheimer's disease from mild cognitive impairment on MRI and positron emission tomography (Feng et al., 2018; Li et al., 2019). The standard pipeline for radiomic analysis is presented in Figure 1.

Within the present study, we hypothesized that cMRI- and qMRI-based radiomic models have a diagnostic value in MS, while qMRI-based features have an advantage in the detection of diffuse damages. The objective of the study was to investigate the ability of radiomic features found in WM, NAWM, and GM, extracted from cMRI and qMRI maps, to distinguish between HCS and MSP. We aimed to compare the diagnostic value of the different image types in different brain tissues. For this, radiomic classification models were developed and tested, and cMRI models were validated on external publicly available datasets.

2 Materials and methods

2.1 Study design

This study was performed on three datasets: dataset 1 (DS1) contains both cMRI (T1w, FLAIR) and four types of qMRI maps (PD, MT, R1, and R2*) of both MSP and HCS, dataset 2 (DS2) contains cMRI (T1W) of HCS, and dataset 3 (DS3) contains cMRI of MSP (T1w, FLAIR), see Table 1. DS2 and DS3 were combined into one validation dataset (DSV) using data selection and additional pre-processing to minimize any mismatch with DS1 regarding demographics and image acquisition parameters. For each participant the same brain tissue segmentation method was applied. DS1 was randomly split and used to train and test multi-channel qMRI models, as well as being used for training cMRI models, while DSV was used to validate the cMRI models. The observations from test set were kept apart from train set and were used only to test the models. For each participant radiomic features were independently extracted from whole WM, NAWM, and GM regions from all available image types. For MSP, WM volume included combined NAWM and focal WM lesions. Since HCS do not have focal WM lesions, for them WM and NAWM volumes are matching.

With the addition of models combining features extracted from all four qMRI maps, a total of 18 models were trained on DS1 (3 ROIs, 5 image types and combination thereof), of which 3 models (3 ROIs, 1 image type) were validated on DSV. All feature selection and model training were performed in the respective training datasets. The testing and/or validation datasets were kept apart and were used only for evaluation purposes. The study design is detailed in Figure 2. For each step, workflow execution times were recorded and the averages reported.

2.2 Data description

Dataset 1 (DS1) is a private dataset consisting of 72 participants, 36 MSP with relapsing-remitting and progressive forms (CHU Liege, Belgium) and 36 HCS (GIGA-CRC in vivo imaging, University of

Liege, Liege, Belgium), acquired within an MS cross-sectional study (local ethic committee approval B707201213806) retrospectively collected between 2013 and 2017 (Lommers et al., 2019). It contains cMRI data (T1w for all the participants and FLAIR only for the MSP) and qMRI maps (PD, MT, R1, R2*, see Figure 3). The details of the MRI protocol are available in (Lommers et al., 2019) and in the Supplementary Table 1. The inclusion criteria were: (1) age between 18 and 65 years, (2) Expanded Disability Status Scale (EDSS) not more than 6.5, (3) no relapse in the previous four weeks, (4) MRI compatibility. Additionally, MS patients with vascular risk factors and comorbidities were excluded to minimize the risk of disturbance caused by potential microvascular lesions. MS status was estimated by CHU Liege neurology specialists, based on McDonald's criteria 2010 (Polman et al., 2011). The detailed demographic data is presented in the Supplementary Table 2. Comparison of the relapsing-remitting MS (RRMS) and progressive MS (PMS) patients in terms of atrophy and qMRI maps values within the different tissue classes is presented in the Supplementary Table 3. DS1 was used for all the exploratory analyses, including feature selection and model parameter tuning. Before the feature selection and subsequent steps, DS1 was randomly split into training and testing sets (80 %/20 %), attempting to maintain distributions of outcome, age, sex, and scanner variables.

Dataset 2 (DS2) is the Calgary-Campinas-359 dataset – an open, multi-vendor, multi-field- strength brain MRI dataset (Souza et al., 2018). It is composed of volumetric T1w images of 359 presumed healthy adults, scanned between 2009 and 2016. In the dataset description, there is no information about the neurological status assessment.

Dataset 3 (DS3) is a subset of the MICCAI 2016 MS lesions segmentation (MSSEG) challenge dataset. The MSSEG challenge dataset contains MRI data for 53 MSP, but only 15 participants from the training subset are publicly available (Commowick et al., 2018; Cotton et al., 2015). The data were acquired no later than 2016 in three different sites in France on four different multi-field multi-vendor scanners with different sequences, including T1w and FLAIR. We used the unprocessed data from DS2 to implement the same image pre-processing protocol for all the datasets.

There are some differences between the DS1 and DSV, the main difference being the different image acquisition equipment and protocols (see Table 1). Other differences are the lack of information about how HCS and MSP status was assessed in DS2 and DS3, and the lack of MS stage of EDSS in DS3, making a comparison between DS1 and DS3 difficult. To minimize those differences and any potential bias, DS2 and DS3 were combined and filtered to match the age range and field strength present in DS1. Within the datasets, there were no incomplete data.

A summary of the datasets is presented in Table 1. The p-values for comparison of age and sex distributions between HCS and MSP groups within development and validation data as well as between development and validation datasets can be found in Supplementary Table 4.

2.3 MRI data pre-processing

All the data processing and analysis hereafter was performed on a system containing 4x 10 core 2.40 GHz Intel Xeon CPU and 64 GB RAM.

The qMRI maps were generated in MATLAB 2017b (The MathWorks Inc., Natick, MA, USA) with the use of the hMRI toolbox v0.2.0 (Tabelow et al., 2019), an extension of SPM12 (URL: http://www.fil.ion.ucl.ac.uk/spm). In the absence of RF sensitivity bias maps acquisition, the radiofrequency field (RF) sensitivity bias was corrected with a unified segmentation approach. The radiofrequency transmit field (B_1^+) bias was corrected using B1 and B0 maps, which were acquired with 3D echo-planar imaging mapping protocols. The B1 data was processed with parameters, which

were identical to the standard default ones. The multiparameter input images included 6 MT-, 8 PD-, and 6 T1-weighted images.

All images within DS1 were reconstructed with a resolution of 1×1×1 mm^3, hence we decided to resample the scans within DS2 and DS3 to the same resolution. We used cubic spline interpolation as it performs well in terms of its Fourier properties, visual image quality, and interpolation errors (Lehmann, Gonner, & Spitzer, 1999).

Following this step, tissue masks for cerebrospinal fluid (CSF), GM, NAWM, and lesions within DS1 were estimated. Tissue segmentation in HCS was performed with a multi-channel unified segmentation protocol (Ashburner & Friston, 2005), using multiple qMRI maps (PD, MT, R2*, R1). It was performed in MATLAB using hMRI for SPM12 with light regularisation (regularisation coefficient 0.001) and 60 mm cut-off for full width at half maximum (FWHM) of Gaussian smoothness of bias. The outputs were tissue probability maps for CSF, GM, and WM, with the voxel values between 0 (background) and 1 (corresponding brain tissue). In order to ensure the inclusion of only the relevant tissue class, binary masks for each tissue were obtained by thresholding the tissue probability maps at a high level of 0.9. For MSP, lesion masks were generated from the combination of T1w and FLAIR images with LST (Schmidt et al., 2012, URL: https://www.applied-statistics.de/lst.html) for SPM12 by the lesion growth algorithm (LGA) and corrected manually by a qualified MS specialist (ELo) when necessary. Multi-channel tissue segmentation was performed using multiple qMRI maps (PD, MT, R2*, R1) with unified segmentation protocol in US-with-Lesion (Phillips & Pernet, 2017, URL: https://github.com/CyclotronResearchCentre/USwLesion), adding an extra lesion tissue class. In DSV, brain tissue segmentation was performed with a single channel (T1w) unified segmentation protocol in MATLAB with SPM12, using T1w images.

After segmentation, the total intracranial volume (TIV) was estimated for each patient as the morphological sum of the CSF, GM, NAWM, and lesion volumes (where applicable). This combined ROI was used for intensity normalization, as described below.

As the magnetic field inside an MRI scanner is not ideally homogeneous and is affected by objects within it, a bias field signal is introduced, degrading image quality as a smooth, low-frequency signal that distorts segmentation results and feature values. To partially correct for this in T1w images, N4 bias field correction (Tustison et al., 2010) was performed in TIV.

As cMRI voxel intensities are expressed in arbitrary units, the Image Biomarker Standardisation Initiative (IBSI) recommends using normalization for raw MR data (Zwanenburg, Leger, Vallières, & Löck, 2016). Therefore, within each T1w scan, the intensities were normalized to arrive at a mean of 0 and a standard deviation of 1. Normalization was performed within the TIV, considering only TIV intensities.

2.4 Radiomic feature extraction and exploration

Radiomic features that quantitatively characterize the ROI, e.g., intensity histogram, simple statistics, and texture (Lambin et al., 2012; Rizzo et al., 2018), were extracted from pre-processed cMRI and qMRI data using PyRadiomics 2.2.0 (van Griethuysen et al., 2017) in python v. 3.7.1. Due to their small volumes, features from lesion ROIs were not extracted, and they were used only as an additional tissue class for brain segmentation. Radiomic features of the following classes were extracted from the original images: FO statistics, Grey Level Co-occurrence Matrix (GLCM) (Haralick, Shanmugam, & Dinstein, 1973), Grey Level Run Length Matrix (GLRLM) (Galloway, 1975), Grey Level Size Zone

Matrix (GLSZM) (Thibault et al., 2013), Neighbouring Grey Tone Difference Matrix (NGTDM) (Amadasun & King, 1989), Grey Level Dependence Matrix (GLDM) (Sun & Wee, 1983). The full list of the extracted features can be found in the Supplementary Table 5. Contrary to oncological radiomic studies, where shape features are usually involved (Lambin et al., 2012; Philippe Lambin et al., 2017; Rizzo et al., 2018), here only first-order and texture features were considered. Many neurodegenerative disorders have reported volumetric brain changes, showing disease-specific patterns in brain substructures (Jakimovski et al., 2020), which were not delineated in the present study. Moreover, WM volumetric atrophy changes are mostly explained with the presence of lesions (Marciniewicz, Podgorski, Sasiadek, & Bladowska, 2019), which also influence first-order and texture features. Therefore, to further reduce the ratio of the number of features versus the number of samples, shape features were excluded. Before grey-level texture matrices were calculated, intensities discretization was performed with a fixed number of bins $N_{bins} = 50$, in line with IBSI recommendations (Zwanenburg et al., 2016). The fixed bin number approach groups voxel intensities before discretization, which additionally harmonizes multi-scanner multi-vendor multi-site data.

No feature harmonization methods, such as ComBat (Johnson, Li, & Rabinovic, 2007), were applied across the different datasets because of the small sample sizes and considerable heterogeneity of scanners and protocols. To speed up feature extraction, the ROI was pre-cropped into a bounding box with 5-voxel-width padding. A separate feature set was calculated for each ROI and image type. An overview of feature sets is presented in Table 2.

Feature analysis was performed in the whole DS1 to describe the data, its results were not included into model building. Statistical tests were performed to gauge diagnostic efficacy in such a small dataset. A univariate Mann-Whitney test was carried out using Bonferroni correction and $p \leq 0.01$ for two-sided hypothesis was considered statistically significant. Point-biserial correlation coefficients r_{pb} and p-values were calculated between radiomic feature values and MS status; a correlation was considered statistically significant if $|r_{pb}| \geq 0.85$ and $p \leq 0.05$. Spearman correlations between the features, and age and the feature ROI volume were computed to gauge the added value of radiomic features compared to age and volumetry, with a $|r_S| > 0.85$ considered highly correlated for each test. Additionally, the univariate area under the receiver operating characteristic curve (AUC) was calculated for each feature.

2.5 Radiomic feature selection

In order to remove redundant and non-informative features, feature reduction and selection were performed on DS1, using the MS status as the binary outcome where applicable. Feature selection was independently carried out for the T1w, PD, MT, R1, and R2* maps to arrive at a subset of N features each, attempting to adhere to published rules of thumb to estimate the optimal number (Hua, Xiong, Lowey, Suh, & Dougherty, 2005; Abu-Mostafa, Magdon-Ismail, & Lin, 2012). We chose the following approach to estimate the number of features $N_{features}=int\ \frac{N_S}{10}$, as outlined in (Abu-Mostafa et al., 2012), where N_S is the number of samples in the minor class.

Since DS1 is relatively small, especially after the train/test split, feature selection as described below was performed 100 times on an extended and balanced subset of 100 participants created by randomly sampling (with replacement) observations from the training set. In each of the 100 iterations, a fixed number N of the highest-ranking features was retained, and at the end, features were ranked according to how often they were selected.

The feature selection pipeline starts with excluding features with a zero or low variance. A feature was considered of low variance if the percentage of its distinct values out of the number of observations was less than 10%, and the ratio of its most frequent values was more than 95/5. Next, features with high inter-correlation were excluded by calculating the pairwise Spearman correlation between all the features. From each pair of highly correlated features (|rS|>0.85), we excluded a feature having the highest correlation on average with all the remaining features. The final selection was performed with recursive feature elimination (Guyon, Weston, Barnhill, & Vapnik, 2002) using random forest classifier (Breiman, 2001) models (100 trees, as recommended by Oshiro, Perez, & Baranauskas, 2012; a number of features to consider when looking for the best split $int(\sqrt{N_{features}})$, where $\sqrt{N_{features}}$ is changing during recursive feature elimination iterations, as recommended by Hastie, Tibshirani, & Friedman, 2009). Random forest (RF) classifiers allow for robust variable importance computation and do not need normalization. Moreover, the number of available features exceeds the number of samples, and a random forest classifier is still able to deal with such data. For each selected feature a distribution map was generated by calculating the feature value within each 26-connected neighbourhood of each voxel within the image ROIs.

2.6 Model training and testing

Models were trained and tested on independent sets of DS1. Observations from the training and testing sets were randomly sampled with a replacement for 100 times, resulting in the creation of extended and balanced training and testing subsets of 100 participants each.

Separate binary classification models were trained on DS1 for different image types: T1w, PD, MT, R1, R2*, and for a combination of features from PD + MT + R1+ R2* (composed qMRI) to investigate the value of each image type and ROI in the estimation of the MS status. For each image type three binary classification models were trained using the same features from each image type and ROI: (i) random forest (RF), (ii) support vector machine (SVM) (Platt, 1999), (iii) logistic regression (LR). For the RF model, the same settings as for the recursive feature elimination were used; for SVM, a radial basis function kernel was used with regularization parameter C=1.0, kernel coefficient $\gamma = 1/(N_{features} \cdot Var(X))$, where $Var(X)$ is the variance of the input features X (since we did not have any a priori expert knowledge about the classification problem and did not perform any empirical validation of model parameters, these are the default parameters for the SVM, keeping a balance between classification accuracy and tolerance to misclassification errors); and for LR, L2 penalty was used, since this regularisation does not lead to high values among the regression coefficients, with dual formulation, as recommended when the amount of observations exceeds the amount of features, and a liblinear solver, which is recommended for small datasets; inverse of regularization strength C=1.0, which is the optimal in terms of balance between accuracy and model complexity. Again, due to the small dataset sizes, DS1 was used as an exploratory dataset.

The models' performances were estimated in terms of the following metrics: sensitivity, specificity, and AUC, with the corresponding 90% confidence intervals (CI); for each model, learning and curves were plotted. Since all the scores were estimated on the data subsets, containing equal numbers of HCS and MSP, the imbalanced data correction was not needed. The best model was selected based on these performance metrics for different ROIs and tissue types, giving the AUC score more weight, excluding models with the median AUC scores below the threshold of 0.7, which is considered an underperforming classification model. In order to select the best model type (RF, SVM, or LR) the number of highest AUC scores was used.

The final models with the original coefficients were subsequently validated on DS2 and DS3. As the combined dataset containing DS2 and DS3 was highly unbalanced regarding the outcome, random sampling (with replacement) was implemented. Therefore, the models were validated on 100 balanced subsets, containing 10 random observations from DS2 as well as 10 random observations from DS3. The models for qMRI were not validated externally due to the unavailability of similar datasets.

To examine the models and methodology for overfitting, a permutation test was performed on DS1. The class labels in both training and testing sets were randomized, maintaining the same distributions as in original sets. Without modifying the pipeline, feature selection was performed, models were trained and tested, and performance metrics were calculated to ascertain whether the pipeline detects patterns in randomly generated outcomes.

3 Results

3.1 Data description and MRI data pre-processing

Participants were drawn from DS2, aiming to match DS1 regarding age and magnetic field strength. Participants with MRI quality, which was not sufficient for robust automatic segmentation, were excluded after a visual check (ELa). Finally, 167 participants were selected from this dataset. Another ten participants were selected from DS3, again trying to match age and field strength distributions with those of DS1. An overview of the resulting feature sets is presented in Table 3. Details of participants' distribution between the train and test sets of DS1 and significance results for comparison of age and sex distributions in the train and test sets can be found in the Supplementary Table 6.

3.2 Radiomic feature extraction and description

For each T1w and qMRI image and ROI combination, 93 features were extracted, resulting in 1395 features per participant. The Mann-Whitney test revealed that 16 % of features (220 features out of 1395) were sampled from significantly different distributions in the HCS and MSP cohorts, mostly originating from WM in all image types but also from NAWM in MT and R2*. In the entire feature set, there was only one feature (R1 first-order minimum in WM) highly correlated with the outcome, no feature highly correlated with age, and 10 features out of 1395 highly correlated with ROI volume. Univariate analysis showed that 28 % of features (395 features out of 1395) had a ROC AUC score >0.75, most of which were obtained from the PD, MT, and R2* maps (see Table 4).

3.3 Radiomic feature selection

In the training set of DS1, on average among all the image types and ROIs, 7 % from the initial feature set were excluded by the low variance step, followed by 79 % exclusion by the high correlation step. The number of features per set kept after each feature selection step is available in the Supplementary Table 7. The RF-based recursive feature elimination using data sampling with replacement yielded the final feature vectors for each ROI and MRI image type. To make the models easier to compare across ROI and MRI image types, the 3 ($N_{features} = \frac{N_S}{10} = \frac{28}{10 \sim 3}$) top ranking features were left in each final feature vector. The list of the selected features can be found in the Supplementary Table 8.

No high correlations were discovered between the selected features, age, and the ROI volume. For the selected features the univariate AUC was below a threshold of 0.7 for PD, MT, and R2* in NAWM and T1w, and PD in GM. A list of the selected features with their univariate ROC AUC scores is presented in Figure 4. Spearman correlation coefficients with age and ROI volume are presented in Supplementary Figure 1.

In the qMRI analysis, the measurements for each participant are calculated as median values of qMRI parameters within voxels of each tissue class. Therefore, traditional quantitative values of the qMRI maps are special cases of the radiomic features. These values were not selected by the applied feature selection method on training set of DS1. Univariate ROC AUC scores and Spearman correlation coefficients with the selected features from the same ROI and image type are presented in the Supplementary Table 9.

Some of the radiomic features, primarily representing first-order statistics, are self-explanatory. At the same time, texture features are not very interpretable. For the better interpretability, for the best features in each ROI and image type, the saliency maps were obtained by calculation of the feature value in the neighborhood of each voxel. The examples of the normalized saliency maps are presented on Figure 5. Even though all the voxels from the ROI contribute to the radiomic feature value, saliency maps can show which areas of the ROI increase or decrease the total value. For example, for R1 first-order 10-Percentile, the total feature value is highly increased by the voxels in the center of the ROI, whereas saliency map (I) values at the border with the GM are relatively low, because this feature represents a statistical characteristic of the intensity distribution. R2* GLDM Low Gray Level Emphasis in GM is a texture feature, defined by the spatial combinations of the voxels with specific intensities, so its map (O) has no evident spatial clustering of the values.

3.4 Models training and testing

According to the Delong test with use of the Bonferroni correction, different ML models had significantly different ($p \leq 0.01$) AUC scores in all the cases, with the exception of MT and qMRIcomb in WM, R1 in NAWM, and PD in GM (p-values for AUC comparison can be found in the Supplementary Table 10, performance metrics in the Supplementary Table 11). Among all the ROI and image types, in most cases, the median values of the RF classifier performance scores dropped below a threshold of 0.7. Having the highest number of top AUC values, the LR model was selected. Results from the LR model will be shown in the main body of the text, while the regression coefficients for the final models are shown in the Table 12. Performance metrics are presented in Table 5.

Models using features extracted from WM achieved the best classification performance with the best performance achieved by the MT data. There were no statistical differences ($p \leq 0.01$) in AUC scores obtained for WM in MT, R1, and qMRIcomb (p-values for AUC comparison can be found in the Supplementary Table 13). The highest median performance across all metrics was achieved with the MT model, all of which yielded a value of 1.00. The T1w model performed generally lower than MT and combined qMRI models, but outperformed the PD model in median specificity, the R1 model in median sensitivity, and the R2* model in median AUC.

In NAWM there were no significant differences in AUC scores obtained for R2* and qMRIcomb models. The highest overall performance was achieved with the R1 model. The PD model yielded a median specificity of 0.00 (no true negatives were achieved). The T1w model performed generally poorer than the MT and R1 models, but better than the PD, R2*, and qMRIcomb models.

In GM there were no significant differences in AUC scores obtained for MT and R1, and R2* and qMRIcomb. The highest overall performance was achieved with the MT-based model, which yielded median AUC of 0.88.

The permutation test results showed a significant ($p \leq 0.01$) drop in AUC for all the models, except for PD and R2* in WM and NAWM, and T1w in GM. The full results obtained with the permutation

test for different models and permutation test p-values can be found in the Supplementary Table 14 and Supplementary Table 15, correspondingly.

Classification performance metrics for T1w models using the WM, NAWM, and GM validated on the external DSV are presented in Table 6.

The actual clinical models always contain demographic and clinical information. Nevertheless, within this study, the clear utility of the imaging features was investigated. Univariate ROC AUC scores for the demographic and clinical variables, such as age, brain parenchymal fraction, GM fraction, WM fraction, Z-scores for the motor and cognitive tests are presented in Supplementary Table 16.

3.5 TRIPOD statement and Radiomics quality assurance

This study was evaluated with the "Radiomics Quality Score" – RQS (Philippe Lambin et al., 2017), which yielded a final result of 39%. Likewise, we evaluated it with the "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis" – TRIPOD (Collins, Reitsma, Altman, & Moons, 2015) checklist score, which was in a range of 0.71-0.77. The RQS and TRIPOD checklists are presented in the Supplementary Table 17 and 18.

4 Discussion

In this exploratory brain tissue MRI and qMRI radiomics study based on a unique dataset, we report on several hypothesis-generating findings for HCS vs. MSP classification. We aimed to investigate the diagnostic utility of the new MRI image types promising in the cross-center studies. Previous studies on radiomics in MS were focused on T2w cMRI data and aimed to distinguish between MS and neuromyelitis optica spectrum disorder (Y. Liu et al., 2019; Ma et al., 2019) without external validation, hence the importance of this work.

While focal WM lesions are the visible part to the disease on cMRI, it was reported that diffuse MS-related pathological changes might appear in normal appearing brain tissue. These changes are partly independent from focal lesions and are detected in the earlier stage of the disease although they predominate in the late phase of the disease especially in the progressive phenotype (Lassmann, 2018). Therefore, to investigate the sensitivity of the radiomic features to the diffuse changes, an analysis of NAWM and GM was performed. Because our MS cohort include both RRMS and PMS patients with a rather long disease duration, their radiomic features in NAWM might differ from those in early MS patients. Nevertheless, we believe our results justify further studies involving early MS cases.

Of the three machine learning models (RFC, SVM, and LR) tested, LR was the most stable with median AUC, sensitivity, and specificity, all exceeding a value of 0.7 while achieving the highest performance in AUC. LR outperforming the other models could be due to the small number of observations, where the simplest models might perform best since they are less likely to overfit. The selected radiomic features were not correlated with age and volume, which indicates that radiomics could provide additional information to those simple variables.

The best LR model performance concerning tissue type was achieved using features extracted from WM. This was expected since focal WM lesions (plaques) in MSP's WM affect the intensities distribution (Trip & Miller, 2005). In NAWM classification, which is more challenging, good classification is achieved not only with MT and R1 maps but also with T1w data. This result was not expected since this MRI sequence is not sensitive to pathological NAWM changes, as reported in (Trip & Miller, 2005; Reitz et al., 2017). These observations are explained by the fact that qMRI voxel values

have physical meaning, reflecting the water and myelin contents (Nikolaus Weiskopf et al., 2013). Furthermore, the qMRI map generation pipeline contains image co-registration and B0 and B1 fields correction steps, leading to interpolation and, therefore, smoothing of the qMRI map. Moreover, T1w images have a higher spatial resolution, leading to more detailed texture analysis. In GM, the T1w-based model underperforms, as it was expected, according to previous publications (Trip & Miller, 2005; Reitz et al., 2017).

Amongst the image types, the best performance was achieved with MT maps, which parameter strongly correlates with histologically measured myelin content (Schmierer, Scaravilli, Altmann, Barker, & Miller, 2004). This corroborates the findings of Lommers et al., 2019, where statistical tests showed the considerable differences between HCS and MSP. In WM, the MT model demonstrated a median AUC, sensitivity, and specificity of 1.00, which means that all the testing observations were classified correctly. As far as testing observations did not enter model training, we can conclude that in our relatively small dataset, focal WM lesions (plaques) presence makes the selected MT features distinctive from the ones extracted from the healthy brain. PD maps showed the most inferior performance, with at least one of the performance metrics crossing below a value of 0.7 in each tissue type. This could be due to the potential residual T2* weighting, as mentioned previously (Lommers et al., 2019). The results obtained with T1w and R1 data were significantly different, although both these image types represent longitudinal relaxation. The main difference between them is that T1w demonstrates the relative level of longitudinal relaxation at some moment, expressed in arbitrary units. In contrast, the R1 map represents the actual physical property of the tissue and is expressed in standardized physical units (Hz). Furthermore and as already discussed, reconstruction of the qMRI images, unlike for T1w data, is always performed with the correction of instrumental biases and receive fields (Tabelow et al., 2019).

Although the T1w models are non-quantitative, they outperformed some of the qMRI models in WM and NAWM, yet had the poorest performance in GM. Among all the T1w models, the WM model yielded the highest median AUC of 0.74 on the testing set of the development dataset. On the external validation, T1w-based models all showed poor performance. Nevertheless, among these models, the best performance was achieved in WM, mainly due to focal WM lesions, which are easily captured in the radiomic analysis. In NAWM and GM, the differences between HCS and MSP are presented on the microstructural level. The T1w data is expressed in arbitrary units, and it is not consistent enough to detect these changes within different scanners and centers. As the T1w-based model in GM underperformed on the testing data, a good performance on the validation dataset was not expected. Thus, even though T1w data can perform well on the development dataset, its application is challenging for multi-centric studies. The explanation can be due to differences in imaging data, lack of sensitivity of T1w contrast for these applications, low predictive ability of the corresponding features, and their susceptibility to data effects. Additionally, we suspect a bias that can be introduced by the clinical differences in DS1, DS2, and DS3. Whereas MS status assessment details, EDSS, and MS stage are known for DS1, there is no such information about the participants from DS3, and there is no information about the tests carried out for DS2 participants to determine them as HCS.

Strengths of the current study include the use of the unique quantitative and reproducible imaging data, the use of an external validation open-source data, and in-depth investigation of the features in traditionally challenging tissues such as NAWM and GM, which can have potential in early MS diagnosis.

This study has some limitations too. The first stems from the small number of observations in the DS1. Consequently, for external validation, we excluded participants, which did not correspond to participants from DS1 in terms of age or MRI magnetic field strength. Also, all participants with insufficient MRI data quality rendering it unsuitable for robust automatic brain tissue segmentation were excluded, introducing more bias. Another limitation is related to the uniqueness of qMRI data, meaning there are no available similar qMRI brain datasets for external validation, especially for MSP. However, it was reported that qMRI is reproducible between different scanner models, and multi-center studies can be expected (R.-M. Gracien et al., 2020). The third limitation is the absence of data harmonization performed across datasets involved in this study. It results in non-uniformity of non-quantitative MRI data between datasets and thus leading to model performance degradation. The following limitation is related to the analysis of only HCS and MSP data. Although the exploratory analysis of the features demonstrated that some had very high univariate AUC scores (> 0.99), considering the absence of data for other neurodegenerative diseases and relatively small amount of observations, specification of the MS radiomic signature is needed. Thus, analysis of other neurodegenerative disorders is needed to distinguish between different diagnoses. The fifth limitation is related to the cohort of the patients in DS1. Our aim was to achieve MS diagnosis at an earlier time point, but we used the data from both RRMS and PMS patients with a rather long disease duration. Nevertheless, RRMS and PMS patients did not significantly differ from each other but PMS patients tend to have more pronounced alterations in NAWM as well as more tissue loss. The next limitation pertains to the cMRI sequence analyzed in this study: even though focal WM lesions are noticeable on T1w, this image type is not the leading one in MS investigation. Among cMRI modalities, T2w, FLAIR, and contrast-enhanced T1w provide appropriate contrast. These modalities were not available for all the participants of DS1 (with qMRI acquisition): FLAIR scans were available for MSP only. Therefore, analysis of another cMRI and qMRI could be a subject of future research. Finally, different brain segmentation approaches were used for DS1 and external validation data. Even though the same method was implemented for all the MRIs, segmentation in DS1 was performed with qMRI data, segmentation for external validation was performed with cMRI data. It could affect the values of radiomic features, as cMRI-based segmentation leads to inaccurate delineation of deep GM regions (Nikolaus Weiskopf et al., 2013; Lommers et al., 2019).

Within the present study, we used standard open source tools for data pre-processing and analysis. Thus, the diagnostic support workflow execution times obtained within this study are indicative. Moreover, they strongly depend on the used hardware, software, original medical image parameters, pre-processing and analysis settings, and radiomic features, composing the final signature. We did not implement any optimization of computational resources consumption; therefore, the obtained execution times represent the upper bound of a workflow duration. Within the present study, cMRI- and qMRI-based workflows took approximately 26 and 38 minutes per participant, excluding the image acquisition time. This difference is due to the relatively long time of qMRI maps reconstruction. This shows that cMRI workflow can be implemented into the brain scanning protocols as a screening for WM abnormalities. The qMRI workflow requires a particular scanning protocol (Nikolaus Weiskopf et al., 2013) and a relatively long analysis time. Therefore, it can be implemented for diagnostic support for patients with suspicious medical evidence.

This study indicated the potential of cMRI and qMRI radiomics in MS-related biomarkers development. The novelty of this work is in the combination of the two MRI techniques and the attempt to overcome the challenge of arbitrary units in MRI we examined the utility of radiomics in qMRI.

In differentiating between MSP and HCS, qMRI showed the advantage over cMRI in NAWM and GM regions. Therefore, the application of qMRI is promising in early MS diagnosis. We believe that qMRI

radiomic signatures can contribute to multi-center studies, as indicated in the previous works (Nikolaus Weiskopf et al., 2013; N. Weiskopf et al., 2015; Tabelow et al., 2019; Lommers et al., 2019). For this, the reproducibility of qMRI features is to be investigated in the future. T1w WM analysis could potentially be applied for a rapid check of cMRI for WM abnormalities. For research purposes, 7 T MRI is often applied to study NAWM and GM (Treaba et al., 2019; Zurawski et al., 2020), but it is not widely used in clinical practice. We believe that 7 T MRI radiomic analysis is a potential research field in MS diagnosis.

Our next step is to validate those findings in a prospective qMRI study and test the hypothesis that those signatures are sensitive to neurodegenerative changes in early RRMS and have a diagnostic value for subjects at risk (e.g., clinically isolated syndrome).

5 Conclusion

This study demonstrates that brain cMRI and qMRI radiomic features have the potential to distinguish between MSP and HCS. In NAWM and GM analysis, having potential in early automated diagnosis, stable results are achieved with qMRI-based data. This is a proof of concept clinical study demonstrating a strong signal in brain imaging, but further research is needed to develop and approve radiomic signatures for MS.

Nevertheless, future large-scale studies should evaluate the reproducibility and generalizability of the proposed method and create an MS-specific radiomic signature. Because of fully automated pipeline and imaging data quantification, the proposed approach shows its potential in relevance to timesaving and reproducibility in MS diagnosis.

6 Conflict of Interest

PL reports — within and outside the submitted work, grants/sponsored research agreements from Radiomics SA, ptTheragnostic/DNAmito, Health Innovation Ventures. He received an advisor/presenter fee and/or reimbursement of travel costs/consultancy fee and/or in kind manpower contribution from Radiomics SA, BHV, Merck, Varian, Elekta, ptTheragnostic, BMS and Convert pharmaceuticals. Dr Lambin has minority shares in the company Radiomics SA, Convert pharmaceuticals, MedC2 and LivingMed Biotech, he is co-inventor of two issued patents with royalties on radiomics (PCT/NL2014/050248, PCT/NL2014/050728) licensed to Radiomics SA and one issued patent on mtDNA (PCT/EP2014/059089) licensed to ptTheragnostic/DNAmito, one non issued patent on LSRT (PCT/ P126537PC00) licensed to Varian Medical, three non-patented invention (softwares) licensed to ptTheragnostic/DNAmito, Radiomics SA and Health Innovation Ventures and three non-issues, non licensed patents on Deep Learning-Radiomics (US P125078US00, PCT/NL/2020/050794, n° N2028271). He confirms that none of the above entities or funding was involved in the preparation of this paper.

HW has (minority) shares in the company radiomics.bio.

7 Author Contributions

ELa, HW, CP, ES, and PL contributed to study conceptualization. PL and ES acquired the funding. ELo, PM, and CP performed data acquisition and curation. ELa, HW, PL, CP, AC, a nd ES developed methodology. ELa, HW, and CP performed analysis. ELa, HW, and CP wrote the original draft. All

the authors contributed to manuscript revision, read, and approved the submitted version. The supervision was performed by HW, CP, PL, ES.

8 Funding

ELa is in a Maastricht-Liege University Imaging Valley PhD program. CP is supported by the Fonds de la Recherche Scientifique (F.R.S.-FNRS, Belgium). Authors acknowledge financial support from the European Union's Horizon 2020 research and innovation programme under grant agreement: MSCA-ITN-PREDICT n° 766276, CHAIMELEON n° 952172 and EuCanImage n° 952103. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

9 Data Availability Statement

DS2 and DS3 are public datasets, the accession details can be found in (Li et al., 2019), and (Souza et al., 2018). DS1 MRI data cannot be shared publicly. The code to perform the analysis and radiomic features values are publically available from GitHub URL: https://github.com/CyclotronResearchCentre/brain-tissue-radiomics-on-clinical-and-quantitative-MRI-for-MS. The details on the packages, with the indication of versions and functions used, can be found in the Supplementary Table 19.

10 References

Lassmann, H. (2018). Multiple sclerosis pathology. *Cold Spring Harbor perspectives in medicine, 8*(3), a028936.

Chen, J., Taylor, B., Palmer, A. J., Kirk-Brown, A., van Dijk, P., Simpson, S., Jr., . . . van der Mei, I. (2019). Estimating MS-related work productivity loss and factors associated with work productivity loss in a representative Australian sample of people with multiple sclerosis. *Mult Scler, 25*(7), 994-1004. doi:10.1177/1352458518781971

Wallin, M. T., Culpepper, W. J., Nichols, E., Bhutta, Z. A., Gebrehiwot, T. T., Hay, S. I., . . . Naghavi, M. (2019). Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology, 18*(3), 269-285.

Kobelt, G., Thompson, A., Berg, J., Gannedahl, M., Eriksson, J., Group, M. S., & Platform, E. M. S. (2017). New insights into the burden and costs of multiple sclerosis in Europe. *Multiple Sclerosis Journal, 23*(8), 1123-1136.

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., . . . Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol, 17*(2), 162-173. doi:10.1016/S1474-4422(17)30470-2

Kaunzner, U. W., & Gauthier, S. A. (2017). MRI in the assessment and monitoring of multiple sclerosis: an update on best practice. *Ther Adv Neurol Disord, 10*(6), 247-261. doi:10.1177/1756285617708911

Oh, J., Vidal-Jordana, A., & Montalban, X. (2018). Multiple sclerosis: clinical aspects. *Curr Opin Neurol, 31*(6), 752-759. doi:10.1097/WCO.0000000000000622

Trip, S. A., & Miller, D. H. (2005). Imaging in multiple sclerosis. *J Neurol Neurosurg Psychiatry, 76 Suppl 3*(suppl 3), iii11-iii18. doi:10.1136/jnnp.2005.073213

583 Filippi, M., Brück, W., Chard, D., Fazekas, F., Geurts, J. J., Enzinger, C., . . . Rovira, À. (2019).
584 Association between pathological and MRI findings in multiple sclerosis. *The Lancet*
585 *Neurology, 18*(2), 198-210.

586 Zivadinov, R., & Leist, T. P. (2005). Clinical-magnetic resonance imaging correlations in multiple
587 sclerosis. *J Neuroimaging, 15*(4 Suppl), 10S-21S. doi:10.1177/1051228405283291

588 Griffin, C. M., Chard, D. T., Parker, G. J., Barker, G. J., Thompson, A. J., & Miller, D. H. (2002).
589 The relationship between lesion and normal appearing brain tissue abnormalities in early
590 relapsing remitting multiple sclerosis. *Journal of neurology, 249*(2), 193-199.

591 Yoo, Y., Tang, L. Y., Brosch, T., Li, D. K., Kolind, S., Vavasour, I., . . . Tam, R. C. (2018). Deep
592 learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish
593 between multiple sclerosis patients and healthy controls. *NeuroImage: Clinical, 17*, 169-178.

594 Bonnier, G., Roche, A., Romascano, D., Simioni, S., Meskaldji, D., Rotzinger, D., . . . Granziera, C.
595 (2014). Advanced MRI unravels the nature of tissue alterations in early multiple sclerosis.
596 *Ann Clin Transl Neurol, 1*(6), 423-432. doi:10.1002/acn3.68

597 Treaba, C. A., Granberg, T. E., Sormani, M. P., Herranz, E., Ouellette, R. A., Louapre, C., . . .
598 Mainero, C. (2019). Longitudinal Characterization of Cortical Lesion Development and
599 Evolution in Multiple Sclerosis with 7.0-T MRI. *Radiology, 291*(3), 740-749.
600 doi:10.1148/radiol.2019181719

601 Davda, N., Tallantyre, E., & Robertson, N. P. (2019). Early MRI predictors of prognosis in multiple
602 sclerosis. *J Neurol, 266*(12), 3171-3173. doi:10.1007/s00415-019-09589-2

603 Weiskopf, N., Suckling, J., Williams, G., Correia, M. M., Inkster, B., Tait, R., . . . Lutti, A. (2013).
604 Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center
605 validation. *Frontiers in neuroscience, 7*, 95.

606 Weiskopf, N., Mohammadi, S., Lutti, A., & Callaghan, M. F. (2015). Advances in MRI-based
607 computational neuroanatomy: from morphometry to in-vivo histology. *Curr Opin Neurol,*
608 *28*(4), 313-322. doi:10.1097/WCO.0000000000000222

609 Tabelow, K., Balteau, E., Ashburner, J., Callaghan, M. F., Draganski, B., Helms, G., . . . Phillips, C.
610 (2019). hMRI–A toolbox for quantitative MRI in neuroscience and clinical research.
611 *NeuroImage, 194*, 191-210.

612 Gracien, R.-M., Maiworm, M., Brüche, N., Shrestha, M., Nöth, U., Hattingen, E., . . . Deichmann, R.
613 (2020). How stable is quantitative MRI?–Assessment of intra-and inter-scanner-model
614 reproducibility using identical acquisition sequences and data analysis programs.
615 *NeuroImage, 207*, 116364.

616 Reitz, S. C., Hof, S. M., Fleischer, V., Brodski, A., Groger, A., Gracien, R. M., . . . Klein, J. C.
617 (2017). Multi-parametric quantitative MRI of normal appearing white matter in multiple
618 sclerosis, and the effect of disease activity on T2. *Brain Imaging Behav, 11*(3), 744-753.
619 doi:10.1007/s11682-016-9550-5

620 Lommers, E., Simon, J., Reuter, G., Delrue, G., Dive, D., Degueldre, C., . . . Maquet, P. (2019).
621 Multiparameter MRI quantification of microstructural tissue alterations in multiple sclerosis.
622 *Neuroimage Clin, 23*, 101879. doi:10.1016/j.nicl.2019.101879

623 Hagiwara, A., Hori, M., Yokoyama, K., Nakazawa, M., Ueda, R., Horita, M., . . . Aoki, S. (2017).
624 Analysis of White Matter Damage in Patients with Multiple Sclerosis via a Novel In Vivo

MR Method for Measuring Myelin, Axons, and G-Ratio. *AJNR Am J Neuroradiol, 38*(10), 1934-1940. doi:10.3174/ajnr.A5312

Andica, C., Hagiwara, A., Hori, M., Nakazawa, M., Goto, M., Koshino, S., . . . Aoki, S. (2018). Automated brain tissue and myelin volumetry based on quantitative MR imaging with various in-plane resolutions. *Journal of Neuroradiology, 45*(3), 164-168.

Saccenti, L., Andica, C., Hagiwara, A., Yokoyama, K., Takemura, M. Y., Fujita, S., . . . Hori, M. (2019). Brain tissue and myelin volumetric analysis in multiple sclerosis at 3T MRI with various in-plane resolutions using synthetic MRI. *Neuroradiology, 61*(11), 1219-1227.

Hagiwara, A., Hori, M., Yokoyama, K., Takemura, M. Y., Andica, C., Tabata, T., . . . Aoki, S. (2017). Synthetic MRI in the Detection of Multiple Sclerosis Plaques. *AJNR Am J Neuroradiol, 38*(2), 257-263. doi:10.3174/ajnr.A5012

Lommers, E., Guillemin, C., Reuter, G., Fouarge, E., Delrue, G., Collette, F., . . . Phillips, C. (2021). Voxel-Based quantitative MRI reveals spatial patterns of grey matter alteration in multiple sclerosis. *Hum Brain Mapp, 42*(4), 1003-1012. doi:10.1002/hbm.25274

Gracien, R. M., Reitz, S. C., Hof, S. M., Fleischer, V., Zimmermann, H., Droby, A., . . . Klein, J. C. (2016). Assessment of cortical damage in early multiple sclerosis with quantitative T2 relaxometry. *NMR Biomed, 29*(4), 444-450. doi:10.1002/nbm.3486

Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R. G., Granton, P., . . . Dekker, A. (2012). Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer, 48*(4), 441-446.

Lambin, P., Leijenaar, R. T., Deist, T. M., Peerlings, J., De Jong, E. E., Van Timmeren, J., . . . Jochems, A. (2017). Radiomics: the bridge between medical imaging and personalized medicine. *Nature reviews Clinical oncology, 14*(12), 749-762.

Rogers, W., Thulasi Seetha, S., Refaee, T. A. G., Lieverse, R. I. Y., Granzier, R. W. Y., Ibrahim, A., . . . Lambin, P. (2020). Radiomics: from qualitative to quantitative imaging. *Br J Radiol, 93*(1108), 20190948. doi:10.1259/bjr.20190948

Prasanna, P., Patel, J., Partovi, S., Madabhushi, A., & Tiwari, P. (2017). Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. *European radiology, 27*(10), 4188-4197.

Coroller, T. P., Grossmann, P., Hou, Y., Rios Velazquez, E., Leijenaar, R. T., Hermann, G., . . . Aerts, H. J. (2015). CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol, 114*(3), 345-350. doi:10.1016/j.radonc.2015.02.015

Lambin, P., Zindler, J., Vanneste, B. G., De Voorde, L. V., Eekers, D., Compter, I., . . . Walsh, S. (2017). Decision support systems for personalized and participative radiation oncology. *Adv Drug Deliv Rev, 109*, 131-153. doi:10.1016/j.addr.2016.01.006

van Timmeren, J. E., Leijenaar, R. T., van Elmpt, W., Reymen, B., Oberije, C., Monshouwer, R., . . . Lambin, P. (2017). Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. *Radiotherapy and Oncology, 123*(3), 363-369.

Liu, Z., Wang, Y., Liu, X., Du, Y., Tang, Z., Wang, K., . . . Tian, J. (2018). Radiomics analysis allows for precise prediction of epilepsy in patients with low-grade gliomas. *Neuroimage Clin, 19*, 271-278. doi:10.1016/j.nicl.2018.04.024

667 Liu, Y., Dong, D., Zhang, L., Zang, Y., Duan, Y., Qiu, X., . . . Li, K. (2019). Radiomics in multiple
668 sclerosis and neuromyelitis optica spectrum disorder. *Eur Radiol, 29*(9), 4670-4677.
669 doi:10.1007/s00330-019-06026-w

670 Ma, X., Zhang, L., Huang, D., Lyu, J., Fang, M., Hu, J., . . . Ma, L. (2019). Quantitative radiomic
671 biomarkers for discrimination between neuromyelitis optica spectrum disorder and multiple
672 sclerosis. *Journal of Magnetic Resonance Imaging, 49*(4), 1113-1121.

673 Feng, F., Wang, P., Zhao, K., Zhou, B., Yao, H., Meng, Q., . . . Liu, Y. (2018). Radiomic Features of
674 Hippocampal Subregions in Alzheimer's Disease and Amnestic Mild Cognitive Impairment.
675 *Front Aging Neurosci, 10*, 290. doi:10.3389/fnagi.2018.00290

676 Li, Y., Jiang, J., Lu, J., Jiang, J., Zhang, H., & Zuo, C. (2019). Radiomics: a novel feature extraction
677 method for brain neuron degeneration disease using 18F-FDG PET imaging and its
678 implementation for Alzheimer's disease and mild cognitive impairment. *Therapeutic*
679 *advances in neurological disorders, 12*, 1756286419838682.

680 Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., . . . Wolinsky, J. S.
681 (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria.
682 *Ann Neurol, 69*(2), 292-302. doi:10.1002/ana.22366

683 Souza, R., Lucena, O., Garrafa, J., Gobbi, D., Saluzzi, M., Appenzeller, S., . . . Lotufo, R. (2018). An
684 open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available
685 skull stripping methods agreement. *NeuroImage, 170*, 482-494.
686 doi:10.1016/j.neuroimage.2017.08.021

687 Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., . . . Ferré, J.-C. (2018).
688 Objective evaluation of multiple sclerosis lesion segmentation using a data management and
689 processing infrastructure. *Scientific reports, 8*(1), 1-17.

690 Cotton, F., Kremer, S., Hannoun, S., Vukusic, S., Dousset, V., & Imaging Working Group of the
691 Observatoire Francais de la Sclerose en, P. (2015). OFSEP, a nationwide cohort of people
692 with multiple sclerosis: Consensus minimal MRI protocol. *J Neuroradiol, 42*(3), 133-140.
693 doi:10.1016/j.neurad.2014.12.001

694 Lehmann, T. M., Gonner, C., & Spitzer, K. (1999). Survey: interpolation methods in medical image
695 processing. *IEEE Trans Med Imaging, 18*(11), 1049-1075. doi:10.1109/42.816070

696 Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage, 26*(3), 839-851.
697 doi:10.1016/j.neuroimage.2005.02.018

698 Schmidt, P., Gaser, C., Arsic, M., Buck, D., Forschler, A., Berthele, A., . . . Muhlau, M. (2012). An
699 automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple
700 Sclerosis. *NeuroImage, 59*(4), 3774-3783. doi:10.1016/j.neuroimage.2011.11.032

701 Phillips, C., & Pernet, C. (2017). Unifying lesion masking and tissue probability maps for improved
702 segmentation and normalization.

703 Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C.
704 (2010). N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging, 29*(6), 1310-1320.
705 doi:10.1109/TMI.2010.2046908

706 Zwanenburg, A., Leger, S., Vallières, M., & Löck, S. (2016). Image biomarker standardisation
707 initiative. arXiv. *arXiv preprint arXiv:1612.07003*.

Rizzo, S., Botta, F., Raimondi, S., Origgi, D., Fanciullo, C., Morganti, A. G., & Bellomi, M. (2018). Radiomics: the facts and the challenges of image analysis. *European radiology experimental, 2*(1), 1-8.

van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., . . . Aerts, H. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res, 77*(21), e104-e107. doi:10.1158/0008-5472.CAN-17-0339

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and Cybernetics*(6), 610-621.

Galloway, M. M. (1975). Texture analysis using gray level run lengths. *Computer graphics and image processing, 4*(2), 172-179.

Thibault, G., Fertil, B., Navarro, C., Pereira, S., Cau, P., Levy, N., . . . Mari, J.-L. (2013). Shape and texture indexes application to cell nuclei classification. *International Journal of Pattern Recognition and Artificial Intelligence, 27*(01), 1357002.

Amadasun, M., & King, R. (1989). Textural features corresponding to textural properties. *IEEE Transactions on systems, man, and Cybernetics, 19*(5), 1264-1274.

Sun, C., & Wee, W. G. (1983). Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing, 23*(3), 341-352.

Jakimovski, D., Bergsland, N., Dwyer, M. G., Hagemeier, J., Ramasamy, D. P., Szigeti, K., . . . Zivadinov, R. (2020). Long-standing multiple sclerosis neurodegeneration: volumetric magnetic resonance imaging comparison to Parkinson's disease, mild cognitive impairment, Alzheimer's disease, and elderly healthy controls. *Neurobiol Aging, 90*, 84-92. doi:10.1016/j.neurobiolaging.2020.02.002

Marciniewicz, E., Podgorski, P., Sasiadek, M., & Bladowska, J. (2019). The role of MR volumetry in brain atrophy assessment in multiple sclerosis: A review of the literature. *Adv Clin Exp Med, 28*(7), 989-999. doi:10.17219/acem/94137

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics, 8*(1), 118-127.

Hua, J., Xiong, Z., Lowey, J., Suh, E., & Dougherty, E. R. (2005). Optimal number of features as a function of sample size for various classification rules. *Bioinformatics, 21*(8), 1509-1515.

Abu-Mostafa, Y., Magdon-Ismail, M., & Lin, H. (2012). Learning from data vol. 4: AMLBook New York. *NY, USA*.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning, 46*(1), 389-422.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). *How many trees in a random forest?* Paper presented at the International workshop on machine learning and data mining in pattern recognition.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*: Springer Science & Business Media.

Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers, 10*(3), 61-74.

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. *Circulation, 131*(2), 211-219.

Schmierer, K., Scaravilli, F., Altmann, D. R., Barker, G. J., & Miller, D. H. (2004). Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. *Ann Neurol, 56*(3), 407-415. doi:10.1002/ana.20202

Zurawski, J., Tauhid, S., Chu, R., Khalid, F., Healy, B. C., Weiner, H. L., & Bakshi, R. (2020). 7T MRI cerebral leptomeningeal enhancement is common in relapsing-remitting multiple sclerosis and is associated with cortical and thalamic lesions. *Multiple Sclerosis Journal, 26*(2), 177-187.

Table 1 – Datasets summary details (μ - average, σ - standard deviation, M – male, F - female)

	Dataset 1	Dataset 2	Dataset 3
Dataset	Private CHU, Liege	CC-359	MICCAI 2016 MSSEG challenge (training subset)
Participants	MSP (15 relapsing-remitting, 21 progressive), HCS (36)	HCS (359)	MSP (15)
Age, $\mu \pm \sigma$ [years]	45.8 ± 12.1	52.7 ± 7.3	40.5 ± 10.8
Sex, M/F	0.76	0.96	1.00
Image types	T1w, PD, MT, R1, R2*, FLAIR	T1w	T1w, FLAIR
Sites	CHU (Liege, Belgium); GIGA-CRC in vivo imaging, University of Liege (Liege, Belgium)	Campinas (Sao Paulo, Brazil); Calgary (Alberta, Canada)	CHU Rennes (Rennes, France); CHU Lyon (Lyon, France)
Equipment	3 T Siemens Magnetom Allegra (37); 3 T Siemens Magnetom Prisma (35)	3 T and 1.5 T Siemens (120), Philips (119), GE Healthcare (120) MRI scanners	3 T Siemens Magnetom Verio (5); 1.5 T Siemens Magnetom Aera (5); 3 T Philips Ingenia (5)
Protocol	MRI protocol with FLASH sequences	3D MP-RAGE (Philips, Siemens), comparable 3D T1w spoiled gradient echo sequence (GE Healthcare)	Sagittal 3D FLAIR, sagittal 3D T1w
Matrix	256×224	224×224 240×240 256×256	256×256 (Siemens) 336×336 (Philips)
Slices	176	164 - 224	176 (Siemens) 200 (Philips)
Voxel resolution [mm^3]	1×1×1	1×1×1 (Siemens)	1.08×1.08×0.9 (1.5 T Siemens) 1×1×1 (3 T Siemens) 0.74×0.74×0.85 (Philips)

Table 2 – Overview of independent features sets per participant

ROI	**Image type**	
WM (for MSP, NAWM + focal WM lesions)	cMRI	T1w
NAWM GM	qMRI	PD MT R1 R2*

In total, 3 ROIs	In total, 5 image types

Table 3 – Datasets summary details for included participants (μ - average, σ - standard deviation, M – male, F - female)

	Dataset 1	Dataset 2	Dataset 3
Participants	MSP (15 relapsing-remitting, 21 progressive), HCS (36)	HCS (167)	MSP (10)
Equipment	3 T Siemens Magnetom Allegra (37); 3 T Siemens Magnetom Prisma (35)	3 T Siemens (53), Philips (54), GE Healthcare (60) MRI scanners	3 T Siemens Magnetom Verio (5); 3 T Philips Ingenia (5)
Age, $\mu \pm \sigma$ [years]	45.8 ± 12.1	52.7 ± 7.3	40.5 ± 10.8
Sex, M/F	0.76	0.96	1.00

Table 4 – Number of features out of 1395 with age, volume, and outcome correlations having an $|r_S|$ > 0.85, as well as univariate AUC > 0.75 and corrected $p_{Mann\text{-}Whitney}$ < 0.01

	ROI	T1w	PD	MT	R1	R2*
$\|r_S^{age}\| > 0.85$	WM	0	0	0	0	0
	NAWM	0	0	0	0	0
	GM	0	0	0	0	0
$\|r_S^{volume}\| > 0.85$	WM	0	3	1	1	0
	NAWM	0	3	1	1	0
	GM	0	0	0	0	0
$\|r_{pb}^{outcome}\| > 0.85$	WM	0	0	0	1	0
	NAWM	0	0	0	0	0
	GM	0	0	0	0	0
$AUC_{univar} > 0.75$	WM	13	62	21	45	52
	NAWM	8	28	57	9	37
	GM	3	7	26	5	22
$p_{Mann-Whitney}^{Bonferroni} < 0.01$	WM	9	41	10	37	7
	NAWM	0	12	42	5	26
	GM	1	0	18	2	10

Table 5 – LR model performances on testing data showing the median (90% CI) for each image and tissue type (ROI) (median values above 0.7 for all the performance metrics for the same model are highlighted with bold font)

ROI	Image	AUC	Sensitivity	Specificity
WM	T1w	**0.74** (0.66, 0.82)	**0.76** (0.67, 0.86)	**0.72** (0.59, 0.82)
	PD	0.64 (0.58, 0.71)	1.00 (1.00, 1.00)	0.28 (0.17, 0.42)
	MT	**1.00** (1.00, 1.00)	**1.00** (1.00, 1.00)	**1.00** (1.00, 1.00)
	R1	0.82 (0.76, 0.88)	0.64 (0.52, 0.75)	1.00 (1.00, 1.00)
	R2*	**0.73** (0.63, 0.83)	**0.76** (0.62, 0.86)	**0.72** (0.58, 0.84)
	$qMRI_{comb}$	**0.93** (0.88, 0.97)	**1.00** (1.00, 1.00)	**0.86** (0.77, 0.94)
NAWM	T1w	**0.73** (0.66, 0.82)	**0.76** (0.64, 0.87)	**0.70** (0.59, 0.81)
	PD	0.37 (0.30, 0.44)	0.74 (0.60, 0.87)	0.00 (0.00, 0.00)
	MT	**0.81** (0.74, 0.89)	**0.76** (0.64, 0.87)	**0.86** (0.77, 0.94)
	R1	**0.87** (0.80, 0.93)	**0.88** (0.77, 0.98)	**0.86** (0.77, 0.94)
	R2*	0.66 (0.56, 0.76)	0.76 (0.64, 0.87)	0.56 (0.40, 0.72)
	$qMRI_{comb}$	0.74 (0.67, 0.81)	0.62 (0.48, 0.77)	0.86 (0.77, 0.94)
GM	T1w	0.41 (0.32, 0.52)	0.26 (0.16, 0.40)	0.56 (0.43, 0.71)
	PD	0.69 (0.61, 0.79)	0.51 (0.38, 0.66)	0.86 (0.77, 0.94)
	MT	**0.88** (0.82, 0.94)	**0.76** (0.64, 0.87)	**1.00** (1.00, 1.00)
	R1	0.82 (0.75, 0.87)	0.64 (0.50, 0.74)	1.00 (1.00, 1.00)
	R2*	**0.73** (0.65, 0.83)	**0.76** (0.64, 0.87)	**0.71** (0.58, 0.84)
	$qMRI_{comb}$	**0.81** (0.73, 0.88)	**0.76** (0.64, 0.87)	**0.84** (0.77, 0.95)

Table 6 – LR model performances on external validation DSV showing the median (90% CI) for each tissue type for T1w images

ROI	AUC	Sensitivity	Specificity
WM	0.65 (0.55, 0.85)	0.30 (0.10, 0.70)	1.00 (0.90, 1.00)
NAWM	0.60 (0.55, 0.95)	0.20 (0.10, 1.00)	1.00 (0.90, 1.00)
GM	0.45 (0.15, 0.45)	0.90 (0.10, 0.90)	0.00 (0.00, 0.30)

Figure 1 – Radiomics pipeline: a) medical imaging and segmentation, b) feature extraction, c) feature selection, d) modelling.

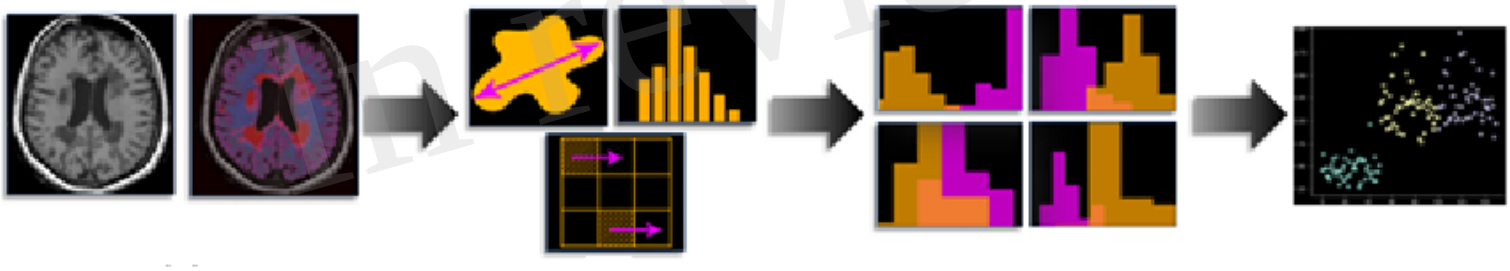
Figure 2 – Study design.

Figure 3 – Example of MRI data presented in DS1: T1w is a clinical MRI, expressed in arbitrary units; PD is linked to free water proportion, expressed in percentage; MT is linked to axonal myelination, expressed in percentage; R1 is linked to axonal myelination, expressed in Hz; R2* is linked to axonal myelination and iron accumulation, expressed in Hz.

Figure 4 – Univariate ROC AUC scores of the selected features (FO – first-order, LDHGLE – Large Dependence High Gray Level Emphasis, SDLGLE – Small Dependence Low Gray Level Emphasis, LAHGLE – Large Area High Gray Level Emphasis, MAD – Mean Absolute Deviation, LGLE – Low Gray Level Emphasis).

Figure 5 – Normalized saliency maps for the best selected features for each ROI and image type highlight the areas with the highest feature values: (A) T1w GLCM Cluster Shade in WM, (B) PD first-order Skewness in WM, (C) MT first-order Minimum in WM, (D) R1 first-order Kurtosis in WM, (E) R2* GLCM Cluster Shade in WM, (F) T1w GLCM Cluster Shade in NAWM, (G) PD GLDM Large Dependence High Gray Level Emphasis in NAWM, (H) MT GLDM Large Dependence High Gray Level Emphasis in NAWM, (I) R1 first-order 10-Percentile in NAWM, (J) R2* GLCM Imc2 in NAWM, (K) T1w first-order 10-Percentile in GM, (L) PD first-order 10-Percentile in GM, (M) MT GLDM Small Dependence Low Gray Level Emphasis in GM, (N) R1 first-order Minimum in GM, (O) R2* GLDM Low Gray Level Emphasis in GM; image resolution 1x1 mm^2

Figure 1.TIF



(A) *(B)* *(C)* *(D)*

Figure 2.TIF

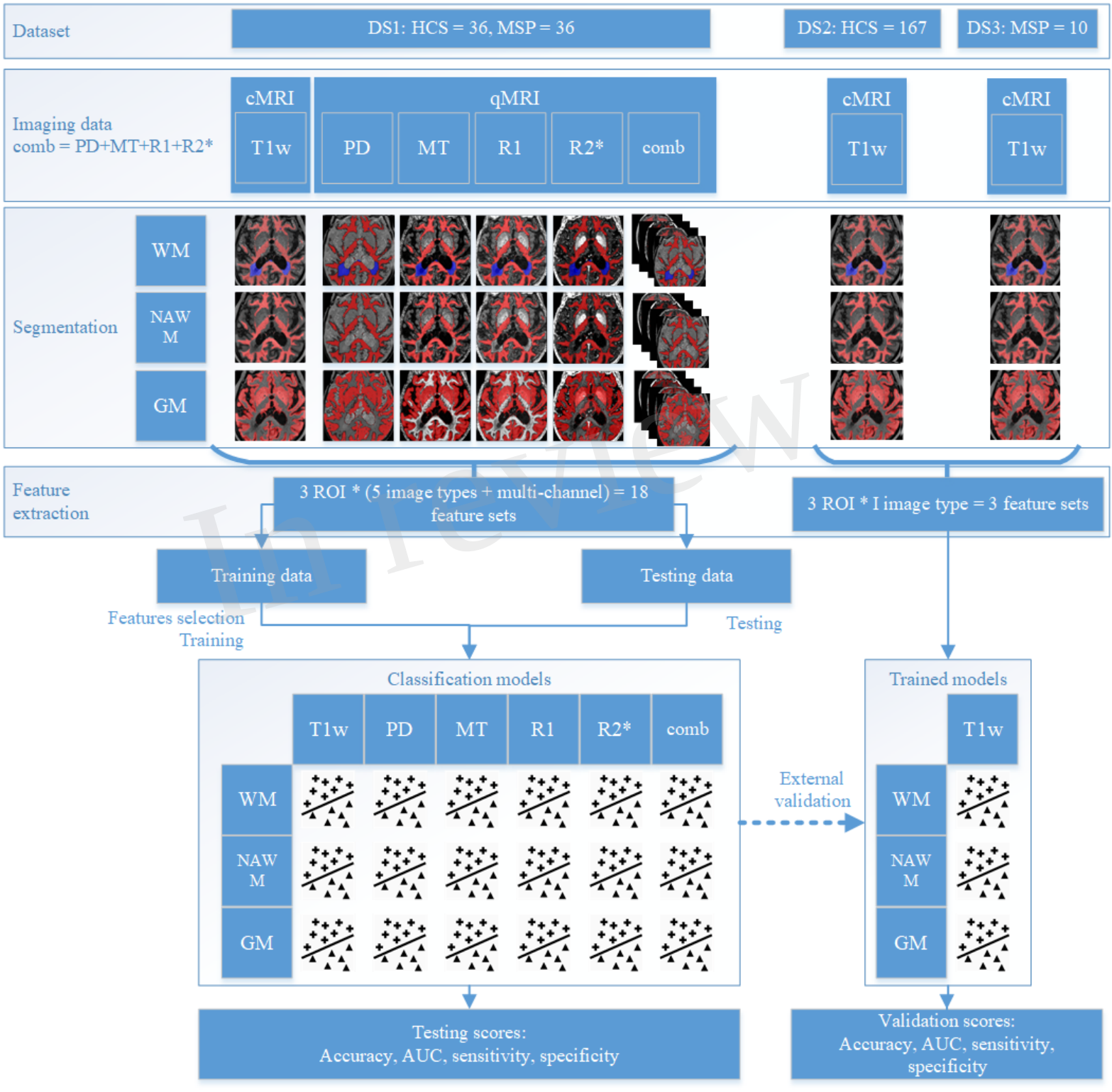

Dataset
DS1: HCS = 36, MSP = 36
DS2: HCS = 167
DS3: MSP = 10
Imaging data
comb = PD+MT+R1+R2*
cMRI
T1w
qMRI
PD
MT
R1
R2*
comb
cMRI
T1w
cMRI
T1w
Segmentation
WM
NAWM
GM
Feature extraction
3 ROI * (5 image types + multi-channel) = 18 feature sets
3 ROI * I image type = 3 feature sets
Training data
Testing data
Features selection
Training
Testing
Classification models
T1w
PD
MT
R1
R2*
comb
WM
NAWM
GM
External validation
Trained models
T1w
WM
NAWM
GM
Testing scores:
Accuracy, AUC, sensitivity, specificity
Validation scores:
Accuracy, AUC, sensitivity, specificity

Figure 3.TIF

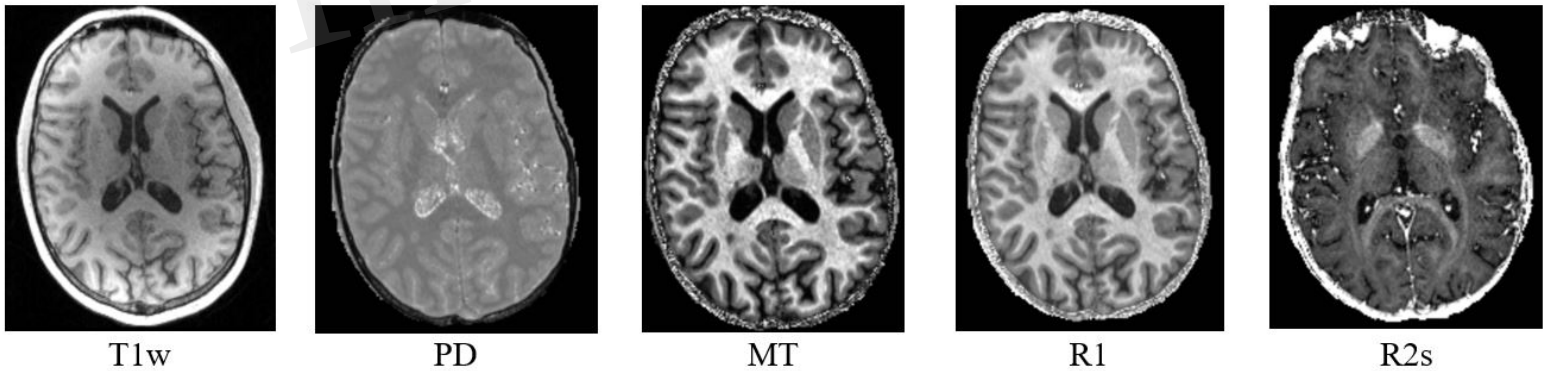


Figure 4.TIF

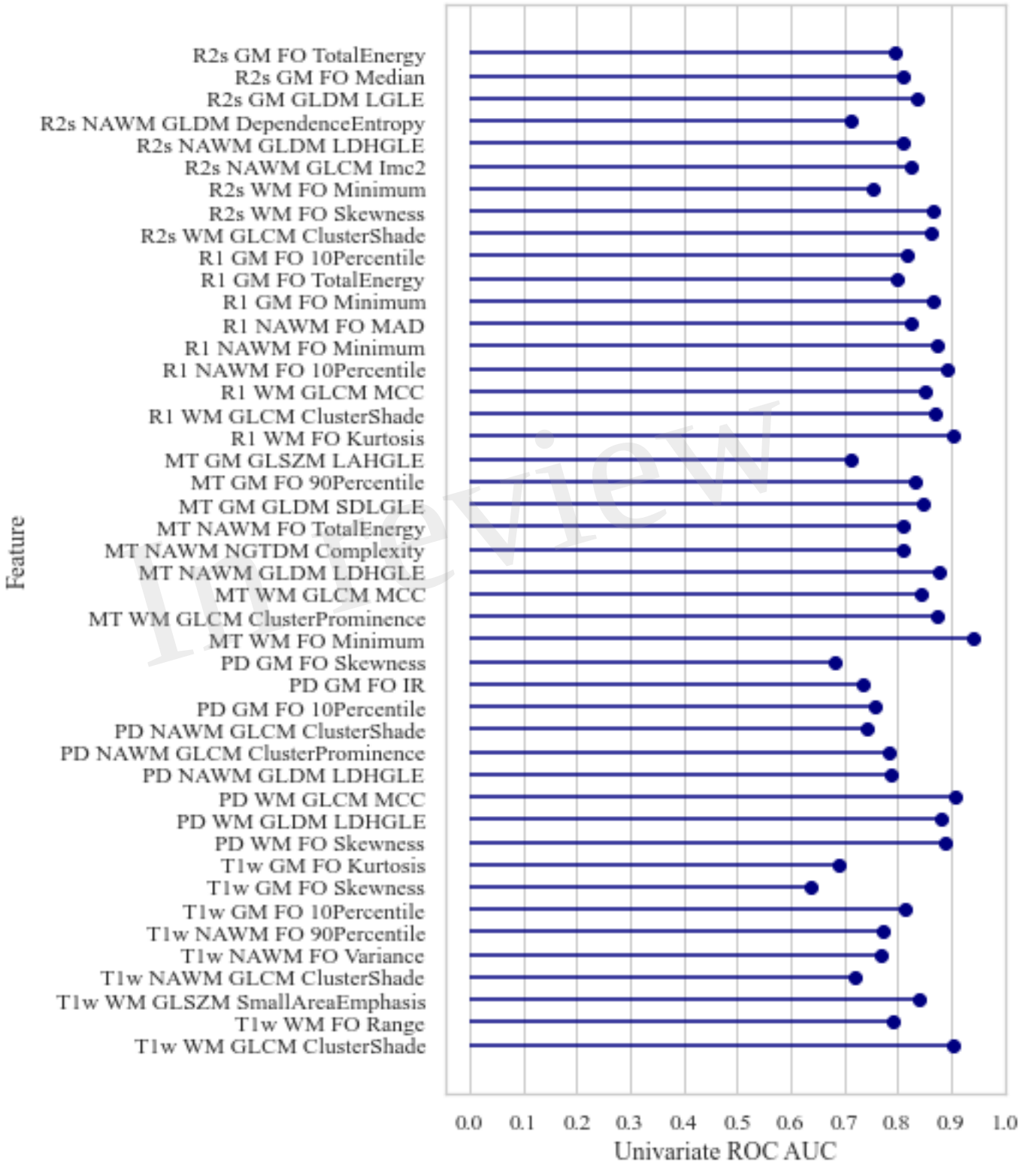


Figure 5.TIF