Technische Universität München
TUM School of Medicine and Health

Unlocking the Potential of Conventional MRI for MS Research: Automated Lesion Detection and Intensity Analysis

Tun Wiltgen

Vollständiger Abdruck der von der TUM School of Medicine and Health der Technischen Universität München zur Erlangung eines

Doctor of Philosophy (Ph.D.)

genehmigten Dissertation.

Vorsitz: apl. Prof. Dr. Christian Sorg

Betreuer: apl. Prof. Dr. Mark Mühlau

Prüfende der Dissertation:

1. Prof. Dr. Benedikt Wiestler
2. Prof. Dr. Valentin Riedl

Die Dissertation wurde am 12.08.2025 bei der TUM School of Medicine and Health der Technischen Universität München eingereicht und durch die TUM School of Medicine and Health am 07.10.2025 angenommen.

Contents

Abstract

In multiple sclerosis (MS), magnetic resonance imaging (MRI) is a key element in diagnosing, monitoring, and better understanding the disease. In clinical routine, conventional MRI sequences are mainly used for manual assessment of demyelinating lesions, which is a time-consuming and rater-dependent task. Therefore, recent research has focused on developing automated lesion segmentation tools to optimize the clinical workflow and enable fast and accurate segmentation of large datasets. Beyond these applications, conventional MRI has been limited due to arbitrary scales and poor interpretability of image intensities. Consequently, quantitative MRI (qMRI) sequences have been developed to quantify biophysical tissue properties and help understand the pathology in MS. However, qMRI is not widely available as the acquisition is lengthy and not yet part of the standard clinical routine. This thesis aimed to address these challenges and explore the unexploited potential of conventional MRI. One goal was to develop a holistic, robust, and open-source deep learning-based lesion segmentation tool to optimize the lesion detection and segmentation process. Another objective was to leverage the largely overlooked information content of conventional T1-weighted (T1w) image intensities to overcome the limited availability of qMRI and develop novel strategies for gaining new insights into MS. This was addressed by developing intensity scaling methods that avoid cerebral reference regions, acknowledging the diffuse nature of MS pathology in the entire brain.

The lesion segmentation tool based on artificial intelligence (LST-AI) presented in this thesis was developed as an ensemble of three 3D U-Nets. As a holistic tool, LST-AI includes all required processing steps and generates binary lesion masks and corresponding lesion statistics. The 3D U-Nets were trained on a large in-house dataset, and the ensemble network was validated on multiple external datasets to confirm robust performance across images from different domains. The lesion segmentation performance of LST-AI was assessed on the voxel and lesion level and was compared to benchmark methods. Regarding intensity scaling, a systematic review was conducted to summarize approaches that do not rely on brain reference regions. Eligible studies were screened and included if they assessed the effects of intensity scaling on variance reduction and the relation between conventional MRI and quantitative imaging, histology, or clinical parameters. Based on the insights from the systematic review, multiple intensity scaling methods were applied to and validated on conventional T1w images. First, a data-driven reference region selection was conducted, followed by technical and biological validation of scaling methods. Technical validation assessed the inter-image variance of intensity-scaled images and their correlation with quantitative R1 relaxation maps. Biological validation included group comparison of normal-appearing white matter (NAWM) intensities between low (≤ 3) and high (> 3) Expanded Disability Status Scale (EDSS) scores, evaluating the sensitivity of scaled conventional T1w image intensities towards disease-related effects. Finally, the most promising intensity scaling method was implemented to analyze relations of intensity patterns with EDSS scores and Symbol Digit Modalities Test (SDMT) scores. Multiple linear regression was used to determine whether intensity-based features from conventional T1w images could provide a new source of information extending beyond traditional features, such as age, sex, disease duration, and lesion volume.

LST-AI demonstrated superior performance across all datasets, segmenting lesions with higher accuracy and sensitivity than benchmark methods. Notably, the purpose-built architecture and holistic design led to a strong performance that was maintained across different datasets, confirming its robustness and generalizability as a widely applicable tool for MS lesion segmentation. Regarding intensity scaling, the systematic review revealed that most studies avoiding cerebral reference regions relied on T1w/T2-weighted ratios and only a few studies analyzed intensity-scaled T1w or T2-weighted images individually.

Studies reported consistent variance reduction and correlations with quantitative imaging, clinical data, and histology. However, no clearly superior method could be identified, with a striking methodological heterogeneity across studies. The review thus underscored both the potential of conventional MRI intensities and methodological gaps, which were addressed in this thesis. The data-driven selection process identified temporal fatty tissue and temporal muscle as suitable reference regions. Among intensity scaling methods, the most consistent variance reduction was achieved by linear scaling with fatty tissue, which also led to an increased correlation of scaled T1w image intensities with R1 values compared to unprocessed or bias-corrected T1w images. Regarding biological validation, statistically significant differences between low and high EDSS groups were detected by mean NAWM intensities of fat-scaled T1w images and notably by coefficients of variation of NAWM intensities. Thus, conventional T1w image intensities contain valuable information on disease severity. This was further underlined by the results from correlation analyses with clinical scores, since the mean lesion intensities and coefficients of variation of NAWM in fat-scaled T1w images significantly correlated with EDSS and SDMT.

Overall, the presented results underline that conventional MRI, combined with appropriate processing tools, can serve as a valuable source of information for clinical and research applications. Accurate and robust lesion segmentation, which is essential for reliable diagnosis and monitoring of MS, was provided through the introduction and validation of LST-AI. As a largely unrecognized but complementary source of information, the intensities of conventional T1w images were found to contain biologically relevant information that extends beyond traditional biomarkers. In summary, the tools developed in this thesis enable new applications relying on conventional MRI to improve clinical assessment and realize large-scale studies to advance the understanding of disease mechanisms in MS.

Publications related to this thesis and their quotations

The results in this thesis have been published in three articles in peer-reviewed scientific journals:

1. [1] **Wiltgen T.**, McGinnis J., Schläger S., Kofler F., Voon C., Berthele A., Bischl D., Grundl L., Will N., Metz M., Schinz D., Sepp D., Prucker P., Schmitz-Koep B., Zimmer C., Menze B., Rückert D., Hemmer B., Kirschke J., Mühlau M. and Wiestler B. (2024). LST-AI: A deep learning ensemble for accurate MS lesion segmentation. NeuroImage: Clinical, 42, 103611, doi: 10.1016/j.nicl.2024.103611

2. [2] **Wiltgen T.**, Voon C., Van Leemput K., Wiestler B., Mühlau M. (2024). Intensity scaling of conventional brain magnetic resonance images avoiding cerebral reference regions: A systematic review. PloS one, 19(3), e0298642, doi: 10.1371/journal.pone.0298642

3. [3] **Wiltgen T.**, McGinnis J., Berg R., Voon C., Puonti O., Giglhuber K., Ganter C., Zimmer C., Hemmer B., Wiestler B., Kirschke J., Preibisch C., Mühlau M. (2025). Towards quantitative intensity analysis of conventional T1-weighted images in multiple sclerosis. NeuroImage, 318, 121395, doi: 10.1016/j.neuroimage.2025.121395

Where applicable, selected tables, figures, and text from these publications have been directly cited in this thesis and are identified by quotation marks and the respective reference. In the quoted text, references, including those to figures and tables, have been modified to align with this thesis, without specific indication of the change. To minimize editing of direct quotations, repeated introductions of abbreviations have been preserved. Additionally, full terms have been added in quotations where abbreviations had not yet been introduced in the thesis.

Acronyms

ANTs Advanced Normalization Tools
ASD average surface distance
BBB blood-brain barrier
CIS clinically isolated syndrome
CNN convolutional neural networks
CNS central nervous system
CoV coefficient of variation
CSF cerebrospinal fluid
DSC Dice similarity coefficient
DWI diffusion-weighted imaging
EDSS Expanded Disability Status Scale
FLAIR fluid-attenuated inversion recovery
FN false negatives
FP false positives
GM gray matter
ICBM International Consortium for Brain Mapping
IT infratentorial
JC juxtacortical
MNI Montreal Neurological Institute
MPM multiparameter mapping
MPRAGE magnetization-prepared rapid gradient echo
MRI magnetic resonance imaging
MT magnetization transfer
MTR magnetization transfer ratio
MWF myelin water fraction
NAWM normal-appearing white matter
PDw proton density-weighted
PPMS primary progressive multiple sclerosis
PPV positive predictive value
PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PV periventricular
RF radio frequency
RRMS relapsing-remitting multiple sclerosis
SC subcortical
SDMT Symbol Digit Modalities Test
SPM12 Statistical Parametric Mapping version 12
SPMS secondary progressive multiple sclerosis
T1w T1-weighted
T2w T2-weighted
TP true positives
WM white matter

List of Figures

List of Tables

1 Introduction

Multiple sclerosis: overview

Multiple sclerosis (MS) is a chronic, neurodegenerative autoimmune disease affecting the central nervous system (CNS) through inflammatory demyelination in the brain and spinal cord that can lead to neurological deficits and progression of clinical disability [4]. Although it is the most frequent chronic disease of the CNS in young adults, the causes of MS are still unknown. Over the years, extensive research has improved the understanding of MS. Its development is likely influenced by history of infections (e.g., Epstein–Barr virus), lifestyle (e.g., smoking), environmental factors (e.g., vitamin D), and genetic risk factors [4–7]. The course of MS is very heterogeneous. The initial phase is most frequently characterized by a first episode of neurological deficits caused by demyelinating lesions in the brain, optic nerve, brainstem, or spinal cord. The neurological deficits are often partially or completely reversible and can resolve over time (remission). The initial clinical presentation is recognized as clinically isolated syndrome (CIS), and is typically followed by further episodes (relapses) which lead to a confirmed diagnosis of relapsing-remitting multiple sclerosis (RRMS). Such episodes usually manifest themselves through symptoms including (but not limited to) visual loss, sensory discomfort (e.g., numbness, tingling), motor manifestations such as paresis or spasticity, impairment of ocular movement, cognitive impairment, or fatigue [4]. The definite diagnosis of MS according to the 2017 revised McDonald criteria requires demonstration of dissemination in space, meaning that two different locations of the CNS are affected, and dissemination in time, meaning that two or more demyelinating events happen at multiple points in time [8]. Dissemination in space can be demonstrated either with clinical evidence (episodes with symptoms indicating involvement of different sites in the CNS) or with identification of lesions located in two of the four typical regions in the CNS (periventricular (PV), juxtacortical (JC), infratentorial (IT), and spinal cord) using magnetic resonance imaging (MRI) [8, 9]. Dissemination in time can be determined with two episodes at different points in time (with at least partial resolution), which can present itself by two separate symptomatic episodes, through simultaneous presence of gadolinium-enhancing and non-enhancing lesions in postcontrast MRI, indicating new active lesions and older inactive lesions, or through the presence of new lesions in follow-up MRI scans [8, 9]. Alternatively, the presence of oligoclonal bands in cerebrospinal fluid (CSF) can substitute for dissemination in time based on clinical and MRI findings, as this suggests ongoing inflammatory processes [4]. In later stages, patients frequently experience fewer or no relapses but a continuous progression of disability; this phase of MS is called secondary progressive multiple sclerosis (SPMS) [4–6, 9]. About 10 percent of people with MS experience a continuous progression of disability without relapses from onset; this type of MS is called primary progressive multiple sclerosis (PPMS). The extent of disease severity and clinical disability is regularly assessed in routine follow-up examinations using established scales and tests, such as the Expanded Disability Status Scale (EDSS) and the Symbol Digit Modalities Test (SDMT). The EDSS is a non-linearly increasing scale that ranges from 0 (normal neurological examination) to 10 (death due to MS) and indicates how severely functional systems are affected by MS [4, 10]. The SDMT score measures cognitive impairment through assessing information processing speed. Scores range from 0 to 110, reflecting the number of correct symbol-digit pairings completed within 90 seconds, with lower scores indicating greater impairment [11].

The exact mechanisms of the disease are very complex, largely unclear, and the subject of ongoing research. The hallmark of MS is the development of acute inflammatory demyelinating sites in the white matter (WM) in the brain or spinal cord, known as lesions. The formation of a lesion represents the occurrence of an inflammatory process that starts with the disruption of the blood-brain barrier (BBB), likely due to a dysregulated peripheral adaptive immune system that triggers the activation of pro- and anti-

inflammatory cells [4, 6, 7, 12, 13]. Yet, the innate immune system also plays a role in MS. A recent study found that, in the brain, lesion formation is accompanied by microglia accumulation and activation, causing the expression of pro- and anti-inflammatory cytokines and molecules involved in immune regulation [12]. Once the BBB has been disrupted, different immune cell types infiltrate the CNS, leading to an amplification of the immune response and generating a volatile environment of inflammation, which causes injury of olygodendrocytes, demyelination, and neuronal and axonal damage [4–7, 12]. After the formation of a demyelinating lesion during a relapse, partial remyelination of axons is possible, which is usually accompanied by symptom recovery [4, 7]. The process includes undifferentiated oligodendrocyte precursors that surround the lesion and start rebuilding myelin sheaths around naked axons [4, 7]. The repaired myelin is typically shorter and thinner, but can still restore conduction of action potentials and ensure axon survival to some extent [5, 6]. The incidence and the extent of remyelination are very heterogeneous and likely depend on the extent of the inflammation, age, disease duration, lesion location, presence of oligodendrocyte progenitor cells, axonal integrity, and genetic factors [4–7]. In the absence of remyelination, the tissue injury remains and causes permanent loss of axons and neurodegeneration, leading to brain atrophy and accumulation of clinical disability [4, 5, 7]. Notably, the disease pathology involves two important additional mechanisms. They are commonly attributed to the progressive phase of MS and characterized by lesion formation in the cortical gray matter (GM) and neurodegeneration in the non-lesioned WM, the normal-appearing white matter (NAWM) [4, 7, 12, 14–16]. Cortical lesions differ from WM lesions in that they show less BBB breakdown, lower degree of inflammation, and more efficient remyelination, potentially indicating differing lesion formation processes compared to WM lesions [4, 14, 15]. Studies investigating NAWM in MS have shown that diffuse inflammatory processes also take place in non-lesioned WM, leading to axonal degeneration and decreased fiber density [13, 14, 17–19]. Specifically, the inflammation is characterized by microglia activation and formation of microglia nodules that can mediate axonal damage [12, 14, 19]. Studies investigating the NAWM based on histology revealed that little or no demyelination accompanies the axonal degeneration in the NAWM [12–14, 17–19], suggesting that the diffuse inflammatory activity in the NAWM might be an early indicator for the formation of a new lesion [12], and involve separate mechanisms that are independent of demyelinating WM lesion formation [4, 14, 19]. In line with this assumption, diffuse tissue injury in the NAWM has also been observed to be more pronounced in the progressive phase of MS, which is marked by a transition from acute inflammatory relapses characterized by WM lesions to chronic inactive lesions and diffuse neuro-axonal degeneration with neuronal loss and progressive disability accumulation [4–7, 14]. Hence, the roles of demyelinating WM lesions and diffuse NAWM injury in MS evolve throughout the disease course and require further investigation to determine how they interact, to what degree they depend on each other, and to what extent their underlying mechanisms are independent.

Multiple sclerosis and magnetic resonance imaging

In general, the diagnostic criteria of CIS, RRMS, SPMS, and PPMS include clinical, CSF, and MRI findings [8, 9]. The latter is routinely used in the clinical setting and has become an established and valuable tool for diagnosing and monitoring MS, positioning the identification and assessment of demyelinating lesions in MRI as a central aspect in routine clinical workflow. Currently, this requires a neuroradiologist to manually identify the lesions based on T1-weighted (T1w) and T2-weighted (T2w) images, in which demyelinating lesions appear hypo- or isointense and hyperintense, respectively. With the administration of gadolinium-based contrast agent, active lesions can be identified as hyperintense in T1w images because of the contrast uptake due to the increased permeability of the BBB [9]. This presents a helpful tool for diagnosis and clinical follow-up, because active lesions indicate ongoing disease activity through acute inflammation, providing valuable information for treatment strategies. However, up to date, routine MRI hardly provides information beyond mere WM lesion identification. Many studies aim to exploit the potential of MRI through new, insightful, and advanced imaging techniques and image processing tools.

One established approach is MRI-based morphometry, which quantifies structural characteristics and volumes of brain regions to find correlates of disease progression. For example, brain or cortical vol-

ume changes have been used to measure atrophy, representing neuroaxonal degeneration, and provide a valuable reference for clinical studies to monitor the progression of MS in the brain [20–22]. Similarly, the lesion number and the total lesion volume are frequently considered complementary indicators of disease progression in MS [5, 23–25]. For reliable lesion number and volume calculation, lesions must be accurately segmented. However, the manual segmentation of lesions is time-consuming and may exhibit considerable variability between neuroradiologists. Therefore, a new field of research has grown in the past two decades, aiming at developing robust and accurate automated lesion segmentation tools, primarily based on conventional MRI sequences that are part of the routine examinations in MS (i.e., T1w, T2w, and fluid-attenuated inversion recovery (FLAIR)) [26–43]. Early automated segmentation tools were mainly based on statistical methods and traditional machine learning models, involving engineered and manually selected features [26]. Over time, the research field has undergone a paradigm shift toward applying more advanced deep learning-based methods, especially convolutional neural networks (CNN), which demonstrate superior segmentation accuracy compared to previous methods [27, 31–35, 37, 39]. However, the development of such tools often comes with notable challenges. Large training datasets are required to avoid overfitting of deep learning models, which have a high architectural complexity with a large number of model parameters that enable them to learn features automatically without a manual procedure. In addition, models should be evaluated on external validation datasets to ensure their generalizability to data from different centers with different scanners and protocols, and to reliably assess the performance on heterogeneous datasets without re-training the models. Finally, these tools should guarantee robust segmentation performance and be easily applicable and publicly available to guarantee deployment and transferability. Within the context of this thesis, a new lesion segmentation tool based on artificial intelligence (LST-AI) was developed, addressing these main ongoing challenges in deep learning-based automated lesion segmentation in MS.

Another extensively researched approach in recent years is the application of quantitative MRI (qMRI), which aims to quantify specific biophysical tissue properties such as myelin content, iron content, or fiber density [44, 45]. In this regard, conventional MRI is limited, since it is optimized for good visual contrast, which leads to arbitrary intensity scales due to multiple acquisition-specific factors, including hardware settings (e.g., coil sensitivity, manufacturer, signal amplifier gain), scanner software, and anatomy and positioning of the patient. Consequently, intensity scales of conventional MRI are affected by technically induced inter-image variability, undermining the direct comparison of intensities. Various qMRI techniques have been developed to overcome these obstacles and have shown great potential in their respective field of application. The most established techniques include magnetization transfer (MT) imaging [46–52], diffusion-weighted imaging (DWI) [53–57], and quantitative susceptibility mapping (QSM) [58]. In addition, one of the most established and contributing techniques has been relaxometry, which measures the absolute relaxation times T1 and T2 of the underlying tissue. It helps in understanding tissue pathology as the relaxation times of brain tissues are sensitive to the composition of the microstructural environments, such as water, iron, and myelin content [44, 59–63]. Specifically, the relation between relaxation rates R1 (= 1/T1) and MS pathology has been thoroughly investigated, even using histology, revealing associations of R1 values in lesions and NAWM with myelin content, axonal density, and tissue integrity [44, 49, 50, 59, 64–73]. In addition, the R1 values have been used to investigate the predictive power of early intensity patterns and their correlation with clinical disability [59, 69, 74–76]. Hence, in addition to morphometry, intensity-based analysis using qMRI can provide valuable information on the underlying mechanisms causing disease progression. However, qMRI has not become standard in clinical practice due to the lengthy acquisition process and the frequent need for off-line analysis. Consequently, it is not frequently used in large cohorts, which would, however, be important to reliably detect unknown but potentially important associations between imaging biomarkers and disease progression.

Given these challenges, and considering that T1w images are widely available, recent research has explored methods to estimate quantitative images from conventional T1w MRI by applying different intensity scaling techniques [77–90]. Applying intensity scaling aims to remove the technically induced inter-image

variance while preserving the biologically induced variance, which can reveal tissue characteristics relevant to disease pathology. Some studies rely on linear scaling factors extracted from reference regions with consistent tissue composition across subjects, where T1w signal variations primarily reflect technical rather than biological factors. By normalizing these reference signals, technical variance can effectively be modeled and corrected for without obscuring disease-related biological variations in regions of interest (e.g., in WM). In the specific case of MS, which affects the integrity of virtually all brain tissue compartments, extracerebral reference regions should be considered, since they are less likely to be directly affected by the disease. In particular, fatty tissues, temporal muscle, and eyes have been chosen as extracerebral reference regions in previous studies [77, 81, 83, 84, 86, 90]. Unfortunately, only a few studies have assessed multiple intensity scaling methods [80, 81] and validated the information content of intensity-scaled conventional T1w images through comparison with qMRI, such as the magnetization transfer ratio (MTR) or DWI [77, 83, 84]. Another method has been increasingly applied in recent studies, which aims to model the T1-dependency of conventional T1w signals. These studies use the corresponding signal equation or statistical methods to estimate T1 values from conventional T1w image intensities and validate the resulting images against actual T1 or R1 maps [78, 82, 87–89].

Overall, the potential of conventional MRI remains to be fully exploited in MS, at times when research on accurate and automated lesion segmentation tools has become increasingly popular, and utilizing intensity-based information of conventional T1w images has become a largely accepted approach.

Thesis objectives and contribution

The goal of this thesis was to leverage advanced image processing techniques to unlock the untapped potential of conventional MRI in MS. First, LST-AI was developed, combining conventional MRI and new advances in deep learning medical image segmentation to efficiently detect and accurately segment WM lesions, which is crucial in diagnosing and monitoring MS. We aimed to develop and validate a holistic tool comprising all required processing steps, requiring no re-training and using only conventional raw T1w and FLAIR image pairs as input. Second, a systematic review was conducted on intensity scaling to assess existing methods that avoid cerebral reference regions, which is important for potential applications in MS. It highlights the validity of existing methods, the biological information they can provide, and identifies methods that show potential in generating images with similarities to qMRI. Third, multiple intensity scaling methods were validated with qMRI (R1 maps) and compared to each other to identify suitable techniques that can leverage the informative wealth of conventional T1w image intensities. In a final step, LST-AI and intensity scaling were combined to analyze the correlation of conventional T1w image intensities, extracted from lesions and NAWM, with clinical scores.

2 Data and methods

Multiple datasets were used and will be presented in this section. Regarding image processing, the steps included in the LST-AI lesion segmentation tool differed from those employed in the validation of intensity scaling methods and clinical correlation analysis; therefore, the specific image processing steps will be described separately. Finally, the statistical analysis methods used for the evaluation of LST-AI, for the systematic review on intensity scaling, for the validation of intensity scaling methods, and for the clinical correlation analysis will be described.

2.1 Datasets

First, the datasets used for the training and testing of LST-AI will be presented. Second, the characteristics of datasets included in the intensity scaling projects are described, including data from the literature for the systematic review and in-house data for the validation of intensity scaling methods. Lastly, a subset of this in-house dataset will be described, which was used for the clinical correlation analysis. The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans [91] was respected; all studies were approved by the local ethics committee.

2.1.1 LST-AI

"For the training set, we used an in-house dataset consisting of 491 paired 3D Fluid-Attenuated Inversion Recovery (FLAIR) and 3D T1-weighted (T1w) images acquired on a 3.0 T Achieva scanner (Philips Medical Systems, Best, The Netherlands) to train both our proposed LST-AI segmentation model and the nnUNet baseline. [...] The test set includes four publicly available datasets: (i) msisbi: ISBI 2015 training data [92] (https://smart-stats-tools.org/lesion-challenge-2015); (ii) msljub: dataset published by Laboratory of Imaging Technologies [93] (https://lit.fe.uni-lj.si/en/research/resources/3D-MR-MS/); (iii) mssegtest: MICCAI 2016 challenge test dataset [94] (https://shanoir.irisa.fr/shanoir-ng/welcome) and (iv) mssegtrain: MICCAI 2016 challenge training dataset [94] (https://shanoir.irisa.fr/shanoir-ng/welcome). One case (msseg-test-center07-08) was removed from the mssegtest dataset because it included incorrect ground truth data. In total, the test set consists of 103 images from 87 subjects (note that the publicly available ISBI dataset is a longitudinal dataset). Further characteristics of the datasets, including the number of lesions, the total lesion volume, and the mean lesion volume are provided in Table 2.1. Details on image acquisition are provided in Table 2.2." [1]

Table 2.1

"Characteristics of the datasets. One in-house (training) dataset was used, as well as the public datasets msisbi from the ISBI 2015 challenge [92], msljub published by the Laboratory of Imaging Technologies [93], and mssegtest and mssegtrain which are the testing and training datasets from the MICCAI 2016 challenge, respectively [94]" [1]. The table was reproduced from Wiltgen et al. [1].

Dataset	#subjects	#scans	age (years) mean ± sd	female/ male	diagnosis (number of images)	number of lesions i) mean ± sd ii) median (IQR)	total lesion volume (mm^3) i) mean ± sd ii) median (IQR)	mean lesion volume (mm^3) i) mean ± sd ii) median (IQR)	publication	link
in-house training[α]	491	491	34.3 ± 9.5	330/161	RRMS (422) CIS (66) ON (3)	i) 25.54 ± 30.59 ii) 15.0 (6.0–33.0)	i) 3492.96 ± 7300.31 ii) 1244.0 (419.5–3767.5)	i) 150.81 ± 387.97 ii) 73.5 (47.0–130.8)	N/A	N/A
msisbi	5	21	43.5 ± 10.3	4/1	RRMS (4) PPMS (1)	i) 45.95 ± 20.92 ii) 41.0 (34.0–47.0)	i) 12889.76 ± 11095.38 ii) 7354.0 (3678.0–18425.0)	i) 255.25 ± 168.73 ii) 175.1 (119.1–371.3)	[92]	(1)
msljub	30	30	39.3 ± 10.1	23/7	RRMS (24) SPMS (2) PRMS (1) CIS (2) Unspecified (1)	i) 111.23 ± 106.68 ii) 92.0 (31.25–125.0)	i) 17336.87 ± 16115.41 ii) 14046.5 (1758.0–28430.25)	i) 178.47 ± 170.36 ii) 117.9 (49.1–208.0)	[93]	(2)
mssegtest	37	37	46.8 ± 10.3	29/8	N/A	i) 44.89 ± 42.11 ii) 29.0 (13.0–64.0)	i) 12672.73 ± 15099.75 ii) 7348.0 (1453.0–17271.0)	i) 275.8 ± 272.36 ii) 190.9 (120.4–328.4)	[94]	(3)
mssegtrain	15	15	41.6 ± 9.8	8/7	N/A	i) 41.67 ± 30.21 ii) 39.0 (18.0–56.5)	i) 20729.87 ± 20606.48 ii) 12366.0 (3783.0–33198.5)	i) 643.33 ± 904.95 ii) 237.1 (125.1–752.3)	[94]	(3)

Abbreviations: CIS: clinically isolated syndrome, IQR: interquartile range, N/A: not applicable/available, ON: optic neuritis, PPMS: primary progressive multiple sclerosis, RRMS: relapsing-remitting multiple sclerosis, sd: standard deviation, SPMS: secondary progressive multiple sclerosis.
(1) https://smart-stats-tools.org/lesion-challenge-2015.
(2) https://lit.fe.uni-lj.si/en/research/resources/3D-MR-MS/.
(3) https://shanoir.irisa.fr/shanoir-ng/welcome.
[α] For stringency, all patients were reclassified according to the 2017 McDonald criteria [8].

Table 2.2

MRI acquisition settings of the datasets used for training and testing of LST-AI. The table was reproduced from Wiltgen et al. [1].

Dataset	scanner	field strength	sequence	voxel size	#scans
in-house training	Achieva, Philips Medical Systems	3.0 T	T1w: TR = 9 ms, TE = 4 ms, FA = 8° (MPRAGE) FLAIR: TR = 10000 ms, TE = 140 ms, TI = 2750 ms	1x1x1 mm^3 0.9x0.9x1.5 mm^3	491
msisbi	Philips Medical Systems	3.0 T	T1w: TR = 10.3 ms, TE = 6 ms, FA = 8° (MPRAGE) FLAIR: TE = 68 ms, TI = 835 ms	0.82x0.82x1.17 mm^3 0.82x0.82x2.2 mm^3	21
msljub	Siemens Magnetom Trio	3.0 T	T1w: TR = 2000 ms, TE = 20 ms, TI = 800 ms, FA = 120° (turbo inversion recovery magnitude) FLAIR: TR = 5000 ms, TE = 392 ms, TI = 1800 ms, FA = 120°	0.42x0.42x3.3 mm^3 0.47x0.47x0.8 mm^3	30
mssegtest	Siemens Verio	3.0 T	T1w: TR = 1900 ms, TE = 2.26 ms, FA = 9° FLAIR: TR = 5000 ms, TE = 400 ms, TI = 1800 ms, FA = 120°	1x1x1 mm^3 0.5x0.5x1.1 mm^3	10
	General Electrics Discovery	3.0 T	T1w: TR = [7.5,8] ms, TE = 3.2 ms, FA = 10° FLAIR: TR = 9000 ms, TE = [140,145] ms, TI = [2355,2362] ms, FA = 90°	0.47x0.47x0.6 mm^3 0.47x0.47x0.9 mm^3	8
	Siemens Aera	1.5 T	T1w: TR = 1860 ms, TE = 3.37 ms, FA = 15° FLAIR: TR = 5000 ms, TE = 336 ms, TI = 1800 ms, FA = 120°	1.08x1.08x0.9 mm^3 1.03x1.03x1.25 mm^3	9
	Ingenia, Philips Medical Systems	3.0 T	T1w: TR = 9.4 ms, TE = 4.3 ms, FA = 8° FLAIR: TR = 5400 ms, TE = 360 ms, TI = 1800 ms, FA = 90°	0.74x0.74x0.85 mm^3 0.74x0.74x0.7 mm^3	10
mssegtrain	Siemens Verio	3.0 T	T1w: TR = 1900 ms, TE = 2.26 ms, FA = 9° FLAIR: TR = 5000 ms, TE = 400 ms, TI = 1800 ms, FA = 120°	1x1x1 mm^3 0.5x0.5x1.1 mm^3	5
	Siemens Aera	1.5 T	T1w: TR = 1860 ms, TE = 3.37 ms, FA = 15° FLAIR: TR = 5000 ms, TE = 336 ms, TI = 1800 ms, FA = 120°	1.08x1.08x0.9 mm^3 1.03x1.03x1.25 mm^3	5
	Ingenia Philips Medical Systems	3.0 T	T1w: TR = 9.4 ms, TE = 4.3 ms, FA = 8° FLAIR: TR = 5400 ms, TE = 360 ms, TI = 1800 ms, FA = 90°	0.74x0.74x0.85 mm^3 0.74x0.74x0.7 mm^3	5

Abbreviations: FA: flip angle, FLAIR: fluid-attenuated inversion recovery, MPRAGE: magnetization-prepared rapid gradient echo, T1w: T1-weighted, TE: echo time, TI: inversion time, TR: repetition time

2.1.2 Intensity scaling

Intensity scaling: systematic review

The dataset used in the systematic review resulted from the selection process illustrated in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 [95] flow chart in Figure 2.1. First, we "conducted the literature search on 1 November 2022 on the following web databases: Scopus, PubMed, and Web of Science. Our search settings included combinations of the terms 'magnetic resonance', 'imag*', 'intensit*', 'normali*', 'brain*', and combinations of synonyms thereof. The exact settings for each database are presented in Table 2.3" [2]. Second, duplicates were removed, and all studies from the literature search (n = 3825) underwent screening. This process comprised two steps: 1) title and abstract screening and 2) full-text screening. In both steps, "studies were included if they met the following inclusion criteria: human study; conventional T1w or T2w brain MRI; intensity scaling avoiding cerebral reference regions; comparison of scaled intensities with another method, quantitative imaging or MRI measures related to tissue microstructure, or histology; description of scaling method; accessible scientific paper; extractable results" [2]. The studies were screened by two researchers to establish their eligibility for inclusion, and any disagreements were resolved through discussion. Hence, if studies seemingly fulfilled the criteria during title and abstract screening, they proceeded to the next step, namely full-text screening, where a more detailed screening was conducted. Of 3825 studies, 245 advanced to the full-text screening, of which 24 were finally included in the systematic review. Among the included studies, "seven different scaling methods were applied. Most studies implemented the method proposed by Ganzetti et al. [83], which uses the median intensity values of extracerebral reference regions (eyeball and temporal muscle) to scale T1w and T2w image intensities before calculating the T1w/T2w-ratio [...]. Some studies also calculated T1w/T2w-ratio, but they did not scale T1w and T2w images beforehand. The studies by Brown et al. [77] and Gilmore et al. [84] scaled intensities by dividing the conventional MR-image intensities by the median intensity of orbital fat and by the mean intensity of subcutaneous fat, respectively. Soun et al. [96] followed the method of Ganzetti et al. [83] but they used the masseter muscle instead of the temporal muscle. The study by Loizou et al. [97] used the matlab function 'gscale' and histogram normalization" [2]. A summary of the study characteristics is provided in Table 2.4.

Table 2.3
The table shows "a detailed overview of the search settings applied in each database during the literature search. Scopus: TITLE-ABS-KEY indicates that any of the words mentioned thereafter should be included in the title, abstract, or keywords; W/3 indicates that the two words before and after should be located within a 3-word span; OR and AND indicate logical operators; * indicates that the word can have further characters in that specific place. PubMed: every word is followed by [Title/Abstract], which indicates that the word has to appear in the title or the abstract; OR and AND and * have the same role as in Scopus; Web of Science: search settings are defined separately for the title, the abstract, the author keywords, and the keywords plus field" [2]. The table was reproduced from Wiltgen et al. [2].

database	Search settings
Scopus	TITLE-ABS-KEY (((magnetic W/3 resonan*) OR t1* OR mr*) AND imag* AND intensit* AND (calibr* OR quantif* OR normali* OR standardi*) AND (brain* OR head* OR cerebr*))
PubMed	(magnetic resonan*[Title/Abstract] OR t1*[Title/Abstract] OR mr*[Title/Abstract]) AND imag*[Title/Abstract] AND intensit*[Title/Abstract] AND (calibr*[Title/Abstract] OR quantif*[Title/Abstract] OR normali*[Title/Abstract] OR standardi*[Title/Abstract]) AND (brain*[Title/Abstract] OR head*[Title/Abstract] OR cerebr* [Title/Abstract])
Web of Science	(TS = (((magnetic W/3 resonan*) OR t1 OR t1w OR mr OR mri) AND imag* AND intensit* AND (calibr* OR quantif* OR normali* OR standardi*) AND (brain* OR head* OR cerebr*))) OR (AB = (((magnetic W/3 resonan*) OR t1 OR t1w OR mr OR mri) AND imag* AND intensit* AND (calibr* OR quantif* OR normali* OR standardi*) AND (brain* OR head* OR cerebr*))) OR (AK = (((magnetic W/3 resonan*) OR t1 OR t1w OR mr OR mri) AND imag* AND intensit* AND (calibr* OR quantif* OR normali* OR standardi*) AND (brain* OR head* OR cerebr*))) OR (KP = (((magnetic W/3 resonan*) OR t1 OR t1w OR mr OR mri) AND imag* AND intensit* AND (calibr* OR quantif* OR normali* OR standardi*) AND (brain* OR head* OR cerebr*)))

Abbreviations: ABS/AB: abstract, AK: author keywords, KP: keyword plus field, TS: title

Figure 2.1
The PRISMA flowchart illustrates the literature screening process applied in the systematic review, indicating the number of included/excluded studies at each stage. Ultimately, 24 studies met the inclusion criteria and were incorporated into the systematic review. The figure was reproduced from Wiltgen et al. [2].

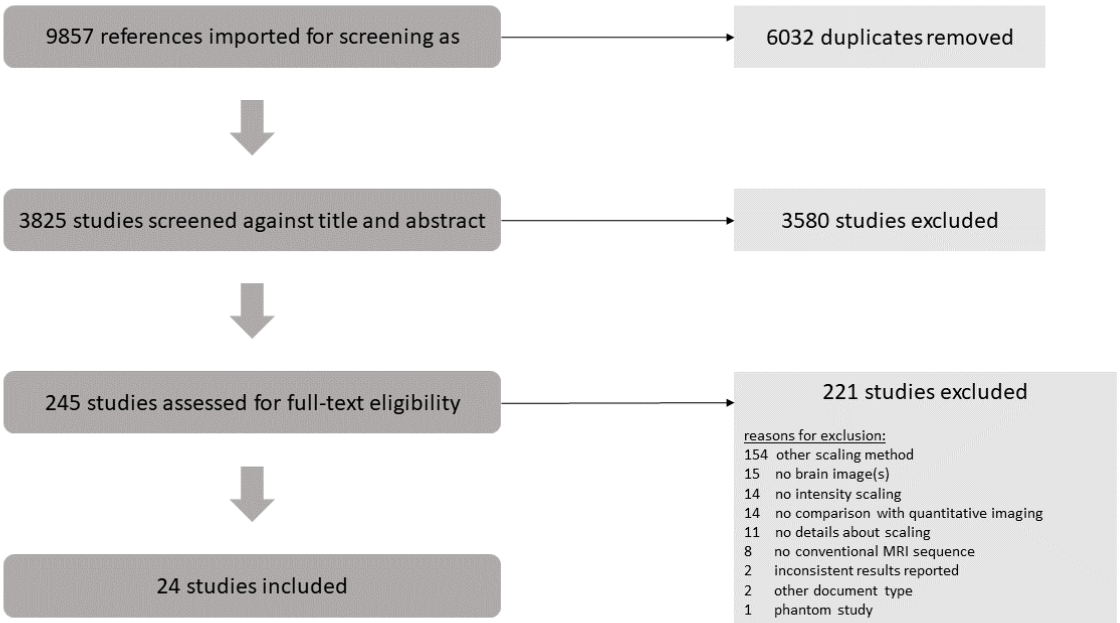


Table 2.4
Characteristics of the studies included in the systematic review. The table was reproduced from Wiltgen et al. [2].

Study	Data Source	#Subjects	Age	Health Status/Disorder	Imaging Modalities	Reference Region(s)	Scaling Method
Arshad et al. 2017 [98]	In-house	20	Mean = 45.9 years	HC	T1w, T2w, MWF, geomT2IEW	eyeballs, temporal muscle	sI(Ganzetti)
Brown et al. 2020 [77]	In-house	179	5-20 years	HC, MS, monophasic demyelinating disorders	T1w, MTR	orbital fat	sI(Brown)
Cappelle et al. 2022 [99]	In-house	207	Mean = 40.65 years	MS	T1w, T2w, FLAIR, MTR	eyeballs, temporal muscle	sI(Ganzetti) (T1w/T2w and T1w/FLAIR
Ganzetti et al. 2014 [83]	(a) (b)	63	Mean = 31.97 years	HC	T1w, T2w, FLAIR, MTR, FA	eyeballs, temporal muscle	sI(Ganzetti)
Gilmore et al. 2007 [84]	In-house	47	Mean = 39.7 weeks gestational age	HC	T1w, T2w, DWI (FA, MD)	subcutaneous fat	sI(Gilmore)
Hagiwara et al. 2018 [100]	In-house	20	Mean = 55.3 years	HC	SyMRI Data (T1w, T2w, MVF, MTsat)	eyeballs, temporal muscle	sI(Ganzetti)
Hannoun et al. 2022 [101]	In-house	52	mean = 36.1 years	HC, MS	T1w, T2w, DWI	eyeballs, temporal muscle	sI(Ganzetti)
Loizou et al. 2009 [97]	In-house	32	Mean = 30.75 years	HC, MS	T2w	CSF, air from sinuses	Matlab function gscale, Histogram Normalization
Luo et al. 2019 [102]	(c)	252	mean = 73.76 years	HC, preclinical AD, prodromal AD, AD dementia	T1w, T2w, FLAIR, PET imaging	eyeballs, temporal muscle	sI(Ganzetti)
Nakamura et al. 2017 [103]	In-house	6	Mean = 62.5 years	MS (post-mortem)	T1w, T2w, MTR		T1w/T2w-ratio
Pareto et al. 2020 [104]	In-house	22	Mean = 41.13 years	MS	T1w, T2w, MTR		T1w/T2w-ratio
Preziosa et al. 2021 [105]	In-house	25	Mean = 67.0 years	HC, MS	T1w, T2w, FLAIR (histology)	eyeballs, temporal muscle	sI(Ganzetti)
Righart et al. 2017 [106]	In-house	In-vivo: 248, Post-mortem: 9	Mean (in-vivo) = 30.8 years Mean (post-mortem) = 65.9 years	HC, MS, MS (post-mortem)	T1w, T2w, FLAIR (histology)		T1w/T2w-ratio
Saccenti et al. 2020 [107]	In-house	21	Mean = 37.9 years	MS	SyMRI Data (T1w, T2w MVF, MTsat) DWI	eyeballs, temporal muscle	sI(Ganzetti)
Sanada et al. 2022 [108]	(d)	163		Lower-grade Glioma	T1w, T2w, T1-relaxometry, T2-relaxometry	eyeballs, temporal muscle	sI(Ganzetti)
Shams et al. 2019 [109]	In-house	17	Mean = 24.7 years	HC	T1w, T2w, T1-relaxometry		T1w/T2w-ratio
Shim et al. 2022 [110]	In-house	10	Mean = 32.7 years	HC	T1w, T2*w, T1-relaxometry, T2-relaxometry		T1w/T2*w-ratio
Soun et al. 2017 [96]	In-house	10	Mean = 38.7 weeks gestational age	HC	T1w, T2w, FLAIR, DWI	eyeballs, masseter muscle	sI(Soun)
Uddin et al. 2018 [111]	In-house	10	Mean = 57.1 years	MS	T1w, T2w, GRASE (MWF)	eyeballs, temporal muscle	sI(Ganzetti)
Uddin et al. 2019 [112]	In-house	31	Mean = 29.6 years	HC	T1w, GRASE (T2w, MWF), DWI	eyeballs, temporal muscle	sI(Ganzetti)
Vandewouw et al. 2019 [113]	In-house	56	Mean = 4.3 years	4-year-old children (born very preterm, born full term)	T1w, T2w, MTR		T1w/T2w-ratio
Yamamoto et al. 2022 [114]	In-house	34	Mean = 63.5 years	Glioblastoma	T1w, T2w, T1-relaxometry, T2-relaxometry, $[11C]$Met-PET	eyeballs, temporal muscle	sI(Ganzetti)
Yasuno et al. 2017 [115]	In-house	38	Mean = 70.4 years	HC	T1w, T2w, PET imaging	eyeballs, temporal muscle	sI(Ganzetti)
Zheng et al. 2022 [116]	In-house	9	Mean = 64.7 years	MS (post-mortem)	T1w, T2w, MTR (histology)		T1w/T2w-ratio

Abbreviations: AD: Alzheimer's disease, CSF: cerebrospinal fluid, DWI: diffusion-weighted imaging, FA: fractional anisotropy, FLAIR: fluid-attenuated inversion recovery, geomT2IEW: geometric-mean of the intra-/extracellular water T2, HC: healthy control, MD: mean diffusivity, Met: methionine, MS: multiple sclerosis, MTR: magnetization transfer ratio, MTsat: magnetization transfer saturation, MWF: myelin water fraction, PET: positron emission tomography, sI: scaled intensity, SyMRI: synthetic magnetic resonance imaging, T1w: T1-weighted, T2w: T2-weighted
(a) IXI database of the Imperial College London
(http://biomedic.doc.ic.ac.uk/brain-development/index.php?n=Main.Dataset)
(b) KIRBY21 database of the Kirby Research Center for Functional Brain Imaging in Baltimore
(http://mri.kennedykrieger.org/databases.html)
(c) Alzheimer's disease Neuroimaging Initiative database
(adni.loni.usc.edu)
(d) In-house & Cancer Imaging Archive/Cancer Genome Atlas low-grade glioma collection dataset
(https://www.cancerimagingarchive.net/)
$sI(Brown) = \frac{I}{median(orbital fat)}$
$sI(Ganzetti) = \frac{X_R - Y_R}{X_S - Y_S} * I + \frac{X_S Y_R - X_R Y_S}{X_S - Y_S}$ with X: median(temporal muscle), Y: median(eyeball), R: template, S: subject
$sI(Gilmore) = \frac{I}{mean(subcutaneous fat)}$
$sI(Soun) = \frac{X_R - Y_R}{X_S - Y_S} * I + \frac{X_S Y_R - X_R Y_S}{X_S - Y_S}$ with X: mean(masseter muscle), Y: mean(eyeball), R: template, S: subject

Intensity scaling: method validation

An in-house cohort was used to develop and validate intensity scaling methods. "In the total cohort, 1078 image data sets of 701 subjects with confirmed radiologically isolated syndrome (RIS), clinically isolated syndrome (CIS), or MS diagnosis according to the McDonald criteria (Thompson et al., 2018) from our in-house database were included. Further inclusion criteria were MRI acquisition on the same scanner and 18 years or older. The total cohort was used to identify valid reference regions." [3].

"A subset of 597 image data sets of 597 subjects (validation cohort) was used for the technical and biological validation of intensity scaling methods. It only included people with CIS or MS whose EDSS scores were acquired within 31 days before or after the MRI scan. When multiple image sessions were available for one subject, we selected the most recent data" [3]. This selection was motivated by the generally low EDSS scores in the cohort; using the latest scans raised the chance of including higher EDSS scores, and, hence, a broader range of disease severity, likely increasing the sensitivity to disease-related effects. "In the context of biological validation, the validation cohort was divided into two groups: low ($\leq$3) and high ($>$3) EDSS [117]" [3]. Characteristics of the total and validation cohorts, as well as the two EDSS subgroups, are presented in Table 2.5.

All images, including multiparameter mapping (MPM), were acquired on a Philips 3.0 T Ingenia scanner (Philips Medical Systems, Best, The Netherlands). The standard imaging protocol included a 3D T1w magnetization-prepared rapid gradient echo (MPRAGE) sequence and a 3D FLAIR sequence. MPM data was acquired with a 3D radio frequency (RF)-spoiled multi-echo gradient-echo sequence, using a Compressed SENSE acceleration factor of 6 [118]. The MPM acquisition comprised three main protocols: 1) T1w, 2) proton density-weighted (PDw), and 3) MT-weighted (MTw). Images were acquired at six echo times (TE: 2.4ms, 4.8ms, 7.2ms, 9.6ms, 12.0ms, and 14.4ms). In addition, B1 mapping for voxel-wise bias field correction was performed and required two measurements with different repetition times (TR: 30ms and 150ms). Table 2.6 contains detailed information on the acquisition parameters of each MRI sequence.

2.1.3 Clinical correlation analysis

After validation of the intensity scaling methods, the most promising method was implemented to investigate the role of intensity features extracted from conventional MRI images in explaining disease severity in terms of EDSS and SDMT. To this end, a clinical cohort was defined as a subset of the intensity scaling validation cohort, only including cases with available SDMT scores acquired at most 31 days before or 31 days after the scan. Note that the availability of EDSS scores was already a criterion for the validation cohort. Moreover, only cases with no new lesions compared to their previous scan were included in the clinical cohort to ensure that the intensities are only extracted from stable lesions. New lesions often reflect acute inflammation and, to some degree, involve temporary tissue change that can (at least partially) be transient. Therefore, they can obstruct a reliable assessment of the extent of permanent tissue damage, the accumulation of which can lead to clinical decline. The resulting clinical cohort consisted of 292 images from 292 people with CIS or MS. The cohort characteristics are presented in Table 2.5. Since the clinical cohort is a subset of the validation cohort, image acquisition parameters (Table 2.6) and image processing steps (section 2.2.2), including intensity scaling, were identical to the validation cohort.

Table 2.5
Characteristics of the datasets used for intensity scaling validation and clinical correlation analysis.

Dataset	#subjects (#scans)	age (years) mean ± sd	female/male	diagnosis (number of images)	number of lesions i) mean ± sd ii) median (IQR)	total lesion volume (mm^3) i) mean ± sd ii) median (IQR)	EDSS median (IQR)	SDMT mean ± sd
Total								
All images	701 (1078)	39.90 ± 11.03	682/396	RRMS (960) SPMS (22) PPMS (40) CIS (39) RIS (17)	i) 42.60 ± 43.39 ii) 30.0 (16.0–54.0)	i) 6173.42 ± 9672.11 ii) 2543.7 (859.8–7098.8)	1.5 (0.0-2.0)	N/A
Validation								
All images	597 (597)	40.48 ± 11.12	376/221	RRMS (541) SPMS (14) PPMS (20) CIS (22)	i) 43.75 ± 41.68 ii) 30.0 (16.0-55.0)	i) 6130.28 ± 8946.79 ii) 2667.1 (866.1-7121.7)	1.5 (0.0-2.0)	N/A
Low EDSS	526 (526)	39.47 ± 10.61	330/196	RRMS (499) SPMS (0) PPMS (5) CIS (22)	i) 40.24 ± 39.74 ii) 29.0 (16.0-49.0)	i) 5175.88 ± 7646.61 ii) 2309.3 (781.2-5995.1)	1.0 (0.0-2.0)	N/A
High EDSS	71 (71)	47.94 ± 12.04	46/25	RRMS (42) SPMS (14) PPMS (15) CIS (0)	i) 69.76 ± 46.66 ii) 64.0 (31.0-103.0)	i) 13200.90 ± 13617.76 ii) 7752.8 (3499.9-18147.6)	4.5 (3.5-6.0)	N/A
Clinical								
All images	292 (292)	41.17 ± 10.47	194/98	RRMS (272) SPMS (4) PPMS (3) CIS (13)	i) 40.27 ± 41.46 ii) 30.0 (15.0–51.0)	i) 5345.52 ± 7818.0 ii) 2006.0 (728.5–6674.5)	1.0 (0.0-2.0)	62.43 ± 12.81

Abbreviations: CIS: clinically isolated syndrome, EDSS: Expanded Disability Status Scale, IQR: interquartile range, N/A: not applicable/available, PPMS: primary progressive multiple sclerosis, RIS: radiologically isolated syndrome, RRMS: relapsing-remitting multiple sclerosis, sd: standard deviation, SDMT: Symbol Digit Modalities Test, SPMS: secondary progressive multiple sclerosis

Table 2.6
MRI acquisition settings of T1w, FLAIR, and MPM images (including T1w, PDw, MTw, and B1 map) that were used for intensity scaling and clinical correlation analyses. The table was reproduced from Wiltgen et al. [3].

	3D T1w (MPRAGE)	3D FLAIR	Multiparameter mapping			
			T1w	PDw	MTw	B1 map
Reconstruction voxel size	0.75x0.75x0.75 mm^3	0.75x0.75x0.75 mm^3	1x1x1 mm^3	1x1x1 mm^3	1x1x1 mm^3	1x1x2.5 mm^3
Acquisition voxel size	1x1x1 mm^3	1x1x1 mm^3	1x1x1 mm^3	1x1x1 mm^3	1x1x1 mm^3	3.5x3.5x5.0 mm^3
TR	9 ms	4800 ms	18 ms	18 ms	48 ms	30/150 ms
TE	4 ms	shortest	6 echoes first: 2.4 ms echo spacing: 2.4 ms	6 echoes first: 2.4 ms echo spacing: 2.4 ms	6 echoes first: 2.4 ms echo spacing: 2.4 ms	shortest
TI	1000 ms	1650 ms	/	/	/	/
Flip angle	8°	90°	25°	4°	6°	60°
Acquisition mode	turbo field echo	turbo spin echo	fast field echo	fast field echo	fast field echo	fast field echo

Abbreviations: FLAIR: fluid-attenuated inversion recovery, MPRAGE: magnetization-prepared rapid gradient echo, MTw: magnetization transfer-weighted, PDw: proton density-weighted, T1w: T1-weighted, TE: echo time, TI: inversion time, TR: repetition time

2.2 Image processing

This section provides a description of all image processing steps. First, it presents a detailed overview of the processing pipeline integrated in LST-AI, along with the processing steps involved in model training and evaluation. Second, it outlines the processing steps associated with the validation and application of intensity scaling methods.

2.2.1 LST-AI: pipeline, ensemble network, and benchmark

LST-AI is a holistic deep learning-based tool for automated lesion segmentation comprising sequential application of numerous processing steps to ensure seamless application and robust segmentation output. Furthermore, the structure of the network and the training strategy were designed to consistently achieve good performance of the tool on unseen data from other centers and scanners. The preprocessing pipeline, as well as the network structure and training strategy, are described in the following paragraphs.

Preprocessing pipeline

The implementation of LST-AI comprises numerous preprocessing steps that are applied before the images are used as input in the ensemble network, which generates the lesion masks. As input, LST-AI always requires a pair of 3D T1w and 3D FLAIR images. "Considering the controversy surrounding the role of bias field correction in CNN-based architectures [119, 120], it was not included in the preprocessing pipeline of LST-AI." [1]. The following preprocessing steps are implemented in LST-AI (see Figure 2.2):

1. The T1w and FLAIR images are both registered to a template with the Greedy command line tool (https://github.com/pyushkevich/greedy) [121, 122] using the following commands:
   `greedy -d 3 -a -dof 6 -m NMI -ia-image-centers -n 100x50x10 -i <template> <original image> -o <affine> -threads <number of threads>`
   was used to calculate the transformation and
   `greedy -d 3 -rf <template> -rm <original image> <registered image> -r <affine> -threads <number of threads>`
   was used to apply the transformation to the image. In the first command line, `-d 3` defines the dimensionality (3D), `-a` specifies that affine registration is used, `-dof 6` sets the number of degrees of freedom to 6, `-m NMI` sets the metric used for registration to normalized mutual information, `-ia-image-centers` initializes the affine matrix based on matching image centers, `-n 100x50x10` specifies the number of iterations per resolution level in a coarse-to-fine approach (i.e., 100 iterations at low resolution, 50 at intermediate resolution, and 10 at high resolution), `-i <template> <original image>` defines the input composed of the paths of the template and the image in original space that is registered to the template, `-o <affine>` specifies the output path of the affine transformation, and `-threads <number of threads>` defines the number of threads that are used for parallel processing. In the command line that applies the transformation, `-rf <template>` defines the path of the reference image to which the moving image is registered, `-rm <original image> <registered image>` defines the paths of the moving image and the resulting registered image, and `-r <affine>` reads the path of the affine transformation that was calculated in the preceeding step and is applied to register the image to the template. Specifically, the template used in this process was the Montreal Neurological Institute (MNI) International Consortium for Brain Mapping (ICBM)152 nonlinear atlas version 2009 template (https://www.mcgill.ca/bic/neuroinformatics/brain-atlases-human), which was created by the MNI using data from the ICBM project. "This atlas registration both ensures a consistent voxel resolution (1x1x1 mm^3) and image orientation, preprocessing steps well established for deep learning segmentation models [123, 124]" [1].

2. MNI-registered T1w images are skull-stripped with the deep learning-based HD-BET brain extraction tool version 1.0 (https://github.com/MIC-DKFZ/HD-BET) [125]. This step is commonly applied in neuroimaging studies to guarantee standardized input data, avoid noise from structures outside of the brain, and increase the quality of subsequent analysis [125]. HD-BET was implemented in LST-AI using the following command line:
   `hd-bet -i <image> -device <device> -mode accurate -tta 1 -o <skull-stripped image>`

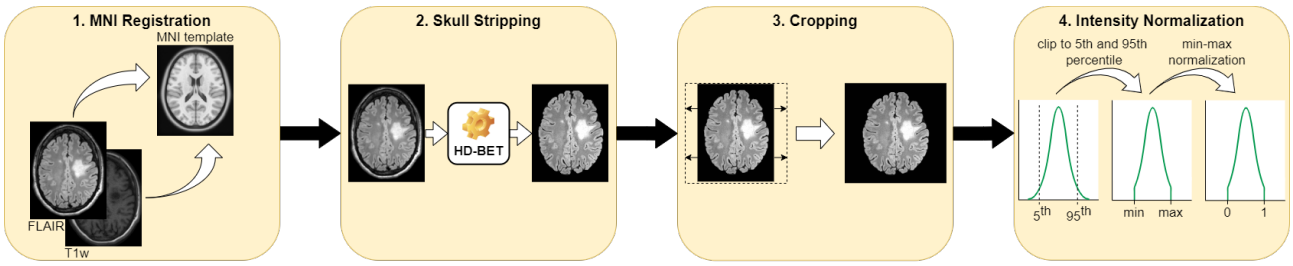
In the command line, `-i <image>` reads the input image (in this case, the MNI-registered T1w image), `-device <device>` indicates the device to use for computation (either a GPU device number or 'cpu'), `-mode accurate` specifies that the accurate operation mode should be used instead of the fast mode, `-tta 1` indicates that test time data augmentation is used (0 should be selected if 'cpu' is defined as device), and `-o <skull-stripped image>` defines the path of the output image which is skull-stripped. In addition to the skull-stripped T1w image, this process also generates a binary brain mask. This mask is then applied to the MNI-registered FLAIR image, resulting in a skull-stripped FLAIR image.

3. The shapes of the 3D MNI-registered and skull-stripped images are modified so that they satisfy the requirements of the 3D U-Nets constituting the ensemble network of LST-AI. Specifically, the 3D U-Nets necessitate input images with dimensions of (192,192,192), requiring T1w and FLAIR images to be cropped accordingly.

4. Intensities of the T1w and FLAIR images are normalized to [0,1] to guarantee that intensity scales of input images are always at a similar scale, making the predictions more robust. This is realized by clipping intensities to the 5^{th} and the 95^{th} percentile, followed by min-max normalization.

After passing through the preprocessing pipeline, the pairs of MNI-registered, skull-stripped, cropped, and intensity-normalized T1w and FLAIR images are used as input for the ensemble network.

Figure 2.2
Preprocessing workflow implemented in LST-AI. Images are first registered to the MNI ICBM152 nonlinear atlas version 2009 template using Greedy, then skull-stripped using HD-BET, followed by cropping to manipulate the shape of the images according to input requirements of the 3D U-Nets, and, finally, intensities are normalized to [0,1].
Abbreviations: FLAIR: fluid-attenuated inversion recovery, MNI: Montreal Neurological Institute, T1w: T1-weighted

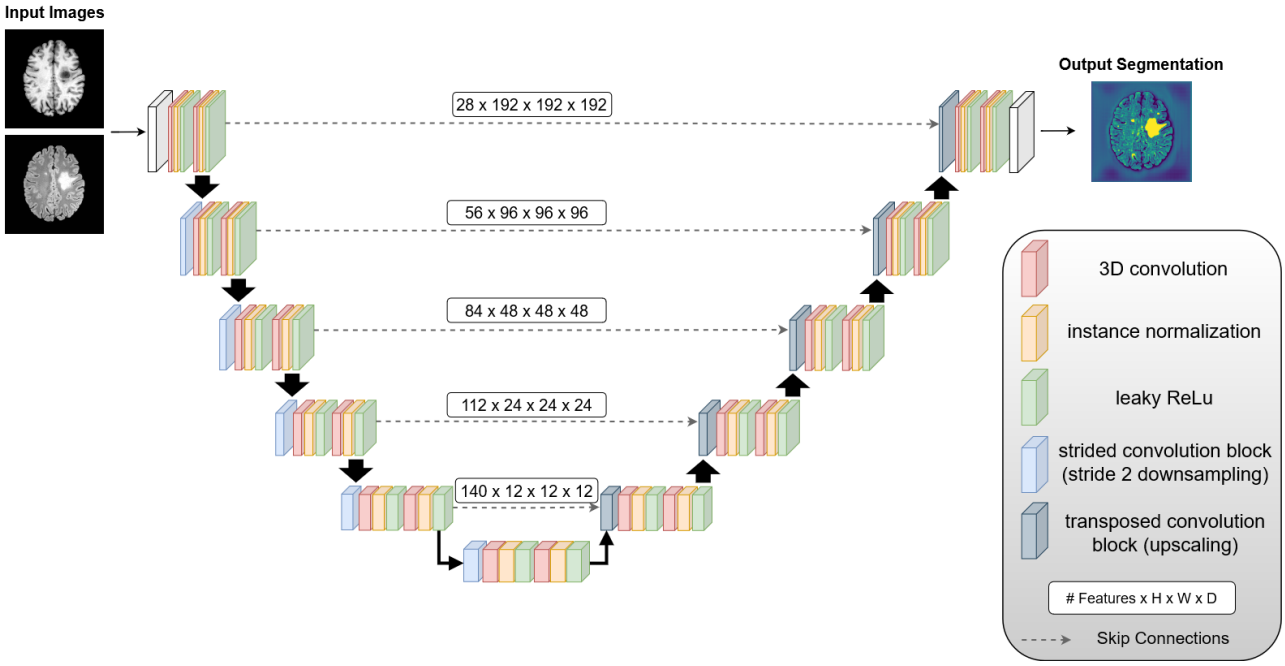


Ensemble network for lesion segmentation

"With respect to the model architecture, LST-AI is based on an ensemble of three 3D U-Nets. Each U-Net is built upon the 3D U-Net [126] architecture and inspired by nnUNet [30]" [1]. Using an ensemble of networks knowingly decreases performance fluctuations compared to single 3D U-Net applications and, therefore, improves generalization across various imaging protocols and centers [27]. Each 3D U-Net in the ensemble "is composed of 5 encoder and 5 decoder blocks. Each of these blocks is built from two convolution blocks (3D convolution, instance normalization, leaky ReLU [rectified linear unit] activation) and skip connections between respective encoder and decoder blocks (see Figure 2.3). In encoder blocks, downsampling is implemented via strided convolutions with stride 2, while transposed convolutions are used for upscaling in decoder blocks. Following the architectural choices in nnUNet [30], we employ deep supervision layers in the training with the intuition of allowing gradients to flow deeper into the networks'

layers [127]. The number of deep supervision layers differed for the three U-Nets: one U-Net included one deep supervision layer and the two other U-Nets included two deep supervision layers to allow for some variability in the ensemble predictions." [1]. The architecture of the 3D U-Nets is illustrated in Figure 2.3.

Figure 2.3
"Architecture of the 3D U-Nets which constitute the ensemble network of LST-AI. They comprise two channels (one for T1w images and one for FLAIR images) and consist of 5 encoder and 5 decoder blocks. Strided convolutions (stride 2) are used for downsampling and transposed convolutions are used for upscaling. Encoder and decoder blocks are connected via skip connections" [1]. The figure was reproduced from Wiltgen et al. [1].
Abbreviations: D: depth, H: height, ReLU: rectified linear unit, W: width



The in-house training dataset (n = 491) was used to train each of the three 3D U-Nets. The training dataset comprised T1w images, FLAIR images, and manually segmented lesion masks that served as ground truth. For training, different combinations of loss functions were used in the deep supervision layers and in the full-resolution output, tailored to the practicality at these different levels. With this in mind, "we used a combination of Tversky loss [128] (with higher penalization of false-negative lesion omissions) and binary cross-entropy in the deep supervision layers, and a combination of Dice loss and binary cross-entropy in the full-resolution output. During training, we randomly chained intensity (random Gaussian noise, random Gaussian smoothing, random gamma adjustment) and geometry augmentations (random flips and crops). Each model was trained for a total of 1000 epochs, using the stochastic gradient descent optimizer (with Nesterov momentum) and a polynomial learning rate decay, starting at 1e-2. This training scheme has been adapted from nnUNet and was shown to generalize well in the medical segmentation decathlon [129]. In total, three training runs were started from scratch to create an ensemble of three models" [1].

The trained models constitute the ensemble network included in LST-AI that is used for inference to generate binary lesion masks. During the application of LST-AI on new data, the pre-processed T1w and FLAIR images are used as input for each of the three trained 3D U-Nets, resulting in three lesion probability maps. Then, the three lesion probability maps are averaged, and a threshold (default is 0.5)

is applied to generate the final binary lesion mask. "As an additional feature, the tool can optionally label lesions according to their location, i.e., periventricular (PV), juxtacortical (JC), subcortical (SC), or infratentorial (IT). To this end, the same MNI ICBM152 nonlinear T1 atlas used above is first registered deformably (using Greedy) to the skull-stripped T1w image in MNI space. The resulting transformation is applied to a manually labeled anatomical mask indicating different brain regions (inter alia: ventricles for PV labeling, infratentorial region for IT labeling, cortex for JC labeling, and subcortical region for SC labeling), which is thereby registered to the skull-stripped T1w image in MNI space. [...] Next, each individual lesion from the binary lesion segmentation map is dilated using a cube as footprint (3x3x3 mm^3), and assigned to the region with which it overlaps by at least one voxel (e.g., if a dilated lesion overlaps with the ventricles of the anatomical mask it is labeled as PV). During this step, lesions are checked to overlap with the four brain regions sequentially so that each lesion can be attributed to only one category. The order of checks is PV, IT, JC, and, finally, SC. By this choice, large lesions overlapping with the inner ventricles and the cortical ribbon are classified as PV (as PV lesions are commonly the largest). In the resulting lesion map, the lesions are labeled according to their location (PV: label = 1, JC: label = 2, SC: label = 3, IT: label = 4)" [1]. The following command is used for the deformable registration of the MNI ICBM152 nonlinear T1 atlas to the skull-stripped T1w image in MNI space:
`greedy -d 3 -m WNCC 2x2x2 -sv -n 100x50x10 <image MNI> <atlas> -o <deformable transformation> -threads <number of threads>`
and the following command is used to apply the transformation to the labeled anatomical mask:
`greedy -d 3 -rf <image MNI> -ri LABEL 0.2vox -rm <anatomical mask> <registered anatomical mask> -r <deformable transformation> -threads <number of threads>`.
In the first command line, `-m WNCC 2x2x2` indicates that the weighted normalized cross-correlation metric with a 2x2x2 patch radius should be used as the image dissimilarity metric, and `-sv` enables the log-Demons deformable registration algorithm. In the second command line, `-ri LABEL 0.2vox` specifies the interpolation mode (here, a specific interpolation mode for labels was used). The remaining inputs have been previously described in step 1 of the preprocessing pipeline. Finally, the binary lesion mask and the labeled lesion mask are also warped from the MNI space back to the original FLAIR space by applying the inverse of the transformation that was calculated during step 1 in the preprocessing pipeline:
`greedy -d 3 -rf <image> -ri LABEL 0.2vox -rm <anatomical mask MNI> <anatomical mask> -r <affine>,-1 -threads <number of threads>`. The addition of `,-1` indicates that the inverse of the affine should be applied.

Ground truth manual segmentation

Training and evaluating automated segmentation tools always require a ground truth to which the output of the segmentation tool can be compared. This usually comes in the form of manually segmented masks. For the training and evaluation of LST-AI, multiple datasets were used that contain manually segmented lesion masks, which were generated with different workflows. In the in-house training dataset, the ground truth was generated by first pre-segmenting the images with the lesion growth algorithm of the lesion segmentation toolbox (LST-LGA) [26], followed by manual correction of experienced neuroradiologists [130]. The 'msisbi' dataset comprised ground truth lesion masks that were manually segmented by two raters without a consensus, which is why we arbitrarily selected the lesion masks of rater 2 [92]. The 'msljub' ground truth segmentation was generated through a consensus of lesion masks from 3 raters that conducted a semi-automated lesion segmentation [93]. The ground truth lesion masks in the 'mssegtest' and 'mssegtrain' datasets are a consensus of manually segmented lesion masks from seven raters [94].

Benchmark lesion segmentation tools

"Evaluation of the performance of the proposed tool is realized through comparison to other publicly available lesion segmentation methods. This includes the widely used LST version 3.0.0

(https://www.applied-statistics.de/lst.html) with its lesion growth algorithm (LGA) [26] and lesion prediction algorithm (LPA) [36], to which our proposed tool presents a complementary, AI-based lesion segmentation method. Additionally, a trained nnUNet, and the recently published SAMSEG [Sequence Adaptive Multimodal SEGmentation] lesion segmentation tool implemented in Freesurfer version 7.3.2 [37]" [1]. The preprocessing steps that are required before applying these tools can differ from the steps implemented in LST-AI. Therefore, the implementations are briefly described for each tool individually:

- **LST-LGA** [26]: Original T1w and FLAIR images were first registered into MNI-space (as described in step 1 of the 'Preprocessing Pipeline' paragraph). Next, a denoising filter was applied to the images using CAT12 (version 12.8.1) [131] implemented in Statistical Parametric Mapping version 12 (SPM12) (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). The 'Denoising Filter (SANLM)' module included in the volume tools of CAT12 was used with default settings for this purpose. Finally, LST-LGA was applied using an initial threshold (kappa value) of 0.3, and a threshold of 0.5 was applied to the lesion probability maps to generate binary lesion maps.

- **LST-LPA** [36]: Original FLAIR images were first registered into MNI-space (as described in step 1 of the 'Preprocessing Pipeline' paragraph). Next, a denoising filter was applied to the images using CAT12 (version 12.8.1) [131] implemented in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). The 'Denoising Filter (SANLM)' module included in the volume tools of CAT12 was used with default settings for this purpose. Finally, LST-LPA was applied, and a threshold of 0.5 was applied to the lesion probability map to generate binary lesion maps.

- **nnUNet** [30]: Original T1w and FLAIR images were registered into MNI-space and skull-stripped (as described in steps 1 and 2 of the 'Preprocessing Pipeline' paragraph). No further processing steps were required since nnUNet was introduced as "an innovative framework that automates the selection of hyperparameters and data augmentation techniques based on the specific dataset employed" [1]. The training data was formatted according to nnUNet's convention, and training of the model was conducted using 1000 epochs and a five-fold cross-validation approach.

- **SAMSEG** [37]: Original T1w and FLAIR images were first registered into MNI-space (as described in step 1 of the 'Preprocessing Pipeline' paragraph). Following the developers' recommendations, no further preprocessing steps were applied. SAMSEG was implemented using the lesion segmentation extension, resulting in the segmentation of multiple regions across the whole brain, including lesions. The binary lesion mask is obtained by only selecting voxels with the lesion-specific label.

2.2.2 Intensity scaling: qMRI, preprocessing, and scaling

Multiparameter mapping

MPM included the acquisition of images in three specific multi-echo protocols (T1w, PDw, and MTw) and one B1 mapping protocol described in 2.1.2. Using the hMRI [132] toolbox (version: v0.1.3-dev; RRID: SCR_017682; https://github.com/tleutritz-cbs/hMRI-toolbox) included in the SPM12 framework (https://www.fil.ion.ucl.ac.uk/spm/software/spm12), these images were combined to calculate quantitative parameter maps, such as the R1 longitudinal relaxation rate maps. This process included multiple steps that are implemented in the hMRI toolbox. Using the 'Auto-Reorient' application in the hMRI toolbox, all images were first re-oriented to the MNI space such that the origin is set at the anterior commissure [132]. Next, the 'Create hMRI maps' module was applied. Based on approximations of the Ernst signal equation in combination with the variable flip angle approach, it uses the multi-echo T1w, PDw, and MTw data to calculate the quantitative R2*, R1, PD, and MT saturation (MT-

sat) maps [132–134]. The module also conducts receiver RF sensitivity bias correction using the unified segmentation approach [135] and transmit field bias correction using the actual flip angle imaging (AFI) method [136]; the latter is based on the two B1 mapping images [132]. The resulting R1 maps can contain residual bias because of imperfect spoiling, which was accounted for by an additional processing step: imperfect spoiling correction (ISC) [134] with an RF spoiling phase increment of 150°.

Preprocessing

The development and application of different intensity scaling methods comprised multiple preprocessing steps. Accurate WM lesion segmentation was conducted with T1w and FLAIR images using LST-AI [1], which was introduced in 2.2.1. "Sequence Adaptive Multimodal SEGmentation (SAMSEG) software integrated into Freesurfer version 7.3.2 [37, 137, 138] was used to segment WM, cortical gray matter (GM), cerebellum GM and WM, accumbens, amygdala, caudate, hippocampus, putamen, and thalamus. The inputs were pairs of registered T1w and FLAIR images, co-registered using 'mri_coreg' and 'mri_vol2vol' integrated into Freesurfer" [3].

In addition, since the focus was on intensity scaling methods avoiding cerebral reference regions, candidate reference regions outside of the brain "were segmented using SAMSEG combined with an atlas comprising extracerebral tissues, similar to the approach by Puonti et al. [139].[...] The extracerebral tissues included cancellous bone, cortical bone, eyes, optic nerves, rectus muscles, and skin. In addition, temporal fatty tissue and temporal muscle were segmented as they are distinctly visible in conventional T1w MRI, and previous studies have also used fatty tissue and temporal muscle for scaling [77, 83, 84, 86, 90]" [3]. Therefore, both structures were manually segmented in 166 T1w images using 3D Slicer version 5.6.2 (https://www.slicer.org/) [140] to train 3D nnUNet models [30]. "Regarding the temporal fatty tissue, the entire connected hyperintense structure of the temporal fatty tissue was segmented to include enough voxels that enable the reliable calculation of mean intensities. Regarding the temporal muscle, the connected structure was segmented starting at a lower limit defined by the slice on which the temporal muscle approximately extends from the eyes to the ears and ending at an upper limit defined by the slice on which the frontal horn of the lateral ventricles appears. Despite inter-subject variations of these upper and lower limits, the temporal muscle region is large and includes a great number of voxels, independent of these variations, and enables the reliable calculation of mean intensities" [3, Supplementary materials]. The manually segmented dataset was used to train two separate nnUNet 3D full-resolution models for 1000 epochs with 3-fold cross-validation. Trained models were used to apply inference on T1w images of the entire cohort, and, if necessary, the generated masks were manually corrected with 3D Slicer.

Technical validation of scaling methods included the comparison of T1w images with R1 maps. Therefore, T1w images were rigidly registered to R1 maps (which were already registered to the MNI space by the hMRI toolbox) using the Greedy command line tool. The T1w-to-MNI transformation matrix was saved for the registration of other imaging files into MNI space. The binary lesion mask generated by LST-AI is defined in the FLAIR image space and was registered to the T1w image using the 'mri_vol2vol' functionality in Freesurfer in combination with the FLAIR-to-T1w transformation matrix that was generated in the previously mentioned SAMSEG implementation. Subsequently, the binary lesion mask was brought into MNI space by applying the T1w-to-MNI transformation matrix with the Greedy command line tool. Similarly, the T1w-to-MNI transformation matrix was used to warp intensity-scaled T1w images as well as brain and reference region segmentation masks to the MNI space.

Further processing steps included N4 bias field correction [141] of T1w images using the Advanced Normalization Tools (ANTs) software (https://github.com/ANTsX/ANTs), skull-stripping of T1w images using the HD-BET brain extraction tool, and creating a binary NAWM mask. For the latter, the binary lesion mask generated with LST-AI was subtracted from the WM mask generated with SAMSEG

to remove any lesion voxels from the WM mask. Finally, the mask was eroded using a 3x3x3 cube as a structuring element to create a conservative segmentation of NAWM.

Scaling methods and signal modeling

Multiple intensity scaling approaches were implemented, validated, and compared to each other. Their purpose is to transform conventional T1w image intensities into quantifiable units, enhancing inter-image comparability beyond what raw T1w images offer. Generally, the inter-image variance can be influenced by technical and biological factors. Intensity scaling aims to remove the former while preserving the latter. The methods applied in this project included MRI signal modeling, linear scaling with a scaling factor derived from reference regions, and z-score standardization.

- **MRI signal modeling**
  In theory, the signal that is produced by an MPRAGE sequence can be described by a signal equation that is dependent on the sequence parameter settings and relates the physical properties of the underlying tissue, e.g., T1 relaxation times, to the measured signal. As shown in the sequence diagram in Figure 2.4, the MPRAGE sequence consists of multiple read-out blocks that start with a 180° inversion pulse, followed by an inversion time TI, a rapid gradient echo (RAGE) block with RF pulses with a flip angle α, a repetition time TR between RF pulses, and a recovery period T_{rec} until the next MPRAGE read-out block starts. The signal of the i^{th} read-out block for a voxel with longitudinal and effective transverse relaxation times T1 and T2*, respectively, is given by [142–146]:

$$S_i = \sin\alpha * M_i^N * e^{-\frac{TE}{T2*}} \tag{2.1}$$

  where M_i^N is the longitudinal magnetization after the N^{th} RF pulse in the RAGE block:

$$M_i^N = M_0(1-e^{-\frac{TR}{T1}})\frac{1-(\cos(\alpha)e^{-\frac{TR}{T1}})^N}{1-\cos(\alpha)e^{-\frac{TR}{T1}}} + \left(M_0(1-e^{-\frac{TI}{T1}}) - M^{eq}e^{-\frac{TI}{T1}}\right)(\cos(\alpha)e^{-\frac{TR}{T1}})^N \tag{2.2}$$

  and M^{eq} is the steady-state equilibrium magnetization:

$$M^{eq} = M_0\frac{1-e^{-\frac{T_{rec}}{T1}} + (1-e^{-\frac{TR}{T1}})\frac{1-(\cos(\alpha)e^{-\frac{TR}{T1}})^N}{1-\cos(\alpha)e^{-\frac{TR}{T1}}}e^{-\frac{T_{rec}}{T1}} + (1-e^{-\frac{TI}{T1}})(\cos(\alpha)e^{-\frac{TR}{T1}})^N e^{-\frac{T_{rec}}{T1}}}{1+(\cos(\alpha)e^{-\frac{TR}{T1}})^N e^{-\frac{TI}{T1}}e^{-\frac{T_{rec}}{T1}}} \tag{2.3}$$

  From the MPRAGE sequence settings, the echo time TE, α, TR, T_{rec}, TI, and the N^{th} phase encoding step sampling the center of k-space are known. "M_0 represents the acquisition-specific equilibrium magnetization, which depends on the spin density (ρ) and on technical properties (k) such as receiver amplifier gain and coil design. This can be expressed as: $M_0 = k * \rho$. For simplification, we relied on the MPRAGE sequence's characteristics, which leverage strong T1 contrast and minimize the impact of ρ differences across tissues, allowing for the approximation of $\rho \approx constant$ without compromising contrast accuracy [145]. Consequently, we consider M_0 as uniform across the entire image. Based on equations [2.1,2.2,2.3], the theoretically expected T1w signal could be calculated if the equilibrium magnetization M_0, relaxation times (T1, T2*), and imaging parameters (TE, α, TR, T_{rec}, and TI) were known. We neglected the influence of T2* relaxation and flip angle inhomogeneity because TE is sufficiently short (TE = 4ms) and we applied

bias field correction of T1w images" [3].

Since the goal was to estimate T1 values from the acquired conventional T1w image intensities, M_0 had to be approximated so that a direct relation between T1 values and T1w image intensities could be derived. "Since equation 2.1 is a formally non-invertible function, M_0 was numerically estimated using pairs of T1 and T1w values. First, for each reference region, a fixed T1 value was determined by extracting the median T1 value from that region in each individual R1 map (T1 = 1/R1) and then computing the median of these values across all data sets (n = 1078). Second, median T1w intensities were extracted from reference regions for each bias-corrected T1w image" [3]. In this manner, pairs of fixed T1 values and image-specific T1w intensities were formed for each reference region. Assuming that these data pairs follow the signal equation, each image's M_0 was estimated by fitting the signal equation to these data pairs using orthogonal distance regression. For each image, a 1-to-1 mapping was then calculated between T1 values and T1w image intensities based on the estimated M_0 constant. T1 values between 200ms (lower than fatty tissue) and 4500ms (approximating CSF) [147] were sampled to calculate corresponding T1w intensities. "The 1-to-1 mapping was then used to generate estimated R1 maps, where each voxel was assigned the inverse T1 value (R1 = 1/T1) based on its actual T1w intensity" [3].

- **Linear scaling**
  Linear scaling aims to remove the technically induced inter-image variance, thereby converting image intensities to the same intensity scale across images, making them comparable and absolute intensity values more interpretable. In the specific application of linear scaling, the assumption is that the technically induced variance can be modeled by a linear scaling factor [77, 145]. Ideally, this scaling factor can be derived from reference regions that exhibit only technical and no or very little biological variance. Regions outside the brain are not directly affected by MS pathology, potentially allowing for accurate modeling of the scaling factor to correct for the technical inter-image variance. Implementation was realized by dividing the image intensities by a scaling factor, in this case, the median intensity of a reference region:

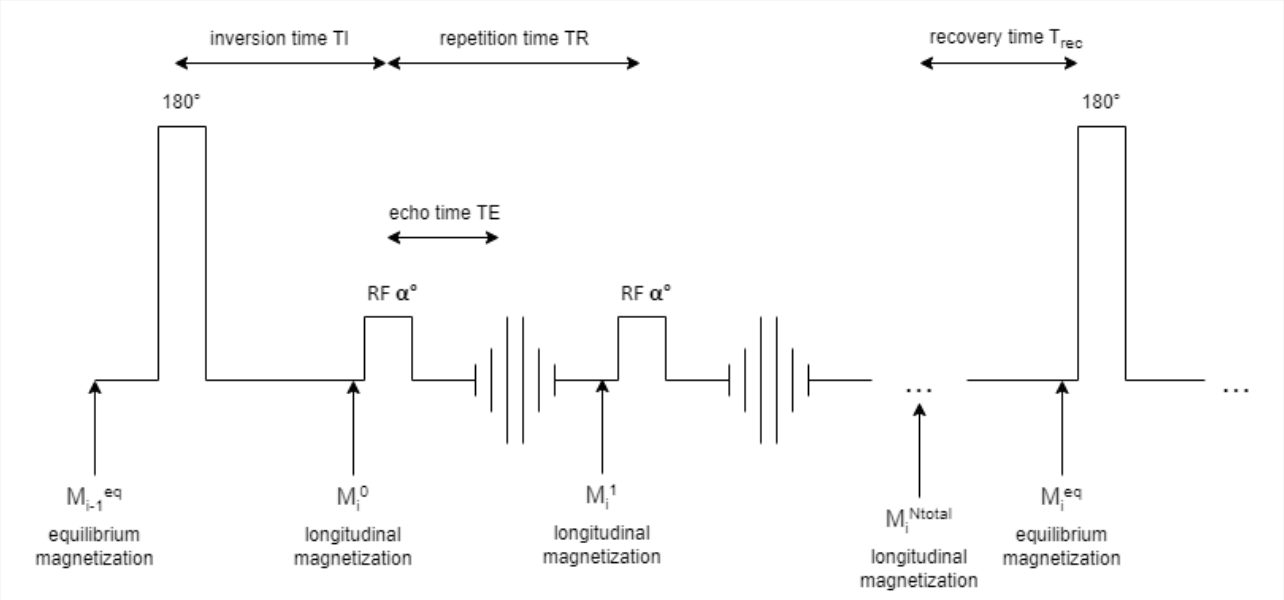
$$T1w_{scaled} = \frac{T1w}{scaling\ factor} \tag{2.4}$$

- **Z-score standardization**
  Z-score standardization is a frequently used method that aims to harmonize datasets to enable their direct comparison. "Z-score standardization is popular in medical image processing and eliminates intensity scale discrepancies by centering the data around zero and transforming it to units of standard deviation" [3]:

$$T1w_{z-score} = \frac{T1w - mean}{standard\ deviation} \tag{2.5}$$

  "We calculated the mean intensity and standard deviation within the binary brain masks generated with HD-BET. The brain masks cover cerebral and cerebellar WM and GM, and CSF" [3].

Figure 2.4
MPRAGE sequence diagram. "The MPRAGE sequence starts with the image acquisition once the steady-state equilibrium magnetization M^{eq} is reached. One block begins with a 180° inversion pulse, followed by an inversion time TI. Next, the RAGE block, consisting of N_{total} phase encoding steps, starts. Each phase encoding step includes the application of an RF pulse with flip angle α, and a gradient readout phase after the echo time TE has passed. Two consecutive RF pulses are separated by the repetition time TR. After the RAGE block, the steady-state equilibrium magnetization is reached again after the recovery time T_{rec}. Then, the next MPRAGE block starts with applying the next inversion pulse" [3, Supplementary materials]. The figure was reproduced from Wiltgen et al. [3, Supplementary materials].
Abbreviations: MPRAGE: magnetization-prepared rapid gradient echo, RAGE: rapid gradient echo, RF: radio frequency, TE: echo time, TI: inversion time, TR: repetition time, T_{rec}: recovery time



2.3 Statistical analysis

This section presents all statistical analysis methods applied in the projects. It starts with the evaluation analysis of LST-AI's segmentation performance. Next, strategies used to summarize data extracted during the systematic review are described. Finally, the different analysis methods implemented in the development and validation of intensity scaling, including clinical correlation analysis, are provided.

2.3.1 LST-AI: evaluation

The lesion segmentation performance of LST-AI was analyzed on the external datasets and compared to results obtained with the benchmark methods. The lesion segmentation accuracy and the lesion detection were investigated to evaluate LST-AI.

The lesion segmentation accuracy was assessed using the animaSegPerfAnalyzer tool integrated in the Anima toolbox (https://anima.irisa.fr/). It takes the ground truth lesion mask and the automatically generated lesion mask as input, and calculates evaluation metrics that quantify the agreement of the latter with the ground truth. Each metric is calculated per image and is then averaged across images to obtain overall scores. The metrics provided by animaSegPerfAnalyzer are mainly based

on true positives (TP) (lesion voxels truly predicted as lesion voxels), false positives (FP) (non-lesion voxels predicted as lesion voxels), and false negatives (FN) (lesion voxels not predicted as lesion voxels), and include the following parameters:

- **Dice similarity coefficient (DSC)**: The DSC quantifies the spatial overlap of the automatically generated lesion mask with the ground truth and ranges from 0 (no overlap) to 1 (perfect overlap):

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{2.6}$$

- **Positive predictive value (PPV)**: The PPV quantifies how many of the positive predictions are TP and ranges from 0 (no predicted lesion voxel is a true lesion voxel) to 1 (all predicted lesion voxels are true lesion voxels):

$$PPV = \frac{TP}{TP + FP} \tag{2.7}$$

- **Sensitivity**: The sensitivity quantifies how many lesion voxels were predicted as such and ranges from 0 (no lesion voxel was predicted as a lesion voxel) to 1 (all lesion voxels were predicted as lesion voxels):

$$sensitivity = \frac{TP}{TP + FN} \tag{2.8}$$

- **Average surface distance (ASD)**: The ASD quantifies the accuracy of the segmentation in that it represents the average distance between the surfaces of the automatically segmented lesions and the ground truth:

$$ASD = \frac{1}{n + n'} \left[\sum_{x=1}^{n} d(x, S') + \sum_{x'=1}^{n'} d(x', S) \right] \tag{2.9}$$

  with: $d(x, S') = min||x - x'||_2$, and "where n and n' are the number of points x and x' on the surface S of the manual segmentation and the surface S' of the automated segmentation, respectively, and d() is the minimal Euclidean distance between a point x on surface S and the surface S'." [1].

"As an additional step, we construct one array by concatenating all images and calculate the DSC across all lesions of all datasets. [...] Thereby, we avoid the per-subject lesion load bias that is introduced when one score is calculated per image. For example, missing a small lesion in an image with only this missed lesion (DSC = 0) would have more weight than missing a similar lesion in an image with many other detected lesions (DSC > 0). We further investigate whether the performance of lesion segmentation varies across brain regions to identify the drivers of the metric values and possible location-dependent variabilities of LST-AI segmentation performance" [1]. Therefore, location-annotated lesion masks were concatenated, and the DSC was calculated for each of the four annotated regions (PV, JC, subcortical (SC), and IT).

The animaSegPerfAnalyzer tool also includes metrics that are calculated on the lesion-level to quantify the lesion detection performance, "this aspect is crucial in MS, since its diagnosis relies on the detection of lesions (and not on the exact measurement of their volume). To this end, we extract the following scores from the animaSegPerfAnalyzer tool: SensL, the lesion detection sensitivity; PPVL, the positive predictive value for lesions; F1 score, a metric which considers both lesion detection sensitivity and positive predictive value for lesions. SensL and PPVL are calculated according to equations 2.8 and 2.7, respectively (on the lesion level rather than on the voxel level)" [1]. The F1 score is also known as the

lesion-wise DSC, and is calculated according to 2.6 (again on the lesion level). More detailed information is provided by the animaDetectedComponents tool from the Anima toolbox. For each individual lesion in the ground truth segmentation, it indicates whether the lesion was detected in the automatically generated lesion mask and provides its volume. This information was used to analyze the lesion detection sensitivity in relation to lesion volume. To this end, the lesions were divided into five volume groups (3-10 mm^3, 11-100 mm^3, 101-1000 mm^3, 1001-10000 mm^3, 10001- mm^3) and the ratio of detected lesions was calculated for each volume group. The detection ratio of LST-AI was then compared to that of the benchmark methods. The animaSegPerfAnalyzer tool and the animaDetectedComponents tool "consider a lesion in the manual segmentation as detected if it overlaps with at least 10% with the lesion voxels in the automatically generated lesion map" [1].

2.3.2 Intensity scaling: systematic review, technical and biological validation

Descriptive analysis in the systematic review

In the systematic review, results were extracted from each study and summarized so that they could be compared between studies. As the reduction of inter-image variance is of high importance in intensity scaling studies, the results for intra- and inter-image variance analysis were reported for each study (if available) using the statistical metrics provided therein. These metrics included coefficient of variation (CoV), standard deviation, ANCOVA to detect differences between datasets, intra- and inter-scan Kullback-Leibler Divergence Distance, and visual histogram inspection. Similarities between (intensity-scaled) conventional MRI images and (semi-)quantitative MRI images were also investigated in the systematic review. Results were reported as Spearman ρ, partial correlation, or R^2 derived from linear and exponential regression. In addition, the significance of relations to demographical, clinical, or histological data was reported, including correlation analyses and group comparisons. "Due to the very heterogeneous nature of scaling and reference methods applied in the included studies, comparing results and combining findings was only possible in descriptive ways. [...] Consequently, it was impossible to aggregate results through a meta-analysis" [2].

Analysis of reference regions

Thorough analysis of candidate reference regions is important to make sure they are appropriate for intensity scaling in the context of the specific use case. "A data-driven approach was used to identify suitable reference regions. As intensity scaling methods aim to convert image intensities into comparable, quantitative units, especially within the brain, the reference regions should not exhibit a higher inter-image variance than brain regions such as the cerebral cortex or WM. Regions with higher variance are expected to contain confounding biological variance, hampering the estimation of technical variance" [3]. To this end, the median intensities of candidate reference regions were extracted from each bias-corrected T1w image (n = 1078), and CoVs of the median intensities were calculated to quantify the inter-image variance of each region. Reference regions were deemed suitable for linear intensity scaling if their CoV was comparable to the CoVs of WM and cortical GM.

Technical validation

Technical validation of intensity scaling methods investigated: 1) the potential of each method to reduce inter-image variance, and 2) the correlation between intensity-scaled T1w images and quantitative R1 maps. While the reduction of variance can be associated with good harmonization of image intensities, the correlation with R1 maps assesses the quantitative nature of image intensities.

Regarding inter-image variance reduction, median intensities were extracted from multiple brain regions (NAWM, cortex, accumbens, amygdala, caudate, cerebellum cortex, cerebellum WM, hippocampus, putamen, thalamus) in each image. Subsequently, the CoVs of median intensities were calculated across images for each brain region. This was realized for each scaling method, and CoVs were compared

between scaling methods and with R1 maps and unprocessed T1w images.

For each scaling method, the correlation between intensity-scaled T1w images and R1 maps was studied using two approaches: "1) calculating the median intensity of each brain ROI [region of interest] per image and computing Pearson's r across all ROIs and images, and 2) computing voxel-wise Pearson's r between R1 maps and intensity-scaled T1w images per data set, then averaging across Pearson's r values" [3].

Biological validation: group comparison

"The difference of each scaling factor between low (n = 526) and high (n = 71) EDSS groups was assessed to ensure their independence from disease severity. Then, for each intensity scaling method, the mean intensity and the CoV were extracted from NAWM and compared between low and high EDSS groups. The two-sided Welch's t-test was used for group comparisons. Cohen's d was calculated to measure the effect size of the group differences (low vs. high EDSS). For the NAWM comparison, we also applied ANCOVA, including age, sex, disease duration, and total lesion volume as covariates" [3].

2.3.3 Regression models for clinical correlation analysis

Multiple linear regression was applied to investigate the contribution of T1w image features in explaining disease severity according to EDSS and SDMT. These features included total lesion volume, mean NAWM intensity, mean lesion intensity, CoV of intensities in the NAWM, and CoV of intensities in lesions. In addition, age, disease duration, and sex were included. Non-normally distributed features were log-transformed to enhance normality. Since sex is a non-numeric binary variable, it was converted into a numerical dummy variable by assigning zeros to males and ones to females. All features, along with EDSS and SDMT scores, were z-score normalized, which is necessary to obtain interpretable standardized regression coefficients (β) in multiple linear regression. Scatter plots were generated to verify linear relationships between features and EDSS/SDMT. Simple correlation between clinical and intensity features was calculated using Pearson's r correlation coefficient. The ultimate goal was to analyze whether intensity measures (i.e., mean and CoV) extracted from lesions and from NAWM can complement features that are commonly used (i.e., age, disease duration, sex, total lesion volume) to explain MS disease severity. Therefore, two baseline models were defined that comprised EDSS or SDMT as the dependent variable and common features as independent variables:

$$dependent\ variable = \underbrace{\beta_0 + \beta_1 * age + \beta_2 * disease\ duration + \beta_3 * sex + \beta_4 * lesion\ volume}_{baseline\ model} \quad (2.10)$$

Next, separate models were calculated for each intensity-based feature by adding the feature as an independent variable to the baseline models:

$$dependent\ variable = baseline\ model + \beta_5 * intensity\ feature \quad (2.11)$$

The intensity-based features were evaluated based on the significance, the standardized regression coefficient (β), and the change in explained variance of the model (R^2) compared to that of the baseline models. A feature was considered to provide complementary information if its inclusion resulted in a significant effect ($p < 0.05$) and an increase in R^2 compared to the baseline models. The R^2 values and the corresponding β values were also compared between intensity features.

3 Results

This section starts with the evaluation of the LST-AI lesion segmentation tool that was developed and applied in this thesis project. Next, the findings of the systematic review are presented, which motivated the development and validation of intensity scaling methods, the results of which are presented thereafter. Finally, the section concludes with the clinical correlation analysis, exploring the potential of features extracted from conventional MRI to explain disease severity.

3.1 LST-AI

LST-AI was evaluated using metrics that quantify the segmentation accuracy on the voxel level and the detection accuracy on the lesion level. Both approaches are important for assessing LST-AI's robustness in measuring lesion volume (voxel level) and lesion number (lesion level).

3.1.1 Segmentation accuracy

Across all datasets, LST-AI yielded a higher average DSC (0.67±0.14) than all benchmark methods (DSC = 0.42 - 0.55). It also outperformed the benchmark methods in terms of sensitivity (LST-AI: 0.66±0.17, benchmark methods: 0.32 - 0.49) and ASD (LST-AI: 0.37±1.12, benchmark methods: 1.35 - 1.46). Hence, more lesion voxels are correctly identified by LST-AI, and lesions are contoured more accurately than with the benchmark methods. In contrast, the PPV was similar or higher for benchmark methods compared to LST-AI. The lesion segmentation evaluation results of LST-AI and the benchmark methods are shown in Table 3.1. Table 3.2 shows the results for each dataset individually. Across all datasets, the LST-AI segmentation performance is consistently excellent as indicated by the evaluation metrics (high DSC: 0.61 - 0.74, high PPV: 0.68 - 0.80, high sensitivity: 0.54 - 0.70, and low ASD: 0.12 - 0.59).

Table 3.1
Lesion segmentation evaluation metrics of LST-AI and benchmark methods. The table includes metrics that quantify segmentation accuracy on the voxel level (DSC, PPV, sensitivity, and ASD) and on the lesion level (F1, SensL, and PPVL). LST-AI outperforms the other methods in all metrics except for PPV and PPVL. "The metrics were calculated for each image in the test datasets, and values were subsequently averaged across all images. The averages are reported as mean ± standard deviation" [1]. The table was reproduced from Wiltgen et al. [1].

Tool	voxel-level				lesion-level		
	DSC	PPV	sensitivity	ASD	F1	SensL	PPVL
LST-AI	0.67±0.14	0.73±0.15	0.66±0.17	0.37±1.12	0.63±0.15	0.70±0.19	0.64±0.20
LST-LGA	0.42±0.22	0.80±0.21	0.32±0.20	1.43±2.81	0.22±0.15	0.20±0.14	0.41±0.26
LST-LPA	0.44±0.22	0.79±0.15	0.34±0.20	1.35±2.21	0.23±0.14	0.25±0.15	0.34±0.23
nnUNet	0.51±0.20	0.90±0.07	0.38±0.18	1.36±4.18	0.46±0.19	0.40±0.21	0.64±0.21
SAMSEG	0.55±0.20	0.72±0.21	0.49±0.19	1.46±4.57	0.38±0.18	0.32±0.18	0.57±0.21

Abbreviations: ASD: average surface distance, DSC: Dice similarity coefficient, PPV: positive predictive value, PPVL: positive predictive value for lesions, SensL: sensitivity for lesions

Table 3.2
Lesion segmentation evaluation metrics of LST-AI for each test dataset. The evaluation metrics show some variability across datasets but remain excellent (high DSC, PPV, sensitivity, F1, SensL, and PPVL; low ASD). "The metrics were calculated for each image in the test datasets, and values were subsequently averaged across all images. The averages are reported as mean ± standard deviation" [1]. The table was reproduced from Wiltgen et al. [1].

Dataset	voxel-wise				lesion-wise		
	DSC	PPV	sensitivity	ASD	F1	SensL	PPVL
All datasets (n=103)	0.67±0.14	0.73±0.15	0.66±0.17	0.37±1.12	0.63±0.15	0.70±0.19	0.64±0.20
msisbi (n=21)	0.61±0.13	0.72±0.11	0.54±0.15	0.41±0.66	0.57±0.12	0.55±0.15	0.61±0.13
msljub (n=30)	0.74±0.10	0.80±0.07	0.70±0.14	0.21±0.88	0.70±0.10	0.62±0.13	0.83±0.11
mssegtest (n=37)	0.65±0.16	0.68±0.19	0.68±0.16	0.59±1.60	0.63±0.17	0.83±0.14	0.55±0.22
mssegtrain (n=15)	0.67±0.16	0.72±0.16	0.67±0.19	0.12±0.24	0.61±0.15	0.77±0.23	0.53±0.09

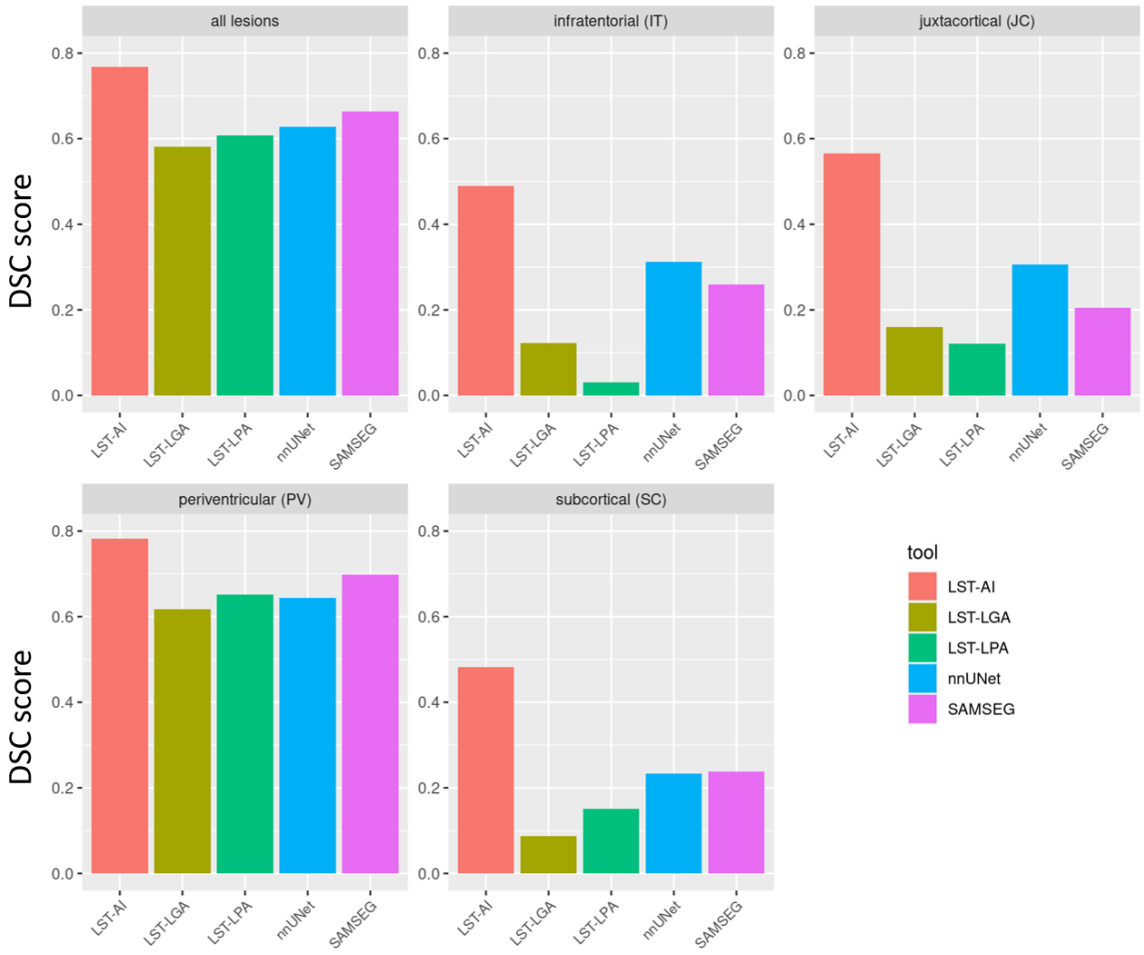
Abbreviations: ASD: average surface distance, DSC: Dice similarity coefficient, PPV: positive predictive value, PPVL: positive predictive value for lesions, SensL: sensitivity for lesions

Similar to the approach of calculating scores per lesion mask and averaging them, the superior lesion segmentation performance was also observed across different brain regions. "In the PV region, LST-AI shows slightly higher [...] DSC scores than the other methods. The difference in terms of [...] DSC scores is more pronounced in the other three regions, with only LST-AI reaching DSC > 0.47 (other methods: DSC = 0.03 – 0.31). Similarly, the highest [...] DSC score within the whole brain is obtained with LST-AI" [1]. In Table 3.3, the DSC scores are shown for each lesion segmentation method and for each brain region (PV, JC, SC, and IT). Figure 3.1 illustrates the comparison of LST-AI with the benchmark methods for all lesions and for lesions in each brain region.

Table 3.3
DSC of concatenated lesion masks for LST-AI and benchmark methods in each brain region. LST-AI outperforms the other methods in all regions. Lesion segmentation is most accurate in the PV region for all methods. The table was reproduced from Wiltgen et al. [1].

tool	Periventricular	Juxtacortical	Subcortical	Infratentorial	Whole brain
LST-AI	0.78	0.57	0.48	0.49	0.77
LST-LGA	0.62	0.16	0.09	0.12	0.58
LST-LPA	0.65	0.12	0.15	0.03	0.61
nnUNet	0.64	0.31	0.23	0.31	0.63
SAMSEG	0.70	0.21	0.24	0.26	0.66

Figure 3.1
DSC of concatenated lesion masks for LST-AI and benchmark methods in the whole brain comprising all lesions and in each brain region individually. LST-AI outperforms the other methods in all regions. Lesion segmentation is most accurate in the PV region for all methods. The figure was reproduced from Wiltgen et al. [1].
Abbreviations: DSC: Dice similarity coefficient



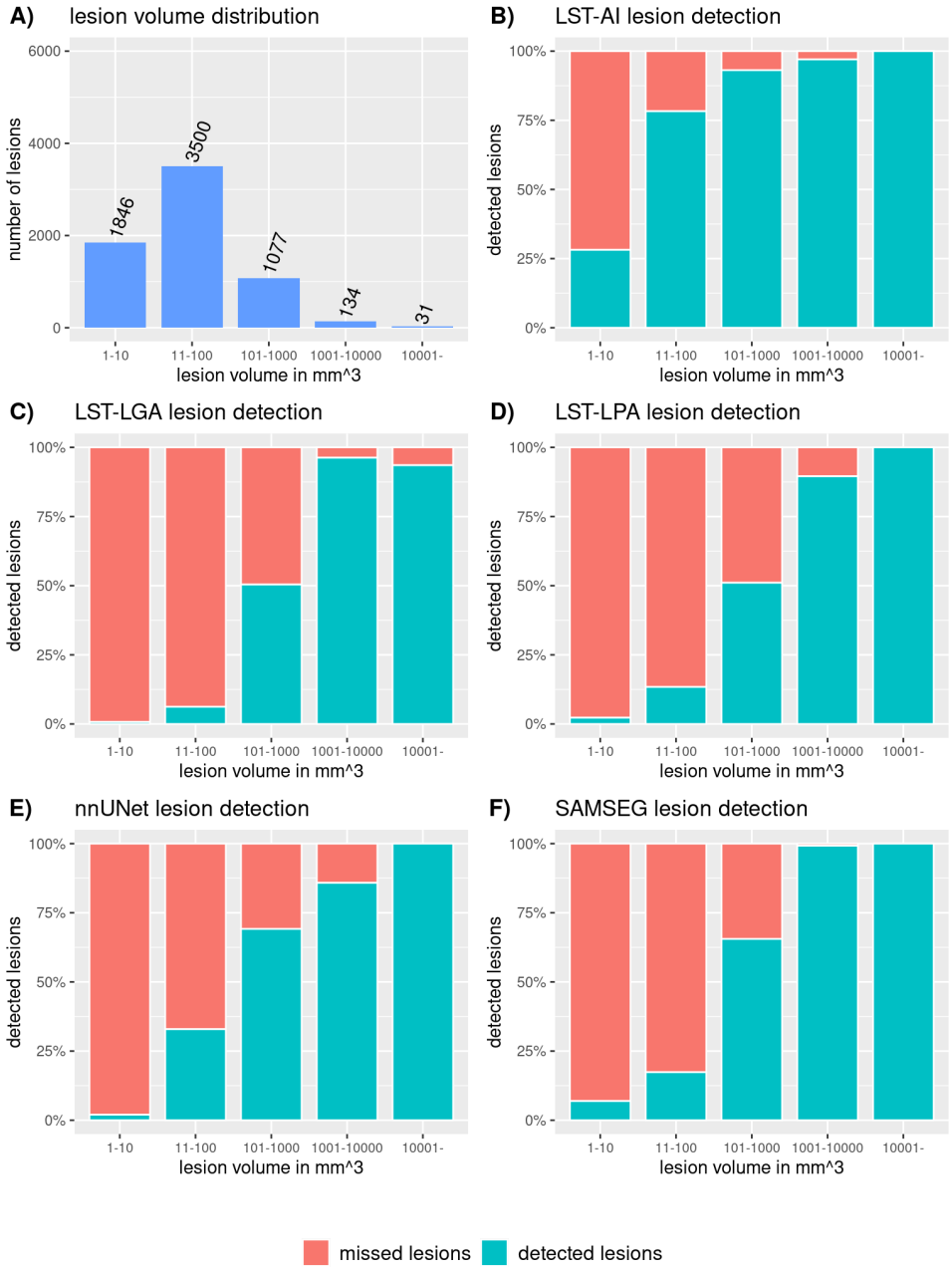
3.1.2 Lesion detection

On the lesion level, again calculating one score per lesion mask and averaging across all lesion masks, LST-AI produces a higher F1 score (0.63±0.15) and higher SensL values (0.70±0.19) than any of the benchmark methods (F1: 0.22 - 0.46, SensL: 0.20 - 0.40). The PPVL of LST-AI (PPVL = 0.64±0.20) was comparable to the nnUNet (PPVL = 0.64±0.21), but higher than all the other benchmark methods (PPVL = 0.34 - 0.57). The largest increase was observed for the sensitivity, indicating that LST-AI detects a larger proportion of true lesions (see Table 3.1). Similar to the voxel level evaluation, the results for each dataset individually show that the lesion detection performance by LST-AI is stable across datasets, confirming its high level of generalization (see Table 3.2).

The efficient lesion detection can also be observed in Figure 3.2, where "we illustrate the accuracy of lesion detection in relation to lesion volume. For this, we divided lesions into groups according to their size: 3–10 mm^3, 11–100 mm^3, 101–1000 mm^3, 1001–10000 mm^3, and larger than 10000 mm^3. Small lesions (>3 mm^3 and <10 mm^3) show a lower detection rate. With increasing lesion size, the detection

rate increases for all methods, with LST-AI showing the steepest incline. Hence, the advantage of LST-AI also applies to small lesions" [1].

Figure 3.2
"These graphs illustrate the proportion of lesions that are detected in each volume group. We divided the lesions into groups according to their volume (on the logarithmic scale): 3–10 mm^3, 11–100 mm^3, 101–1000 mm^3, 1001–10000 mm^3, and larger than 10000 mm^3. A) shows the number of lesions distribution across the volume groups; B)-F) show the lesion detection ratios of LST-AI, LST-LGA, LST-LPA, nnUNet, and SAMSEG for the different lesion volumes. Note, how the detection rate increases with increasing lesion volume for each segmentation, whereby LST-AI yields the highest detection rates. The detection rate is given in %." [1]. The figure was reproduced from Wiltgen et al. [1].



3.2 Intensity scaling

3.2.1 Systematic review

Variance reduction

Four studies analyzed inter-image variance and all reported successful variance reduction of scaled T1w [77], T2w [97], and T1w/T2w-ratio [83, 99] images compared to their non-scaled counterparts. Intensity histograms provided in the study by Brown et al. [77] showed considerably lower inter-image variance after scaling with the median intensity of orbital fat and even more so after scaling with the NAWM intensity (White-Stripe approach [80]). The two studies employing scaled T1w/T2w-ratios could show that scaling with extra-cerebral reference tissues reduced the inter-image CoV in multiple brain regions [99] and eliminated significant intensity differences between datasets [83]. The other studies did not conduct such variance analyses. Although sufficient inter-image variance reduction is a good indicator for intensity scaling performance, it should not be the only quality criterion, as excessive variance reduction may also eliminate important biological variance, which has also been discussed in Brown et al. [77] with regard to the White-Stripe approach.

Relation to quantitative imaging

In total, twenty studies included a comparison of images derived from conventional T1w and T2w images with quantitative imaging. Three of those studies compared scaled T1w image intensities to MTR [77], DWI [84], or myelin water fraction (MWF) [111]. Only MTR showed good agreement with scaled T1w image intensities (R^2 = 0.63) [77]. Limited evidence was available on scaled T2w images, as only two studies found significant and weak correlations with DWI ($p<0.05$) [84] and MWF (R^2 = 0.05 - 0.23) [111]. Regarding T1w/T2w-ratio without prior intensity scaling of T1w and T2w images, intensities consistently correlated with MTR (R^2 = 0.19 - 0.77) [103, 104, 113] and with relaxometry (R^2 = 0.48 - 0.82) [109, 110]. Twelve studies implemented the calculation of T1w/T2w-ratios according to Ganzetti et al. [83], which includes intensity scaling of T1w and T2w images, using eyes and temporal muscle, before calculating the ratio. Positive correlations were observed with MTR (no quantitative results available), and negative correlations were found with T1 and T2 relaxation times (T1: R^2 = 0.002 - 0.64, T2: R^2 = 0.07 - 0.59). Weak and inconsistent correlations were reported for DWI [83, 96, 101, 107, 112] and MWF [98, 111, 112].

Relation to clinical and histological data

On the one hand, studies found significant correlations of scaled T1w images and T1w/T2w-ratios with disease duration in MS [77, 104]. Additionally, correlations with the development assessment of four-year-old children [113] and with the EDSS of subjects with MS [104] were observed for T1w/T2w-ratios. These results were similar to those obtained with MTR. On the other hand, scaled T1w/T2w-ratios based on SyMRI data of subjects with MS did not correlate with EDSS or disease duration [107]. However, scaled T1w/T2w-ratios in WM were significantly lower in people with MS compared to healthy controls [101]. Four studies analyzed the relation of T1w/T2w-ratios with histology [103, 105, 106, 116]. Histology revealed significant differences between healthy controls and people with MS regarding neurite density in the cortex, which was also the case for scaled T1w/T2w-ratios [105]. In the same setting, significant correlations were additionally observed between scaled T1w/T2w-ratios and neurite density [105]. Furthermore, T1w/T2w-ratios (without scaling) differed significantly between demyelinated and myelinated cortical tissue [103], correlated with dendrite density [106], and were sensitive, albeit with low specificity and accuracy, towards demyelination [116]. Overall, conventional T1w and T2w seem to be sensitive to various disease-specific biological properties since multiple significant relations were reported with qMRI, clinical, and histological data, particularly in MS. Hence, thorough testing and validation of intensity scaling methods are required in order to reliably exploit this potential of conventional MRI.

3.2.2 Selection of reference regions

For the purpose of identifying reliable reference regions that can be used for intensity scaling, the CoVs (extracted from unscaled bias-corrected T1w images) of multiple candidate reference regions were compared to the CoVs in WM (CoV = 0.130) and cortical GM (CoV = 0.139). Only temporal fatty tissue (CoV = 0.139) and temporal muscle (CoV = 0.139) exhibited similar CoVs of median intensities, all other regions presented larger inter-image variability (CoV = 0.19 - 0.45). Thus, two variants of linear scaling were implemented: 1) using the median intensity of temporal fatty tissue (hereafter referred to as fatty tissue), and 2) using the median intensity of temporal muscle as scaling factors. "Regarding MPRAGE signal modeling, fixed T1 values for fatty tissue (T1 = 378.62 ms) and temporal muscle (T1 = 1491.44 ms) were calculated using all R1 maps (n = 1078). Additionally, we included the eye's vitreous body (T1 = 4838.19 ms), which was the only accurately segmented region in the low T1w intensity range with reasonably low absolute T1w intensity variability (standard deviation: $\pm$425.89 [a.u.]) compared to GM and WM (standard deviation: $\pm$1893.65 [a.u.] and $\pm$3191.54 [a.u.]), providing a stable reference point for orthogonal distance regression." [3]. Figure 3.3 shows an exemplary segmentation of the three reference regions, and a representative fitted MPRAGE signal modeling curve is shown in Figure 3.4.

Figure 3.3
Exemplary segmentation of reference regions. The yellow/orange region represents the vitreous body of the eye (segmented with SAMSEG), the red region represents the temporal fatty tissue (segmented with a 3D nnUNet), and the blue region represents the temporal muscle (segmented with a 3D nnUNet).

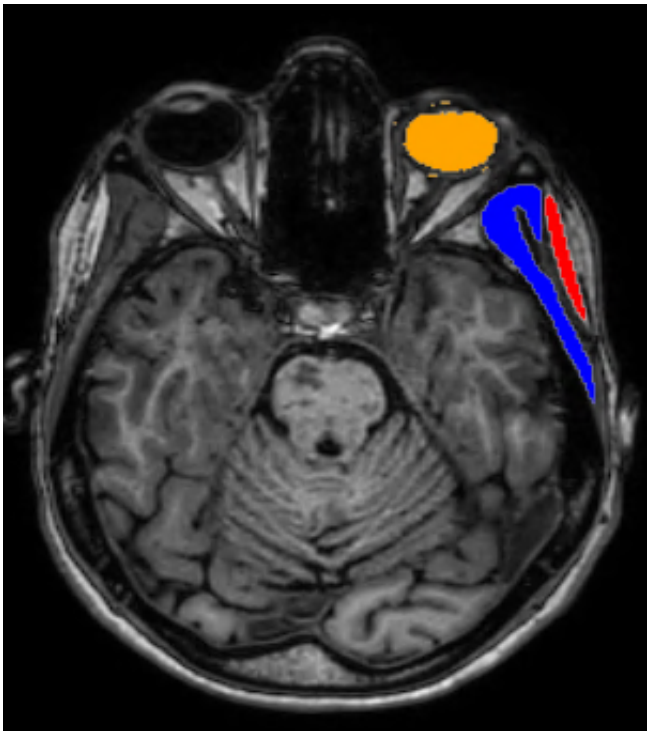
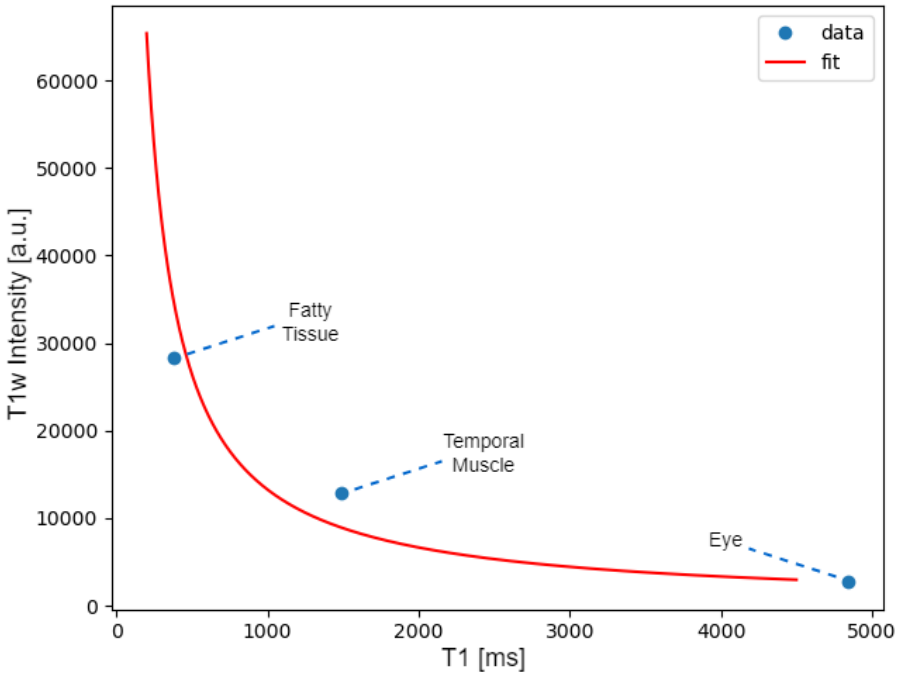


Figure 3.4
MPRAGE signal modeling with reference regions. "The red curve shows how the MPRAGE signal was fitted to data from reference regions (fatty tissue, temporal muscle, and eye). The blue points represent, for each reference region, data points defined by pairs of fixed median intensity extracted from T1 maps and median intensity extracted from a specific bias-corrected T1w image. Notice how each data point provides a reference in a different intensity range, sampling the MRI signal spectrum in the low-, mid-, and high-intensity range" [3]. The fixed T1 values were 378.62 ms for the fatty tissue, 1491.44 ms for the temporal muscle, and 4838.19 ms for the eye. Based on the fitted curve, a 1-to-1 mapping can be established between T1w intensities and T1 values, enabling the estimation of R1 maps from T1w images. The figure was reproduced from Wiltgen et al. [3].
Abbreviations: a.u.: arbitrary units, T1w: T1-weighted

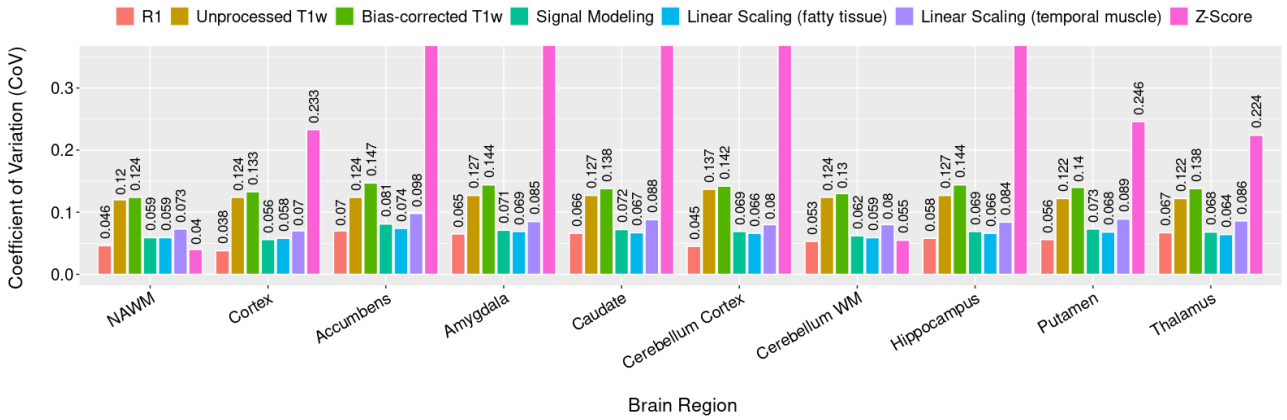


3.2.3 Technical validation

Variance Analysis

In all the brain regions of interest, the lowest CoVs were observed for R1 maps, except for NAWM, where T1w images after z-score standardization yielded a slightly lower CoV. In the other regions of the brain, the CoVs of signal modeling (CoV: 0.056 - 0.081) and linear scaling with fatty tissue (CoV: 0.058 - 0.074) were closest to those of R1 maps (CoV: 0.038 - 0.070). Of all scaling methods, z-score standardization yielded the lowest CoV in cerebellum WM (CoV: 0.055) but, at the same time, produced outliers in the remaining regions (CoV: 0.224 - 3.193). Besides z-score standardization, unprocessed (CoV: 0.120 - 0.137) and bias-corrected T1w images (CoV: 0.124 - 0.147) exhibited the largest CoVs across all regions. Only z-score standardization led to distinct variations in CoVs across brain regions, while all other scaling methods uniformly affect the entire image and are therefore more suitable for analyzing various regions within the brain. The results of all image types are compared and presented for each brain region in Figure 3.5.

Figure 3.5
The CoV of median intensities is illustrated for each brain region, quantifying the inter-image variance. Different scaling methods, as well as unprocessed T1w images and R1 maps, are presented in different colors. "Linear scaling with fatty tissue and signal modeling consistently yields CoVs similar to R1 maps. Linear scaling with temporal muscle results in slightly larger CoVs; Z-score standardization provides the lowest CoVs in WM but the largest CoVs in all other regions (CoVs in the Accumbens, Amygdala, Caudate, Cerebellum Cortex, and Hippocampus are not completely displayed, and range between CoV = 0.6 and CoV = 3.2)" [3]. The figure was reproduced from Wiltgen et al. [3].
Abbreviations: CoV: coefficient of variation, NAWM: normal-appearing white matter, T1w: T1-weighted, WM: white matter



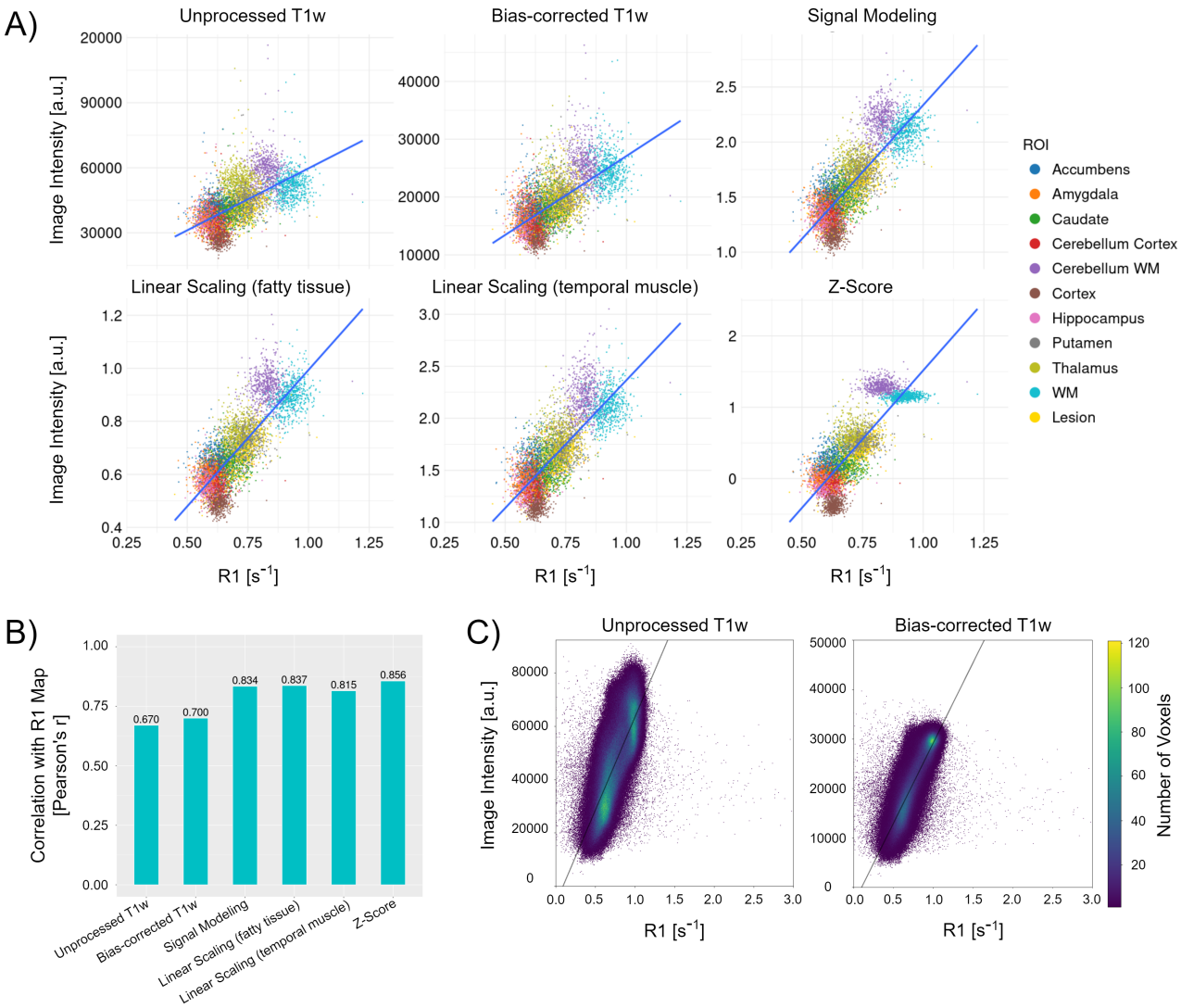
Relation to quantitative imaging

To quantify how accurately the R1 maps can be estimated from conventional T1w images, the correlation between intensity-scaled T1w images and R1 maps was calculated 1) based on median intensities from multiple brain regions, and 2) based on voxel-wise correlations.

Correlations between intensity-scaled T1w images and R1 maps across brain regions were similar for all scaling methods (Pearson's r: 0.815 - 0.856). The highest correlation was observed for z-score standardization (r = 0.856), followed by linear scaling with fatty tissue (r = 0.837), signal modeling (r = 0.834), and linear scaling with temporal muscle (r = 0.815). Unprocessed and bias-corrected T1w images correlated less strongly with R1 maps (r = 0.670 and r = 0.700, respectively). The correlation plots in Figure 3.6A illustrate the relation between median intensities of T1w-derived images and R1 maps across brain regions, and Figure 3.6B shows the corresponding Pearson's r values for each scaling method.

In addition, the average of voxel-wise correlations between intensity-scaled T1w images and the corresponding R1 map was calculated for each scaling method. "The averages of voxel-wise correlation values were similar for all image types (r = 0.84±0.04) except for unprocessed T1w images (r = 0.76±0.05). As expected, linear scaling and z-score standardization preserved the linear relationship of unscaled bias-corrected T1w image intensities with R1 values. This is also approximately true for signal modeling, as the conversion of T1w image intensities into estimated R1 (= 1/T1) values is a near-to-linear transformation being the inversion of the hyperbolic relation between T1 values and T1w signal intensities illustrated in Figure 3.4" [3]. Exemplary voxel-wise correlation plots of an unprocessed T1w image and of a bias-corrected T1w image with the corresponding R1 map are shown in Figure 3.6C. The unprocessed T1w image yields a much larger spread in the correlation plot, illustrating the decreased correlation with the R1 map compared to the bias-corrected T1w image. This observation was consistent across all images.

Figure 3.6
Plots illustrating the correlation between T1w-derived images and quantitative R1 maps. "A) Correlation plots across brain regions. Each dot represents the median R1 value (x-axis) and the median intensity according to the selected image type (y-axis) within a brain region; brain regions are highlighted in different colors. The blue linear regression line shows the relation between the corresponding scaled T1w image intensities and R1 values. B) Pearson's r values associated with the linear regressions from A). The strongest correlation with R1 was observed for z-score standardization (r = 0.86), followed by linear scaling with fatty tissue (r = 0.84) and MRI signal modeling (r = 0.83). C) Example of voxel-wise correlation. The voxel-wise correlation (Pearson's r) with R1 is calculated per image and averaged across all images. The mean correlation with R1 was the same for each image type derived from bias-corrected T1w images (r = 0.84±0.04) and was higher than that of unprocessed T1w images (r = 0.76±0.05). The correlation plot of the unprocessed T1w image shows a much larger spread (y-axis) than that of the bias-corrected T1w image. The correlation plot was virtually the same for all other intensity scaling methods, apart from different scales of the y-axis, since the voxel-wise correlation within an image is invariant under linear transformation (i.e., intensity scaling)" [3]. The figure was reproduced from Wiltgen et al. [3].
Abbreviations: ROI: region of interest, T1w: T1-weighted, WM: white matter



3.2.4 Biological validation

In addition to technical validation, showing reduced inter-image variance and increased similarity with R1 maps after scaling, it is important to confirm that the variance reduction mainly results from eliminating technically induced variance while preserving biologically induced variance, which is essential for measuring disease-related effects. In this context, group comparison of scaling factors revealed that only the standard deviation of whole-brain intensities, which contributes to z-score standardization, was significantly different between the low and high EDSS groups (p = 0.001). Thus, z-score standardization cannot be considered independent of disease severity, as it may remove or even artificially introduce biological variance, and was excluded from further analysis. The scaling factors of all other methods did not exhibit significant differences between the low and high EDSS groups (p = 0.10 - 0.81), and were considered independent of disease-related effects. For biological validation, the mean intensity as well as CoVs of NAWM were compared between the two groups with low and high EDSS (Table 3.4 and Figure 3.7). Moreover, ANCOVA was conducted to determine whether NAWM measures could provide information related to disease progression in addition to the traditional biomarkers age, sex, disease duration, and lesion volume.

Mean intensity

The most significant difference in mean NAWM intensities between the low and high EDSS groups was observed for R1 maps (p = 0.008), with higher values detected in the low EDSS group. The R1 maps also yielded the largest effect size with regard to group differences in mean NAWM intensities (Cohen's d: -0.351). Among all the other image types, only linear scaling with fatty tissue exhibited a significant difference in mean NAWM intensities between the low and high EDSS groups (p = 0.034). Similar to R1 maps, the mean NAWM intensities were higher in the low EDSS group. The difference is also observable in the corresponding effect size (Cohen's d: -0.252). Unprocessed T1w images exhibited a similar effect size (Cohen's d: -0.256), but the difference in mean NAWM intensities was not statistically significant and merely showed a trend (p = 0.07). Differences were insignificant for all other image types (p = 0.183 - 0.322) and effect sizes were low (Cohen's d between -0.157 and 0.13). The differences detected in R1 maps and fat-scaled T1w images were no longer significant when ANCOVA was conducted, controlling for age, sex, disease duration, and lesion volume.

Intensity variability

All image types exhibited significantly higher NAWM CoVs in the high EDSS group compared to the low EDSS group, indicating higher intensity variability in the NAWM in patients with high EDSS. The level of significance (p = 1.64×10^{-7} - 0.003) and effect sizes (Cohen's d: 0.41 - 0.82) were generally higher compared to the mean NAWM group differences (p = 0.008 - 0.32, Cohen's d: -0.35 - 0.15). The most significant results were obtained from bias-corrected T1w images (p = 1.64×10^{-7}), also yielding the largest group difference effect size (Cohen's d: 0.82). Images that were derived from bias-corrected T1w images produced the same results since CoVs are invariant under the applied intensity scaling methods. The difference in NAWM CoVs of bias-corrected T1w images remained significant (p = 0.004) when ANCOVA was applied, using age, sex, disease duration, and lesion volume as covariates. In contrast, differences in NAWM CoVs were less significant for R1 maps (p = 0.003, Cohen's d: 0.41) and unprocessed T1w images (p = 0.003, Cohen's d: 0.43), and they were no longer significant in the ANCOVA (R1: p = 0.06, unprocessed T1w: p = 0.26). Interestingly, the group differences measured with NAWM CoVs of bias-corrected T1w images were larger than any group difference measured with the R1 maps.

Table 3.4
Group comparison of mean and CoV of NAWM intensities between patients with low ($\leq$3) and high ($>$3) EDSS. Only R1 maps and linear scaling with fatty tissue showed significant differences in mean NAWM intensities between the low and high EDSS groups. CoVs were significantly different between the low and high EDSS groups for all image types. Differences in CoVs remained significant in ANCOVA (including age, sex, disease duration, and lesion volume) for bias-corrected T1w images (and images derived from bias-corrected T1w images), but not for R1 maps and unprocessed T1w images. The values are given as the mean ($\pm$standard deviation) of the mean NAWM intensities and of the NAWM CoVs. Bold numbers indicate statistical significance, i.e., p-values < 0.05. The table was reproduced from Wiltgen et al. [3].

	NAWM Mean intensity			NAWM CoV		
	Low EDSS	**High EDSS**	**p-value**	**Low EDSS**	**High EDSS**	**p-value**
R1	0.972 $\pm$ 0.044	0.956 $\pm$ 0.047	**0.0081**	0.070 $\pm$ 0.013	0.076 $\pm$ 0.015	**0.0027**
Unprocessed T1w	53873.9 $\pm$ 6154.9	52025.8 $\pm$ 8156.6	0.0695	0.109 $\pm$ 0.008	0.114 $\pm$ 0.012	**0.0031**
Bias-corrected T1w	25187.8 $\pm$ 3084.7	25610.6 $\pm$ 3391.1	0.3218	0.062 $\pm$ 0.010	0.072 $\pm$ 0.015	**1.636E-07***
Signal modeling	2.155 $\pm$ 0.128	2.136 $\pm$ 0.107	0.1833	0.062 $\pm$ 0.010	0.072 $\pm$ 0.015	**1.636E-07***
Linear scaling (fatty tissue)	0.918 $\pm$ 0.055	0.905 $\pm$ 0.046	**0.0344**	0.062 $\pm$ 0.010	0.072 $\pm$ 0.015	**1.636E-07***
Linear scaling (temporal muscle)	2.174 $\pm$ 0.160	2.197 $\pm$ 0.137	0.1989	0.062 $\pm$ 0.010	0.072 $\pm$ 0.015	**1.636E-07***

Abbreviations: CoV: coefficient of variation, EDSS: Expanded Disability Status Scale, NAWM: normal-appearing white matter, T1w: T1-weighted
*remains significant in ANCOVA with age, sex, disease duration, and lesion volume as covariates

Figure 3.7
Group comparison of mean NAWM (orange) and CoV (blue) of NAWM intensities between patients with low ($\leq$3) and high ($>$3) EDSS. The upper panel shows the effect sizes (Cohen's d) of group differences. The NAWM CoVs exhibit larger effect sizes than mean NAWM intensities in all image types, with NAWM CoVs of bias-corrected T1w images showing the largest Cohen's d value overall (0.818). The middle panel (orange box plots) shows that only R1 maps and linear scaling with fatty tissue yield significant differences in mean NAWM intensities between the low and high EDSS groups. No statistically significant differences were observed in the remaining image types. In contrast, the lower panel (blue box plots) shows that all image types exhibited significant differences in NAWM CoVs, and that bias-corrected T1w images showed the most significant differences.
Abbreviations: CoV: coefficient of variation, EDSS: Expanded Disability Status Scale, NAWM: normal-appearing white matter, T1w: T1-weighted

3.3 Clinical correlation analysis

As linear scaling with temporal fatty tissue showed the most promising results during technical and biological validation, it was combined with LST-AI to extract intensity features from conventional T1w images. These features included the mean and the CoV of intensities from lesions and NAWM and were used in addition to traditional features (age, sex, disease duration, and lesion volume) to explain EDSS and SDMT. All features were z-score normalized (except for the binary feature 'sex'), and their distributions are shown in Figure 3.8. Considering the large sample size (n = 292), slight deviations from normality were acceptable. Lesion volume was the only variable with a severely skewed distribution, which was resolved by applying a log-transformation. The results reported hereafter always refer to the log-transformed lesion volume.

Figure 3.8
Distribution of z-score normalized features. Lesion volume had to be log-transformed before applying the z-score normalization since its distribution was severely skewed.
Abbreviations: CoV: coefficient of variation, NAWM: normal-appearing white matter

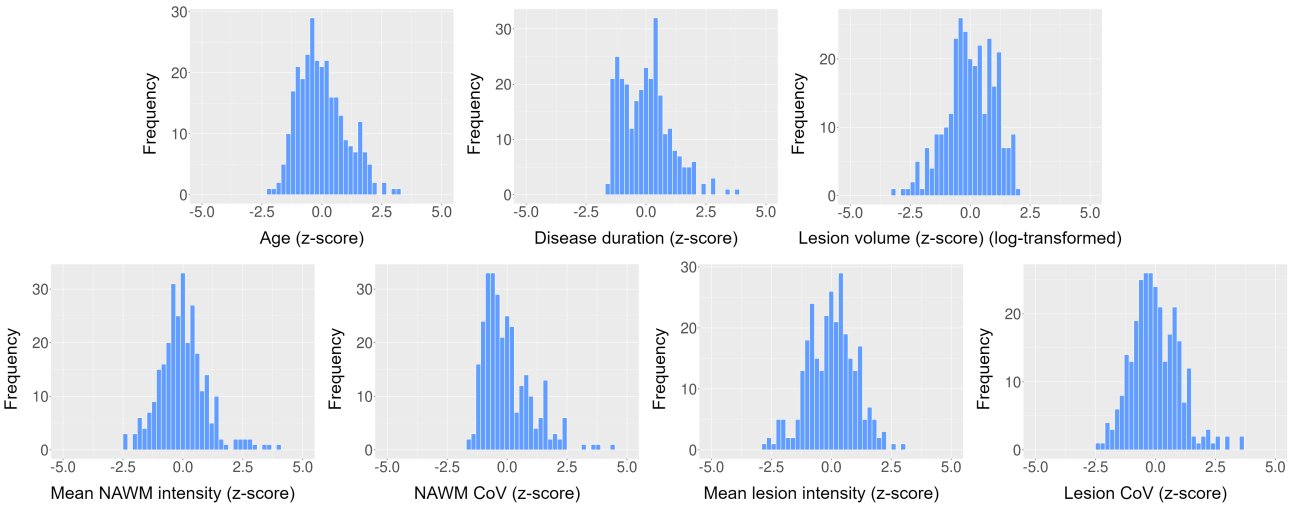
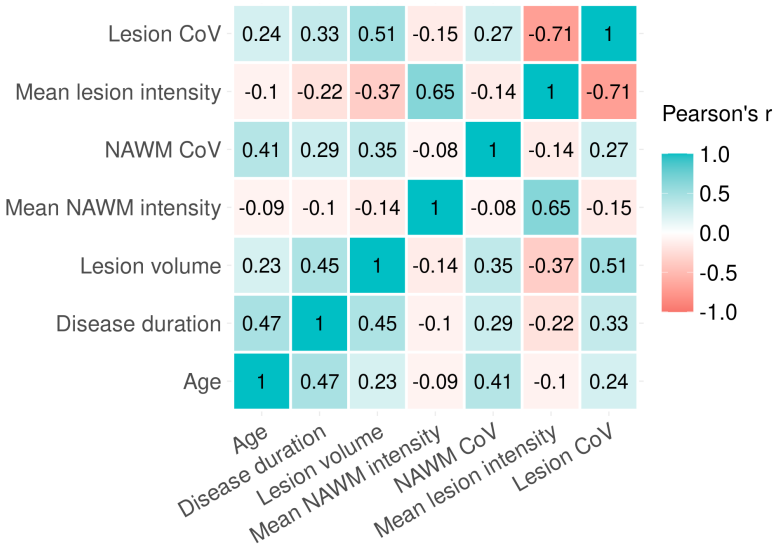


Figure 3.9
Correlation matrix of predictors included in multiple linear regression. Mean lesion intensity strongly correlated with lesion CoV (r = -0.71) and mean NAWM intensity (r = 0.65). Thus, these features should not be included in the same model to avoid collinearity.
Abbreviations: CoV: coefficient of variation, NAWM: normal-appearing white matter



EDSS

The scatter plots (see Figure 3.10) confirmed a linear relation between EDSS and the features, and Pearson's r correlation coefficients revealed a significant correlation between every feature and EDSS scores, except for mean NAWM intensity. The strongest correlation was observed for NAWM CoV (r = 0.367, p = 1.03x10^{-10}) and age (r = 0.365, p = 1.19x10^{-10}), and mean lesion intensity was the only feature with a negative correlation. Hence, increases in age, disease duration, lesion volume, NAWM CoV, and lesion CoV, as well as a decrease in mean lesion intensity, are all associated with greater EDSS scores, indicating more severe disability.

In multiple linear regression, the baseline model included EDSS as the dependent variable and age, disease duration, sex, and lesion volume as independent variables. In all models, the regression coefficients represent the standardized β since features (except for the dummy variable 'sex') and outcome variables were z-score normalized. Detailed results of each model are reported in Table 3.5. The baseline model yielded a coefficient of determination of R^2 = 0.2077. In this model, age (β = 0.26, p = 1.66x10^{-5}) and lesion volume (β = 0.19, p = 1.14x10^{-3}) were the most important features. When intensity features were added to the baseline model, only the CoV of intensities in the NAWM significantly contributed (β = 0.19, p = 0.003), increasing the model's R^2 from 0.2077 to 0.2326. Interestingly, the standardized β of NAWM CoV was greater than the ones associated with sex, disease duration, and lesion volume, indicating greater explanatory power for EDSS scores.

Figure 3.10
The scatter plots show the linear relation between EDSS scores and the features included in the multiple linear regression models. The blue line represents the linear regression line, illustrating the simple correlation between the two respective variables. The corresponding Pearson's r correlation coefficients are also provided in the upper right corner of each plot, along with the p-value indicating the level of significance. All features are significantly correlated with EDSS scores, except for mean NAWM intensity. Abbreviations: CoV: coefficient of variation, EDSS: Expanded Disability Status Scale, NAWM: normal-appearing white matter

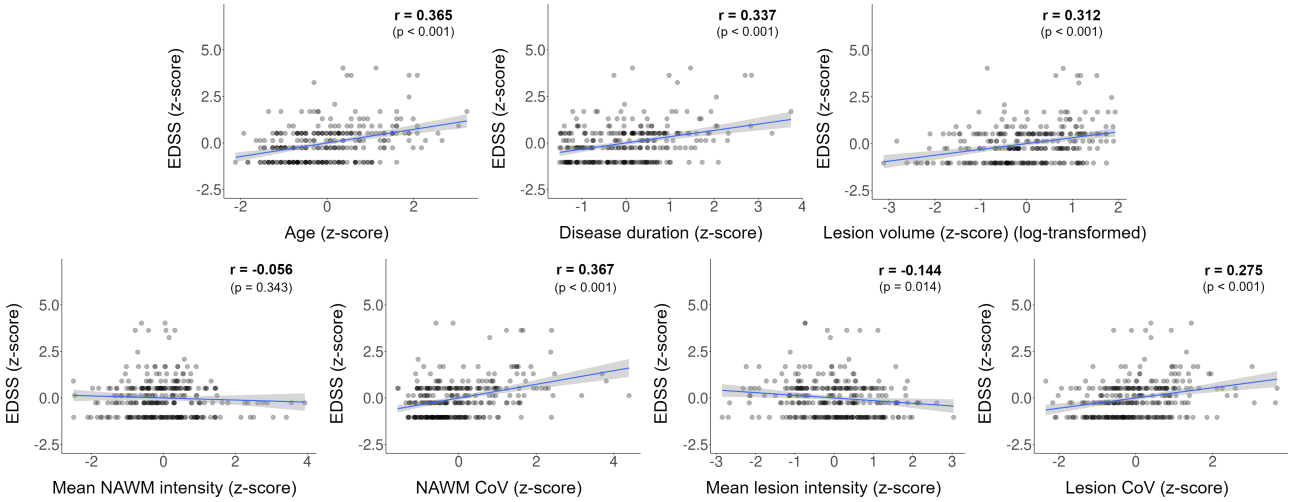


Table 3.5
EDSS multiple linear regression models. Besides the baseline model, including age, disease duration, sex, and lesion volume as independent variables, the table shows the results of multiple linear regression models including intensity features. Only the CoV of NAWM significantly contributed to the multiple linear regression model and considerably improved the baseline model's ability to explain variance in EDSS scores. P-values of features that significantly contributed to the respective model are highlighted in bold, and features that were added to the baseline model are listed below the dashed lines in each case.

Dependent variable	Independent variables	Regression coefficient	p-value	R^2
EDSS	age	0.2605	**1.66E-05**	0.2077
	disease duration	0.1208	0.0640	
	sex	0.0927	0.0796	
	lesion volume	0.1936	**1.14E-03**	
EDSS	age	0.2586	**2.02E-05**	0.2085
	disease duration	0.1197	0.0669	
	sex	0.1041	0.0681	
	lesion volume	0.1896	**1.58E-03**	
	mean(NAWM)	-0.0308	0.5912	
EDSS	age	0.1966	**1.77E-03**	0.2326
	disease duration	0.1211	0.0597	
	sex	0.0546	0.3084	
	lesion volume	0.1441	**1.75E-02**	
	CoV(NAWM)	0.1861	**2.56E-03**	
EDSS	age	0.2614	**1.55E-05**	0.2106
	disease duration	0.1145	0.0803	
	sex	0.1118	**4.64E-02**	
	lesion volume	0.1728	**5.92E-03**	
	mean(lesions)	-0.0615	0.3066	
EDSS	age	0.2505	**3.55E-05**	0.2148
	disease duration	0.1126	0.0841	
	sex	0.0940	0.0748	
	lesion volume	0.1487	**2.29E-02**	
	CoV(lesions)	0.0994	0.1090	

Abbreviations: CoV: coefficient of variation, EDSS: Expanded Disability Status Scale, NAWM: normal-appearing white matter

SDMT

Linear relations were also observed between SDMT and the features, yielding significant correlations for every feature. The strongest correlation with SDMT was observed for age (r = -0.438, p = 4.17×10^{-15}), followed by CoVs of intensities in lesions (r = -0.335, p = 4.28×10^{-9}) and in the NAWM (r = -0.322, p = 1.83×10^{-8}), and lesion volume (r = -0.327, p = 1.05×10^{-8}). Mean intensities extracted from NAWM and lesions were the only positively correlated features. Hence, increases in age, disease duration, lesion volume, NAWM CoV, and lesion CoV, and a decrease in mean NAWM and mean lesion intensity are all associated with lower SDMT scores, indicating decreasing cognitive capacity.

The baseline model with SDMT as the dependent variable yielded a coefficient of determination of R^2 = 0.2584. In this model, age, sex, and lesion volume were significantly contributing features (see Table 3.5). Conversely, disease duration yielded a very low regression coefficient (β = 0.0034) and no statistical significance (p = 0.9574), and removing the feature from the baseline model did not impact the R^2 value. Consequently, to reduce the risk of overfitting and improve interpretability, disease duration

was not included in this baseline model, which only comprised age (β = -0.3871, p = 1.30x10^{-12}), sex (β = 0.1109, p = 0.03), and lesion volume (β = -0.2441, p = 4.43x10^{-6}). Mean lesion intensity (β = 0.1951, p = 6.62x10^{-4}), lesion CoV (β = -0.1649, p = 0.0055), and NAWM CoV (β = -0.1419, p = 0.0176) significantly contributed when added to the baseline model individually. The intensity features extracted from lesions also provided the greatest increase in R^2 (mean lesion intensity: R^2 = 0.2878, lesion CoV: R^2 = 0.2780), improving the baseline model's ability to explain variance in the SDMT score. Since several features showed significant contributions, additional multiple linear regression analysis was conducted by adding multiple intensity features to the baseline model to determine whether these features provide complementary information. Prior to that, a correlation matrix was used to investigate the correlation between features. The correlation matrix in Figure 3.9 shows that the lesion CoV and mean lesion intensity are highly correlated. Consequently, only mean lesion intensity (higher standardized β than lesion CoV) and NAWM CoV were added to the baseline model to avoid collinearity. This step further increased the coefficient of determination (R^2 = 0.2986), with both intensity features significantly contributing to the model (mean lesion intensity: β = 0.1837, p = 0.0013; NAWM CoV: β = -0.1236, p = 0.0363).

Figure 3.11
The scatter plots show the linear relation between SDMT scores and the features included in the multiple linear regression models. The blue line represents the linear regression line, illustrating the simple correlation between the two respective variables. The corresponding Pearson's r correlation coefficients are also provided in the upper right corner of each plot, along with the p-value indicating the level of significance. All features are negatively correlated with SDMT scores, except for mean NAWM intensity and mean lesion intensity, which show a positive correlation with SDMT scores.
Abbreviations: CoV: coefficient of variation, NAWM: normal-appearing white matter, SDMT: Symbol Digit Modalities Test

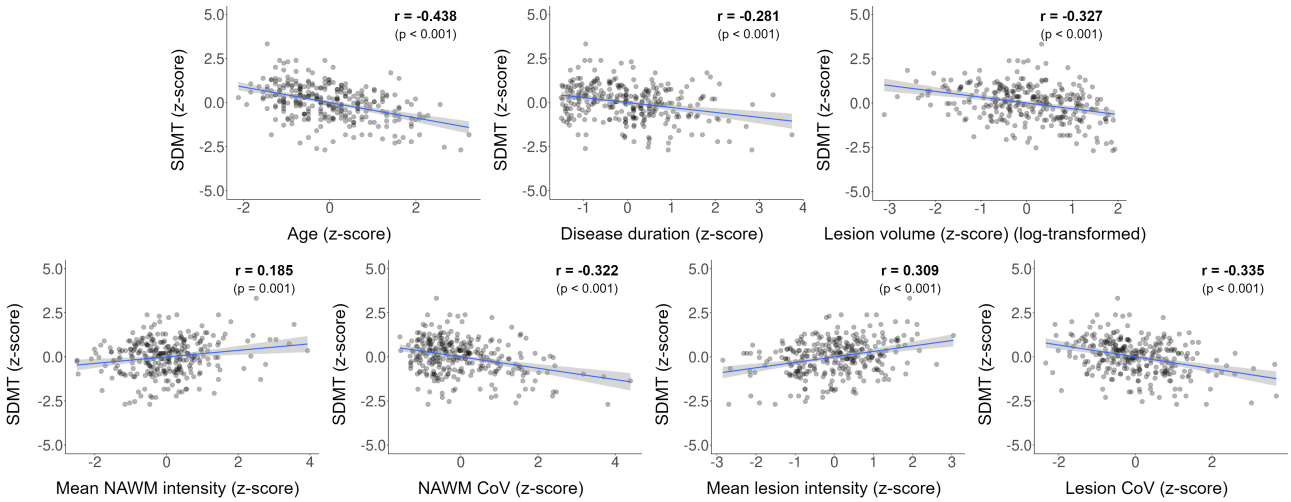


Table 3.6
SDMT multiple linear regression models. Besides the baseline models, including age, disease duration, sex, and lesion volume as independent variables, the table also shows the results of multiple linear regression models including intensity features. The CoV of NAWM, mean lesion intensity, and the CoV of lesions significantly contributed to the multiple linear regression model and considerably improved the baseline model's ability to explain variance in SDMT scores. Adding the mean lesion intensity to the baseline model yielded the most accurate model (R^2 = 0.2878). In addition, a model including both mean lesion intensity and NAWM CoV further improved the model's ability to explain variance in SDMT scores. The lesion CoV was not included to avoid collinearity (it strongly correlated with mean lesion intensity, see Figure 3.9). P-values of features that significantly contributed to the respective model are highlighted in bold, and features that were added to the baseline model are listed below the dashed lines in each case.

Dependent variable	Independent variables	Regression coefficient	p-value	R^2
SDMT	age	-0.3884	**8.10E-11**	0.2584
	disease duration	0.0034	0.9574	
	sex	0.1107	**3.07E-02**	
	lesion volume	-0.2453	**2.31E-05**	
SDMT	age	-0.3871	**1.30E-12**	0.2584
	sex	0.1109	**3.00E-02**	
	lesion volume	-0.2441	**4.43E-06**	
SDMT	age	-0.3802	**3.08E-12**	0.2651
	sex	0.0780	0.154	
	lesion volume	-0.2315	**1.52E-05**	
	mean(NAWM)	0.0896	0.1050	
SDMT	age	-0.3385	**3.66E-09**	0.2728
	sex	0.1400	**7.39E-03**	
	lesion volume	-0.2064	**1.66E-04**	
	CoV(NAWM)	-0.1419	**1.76E-02**	
SDMT	age	-0.3823	**1.00E-12**	0.2878
	sex	0.0511	0.3345	
	lesion volume	-0.1709	**2.26E-03**	
	mean(lesions)	0.1951	**6.62E-04**	
SDMT	age	-0.3653	**1.73E-11**	0.2780
	sex	0.1093	**3.04E-02**	
	lesion volume	-0.1648	**5.46E-03**	
	CoV(lesions)	-0.1649	**5.54E-03**	
SDMT	age	-0.3403	**1.76E-09**	0.2986
	sex	0.0799	0.1420	
	lesion volume	-0.1424	**1.27E-02**	
	mean(lesions)	0.1837	**1.32E-03**	
	CoV(NAWM)	-0.1236	**3.63E-02**	

Abbreviations: CoV: coefficient of variation, NAWM: normal-appearing white matter, SDMT: Symbol Digit Modalities Test

4 Discussion and conclusion

This thesis investigated how the potential of conventional MRI, already abundant in clinical workflows, can be exploited by advanced deep learning techniques to improve WM lesion detection, and by leveraging the informative value in T1w intensity patterns, an underexplored aspect despite the widespread availability of T1w images and inherent biophysical properties. Research in the field of neuroimaging, especially using MRI, has advanced rapidly in recent years, and deep learning-based automated segmentation tools have become increasingly popular as they provide improved segmentation performance compared to earlier methods [32, 92]. Leveraging these advancements, this thesis consists in part of the development and validation of LST-AI, a holistic deep learning-based tool for automated segmentation of WM lesions in MS, aiming to bridge the gap between research and practical clinical application in MS. In addition, intensity scaling of conventional MRI was investigated in this thesis. It is known that advanced qMRI, such as R1 relaxometry, is sensitive to MS-related pathologies in the brain [59, 64, 68–72], but its availability is limited, hampering studies on large datasets. Therefore, substituting qMRI by exploiting the informative wealth of conventional MRI image intensities is very attractive. To this end, a systematic review was conducted on intensity scaling of conventional MRI, summarizing existing methods and identifying similarities between scaled images and advanced qMRI and histology. Building on this, the thesis also involved implementing several intensity scaling methods and validating their effectiveness in converting conventional T1w MRI intensities into quantitative values. Ultimately, LST-AI and the most promising scaling method were combined to assess whether intensity patterns of lesions and NAWM in conventional T1w brain MRIs could provide new insights into disease progression, beyond established biomarkers like lesion volume and number.

LST-AI lesion segmentation

The field of research on automated lesion segmentation in MS has continuously evolved, from the first tools based on statistics [26, 148], through machine learning methods [26, 37, 39], to state-of-the-art deep learning–based models, particularly CNNs [27, 28, 31–35, 38, 40–43, 92]. In the current thesis, we developed and introduced "LST-AI, a new deep learning-based segmentation method for WM lesions in MS. It is built from an ensemble of three 3D U-Nets. Using LST-AI and a pair of T1w and FLAIR MRI images as input, it is possible to accurately segment lesions. We analyze the segmentation performance on multiple external datasets, thereby showing that LST-AI generalizes to data from different centers and scanners without retraining. We also compare our method to benchmark methods for validation and find excellent lesion segmentation performance of LST-AI. In addition, LST-AI can label lesions according to their location, thereby providing further possibilities for lesion characterization in MS" [1]. In terms of segmentation accuracy on the voxel-level (DSC) and lesion detection (F1), LST-AI (DSC = 0.67 and F1 = 0.63) outperformed all benchmark methods (DSC = 0.42 - 0.55 and F1 = 0.22 - 0.46). Given that LST-AI was trained on an in-house dataset, its superior performance on the test dataset, which exclusively consists of external datasets, indicates that it can be applied to new data without retraining the ensemble network. This is an important attribute for utilizing the tool in settings with limited data availability and for potential deployment into clinical workflows, where retraining is not an option. Previous studies already aimed for improved generalizability and evaluated their segmentation tools on some of the publicly available datasets included in our test set. "Valverde et al. [31], have previously optimized retraining on small datasets, as their tool only requires a single case to adapt their model to new datasets. They also validated their method on the ISBI 2015 test dataset and achieved a mean DSC of 0.58 [31]" [1], which is slightly lower than we observed for LST-AI on the ISBI training dataset (DSC = 0.61). Interestingly, the best-performing models in the ISBI 2015 challenge, which were trained on the ISBI 2015 training dataset, achieved DSC scores ranging between 0.50 and 0.68 on the test dataset [34, 149]. However, this might be an unfair comparison,

since "assessing generalizability of segmentation models requires validation on external datasets. This has been done in recent studies, which used different train and test set pairings, including in-house and publicly available data such as ISBI 2015 and MICCAI 2016 data (e.g., train on in-house data and test on MICCAI 2016 data) [28, 37, 39, 41, 150–152]" [1]. Gentile et al. adapted and trained BIANCA, a k-nearest neighbour machine learning tool to segment WM hyperintensities [153], to develop an MS-specific tool BIANCA-MS [39]. When trained on a multi-center dataset (n = 140) and tested on the mssegtest dataset (n = 37), BIANCA-MS yielded a slightly larger DSC score (DSC = 0.7) than LST-AI (DSC = 0.65) on this specific dataset. In the study by Kamraoui et al., a model based on a large group of compact 3D CNNs, the impact of different training and testing datasets on the segmentation performance was also investigated [41]. They show that "the segmentation performance on the ISBI 2015 test dataset drops when models are trained on in-house data (DSC = 0.13 – 0.48) compared to when they are trained on the ISBI training dataset (DSC = 0.64 – 0.67). On the MICCAI 2016 dataset, however, the models trained on the in-house training dataset showed robust and high DSC scores (0.65 – 0.72) [41]" [1]. McKiney et al. suggested DeepSCAN, a 3D-to-2D architecture network, and evaluated the model on the mssegtrain dataset [28]. They found that the DSC score considerably dropped when the model was trained on an out-of-sample dataset (DSC = 59.7) compared to cross-validation using only the mssegtrain dataset (DSC = 75.7). Similarly, the results from the studies by Li et al. [151] and Rakić et al. [152] further report lower DSC scores on out-of-sample test sets compared to utilizing training and testing datasets from the same image domain. "This highlights the impact of differing image domains in train and test sets and the need for validation on multiple test datasets, which can provide a more realistic representation of a model's generalizability. In this study, image domain heterogeneity is simulated by the validation of our method on multiple datasets, which were also part of MS lesion segmentation challenges of the ISBI 2015 conference and the MICCAI 2016 conference [32, 92, 94]. While our model achieves similar scores (mean DSC of 0.61 and 0.65 for ISBI 2015 and MICCAI 2016, respectively) as the top-performing models in both challenges, we want to emphasize that, in contrast to the participating models, our model is not specifically trained on the corresponding training datasets provided in the challenges. These two scores are also close to the inter-rater DSC scores of the expert segmentation used in the challenges (DSC of 0.63 and 0.66 – 0.76 in ISBI 2015 and MICCAI 2016, respectively) [92, 94]" [1]. The study by Kamraoui et al. [41] also shows that the performance on external datasets can significantly depend on the specific selection of the test set. Therefore, LST-AI was validated on multiple publicly available datasets individually, and yielded good segmentation performance throughout (DSC = 0.61 - 0.74). The observation of robust lesion segmentation performance of LST-AI underlines "the good generalization of our model and its reliable application to multicenter data acquired with different scanners and protocols. [...] In contrast, the lower performance of the other methods, e.g., the pre-trained nnUNet, suggests the need for adaptation of these methods through retraining" [1].

A robust performance was also observed on the lesion level, where LST-AI (F1 = 0.63) outperformed the benchmark methods (F1 = 0.22 - 0.46) across all datasets, and performed consistently well in all datasets individually (F1 = 0.57 - 0.70). Compared to LST-AI (PPV = 0.73, PPVL = 0.64), the benchmark methods provided similar or better PPV (0.72 - 0.90) and PPVL (0.34 - 0.64) scores, which might occur at the cost of sensitivity, where LST-AI clearly exceeded all benchmark methods on the voxel-level (LST-AI: 0.66; benchmark: 0.32 - 0.49) and lesion-level (LST-AI: 0.70; benchmark: 0.20 - 0.40). These results reinforce the inclusion of the Tversky loss in the model, which is designed to reduce FN, thereby improving the sensitivity. This is underlined by comparing the median number of FN (in terms of lesion detection) of LST-AI (FN = 4, IQR: 8) with the results obtained with BIANCA-MS by Gentile et al., which, in spite of a larger DSC score, yielded a greater median number of FN (FN = 11, IQR: 18) [39]. Hence, LST-AI consistently detects WM lesions with high sensitivity, which is crucial for the integration of automated lesion segmentation tools in the clinical setting, since diagnosing and monitoring MS significantly rely on the detection of lesions. The improved sensitivity in comparison to existing methods is also reflected in the lesion detection analysis in relation to lesion volume. "Similarly to previous reports by Commowick et al. [32] and Rakić et al. [152], we also found that it is particularly hard to detect small lesions (<10 mm3). Nonetheless,

the steep improvement of lesion detection with increasing lesion size provides a promising perspective for the integration of automated lesion segmentation tools in clinical settings, since it can help clinicians to detect lesions faster and to diagnose and monitor MS more accurately" [1]. Overall, based on the robust performance on external datasets on the voxel-level as well as the lesion-level, we "believe that using an ensemble approach including multiple pre-trained U-Nets translates into robustness against performance variability of individual 3D U-Nets and, therefore, generalizes better across different imaging protocols and centers" [1].

Moreover, LST-AI can label lesions according to their location in specific brain regions that are relevant in MS (i.e., PV, IT, JC, and SC) [8, 154]. The segmentation performance of LST-AI was assessed in each of these regions individually, showing greater DSC scores than any of the benchmark methods in all regions (see Figure 3.1). This is particularly relevant in the JC region, as JC lesions are considered to be specific to MS and strongly linked to clinical disability, but their segmentation has been a challenge in existing methods, as evidenced by the relatively low DSC scores in benchmark methods.

In addition to the segmentation performance, LST-AI provides another advantage in that it was developed in a holistic concept and comprises all required processing steps, generates probability and binary lesion maps, and automatically provides summary data on the total lesion volume and number. The integration of processing steps prevents the application of different tools and software, which could result in degradation of the segmentation output and decrease the robustness of the model's performance on new data. Furthermore, providing LST-AI as a holistic and open-source tool is essential for transparency and deployment, enabling the entire community to apply the tool and actively contribute to continuously improving the tool, thereby bringing advanced automated lesion segmentation closer to clinical integration.

"Our study does not come without limitations. First, our model requires T1w and FLAIR image pairs, which might not always be available. Second, although less pronounced than the benchmark methods, our model still shows decreased lesion detection efficiency with decreasing lesion volumes. Even though the explainability of features learned via CNNs and, more specifically U-Nets have been comparatively well studied, they still lack some interpretability in contrast to methods leveraging manually selected features. In addition, preprocessing is part of the LST-AI toolbox and includes registration to MNI space, which prevents the possible effects of different preprocessing methods on segmentation performance. To handle all segmentation tools under comparison equally, we also followed this strategy to validate our method. Yet this may have lowered the performance of those lesion segmentation tools not inherently operating in MNI space. Finally, the quality of the publicly available datasets used for validation in this study is likely above average; therefore, segmentation performance may be lower for data of lower quality closer to real-life clinical data" [1].

Intensity scaling

Intensity scaling is an attractive approach that potentially opens the possibility to quantitatively analyze conventional MRI. Since neurodegenerative diseases, such as MS, can affect the whole brain, selecting brain tissue as a reference region for intensity scaling can be problematic. In light of this circumstance, the systematic review conducted in this thesis provided an overview of scaling strategies that avoid reference regions within the brain, which can be of particular interest in MS research. In general, the included studies were heterogeneous in the applied intensity scaling techniques, the reference methods (qMRI sequences, histology, other scaling methods), and the brain regions (tumor, cortex, WM,...) that were investigated in the context of varying clinical settings (healthy controls, glioma, MS,...). A consistent observation was that inter-image variance was considerably reduced through intensity scaling. However, it is important to show that intensity scaling does not remove or aggravate biological variance due to inappropriate reference region selection. This was specifically addressed by the studies of Ganzetti et al. [83, 155] and Brown et al. [77], which confirmed the reduction of inter-image variance [77, 83] and highlighted the importance of preserving biological variance through the selection of appropriate reference regions [77, 155]. Hence,

"sparse evidence suggests that the decrease in inter-scan variance with the preservation of considerable biological variance can be achieved without specific cerebral reference regions. This variance reduction can provide many advantages in studies that need to harmonize multi-subject or multi-center images" [2]. In addition, the level of evidence on correlations with qMRI was low, and only relaxometry and MTR, which are sensitive to myelin, axonal counts, and tissue integrity [49–52], consistently correlated with scaled conventional T1w or T2w intensities and with T1w/T2w-ratios [77, 83, 103, 104, 108–110, 113, 114]. Similarly, scaled image intensities and MTR yielded comparable correlations with clinical data such as EDSS [77, 104, 113], showcasing the possible application of intensity scaling in analyzing disease progression. A few studies also compared T1w/T2w-ratios to histology in the GM, which confirmed the notion of conventional MRI as a composite measure of tissue integrity, since the ratio was observed to be sensitive to multiple properties (myelin, neurite density, and dendrite density) [103, 105, 106, 116]. "In summary, intensity scaling of conventional T1w and T2w MR images without cerebral reference regions is feasible, leading to variance reduction while possibly preserving biologically meaningful information, with the potential to allow for signal quantification as a compound measure of tissue integrity. However, the overall level of evidence is low, with numerous open methodological questions" [2]. Particularly, only three studies investigated intensity scaling of conventional T1w images alone, and no evidence could be found on the validation of intensity-scaled T1w images with quantitative T1 maps, two imaging techniques that are intrinsically related through the strong T1-dependency of their signal.

Based on the systematic review of available literature, this research gap was then addressed with an extensive comparison and validation of intensity scaling methods for conventional T1w images, using quantitative R1 maps, acquired at our institution, as reference. Comparing the results obtained with intensity-scaled images to unscaled T1w images highlighted the benefit of intensity scaling. Suitable extracerebral reference regions were identified from eight candidate regions using a data-driven approach. Only temporal fatty tissue and temporal muscle were identified as appropriate reference regions, as their inter-image variance was comparable to WM and GM, which agrees with the use of similar tissues in previous studies [77, 83, 84, 86, 90]. Eyes were used as a third reference region for MRI signal modeling, since their low (absolute) standard deviation makes them well-suited for orthogonal distance regression. Thorough systematic evaluation also ensured independence of disease-related effects of these reference regions, which makes them ideal candidates to model technical variance without including disease-related biological variance. This, however, was not the case for the standard deviation used as a scaling factor in z-score standardization. Hence, z-score standardization may lead to efficient reduction of inter-image variance and good correlation with R1 maps, but the significant group difference of the standard deviation also indicates removal of biological variance. This is a plausible finding, since the standard deviation used in z-score standardization is calculated across the whole brain, and is likely influenced by disease-related effects such as lesion load, atrophy, and NAWM damage. Thus, z-score standardization seems unsuitable for intensity conversion of T1w images in the specific case of MS.

To ensure reliable application of intensity scaling methods, they were validated technically, serving as proof-of-concept, and biologically, substantiating their appropriateness in specific use cases. Due to their intrinsic relation to T1w images and their confirmed sensitivity towards pathology in MS [49, 59, 64, 68–72], R1 maps from MPM were chosen as a reference for validation. Figure 3.5 shows that "linear scaling with fatty tissue and signal modeling yielded CoVs similar to those of R1 maps, achieving the most effective variance reduction across all brain ROIs [...]. Their slightly increased CoVs in comparison to R1 maps in bigger regions (e.g., NAWM and cortex) may be caused by less effective bias field correction in comparison to the measurement-based correction of the low-frequency bias field in MPM. Regarding the correlation with R1 maps, we found that bias correction expectedly increases the voxel-wise correlation compared to unprocessed T1w images (Figure 3.6C), well compatible with the notion that bias correction decreases technical variance in the individual image while ignoring the absolute intensity scale. Interestingly, ROI-based correlations of scaled T1w images with R1 maps (Figure 3.6A and Figure 3.6B) were virtually the same for all scaling methods (R^2 = 0.66 - 0.73), not allowing for a differentiated quality assessment. Brown

et al. found similar correlation values (R^2 = 0.63) between fat-scaled T1w images and MTR, a measurement that is sensitive to similar properties as R1 maps (e.g., density of macromolecules) [77]. Our results indicate that scaling with regions outside the brain can yield correlations with R1 maps comparable to scaling methods based on WM, GM, and CSF, as Hasse et al. [87] reported in healthy subjects (R^2 = 0.72)" [3].

Parallels between fat-scaled images and R1 maps were also observed in the biological validation, as mean NAWM intensities in both images significantly differed between the low and high EDSS groups. The difference in R1 maps is in line with previous studies associating decreased R1 values in NAWM with MS pathology [49, 59, 64, 68–72]. Thus, fat-scaled images may be similarly sensitive towards MS disease progression as R1 maps.

Even stronger evidence was observed based on NAWM CoV, a potential indicator for local alterations within one individual. "The CoV is a scale-independent measure that can be directly derived from bias-corrected T1w images. To our surprise, the group difference between people with low ($\leq$3) and high ($>$3) EDSS in NAWM was most significant when using the CoV instead of the mean values for all image types. Even after including age, sex, disease duration, and lesion volume as covariates, the CoV from bias-corrected T1w images yielded a significant difference. This was not the case in R1 maps since, in contrast to conventional T1w images, R1 maps are optimized for generating reliable quantitative intensities, which comes at the expense of a lower image resolution and tissue contrast. Similar results were found by Johnson et al. [156], with CoVs of the MWF in NAWM differentiating more accurately between healthy controls and MS than the mean" [3]. Overall, technical and biological validation suggest that linear scaling of conventional T1w with fatty tissue as reference region, and CoV can be used to analyze conventional T1w images when no qMRI is available. This is particularly interesting to exploit the large availability of conventional T1w images compared to qMRI, potentially identifying new biomarkers related to disease progression in MS.

"We acknowledge the limitations of our study. The MRI images were acquired on the same scanner and with the same protocol. T1w sequences are optimized for brain tissues; coverage of regions outside the brain may vary between scanners and protocols, and so may their suitability as reference regions. Finally, our study was limited to brain MRI of people with MS. Consequently, we regard a validation of intensity scaling based on temporal fatty tissue in different T1w sequences in a multi-center study, ideally across multiple neuropsychiatric diseases, as a reasonable next step. In the same context, the advantage of using a single reference region (temporal fatty tissue) over multiple reference regions (signal modeling) needs to be confirmed" [3].

Clinical correlation

Inspired by the validation of intensity scaling methods, further analyses were conducted to explore the potential of conventional MRI in explaining disease severity reflected by EDSS and SDMT. While the previous biological validation assessed whether intensity scaling can convert intensities into meaningful values to measure disease effects, this section examined to what extent intensities in lesions and NAWM contribute to explaining disease severity, in addition to established variables such as age, lesion volume, sex, and disease duration.

Simple correlations with EDSS were strongest for age and NAWM CoV, and, generally, the variability of intensities in both NAWM and lesions were more closely related to EDSS scores than global mean intensity measures. Table 3.5 clearly shows that NAWM CoV improved the baseline model's ability to explain EDSS, indicating that the variability of intensities in the NAWM contains information on clinically relevant brain pathologies that is not captured by traditional features. However, the results also indicate a partial overlap of information content with traditional features, since their standardized β values slightly decrease when NAWM CoV is added to the model, and the correlation matrix in Figure 3.9 indicates a weak association between NAWM CoV and age, disease duration, and lesion volume.

SDMT is a measure of cognitive function and is consequently more closely related to brain-specific pathologies compared to the EDSS score, which also captures disability induced by MS-related injuries in the spinal cord [157–159]. This property is also reflected in the simple correlation analysis, which shows that all features extracted from the conventional T1w brain images display linear relations with SDMT and significantly correlate with the score, which was not the case for EDSS. Specifically, intensities extracted from lesions correlate more strongly with SDMT than with EDSS. In the baseline model, disease duration did not significantly contribute and was removed to avoid collinearity. Considering the moderate correlation of disease duration with age and lesion volume, it is likely that these two variables explain the same variance in SDMT as disease duration, rendering disease duration redundant. In contrast, disease duration provided complementary information in the models explaining EDSS, which might reflect disability progression due to spinal cord injury that cannot be captured by features of brain images. When added to the baseline model (R^2 = 0.258), mean lesion intensity provided the greatest additional value (R^2 = 0.288, β = 0.195), showing that lesion intensity could play an important role in quantifying pathologies that induce cognitive decline. In a multiple linear regression model including multiple intensity features, results even show that mean lesion intensity and NAWM CoV provided complementary information, which further improved the model's ability to explain disease severity in terms of SDMT scores (R^2 = 0.299). This suggests that intensities in lesions and in NAWM reflect different disease mechanisms, at least to some degree.

Overall, NAWM CoV provides information on disease severity in addition to traditional variables in both settings (EDSS and SDMT), suggesting that it reflects brain NAWM pathology and disease mechanisms that are not directly linked to lesion formation. This is underlined by the fact that lesion volume, mean lesion intensity, and NAWM CoV seem to contribute complementary information in explaining cognitive disability in terms of SDMT scores. This is in agreement with previous studies that investigated the role of NAWM in MS pathology using MRI as well as histology [4, 12–14, 17–19, 49, 59, 64, 68–72]. The results of these studies point towards diffuse inflammatory processes in NAWM characterized by microglia activation and accumulation, and increased cytokine expression, causing tissue injury, such as axonal loss and decreased fiber density [12–14, 17–19]. Interestingly, the extent of NAWM injury could be associated with disease severity [160, 161], indicating its potential value in disease and treatment monitoring. Moreover, histological studies indicated that demyelination does not strictly accompany axonal degeneration in NAWM [12–14, 17, 19], allowing for two ways of interpreting the pathological mechanism in NAWM: 1) inflammatory activity in NAWM could be an indicator of lesion formation [12], and 2) diffuse NAWM injury is caused by a separate mechanism independent of WM lesion formation [14, 17, 19]. The results in this thesis suggest that the NAWM CoV can provide valuable information on the latter, as it complements lesion volume in explaining cognitive as well as functional disability. In contrast, mean NAWM intensity did not exhibit these attributes, as it poorly correlates with EDSS and SDMT, which could be explained by its reduced sensitivity towards local, rather subtle, alterations of the NAWM. Hence, using only conventional MRI, the results of this thesis concur with the findings by Johnson et al. [156], who observed stronger associations between NAWM CoV and EDSS/SDMT compared to mean NAWM intensities, using quantitative MWF.

Yet, mean (lesion) intensity also seems to play an important role in explaining cognitive disability in terms of SDMT, since it improved the accuracy of the multiple linear regression the most. In general, intensities extracted from lesions significantly contributed only to the models explaining SDMT. Conversely, the increase in EDSS seems to be strongly influenced by the total lesion volume, and the EDSS-relevant information content of lesion intensities does not extend beyond that. This seems plausible, since EDSS is a measure of functional disability, which is also strongly sensitive towards MS-related injuries in the spinal cord [157–159], limiting the value of information extracted from brain MRI. SDMT, on the other hand, is a measure exclusively sensitive to brain pathologies, justifying the increased explanatory power of brain-related features. The significant contribution of mean lesion intensity agrees with the notion that hypointense T1w lesions (black holes) are more destructive, with increased axonal loss and less remyelination [162–164], and can lead to more severe cognitive decline reflected by low SDMT scores. Therefore,

it is important to evaluate not only the lesion load but also the type of lesion, as a higher proportion of hypointense lesions (reflected by a lower mean lesion intensity) indicates more severe disease progression. While black hole lesions have been investigated in previous publications [164–167], studies evaluating the level of hypointensity of black holes are scarce and have so far relied on qMRI [166, 168–170]. The evidence gathered in this thesis suggests that intensity scaling could pave the way for new biomarkers of lesion destructiveness that can be extracted from conventional MRI.

While conventional MRI images were previously used for lesion segmentation to extract lesion volume, qMRI was required to quantify NAWM integrity and lesion characteristics, as conventional MRI was limited to visual assessment of intensities. In this thesis, an automated lesion segmentation tool has been developed and validated, addressing the limitations of previous tools in terms of segmentation accuracy, accessibility, transferability, and applicability, allowing for reliable lesion identification that aims to bridge the gap between research and clinical integration. Moreover, conventional T1w images were shown to contain exploitable information on NAWM pathology, and appropriate intensity scaling techniques were used to convert conventional T1w image intensities into interpretable values that can help measure the level of destructiveness of lesions and describe disease severity.

Conclusion and outlook

Conventional MRI images are part of routine clinical examinations in MS and are essential for the detection of WM lesions, which are pivotal for diagnosis and disease monitoring. The abundant availability of conventional MRI images is attractive not only for the development of deep learning-based applications but also for retrospective analysis of large-scale datasets to identify MS biomarkers. This thesis explored the potential of conventional MRI by developing LST-AI, a novel deep learning-based tool for automated lesion segmentation, and by unlocking new possibilities for large-scale quantitative analysis of conventional T1w images.

The development and validation of LST-AI support the progress toward better automated MS lesion detection, providing accurate lesion segmentation across multiple datasets and outperforming established benchmark methods on both the voxel and lesion levels. Notably, LST-AI does not require retraining when applied to new datasets, a critical attribute for the application in settings where data can be heterogeneous and retraining is not feasible. The holistic design of LST-AI, which takes raw T1w and FLAIR images as input and outputs lesion volume and number, integrates all necessary processing steps and helps bridge the gap between research and clinical integration. Moreover, its open-source availability promotes transparency and continuous improvement through contributions by the community.

The thesis also assessed the value of intensity scaling to transform conventional T1w image intensities into quantitative units, addressing the shortcomings of advanced qMRI techniques, which are often inaccessible due to limited availability. The systematic review highlights that intensity scaling can effectively reduce inter-image variance, facilitating harmonization across multi-center studies without compromising biologically meaningful variance, provided that reference regions are appropriately selected. Pertaining to this, the thesis addressed a key gap in neuroimaging research by systematically evaluating intensity scaling methods for conventional T1w MRI, with the goal of improving the interpretability of image intensities in the context of MS. By validating these methods against quantitative R1 maps, and through both technical and biological validation, it could be shown that intensity scaling using extracerebral reference regions, particularly temporal fatty tissue, can effectively reduce inter-image variance while preserving biologically meaningful signal variation. Notably, the NAWM CoV extracted from bias-corrected T1w images emerged as a sensitive and scale-independent marker of MS pathology, as it significantly distinguished between patient subgroups stratified by EDSS scores.

The clinical relevance of these findings was further underscored by the association of conventional T1w-derived features, specifically NAWM CoV and mean lesion intensity, with functional (EDSS) and cognitive

(SDMT) disability. These features appear to capture complementary aspects of MS-related brain pathology, with NAWM CoV likely reflecting diffuse, non-lesional tissue injury and lesion intensity corresponding to the degree of tissue destruction within focal lesions - aspects that could previously only be analyzed through histology or qMRI. The correlation of lesion hypointensity with lower SDMT scores demonstrates the utility of intensity scaling, which converts lesion intensities into a quantitative measure to investigate the pathological burden of 'black hole' lesions. These quantitative intensity features extend beyond traditional biomarkers and offer new avenues for analyzing and monitoring disease progression. Specifically, the findings in this thesis suggest that lesion and NAWM pathology can occur independently, and that NAWM damage and lesion destructiveness can be quantified using conventional T1w MRI. In addition, transforming conventional T1w images from a qualitative to a more quantitative tool enables large-scale longitudinal studies, which can help in better understanding disease mechanisms, including the role of NAWM damage as a precursor of lesion formation and as a lesion-independent process.

In summary, this work demonstrates that with appropriate processing and analysis, relying on LST-AI, validated intensity scaling methods, and combinations thereof, conventional MRI can be leveraged to provide not only qualitative but also quantitative information. Features such as NAWM CoV and lesion intensity have strong potential to enhance our understanding of disease mechanisms, and enable large-scale and longitudinal studies in settings where advanced qMRI is not available.

Acknowledgements

I would like to express my deepest gratitude to all the people who have supported me throughout this journey, both academically and personally. Completing a PhD is not a solitary endeavour, and I am incredibly fortunate to have been surrounded by so many inspiring, kind, and supportive individuals.

First and foremost, I thank Mark Mühlau, who gave me a chance to return to academia. His unwavering support allowed me to pursue my own research interests, while his thoughtful guidance and endless stream of ideas continuously opened up new directions and projects. Therefore, I am deeply grateful for the trust and freedom he granted me. I also want to sincerely thank my Thesis Advisory Committee. Benedikt Wiestler always made time to meet and support me, no matter how busy his schedule was. His pragmatic approach and friendly nature made every interaction productive and enjoyable, and I particularly appreciated the chance to collaborate with him. Valentin Riedl brought a refreshing perspective to every discussion. His creative and inspiring ideas often challenged my thinking in the best ways possible. Furthermore, this thesis would not have been possible without the collaboration with colleagues from the Department of Neuroradiology. I am especially grateful to Jan Kirschke, whose pleasant and open manner made working together both smooth and rewarding. The collaboration with the department laid the essential groundwork for everything that followed.

To my colleagues in the research group and TUM-NIC, thank you for creating such a special environment. Sophia and Matthias, thank you for welcoming me into the research group, for your patient guidance, and for your thoughtful perspectives. Special thanks to Julian, working with you was not only extremely insightful and helpful but also incredibly fun. Your willingness to support others is inspiring, and I'm truly thankful for our friendship, both in and outside the office. Thanks to Markus, whose presence always brings joy, whether it's productive work and brainstorming new ideas or just mucking about, I always look forward to your company. Thanks to Cui Ci, with whom I shared the ups and downs of the PhD journey, which made the hard times much easier to endure. I would also like to thank Laura, Elisabet, Felix and Felix, Nico, Alyssa, Cristina, and Vanessa, you made every day at work enjoyable, turning the workplace into a uniquely fun and motivating environment. I'll especially treasure our legendary lunch breaks as well as the fun sporty activities and social events.

Most importantly, I thank Lina, whose unwavering support has carried me through this journey. From the first years of our studies to the moment I decided to leave my previous job to pursue this PhD, you were by my side, encouraging me, believing in me, and pushing me when I needed it most. Despite the immense stress of your own work, you never failed to provide me with a safe haven. Your strength and dedication continue to inspire me every day. Finally, I thank my family for their unconditional support. Even though the nuances of my research may not have been entirely clear to them, their fascination and curiosity towards my work gave me renewed motivation and energy. Their frequent visits to Munich gave me a profound sense of stability and a feeling of home.

References

[1] Tun Wiltgen et al. “LST-AI: A deep learning ensemble for accurate MS lesion segmentation”. In: *NeuroImage: Clinical* 42 (2024), p. 103611.

[2] Tun Wiltgen et al. “Intensity scaling of conventional brain magnetic resonance images avoiding cerebral reference regions: A systematic review”. In: *PloS one* 19.3 (2024), e0298642.

[3] Tun Wiltgen et al. “Towards quantitative intensity analysis of conventional T1-weighted images in multiple sclerosis”. In: *NeuroImage* 318 (2025), p. 121395.

[4] Massimo Filippi et al. “Multiple sclerosis”. In: *Nature Reviews Disease Primers* 4.1 (2018), p. 43.

[5] Alan J Thompson et al. “Multiple sclerosis”. In: *The Lancet* 391.10130 (2018), pp. 1622–1636.

[6] Tanja Kuhlmann et al. “Multiple sclerosis progression: time for a new mechanism-driven framework”. In: *The Lancet Neurology* 22.1 (2023), pp. 78–88.

[7] Alastair Compston and Alasdair Coles. “Multiple sclerosis”. In: *The Lancet* 372.9648 (Oct. 25, 2008). Publisher: Elsevier, pp. 1502–1517.

[8] Alan J Thompson et al. “Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria”. In: *The Lancet Neurology* 17.2 (2018), pp. 162–173.

[9] Massimo Filippi et al. “Diagnostic Criteria for Multiple Sclerosis, Neuromyelitis Optica Spectrum Disorders, and Myelin Oligodendrocyte Glycoprotein–immunoglobulin G-associated Disease”. In: *Neuroimaging Clinics* 34.3 (2024), pp. 293–316.

[10] John F Kurtzke. “Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS)”. In: *Neurology* 33.11 (1983), pp. 1444–1444.

[11] Ralph HB Benedict et al. “Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis”. In: *Multiple Sclerosis Journal* 23.5 (2017), pp. 721–733.

[12] Aletta MR Van Den Bosch et al. “Profiling of microglia nodules in multiple sclerosis reveals propensity for lesion formation”. In: *Nature communications* 15.1 (2024), p. 1667.

[13] Thomas Zeis et al. “Normal-appearing white matter in multiple sclerosis is in a subtle balance between inflammation and neuroprotection”. In: *Brain* 131.1 (2008), pp. 288–303.

[14] Alexandra Kutzelnigg et al. “Cortical demyelination and diffuse white matter injury in multiple sclerosis”. In: *Brain* 128.11 (2005), pp. 2705–2712.

[15] Massimiliano Calabrese, Massimo Filippi, and Paolo Gallo. “Cortical lesions in multiple sclerosis”. In: *Nature Reviews Neurology* 6.8 (2010), pp. 438–444.

[16] Massimiliano Calabrese et al. “Cortical lesion load associates with progression of disability in multiple sclerosis”. In: *Brain* 135.10 (2012), pp. 2952–2961.

[17] Natalia M Moll et al. “Multiple sclerosis normal-appearing white matter: Pathology–imaging correlations”. In: *Annals of neurology* 70.5 (2011), pp. 764–773.

[18] Shailender Singh et al. “Microglial nodules in early multiple sclerosis white matter are associated with degenerating axons”. In: *Acta neuropathologica* 125 (2013), pp. 595–608.

[19] Owain W Howell et al. “Activated microglia mediate axoglial disruption that contributes to axonal injury in multiple sclerosis”. In: *Journal of Neuropathology & Experimental Neurology* 69.10 (2010), pp. 1017–1033.

[20] David R Van Nederpelt et al. “Reliability of brain atrophy measurements in multiple sclerosis using MRI: an assessment of six freely available software packages for cross-sectional analyses”. In: *Neuroradiology* 65.10 (2023), pp. 1459–1472.

[21] Loredana Storelli et al. “Methods for brain atrophy MR quantification in multiple sclerosis: application to the multicenter INNI dataset”. In: *Journal of Magnetic Resonance Imaging* 58.4 (2023), pp. 1221–1231.

[22] Jaume Sastre-Garriga et al. “MAGNIMS consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice”. In: *Nature Reviews Neurology* 16.3 (2020), pp. 171–182.

[23] Massimo Filippi et al. “Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis”. In: *Brain* 118.6 (1995), pp. 1601–1612.

[24] LK Fisniku et al. “Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis”. In: *Brain* 131.3 (2008), pp. 808–817.

[25] Devon Oship et al. “Assessment of T2 lesion-based disease activity volume outcomes in predicting disease progression in multiple sclerosis over 10 years”. In: *Multiple Sclerosis and Related Disorders* 67 (2022), p. 104187.

[26] Paul Schmidt et al. “An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis”. In: *Neuroimage* 59.4 (2012), pp. 3774–3783.

[27] Hongwei Li et al. “Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images”. In: *NeuroImage* 183 (2018), pp. 650–665.

[28] Richard McKinley et al. “Simultaneous lesion and brain segmentation in multiple sclerosis using deep neural networks”. In: *Scientific reports* 11.1 (2021), p. 1087.

[29] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer. 2015, pp. 234–241.

[30] Fabian Isensee et al. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. In: *Nature methods* 18.2 (2021), pp. 203–211.

[31] Sergi Valverde et al. “One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks”. In: *NeuroImage: Clinical* 21 (2019), p. 101638.

[32] Olivier Commowick et al. “Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure”. In: *Scientific reports* 8.1 (2018), p. 13650.

[33] Marcos Diaz-Hurtado et al. “Recent advances in the longitudinal segmentation of multiple sclerosis lesions on magnetic resonance imaging: a review”. In: *Neuroradiology* 64.11 (2022), pp. 2103–2117.

[34] Yang Ma et al. “Multiple sclerosis lesion analysis in brain magnetic resonance images: techniques and clinical applications”. In: *IEEE Journal of Biomedical and Health Informatics* 26.6 (2022), pp. 2680–2692.

[35] Chenyi Zeng et al. “Review of deep learning approaches for the segmentation of multiple sclerosis lesions on brain MRI”. In: *Frontiers in Neuroinformatics* 14 (2020), p. 610967.

[36] Quentin Vanderbecq et al. “Comparison and validation of seven white matter hyperintensities segmentation software in elderly patients”. In: *NeuroImage: Clinical* 27 (2020), p. 102357.

[37] Stefano Cerri et al. “A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis”. In: *Neuroimage* 225 (2021), p. 117471.

[38] Pooya Ashtari et al. “New multiple sclerosis lesion segmentation and detection using pre-activation U-Net”. In: *Frontiers in Neuroscience* 16 (2022), p. 975862.

[39] Giordano Gentile et al. “BIANCA-MS: An optimized tool for automated multiple sclerosis lesion segmentation”. In: *Human Brain Mapping* 44.14 (2023), pp. 4893–4913.

[40] Maryam Hashemi, Mahsa Akhbari, and Christian Jutten. “Delve into Multiple Sclerosis (MS) lesion exploration: A modified attention U-Net for MS lesion segmentation in Brain MRI”. In: *Computers in Biology and Medicine* 145 (2022), p. 105402.

[41] Reda Abdellah Kamraoui et al. “DeepLesionBrain: Towards a broader deep-learning generalization for multiple sclerosis lesion segmentation”. In: *Medical Image Analysis* 76 (2022), p. 102312.

[42] Anitha Priya Krishnan et al. “Multi-arm U-Net with dense input and skip connectivity for T2 lesion segmentation in clinical trials of multiple sclerosis”. In: *Scientific Reports* 13.1 (2023), p. 4102.

[43] Francesco La Rosa et al. “Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE”. In: *NeuroImage: Clinical* 27 (2020), p. 102335.

[44] Cristina Granziera et al. “Quantitative magnetic resonance imaging towards clinical application in multiple sclerosis”. In: *Brain* 144.5 (2021), pp. 1296–1311.

[45] Maria A Rocca et al. “Current and future role of MRI in the diagnosis and prognosis of multiple sclerosis”. In: *The Lancet Regional Health–Europe* 44 (2024).

[46] Elizabeth N. York et al. “Quantitative magnetization transfer imaging in relapsing-remitting multiple sclerosis: a systematic review and meta-analysis”. In: *Brain Communications* 4.2 (2022), fcac088.

[47] John G. Sled. “Modelling and interpretation of magnetization transfer imaging in the brain”. In: *NeuroImage* 182 (2018), pp. 128–135.

[48] Steven D. Wolff and Robert S. Balaban. “Magnetization transfer contrast (MTC) and tissue water proton relaxation in vivo”. In: *Magnetic Resonance in Medicine* 10.1 (1989), pp. 135–144.

[49] Klaus Schmierer et al. “Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain”. In: *Annals of Neurology* 56.3 (2004), pp. 407–415.

[50] Klaus Schmierer et al. “Quantitative magnetic resonance of postmortem multiple sclerosis brain before and after fixation”. In: *Magnetic Resonance in Medicine* 59.2 (2008), pp. 268–277.

[51] Cornelia Laule et al. “Magnetic Resonance Imaging of Myelin”. In: *Neurotherapeutics* 4.3 (2007), pp. 460–484.

[52] Jacqueline Tien-Hsiang Chen et al. “Clinically feasible MTR is sensitive to cortical demyelination in MS”. In: *Neurology* 80.3 (2013), pp. 246–252.

[53] Reza Rahmanzadeh et al. “Myelin and axon pathology in multiple sclerosis assessed by myelin water and multi-shell diffusion imaging”. In: *Brain* 144.6 (2021), pp. 1684–1696.

[54] James Kolasinski et al. “A combined post-mortem magnetic resonance imaging and quantitative histological study of multiple sclerosis pathology”. In: *Brain* 135.10 (2012), pp. 2938–2951.

[55] M. Inglese and Maxim Bester. “Diffusion imaging in multiple sclerosis: research and clinical implications”. In: *NMR in Biomedicine* 23.7 (2010), pp. 865–872.

[56] Klaus Schmierer et al. “Diffusion tensor imaging of post mortem multiple sclerosis brain”. In: *NeuroImage* 35.2 (2007), pp. 467–477.

[57] M. Filippi et al. “Diffusion tensor magnetic resonance imaging in multiple sclerosis”. In: *Neurology* 56.3 (2001), pp. 304–311.

[58] Cui Ci Voon et al. “Quantitative susceptibility mapping in multiple sclerosis: a systematic review and meta-analysis”. In: *NeuroImage: Clinical* (2024), p. 103598.

[59] A.L. MacKay et al. “MR Relaxation in Multiple Sclerosis”. In: *Neuroimaging Clinics of North America* 19.1 (2009), pp. 1–26.

[60] Hai-Ling Margaret Cheng et al. “Practical medical applications of quantitative MR relaxometry”. In: *Journal of Magnetic Resonance Imaging* 36.4 (2012), pp. 805–824.

[61] Ibrahim Khormi et al. “MR myelin imaging in multiple sclerosis: A scoping review”. In: *Journal of the Neurological Sciences* 455 (2023), p. 122807.

[62] Alberto Lazari and Ilona Lipp. “Can MRI measure myelin? Systematic review, qualitative assessment, and meta-analysis of studies validating microstructural imaging with myelin histology”. In: *NeuroImage* 230 (2021), p. 117744.

[63] Tobias Leutritz et al. “Multiparameter mapping of relaxation (R1, R2*), proton density and magnetization transfer saturation at 3 T: A multicenter dual-vendor reproducibility and repeatability study”. In: *Human brain mapping* 41.15 (2020), pp. 4232–4247.

[64] Theodoros Ladopoulos et al. “Relaxometry and brain myelin quantification with synthetic MRI in MS subtypes and their associations with spinal cord atrophy”. In: *NeuroImage: Clinical* 36 (2022), p. 103166.

[65] JP Mottershead et al. “High field MRI correlates of myelin content and axonal density in multiple sclerosis: a post-mortem study of the spinal cord”. In: *Journal of neurology* 250 (2003), pp. 1293–1301.

[66] Nikolaus Weiskopf et al. “Quantitative magnetic resonance imaging of brain anatomy and in vivo histology”. In: *Nature Reviews Physics* 3.8 (2021), pp. 570–588.

[67] Riccardo Galbusera et al. “Postmortem quantitative MRI disentangles histological lesion types in multiple sclerosis”. In: *Brain Pathology* 33.6 (2023), e13136.

[68] Emilie Lommers et al. “Multiparameter MRI quantification of microstructural tissue alterations in multiple sclerosis”. In: *NeuroImage: Clinical* 23 (2019), p. 101879.

[69] Francesco Manfredonia et al. “Normal-appearing brain t1 relaxation time predicts disability in early primary progressive multiple sclerosis”. In: *Archives of neurology* 64.3 (2007), pp. 411–415.

[70] Jonathan O’Muircheartaigh et al. “Quantitative neuroimaging measures of myelin in the healthy brain and in multiple sclerosis”. In: *Human Brain Mapping* 40.7 (2019), pp. 2104–2116.

[71] Giuseppe Pontillo et al. “Clinical correlates of R1 relaxometry and magnetic susceptibility changes in multiple sclerosis: a multi-parameter quantitative MRI study of brain iron and myelin”. In: *European Radiology* 33.3 (2023), pp. 2185–2194.

[72] Hugo Vrenken et al. “Whole-brain T1 mapping in multiple sclerosis: global changes of normal-appearing gray and white matter”. In: *Radiology* 240.3 (2006), pp. 811–820.

[73] Jan-Hein TM Van Waesberghe et al. “Axonal loss in multiple sclerosis lesions: magnetic resonance imaging insights into substrates of disability”. In: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 46.5 (1999), pp. 747–754.

[74] James G Harper et al. “Quantitative T 1 brain mapping in early relapsing-remitting multiple sclerosis: longitudinal changes, lesion heterogeneity and disability”. In: *European Radiology* 34.6 (2024), pp. 3826–3839.

[75] Ian J Tagge et al. “Permanent tissue damage in multiple sclerosis lesions is associated with reduced pre-lesion myelin and axon volume fractions”. In: *Multiple Sclerosis Journal* 28.13 (2022), pp. 2027–2037.

[76] Christian Thaler et al. “T1 relaxation times in the cortex and thalamus are associated with working memory and information processing speed in patients with multiple sclerosis”. In: *Frontiers in Neurology* 12 (2021), p. 789812.

[77] Robert A Brown et al. “Deep learning segmentation of orbital fat to calibrate conventional MRI for longitudinal studies”. In: *Neuroimage* 208 (2020), p. 116442.

[78] Amanda F Mejia et al. “Statistical estimation of T1 relaxation times using conventional magnetic resonance imaging”. In: *NeuroImage* 133 (2016), pp. 176–188.

[79] Shachar Moskovich, Oshrat Shtangel, and Aviv A. Mezer. “Approximating R1 and R2: A Quantitative Approach to Clinical Weighted MRI”. In: *Human Brain Mapping* 45 (2024), e70102.

[80] Russell T Shinohara et al. “Statistical normalization techniques for magnetic resonance imaging”. In: *NeuroImage: Clinical* 6 (2014), pp. 9–19.

[81] Kareem A Wahid et al. “Intensity standardization methods in magnetic resonance imaging of head and neck cancer”. In: *Physics and imaging in radiation oncology* 20 (2021), pp. 88–93.

[82] Griffin Young et al. “T1 mapping from routine 3D T1-weighted inversion recovery sequences in clinical practice: comparison against reference inversion recovery fast field echo T1 scans and feasibility in multiple sclerosis”. In: *Neuroradiology* 66.10 (2024), pp. 1709–1719.

[83] Marco Ganzetti, Nicole Wenderoth, and Dante Mantini. “Whole brain myelin mapping using T1-and T2-weighted MR imaging data”. In: *Frontiers in human neuroscience* 8 (2014), p. 671.

[84] John H Gilmore et al. “Early postnatal development of corpus callosum and corticospinal white matter assessed with quantitative tractography”. In: *American Journal of Neuroradiology* 28.9 (2007), pp. 1789–1795.

[85] Jean-Philippe Fortin et al. “Removing inter-subject technical variability in magnetic resonance imaging studies”. In: *NeuroImage* 132 (2016), pp. 198–212.

[86] Rezwan Ghassemi et al. “Normalization of white matter intensity on T1-weighted images of patients with acquired central nervous system demyelination”. In: *Journal of Neuroimaging* 25.2 (2015), pp. 184–190.

[87] Adam Hasse et al. “Application of a novel T1 retrospective quantification using internal references (T1-REQUIRE) algorithm to derive quantitative T1 relaxation maps of the brain”. In: *International Journal of Imaging Systems and Technology* 32.6 (2022), pp. 1903–1915.

[88] Audrey Lavielle et al. “T1 mapping from MPRAGE acquisitions: application to the measurement of the concentration of nanoparticles in tumors for theranostic use”. In: *Journal of Magnetic Resonance Imaging* 58.1 (2023), pp. 313–323.

[89] Audrey Lavielle et al. “Quantitative brain T1 maps derived from T1-weighted MRI acquisitions: a proof-of-concept study”. In: *European Radiology Experimental* 8.1 (2024), p. 109.

[90] Kevin Lee et al. “Early postnatal myelin content estimate of white matter via T1w/T2w ratio”. In: *Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging*. Vol. 9417. SPIE. 2015, pp. 484–490.

[91] World Medical Association et al. “World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects”. In: *Jama* 310.20 (2013), pp. 2191–2194.

[92] Aaron Carass et al. “Longitudinal multiple sclerosis lesion segmentation: resource and challenge”. In: *NeuroImage* 148 (2017), pp. 77–102.

[93] Žiga Lesjak et al. “A novel public MR image dataset of multiple sclerosis patients with lesion segmentations based on multi-rater consensus”. In: *Neuroinformatics* 16 (2018), pp. 51–63.

[94] Olivier Commowick et al. “Multiple sclerosis lesions segmentation from multiple experts: The MICCAI 2016 challenge dataset”. In: *Neuroimage* 244 (2021), p. 118589.

[95] Matthew J Page et al. “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”. In: *bmj* 372 (2021).

[96] Jennifer E Soun et al. “Evaluation of neonatal brain myelination using the T1-and T2-weighted MRI ratio”. In: *Journal of Magnetic Resonance Imaging* 46.3 (2017), pp. 690–696.

[97] Christos P Loizou et al. “Brain MR image normalization in texture analysis of multiple sclerosis”. In: *2009 9th International Conference on Information Technology and Applications in Biomedicine*. IEEE. 2009, pp. 1–5.

[98] Muzamil Arshad, Jeffrey A Stanley, and Naftali Raz. “Test–retest reliability and concurrent validity of in vivo myelin content indices: Myelin water fraction and calibrated T1w/T2w image ratio”. In: *Human brain mapping* 38.4 (2017), pp. 1780–1790.

[99] Sarah Cappelle et al. “T1w/FLAIR ratio standardization as a myelin marker in MS patients”. In: *NeuroImage: Clinical* 36 (2022), p. 103248.

[100] Akifumi Hagiwara et al. “Myelin measurement: comparison between simultaneous tissue relaxometry, magnetization transfer saturation index, and T1w/T2w ratio methods”. In: *Scientific reports* 8.1 (2018), p. 10554.

[101] Salem Hannoun et al. “T1/T2 ratio: A quantitative sensitive marker of brain tissue integrity in multiple sclerosis”. In: *Journal of Neuroimaging* 32.2 (2022), pp. 328–336.

[102] Xiao Luo et al. “Application of T1-/T2-weighted ratio mapping to elucidate intracortical demyelination process in the Alzheimer’s disease continuum”. In: *Frontiers in Neuroscience* 13 (2019), p. 904.

[103] Kunio Nakamura et al. “T1-/T2-weighted ratio differs in demyelinated cortex in multiple sclerosis”. In: *Annals of neurology* 82.4 (2017), pp. 635–639.

[104] D Pareto et al. “Ratio of T1-weighted to T2-weighted signal intensity as a measure of tissue integrity: comparison with magnetization transfer ratio in patients with multiple sclerosis”. In: *American Journal of Neuroradiology* 41.3 (2020), pp. 461–463.

[105] Paolo Preziosa et al. “Neurite density explains cortical T1-weighted/T2-weighted ratio in multiple sclerosis”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 92.7 (2021), pp. 790–792.

[106] Ruthger Righart et al. “Cortical pathology in multiple sclerosis detected by the T 1/T 2-weighted ratio from routine magnetic resonance imaging”. In: *Annals of neurology* 82.4 (2017), pp. 519–529.

[107] Laetitia Saccenti et al. “Myelin measurement using quantitative magnetic resonance imaging: a correlation study comparing various imaging techniques in patients with multiple sclerosis”. In: *Cells* 9.2 (2020), p. 393.

[108] Takahiro Sanada et al. “Correlation of T1-to T2-weighted signal intensity ratio with T1-and T2-relaxation time and IDH mutation status in glioma”. In: *Scientific Reports* 12.1 (2022), p. 18801.

[109] Zahra Shams, David G Norris, and José P Marques. “A comparison of in vivo MRI based cortical myelin mapping using T1w/T2w and R1 mapping at 3T”. In: *PloS one* 14.7 (2019), e0218089.

[110] Jeong-Min Shim et al. “Quantitative myelin-related maps from R1 and T2* ratio images using a single ME-MP2RAGE sequence in 7T MRI”. In: *Frontiers in Neuroanatomy* 16 (2022), p. 950650.

[111] Md Nasir Uddin et al. “Can T1w/T2w ratio be used as a myelin-specific measure in subcortical structures? Comparisons between FSE-based T1w/T2w ratios, GRASE-based T1w/T2w ratios and multi-echo GRASE-based myelin water fractions”. In: *NMR in Biomedicine* 31.3 (2018), e3868.

[112] Md Nasir Uddin et al. “Comparisons between multi-component myelin water fraction, T1w/T2w ratio, and diffusion tensor imaging measures in healthy human brain structures”. In: *Scientific reports* 9.1 (2019), p. 2500.

[113] Marlee M Vandewouw et al. “Altered myelin maturation in four year old children born very preterm”. In: *NeuroImage: Clinical* 21 (2019), p. 101635.

[114] Shota Yamamoto et al. “Prediction and visualization of non-enhancing tumor in glioblastoma via T1w/T2w-ratio map”. In: *Brain Sciences* 12.1 (2022), p. 99.

[115] Fumihiko Yasuno et al. “Use of T1-weighted/T2-weighted magnetic resonance ratio to elucidate changes due to amyloid β accumulation in cognitively normal subjects”. In: *NeuroImage: Clinical* 13 (2017), pp. 209–214.

[116] Yufan Zheng et al. “Sensitivity of T1/T2-weighted ratio in detection of cortical demyelination is similar to magnetization transfer ratio using post-mortem MRI”. In: *Multiple Sclerosis Journal* 28.2 (2022), pp. 198–205.

[117] Emmanuelle Leray et al. “Evidence for a two-stage disability progression in multiple sclerosis”. In: *Brain* 133.7 (2010), pp. 1900–1913.

[118] Ronja C Berg et al. “Multi-parameter quantitative mapping of R1, R2*, PD, and MTsat is reproducible when accelerated with Compressed SENSE”. In: *Neuroimage* 253 (2022), p. 119092.

[119] KB De Raad et al. “The effect of preprocessing on convolutional neural networks for medical image segmentation”. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2021, pp. 655–658.

[120] Bjoern Menze et al. “Analyzing magnetic resonance imaging data from glioma patients using deep learning”. In: *Computerized medical imaging and graphics* 88 (2021), p. 101828.

[121] Paul A Yushkevich et al. “IC-P-174: fast automatic segmentation of hippocampal subfields and medial temporal lobe subregions In 3 Tesla and 7 Tesla T2-Weighted MRI”. In: *Alzheimer’s & Dementia* 12 (2016), P126–P127.

[122] Paul A Yushkevich et al. “User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability”. In: *Neuroimage* 31.3 (2006), pp. 1116–1128.

[123] Florian Kofler et al. “Brats toolkit: translating brats brain tumor segmentation algorithms into clinical and scientific practice”. In: *Frontiers in neuroscience* 14 (2020), p. 125.

[124] Sarthak Pati et al. “Federated learning enables big data for rare cancer boundary detection”. In: *Nature communications* 13.1 (2022), p. 7346.

[125] Fabian Isensee et al. “Automated brain extraction of multisequence MRI using artificial neural networks”. In: *Human brain mapping* 40.17 (2019), pp. 4952–4964.

[126] Özgün Çiçek et al. “3D U-Net: learning dense volumetric segmentation from sparse annotation”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*. Springer. 2016, pp. 424–432.

[127] Liwei Wang et al. “Training deeper convolutional networks with deep supervision”. In: *arXiv preprint arXiv:1505.02496* (2015).

[128] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. “Tversky loss function for image segmentation using 3D fully convolutional deep networks”. In: *International workshop on machine learning in medical imaging*. Springer. 2017, pp. 379–387.

[129] Michela Antonelli et al. “The medical segmentation decathlon”. In: *Nature communications* 13.1 (2022), p. 4128.

[130] Alexander Hapfelmeier et al. “Retrospective cohort study to devise a treatment decision score predicting adverse 24-month radiological activity in early multiple sclerosis”. In: *Therapeutic Advances in Neurological Disorders* 16 (2023), p. 17562864231161892.

[131] Christian Gaser et al. “CAT: a computational anatomy toolbox for the analysis of structural MRI data”. In: *Gigascience* 13 (2024), giae049.

[132] Karsten Tabelow et al. “hMRI–A toolbox for quantitative MRI in neuroscience and clinical research”. In: *Neuroimage* 194 (2019), pp. 191–210.

[133] C Preibisch and R Deichmann. “Influence of RF spoiling on the stability and accuracy of T1 mapping based on spoiled FLASH with varying flip angles”. In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 61.1 (2009), pp. 125–135.

[134] Simon Baudrexel et al. “T1 mapping with the variable flip angle technique: a simple correction for insufficient spoiling of transverse magnetization”. In: *Magnetic resonance in medicine* 79.6 (2018), pp. 3082–3092.

[135] John Ashburner and Karl J Friston. “Unified segmentation”. In: *neuroimage* 26.3 (2005), pp. 839–851.

[136] Vasily L Yarnykh. “Actual flip-angle imaging in the pulsed steady state: a method for rapid three-dimensional mapping of the transmitted radiofrequency field”. In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 57.1 (2007), pp. 192–200.

[137] Bruce Fischl. “FreeSurfer”. In: *Neuroimage* 62.2 (2012), pp. 774–781.

[138] Oula Puonti, Juan Eugenio Iglesias, and Koen Van Leemput. “Fast and sequence-adaptive whole-brain segmentation using parametric Bayesian modeling”. In: *NeuroImage* 143 (2016), pp. 235–249.

[139] Oula Puonti et al. “Accurate and robust whole-head segmentation from magnetic resonance images for individualized head modeling”. In: *Neuroimage* 219 (2020), p. 117044.

[140] Andriy Fedorov et al. “3D Slicer as an image computing platform for the Quantitative Imaging Network”. In: *Magnetic Resonance Imaging* 30 (2012). Quantitative Imaging in Cancer, pp. 1323–1341. DOI: https://doi.org/10.1016/j.mri.2012.05.001.

[141] Nicholas J Tustison et al. “N4ITK: improved N3 bias correction”. In: *IEEE transactions on medical imaging* 29.6 (2010), pp. 1310–1320.

[142] Ralf Deichmann and Axel Haase. “Quantification of T1 values by SNAPSHOT-FLASH NMR imaging”. In: *Journal of Magnetic Resonance (1969)* 96.3 (1992), pp. 608–612.

[143] Ralf Deichmann et al. “Optimization of 3-D MP-RAGE sequences for structural brain imaging”. In: *Neuroimage* 12.1 (2000), pp. 112–127.

[144] Benoît Vernier et al. “MRI contrast enhancement of magnetization prepared steady state sequence: an optimal control framework”. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2021, pp. 1709–1713.

[145] Axel Haase. “Snapshot FLASH MRI. Applications to T1, T2, and chemical-shift imaging”. In: *Magnetic resonance in medicine* 13.1 (1990), pp. 77–89.

[146] Michael Brant-Zawadzki, Gary D Gillan, and Wolfgang R Nitz. “MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence–initial experience in the brain.” In: *Radiology* 182.3 (1992), pp. 769–775.

[147] Jorge Zavala Bojorquez et al. “What are normal relaxation times of tissues at 3 T?” In: *Magnetic resonance imaging* 35 (2017), pp. 69–80.

[148] K. Van Leemput et al. “Automated segmentation of multiple sclerosis lesions by model outlier detection”. In: *IEEE Transactions on Medical Imaging* 20.8 (2001), pp. 677–688.

[149] Huahong Zhang and Ipek Oguz. “Multiple Sclerosis Lesion Segmentation - A Survey of Supervised CNN-Based Methods”. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Ed. by Alessandro Crimi and Spyridon Bakas. Springer International Publishing, 2021, pp. 11–29. ISBN: 978-3-030-72084-1.

[150] Benjamin Billot et al. “Joint Segmentation Of Multiple Sclerosis Lesions And Brain Anatomy In MRI Scans Of Any Contrast And Resolution With CNNs”. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. 2021, pp. 1971–1974.

[151] Xinxin Li et al. “White matter hyperintensities segmentation using an ensemble of neural networks”. In: *Human Brain Mapping* 43.3 (2022), pp. 929–939.

[152] Mladen Rakić et al. “icobrain ms 5.1: Combining unsupervised and supervised approaches for improving the detection of multiple sclerosis lesions”. In: *NeuroImage: Clinical* 31 (2021), p. 102707.

[153] Ludovica Griffanti et al. “BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities”. In: *Neuroimage* 141 (2016), pp. 191–205.

[154] Viola Pongratz et al. “Lesion location across diagnostic regions in multiple sclerosis”. In: *NeuroImage: Clinical* 37 (2023), p. 103311.

[155] Marco Ganzetti, Nicole Wenderoth, and Dante Mantini. “Mapping pathological changes in brain structure by combining T1-and T2-weighted MR imaging data”. In: *Neuroradiology* 57 (2015), pp. 917–928.

[156] Poljanka Johnson et al. “Myelin heterogeneity for assessing normal appearing white matter myelin damage in multiple sclerosis”. In: *Journal of Neuroimaging* 33.2 (2023), pp. 227–234.

[157] Hugh Kearney, David H Miller, and Olga Ciccarelli. “Spinal cord MRI in multiple sclerosis—diagnostic, prognostic and clinical value”. In: *Nature Reviews Neurology* 11.6 (2015), pp. 327–338.

[158] Matthias Bussas et al. “Multiple sclerosis lesions and atrophy in the spinal cord: distribution across vertebral levels and correlation with disability”. In: *NeuroImage: Clinical* 34 (2022), p. 103006.

[159] Markus Lauerer et al. “Prognostic value of spinal cord lesion measures in early relapsing-remitting multiple sclerosis”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 95.1 (2024), pp. 37–43.

[160] A Traboulsee et al. “Disability in multiple sclerosis is related to normal appearing brain tissue MTR histogram abnormalities”. In: *Multiple Sclerosis Journal* 9.6 (2003), pp. 566–573.

[161] Zhaleh Khaleeli et al. “Magnetisation transfer ratio in the normal appearing white matter predicts progression of disability over 1 year in early primary progressive multiple sclerosis”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 78.10 (2007), pp. 1076–1082.

[162] Frederik Barkhof. “MRI in multiple sclerosis: correlation with expanded disability status scale (EDSS)”. In: *Multiple Sclerosis Journal* 5.4 (1999), pp. 283–286.

[163] Frederik Barkhof. “The clinico-radiological paradox in multiple sclerosis revisited”. In: *Current opinion in neurology* 15.3 (2002), pp. 239–245.

[164] MA Sahraian et al. “Black holes in multiple sclerosis: definition, evolution, and clinical correlations”. In: *Acta Neurologica Scandinavica* 122.1 (2010), pp. 1–8.

[165] L Truyen et al. “Accumulation of hypointense lesions (" black holes") on T1 spin-echo MRI correlates with disease progression in multiple sclerosis”. In: *Neurology* 47.6 (1996), pp. 1469–1476.

[166] Christian Thaler et al. “T1-thresholds in black holes increase clinical-radiological correlation in multiple sclerosis patients”. In: *PLoS One* 10.12 (2015), e0144693.

[167] Maria A Rocca, Giancarlo Comi, and Massimo Filippi. “The role of T1-weighted derived measures of neurodegeneration for assessing disability progression in multiple sclerosis”. In: *Frontiers in neurology* 8 (2017), p. 433.

[168] Caterina Lapucci et al. “Degree of microstructural changes within T1-SE versus T1-GE hypointense lesions in multiple sclerosis: relevance for the definition of “black holes””. In: *European Radiology* 30 (2020), pp. 3843–3851.

[169] C Thaler et al. “T1 recovery is predominantly found in black holes and is associated with clinical improvement in patients with multiple sclerosis”. In: *American Journal of Neuroradiology* 38.2 (2017), pp. 264–269.

[170] Simona Schiavi et al. “Non-invasive quantification of inflammation, axonal and myelin injury in multiple sclerosis”. In: *Brain* 144.1 (2021), pp. 213–223.

List of publications

- **Wiltgen T.**, McGinnis J., Berg R., Voon C., Puonti O., Giglhuber K., ... & Mühlau M. (2025). Towards quantitative intensity analysis of conventional T1-weighted images in multiple sclerosis. NeuroImage, 318, 121395, doi: 10.1016/j.neuroimage.2025.121395

- **Wiltgen T.**, McGinnis J., Schläger S., Kofler F., Voon C., Berthele A., ... & Mühlau M. and Wiestler B. (2024). LST-AI: A deep learning ensemble for accurate MS lesion segmentation. NeuroImage: Clinical, 42, 103611, doi: 10.1016/j.nicl.2024.103611

- **Wiltgen T.**, Voon C., Van Leemput K., Wiestler B., Mühlau M. (2024). Intensity scaling of conventional brain magnetic resonance images avoiding cerebral reference regions: A systematic review. PloS one, 19(3), e0298642, doi: 10.1371/journal.pone.0298642

- Gasperi C., **Wiltgen T.**, McGinnis J., Cerri S., Moridi T., Ouellette R., ... & Mühlau M. (2023). A genetic risk variant for multiple sclerosis severity is associated with brain atrophy. Annals of neurology, 94(6), 1080-1085, doi: 10.1002/ana.26807

- Lauerer M., **Wiltgen T.**, Brückner C., Engl C., Giglhuber K., Lambrecht S., ... & Mühlau M. (2025). Predictors of early disability accumulation in newly diagnosed multiple sclerosis: clinical, imaging and cerebrospinal fluid measures. Journal of Neurology, Neurosurgery & Psychiatry, doi: 10.1136/jnnp-2024-335037

- Voon C., **Wiltgen T.**, Wiestler B., Schläger S., Mühlau M. (2024). Quantitative susceptibility mapping in multiple sclerosis: A systematic review and meta-analysis. NeuroImage: Clinical, 42, 103598, doi: 10.1016/j.nicl.2024.103598

- Romahn E. F., **Wiltgen T.**, Bussas M., Aly L., Wicklein R., Noll C., ... & Knier B. (2023). Association of retinal vessel pathology and brain atrophy in relapsing-remitting multiple sclerosis. Frontiers in Immunology, 14, 1284986, doi: 10.3389/fimmu.2023.1284986

- Tahedl M., **Wiltgen T.**, Voon C., Berthele A., Kirschke J. S., Hemmer B., ... & Wiestler B. (2024). Cortical Thin Patch Fraction Reflects Disease Burden in MS: The Mosaic Approach. American Journal of Neuroradiology, 45(1), 82-89, doi: 10.3174/ajnr.A8064

- Tahedl M., **Wiltgen T.**, Voon C., Berthele A., Kirschke J. S., Hemmer B., ... & Wiestler B. (2023). Benefits of a mosaic approach for assessing cortical atrophy in individual multiple sclerosis patients. Brain and Behavior, 13(12), e3327, doi: 10.1002/brb3.3327

- Hess F., McGinnis J., Baki E., **Wiltgen T.**, Müller A., Maegerlein C., ... & Mühlau M. (2025). Predictors and implications of myocardial injury in intracerebral hemorrhage. Clinical Neuroradiology, 1-10, doi: 10.1007/s00062-025-01498-4

Appendix

MPRAGE signal modeling

The MPRAGE signal can theoretically be described by an equation that includes sequence parameters and physical properties of the underlying tissue (i.e., relaxation times). In this thesis, the signal equation was derived based on the technical specifications of the applied sequence and on information from literature [142–146]. The T1w MPRAGE sequence diagram is presented in the main text in Figure 2.4.

In general, the longitudinal magnetization recovery is governed by T1 relaxation and evolves over time according to:

$$M(t) = M(0) * e^{-\frac{t}{T1}} + M_0 * (1 - e^{-\frac{t}{T1}}) \tag{1}$$

Here, M_0 is the acquisition-specific equilibrium magnetization, and $M(0)$ designates the longitudinal magnetization immediately after an RF pulse with flip angle α:

$$M(0) = M_{pre-RF}(0) * cos(\alpha), \tag{2}$$

The acquisition of the MPRAGE image starts once the steady-state equilibrium magnetization is reached, in which the following holds: $M_i^{eq} = M_{i-1}^{eq}$ (for the i^{th} MPRAGE block).

These equations can be used to derive the MPRAGE signal equation. They describe the behavior of the longitudinal magnetization in each part of the MPRAGE blocks. An MPRAGE block starts with a 180° inversion pulse followed by an inversion time TI until the rapid gradient echo (RAGE) block starts. A RAGE block comprises N phase encoding steps, with each step involving a low-angle RF pulse followed by a gradient echo readout. The low-angle RF pulses in the RAGE block are separated by a repetition time TR. The recovery time T_{rec} after the RAGE block allows the longitudinal magnetization to reach the steady-state equilibrium again before the next MPRAGE block starts. The following substitutions are used for better readability:

$$E_{TR} = e^{-\frac{TR}{T1}}, E_{TI} = e^{-\frac{TI}{T1}}, E_{rec} = e^{-\frac{T_{rec}}{T1}}$$

1. The longitudinal magnetization after a 180° inversion pulse and an inversion time TI is given by (use equation (2) in equation (1)):

$$M_i^0 = -M_{i-1}^{eq} * E_{TI} + M_0 * (1 - E_{TI}), \qquad with \quad cos(180°) = -1$$

2. The RAGE block starts with a first phase encoding step applying a low-angle RF pulse (with the flip angle α), followed by the recovery time TR until the next low-angle RF pulse is applied. Again, combining equations (1) and (2), the longitudinal magnetization after the low-angle RF pulse and TR is described by:

$$M_i^1 = M_i^0 * cos(\alpha) E_{TR} + M_0 * (1 - E_{TR})$$

3. Similarly, the longitudinal magnetization after the following phase encoding steps are described by:

$$M_i^2 = M_i^1 * cos(\alpha) E_{TR} + M_0 * (1 - E_{TR}),$$

$$M_i^3 = M_i^2 * cos(\alpha) E_{TR} + M_0 * (1 - E_{TR}),$$

...

4. Recursive integration of these equations leads to the following expression of M_i^3:

$$M_i^3 = M_i^0 * [\cos(\alpha)E_{TR}]^3 + M_0 * (1 - E_{TR}) * \left[[\cos(\alpha)E_{TR}]^2 + \cos(\alpha)E_{TR} + 1\right]$$

5. For the N-th phase encoding step, repeated iteration yields the following equation:

$$M_i^N = M_i^0 * [\cos(\alpha)E_{TR}]^N + M_0 * (1 - E_{TR}) * \sum_{n=0}^{N-1} [cos(\alpha)E_{TR}]^n$$

6. Considering the geometric series, this can be written as:

$$M_i^N = M_i^0 * [\cos(\alpha)E_{TR}]^N + M_0 * (1 - E_{TR}) * \frac{1 - [cos(\alpha)E_{TR}]^N}{1 - cos(\alpha)E_{TR}}$$

7. With M_i^0 from step 1:

$$M_i^N =$$
$$M_0 * (1 - E_{TR}) * \frac{1 - [cos(\alpha)E_{TR}]^N}{1 - cos(\alpha)E_{TR}} - M_{i-1}^{eq} * E_{TI} * [cos(\alpha)E_{TR}]^N + M_0 * (1 - E_{TI}) * [cos(\alpha)E_{TR}]^N$$

8. After the RAGE block, which includes N_{total} phase encoding steps, the steady-state equilibrium magnetization is reached after the recovery time T_{rec}. Consequently, the steady-state at the end of the MPRAGE block is given by:

$$M_i^{eq} =$$
$$M_0 * (1 - E_{TR}) * \frac{1 - [cos(\alpha)E_{TR}]^{N_{total}}}{1 - cos(\alpha)E_{TR}} * E_{rec} - M_{i-1}^{eq} * E_{TI} * [cos(\alpha)E_{TR}]^{N_{total}} * E_{rec}$$
$$+M_0 * (1 - E_{TI}) * [cos(\alpha)E_{TR}]^{N_{total}} * E_{rec} + M_0 * (1 - E_{rec})$$

9. At steady state, the condition $M_i^{eq} = M_{i-1}^{eq}$ holds, yielding:

$$M^{eq} =$$
$$M_0 * \frac{1 - E_{rec} + (1 - E_{TR}) * \frac{1-[cos(\alpha)E_{TR}]^{N_{total}}}{1-cos(\alpha)E_{TR}} * E_{rec} + (1 - E_{TI}) * [cos(\alpha)E_{TR}]^{N_{total}} * E_{rec}}{1 + [cos(\alpha)E_{TR}]^{N_{total}} * E_{TI} * E_{rec}}$$

Modeling the signal of the i^{th} readout pulse $S_i = sin(\alpha) * M_i^N * e^{-\frac{TE}{T2^*}}$ depends on which step of the RAGE block samples the center of the k-space, which encodes the contrast of the image. As a result, when the center is sampled during the n-th step, the magnetization corresponding to that specific step is the one that should be taken into account (i.e., use N=n). The MPRAGE sequence applied to acquire the data in this thesis used a linear cartesian profile order for k-space filling. Therefore, $N = N_{total}/2$ should be used. Further MPRAGE parameters that were used and are required to model the signal are presented in the following table:

Parameter	**Value**
Shot interval	3000 ms
TI	1000 ms
TE	4 ms
TR	9 ms
FA	8°
N_{total}	48
T_{rec}	$shot\ interval - TI - N_{total} * TR = 3000 - 1000 - 48 * 9$ = 1568ms

Using the method described in the main text, the signal equation (S_i) was fitted to the image data. "The success of the fitting process was evaluated by calculating the R^2 and Normalized Root Mean Squared Error (NRMSE) for each image and reporting the mean with the standard deviation and the median with the interquartile range (IQR). We found a mean R^2 value of 0.91 $\pm$ 0.04, and a median of 0.92 (0.90 - 0.94). The NRMSE values showed a mean of 0.12 $\pm$ 0.03, and a median of 0.11 (0.10 - 0.13)" [3, Supplementary materials]. Hence, the fit was considered successful.

mediaTUM: Abstract

In multiple sclerosis, conventional MRI is primarily used for manual assessment of brain abnormalities, with quantitative analysis limited by arbitrary scales and poor interpretability of image intensities. This thesis introduces novel approaches for conventional MRI image analysis, including a deep learning-based tool for accurate and automated lesion segmentation, and intensity scaling methods for quantifying the biologically relevant information content in T1-weighted image intensities. These advances open new possibilities for MS research with conventional MRI.

mediaTUM: Zusammenfassung

In der MS dient die konventionelle MRT meist der manuellen Beurteilung von Hirnveränderungen, wobei quantitative Analysen durch willkürliche Skalen und geringe Interpretierbarkeit der Bildintensitäten eingeschränkt sind. Diese Arbeit stellt neue Ansätze vor, darunter ein Deep-Learning-basiertes Tool zur automatischen Läsionssegmentierung und Skalierungsmethoden zur Quantifizierung biologisch relevanter Informationen in T1-gewichteten Bildintensitäten. Diese Fortschritte eröffnen neue Möglichkeiten für die MS-Forschung mit konventioneller MRT.