NeuroImage: Clinical 36 (2022) 103205

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Cortical lesions, central vein sign, and paramagnetic rim lesions in multiple sclerosis: Emerging machine learning techniques and future avenues

Francesco La Rosa [a,b,c,*], Maxence Wynen [b,d,e,f], Omar Al-Louzi [g,h], Erin S Beck [c,g], Till Huelnhagen [a,f,i], Pietro Maggi [e,j,k], Jean-Philippe Thiran [a,b,f], Tobias Kober [a,f,i], Russell T Shinohara [l,m,n], Pascal Sati [g,h], Daniel S Reich [g], Cristina Granziera [o,p], Martina Absinta [q,r], Meritxell Bach Cuadra [b,f]

[a] Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[b] CIBM Center for Biomedical Imaging, Switzerland
[c] Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[d] ICTeam, UCLouvain, Louvain-la-Neuve, Belgium
[e] Louvain Inflammation Imaging Lab (NIL), Institute of Neuroscience (IoNS), UCLouvain, Brussels, Belgium
[f] Radiology Department, Lausanne University and University Hospital, Switzerland
[g] Translational Neuroradiology Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA
[h] Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, USA
[i] Advanced Clinical Imaging Technology, Siemens Healthcare AG, Lausanne, Switzerland
[j] Department of Neurology, Cliniques universitaires Saint-Luc, Université catholique de Louvain, Brussels, Belgium
[k] Department of Neurology, CHUV, Lausanne, Switzerland
[l] Center for Biomedical Image Computing and Analysis (CBICA), Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA
[m] Penn Statistics in Imaging and Visualization Endeavor (PennSIVE), Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA, USA
[n] Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[o] Translational Imaging in Neurology (ThINk) Basel, Department of Biomedical Engineering, Faculty of Medicine, University Hospital Basel and University of Basel, Switzerland
[p] Neurologic Clinic and Policlinic, MS Center and Research Center for Clinical Neuroimmunology and Neuroscience Basel (RC2NB), University Hospital Basel and University of Basel, Basel, Switzerland
[q] IRCCS San Raffaele Hospital and Vita-Salute San Raffaele University, Milan, Italy
[r] Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

ABSTRACT

The current diagnostic criteria for multiple sclerosis (MS) lack specificity, and this may lead to misdiagnosis, which remains an issue in present-day clinical practice. In addition, conventional biomarkers only moderately correlate with MS disease progression. Recently, some MS lesional imaging biomarkers such as cortical lesions (CL), the central vein sign (CVS), and paramagnetic rim lesions (PRL), visible in specialized magnetic resonance imaging (MRI) sequences, have shown higher specificity in differential diagnosis. Moreover, studies have shown that CL and PRL are potential prognostic biomarkers, the former correlating with cognitive impairments and the latter with early disability progression. As machine learning-based methods have achieved extraordinary performance in the assessment of conventional imaging biomarkers, such as white matter lesion segmentation, several automated or semi-automated methods have been proposed as well for CL, PRL, and CVS. In the present review, we first introduce these MS biomarkers and their imaging methods. Subsequently, we describe the corresponding machine learning-based methods that were proposed to tackle these clinical questions, putting them into context with respect to the challenges they are facing, including non-standardized MRI protocols, limited datasets, and moderate inter-rater variability. We conclude by presenting the current limitations that prevent their broader deployment and suggesting future research directions.

Abbreviations: MS, multiple sclerosis; MRI, magnetic resonance imaging; DL, deep learning; ML, machine learning; CL, cortical lesions; PRL, paramagnetic rim lesions; CVS, central vein sign; WML, white matter lesions; FLAIR, fluid-attenuated inversion recovery; MPRAGE, magnetization prepared rapid gradient-echo; GM, gray matter; WM, white matter; PSIR, phase-sensitive inversion recovery; DIR, double inversion recovery; MP2RAGE, magnetization-prepared 2 rapid gradient echoes; SELs, Slowly evolving/expanding lesions; CNN, convolutional neural network; XAI, explainable AI; PV, partial volume.

* Corresponding author at: Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
E-mail address: francesco.larosa@mssm.edu (F. La Rosa).

https://doi.org/10.1016/j.nicl.2022.103205
Received 13 December 2021; Received in revised form 9 September 2022; Accepted 16 September 2022
Available online 24 September 2022
2213-1582/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Multiple sclerosis (MS) is a chronic inflammatory disease and a common cause of neurological disability in young adults (Reich et al., 2018). Its hallmark is demyelinated white matter lesions (WML) forming in the central nervous system (Reich et al., 2018). These lesions are assessed in-vivo with magnetic resonance imaging (MRI), which is the imaging technique of choice to diagnose MS and monitor the disease over time (Hemond and Bakshi, 2018). The current MRI diagnostic criteria (McDonald criteria) are based on the dissemination in space and time of such lesions (Thompson et al., 2018). Moreover, the quantification of the total lesion volume is important to determine ongoing disease activity and monitor treatment effect over time (Giorgio et al., 2014). Recommended MRI techniques include T2 and T1-weighted inversion recovery sequences, such as fluid-attenuated inversion recovery (FLAIR), and magnetization prepared rapid gradient-echo (MPRAGE) (Wattjes et al., 2021). At common clinical magnetic fields (1.5 T and 3 T), the use of gadolinium-based contrast agents is useful to evaluate patients suspected of MS and monitor disease activity causing breakdown of the blood–brain barrier (Filippi et al., 2019).

As the manual detection of WML is time-consuming and prone to inter-rater variability (Hagens et al., 2019), a myriad of automated or semi-automated approaches have been developed to facilitate this task (Lladó et al., 2012). These methods were initially based primarily on MRI intensity features and probabilistic atlases (Lladó et al., 2012), whereas, more recently, the vast majority use deep learning (DL) approaches (Zeng et al., 2020), without prior feature extraction. Substantial effort is now being made towards reproducibility of the results and open science (Vrenken et al., 2021). Several grand challenges have been organized (Carass et al., 2017; Commowick et al., 2018; Commowick et al., 2021), in which DL-based methods have achieved the best performance, approaching or sometimes even outperforming human readers (Carass et al., 2017; Commowick et al., 2021). WML segmentation methods have been reviewed recently (Zeng et al., 2020; Kaur et al., 2021); the present review thus focuses on machine learning techniques tailored for lesional biomarkers specific to MS that require advanced MRI techniques and have the potential to improve MS diagnosis and prognosis.

One major drawback of the current MS diagnostic criteria is their lack of specificity, as they were proposed to identify patients with a high likelihood of MS rather than distinguish MS from other conditions (Thompson et al., 2018). The lack of specificity of these criteria may lead to misdiagnosis, which remains a persistent problem of MS (Solomon et al., 2019). Multi-center studies have shown a misdiagnosis rate of 18% (Kaisey et al., 2019), often associated with atypical clinical or imaging findings. Improving the diagnostic specificity would prevent harmful consequences for patients (Solomon et al., 2016) and allow clinicians to prescribe the appropriate treatment earlier. In addition, although clinical relapses are often associated with the appearance of new WML, the overall WML burden, which is the most common MRI biomarker examined in clinical routine, is only moderately correlated with disability and poorly predicts transition to progressive disease (Barkhof, 2002). For all these reasons, there is a need for additional biomarkers that are highly specific to MS or correlate with disease progression.

Quantitative MRI, such as relaxometry, myelin imaging, or diffusion MR, provides information related to the microstructural composition and organization of tissues. In MS, quantitative MRI techniques complement conventional MRI techniques by providing insights into disease mechanisms (Granziera et al., 2021). For instance, diffusion tensor imaging and microstructure models of diffusion can help better understand the MS lesion heterogeneity (myelin and axonal damage). Voxel-wise analysis methods allow exploring group-wise differences without the need for prior lesion segmentation (Thaler et al., 2021;16(2): e0245844.). On the contrary, classification methods in this context have been used to cluster different lesion types based on prior lesion segmentation and derived scalar measurements from diffusion-based measurements (FA, MD, NODDI parameters, etc) averaged at the lesion level (Lu et al., 2021; Oladosu et al., 2021; Ye et al., 2020; Martínez-Heras et al., 2020). Further studies, however, are still needed to verify the possible use of these quantitative features for patient stratification.

Recently, advances in MR technology, such as the development of specialized sequences, acceleration of protocols, and the proliferation of ultra-high field MRI, have allowed the imaging of pathologically specific MS lesional biomarkers (Cortese et al., 2019; Ineichen et al., 2021). These include cortical lesions (CL), the central vein sign (CVS), and paramagnetic rim lesions (PRL). Studies have shown that CL and PRL are potential prognostic biomarkers: CL are associated with cognitive impairments, while patients with PRL experience an earlier progression in disability (Calabrese et al., 2010; Absinta et al., 2019). Furthermore, the CVS and PRL have proven to be effective for differentiating MS from mimicking diseases (Ontaneda et al., 2021; Sati et al., 2016; Clarke et al., 2020; Maggi et al., 2020). All three biomarkers, however, require dedicated MRI sequences at high (3 T) or ultra-high (7 T) magnetic fields, and experienced raters for their manual assessment, which can be very time-consuming. As done in the past for WML, various automated or semi-automated methods, mostly based on machine learning (ML), have been developed to facilitate the three aforementioned biomarkers' assessment (see Table 1). Compared to their WML counterparts, however, they face additional challenges, including non-standardized imaging protocols, moderate inter-rater variability when determining ground truth annotations, and smaller datasets. Automated assessment could improve standardization and facilitate large-scale assessment in clinical routine of the aforementioned biomarkers, with clear benefits in terms of MS diagnosis and prognosis.

In this review, we first briefly describe these advanced imaging biomarkers and their imaging requirements and then focus on image processing techniques tailored for their automated segmentation and classification. We conclude with a discussion on current limitations and future lines of research to boost the development of ML approaches in this area and encourage their adoption in MS research and clinical settings.

2. Cortical lesions, paramagnetic lesions, and central vein sign

In this section, we present a brief description of CL, CVS, and PRL, and their respective imaging protocols. In addition to the CVS and PRL, which have emerged as promising MS biomarkers in recent years, we also included CL which, although included in the MS diagnostic criteria, are not yet commonly analyzed in clinical practice. For the sake of completeness, a short description of slowly expanding lesions (SELs) is also provided, although these have not been assessed with ML-based approaches yet.

Cortical lesions (CL) **-** Cortical lesions are a type of MS lesions that involve, at least partially, the cortex and have been classified into three main categories (Calabrese et al., 2010) (see Fig. 1): leukocortical lesions are located at the interface between WM and gray matter (GM) (type I), intracortical lesions are purely in the cortex and do not reach the pial surface (type II), and subpial lesions touch the subpial surface of the brain (type III) and sometimes extend all the way to the white matter (type IV). Cortical demyelination in MS has long been recognized in pathology studies, but only in the last two decades have dedicated sequences on high- and ultra-high field scanners provided in-vivo evidence of cortical damage (Calabrese et al., 2010). Cortical lesions are clinically interestingfor several reasons. First, they have been observed in the early stages of the disease and in all of the major MS phenotypes (Kidd et al., 1999). Second, they are associated with disability (Harrison et al., 2015; Nielsen et al., 2013; Calabrese et al., 2012) and in some studies, their number was associated with cognitive disability more strongly than the number of WML (Harrison et al., 2015; Favaretto et al., 2016). Third, longitudinal studies have linked them with disease progression

(Treaba et al., 2019; Mainero et al., 2015; Scalfari et al., 2018; Calabrese et al., 2013). Fourth, subpial cortical demyelination is highly specific to MS (Junker et al., 2020); CL have been observed in patients with radiologically isolated syndrome (Giorgio et al., 2011), but not in patients with neuromyelitis optica (Sinnecker et al., 2012). Since 2017, CL have been included in the MS diagnostic criteria (Thompson et al., 2018), but their visualization from routine MRI sequences remains difficult. For instance, a postmortem study showed that 3D FLAIR at 3 T could detect about 41% of leukocortical lesions and only 5% of intracortical and subpial lesions (Geurts et al., 2005). This supports the need for specialized MRI techniques (see Fig. 2) such as the phase-sensitive inversion recovery (PSIR), double inversion recovery (DIR), and magnetization-prepared 2 rapid gradient echoes (MP2RAGE) (Filippi et al., 2019; Müller et al., 2022). However, these sequences are still relatively insensitive to CL at 1.5 T and 3 T (Müller et al., 2022; Kilsdonk et al., 2016; Beck et al., 2020). Ultra-high field MRI, with its higher signal-to-noise ratio and increased susceptibility effects, has proven to be a powerful tool for increasing the sensitivity to CL, especially for intracortical and subpial lesions (Madsen et al., 2021; Beck et al., 2022; Maranzano et al., 2019). Even with the most sensitive methods, however, CL are small and often subtle, making manual segmentation extremely time consuming and subject to relatively low inter-rater reliability (Harrison et al., 2015; Faizy et al., 2017).

Central vein sign (CVS) **-** Recently, studies have suggested that an MRI-detectable central vein inside MS lesions might be evidence of pathological processes specific to MS (see Fig. 3) (Maggi et al., 2018; Solomon et al., 2016). This marker, referred to as the "central vein sign," has gained attention in recent years, as it could help to differentiate MS from mimicking diseases (Sati et al., 2016; Sinnecker et al., 2019; Ciotti et al., 2022; Sparacia et al., 2018; Tranfa et al., 2022). Small cerebral veins can be detected with susceptibility-based MRI sequences, taking advantage of the magnetic properties of venous blood that is rich in deoxyhemoglobin (Haacke et al., 2009; Mittal et al., 2009). The CVS can be reliably observed across different T2* sequences at 3 T, although the sensitivity depends on the sequence considered (Samaraweera et al., 2017). To obtain the best detection sensitivity for the CVS, optimized MRI acquisitions have been proposed (T2*-weighted acquired with 3D-segmented echo-planar-imaging or T2*w 3D-*EPI* (Sati et al., 2014), combined T2-FLAIR and T2*, also called FLAIR* (Sati et al., 2012), and susceptibility-based sequence, called SWAN-Venule (Gaitán et al., 2020). These sequences were shown to provide superior CVS detection compared to clinical acquisitions at 1.5 T and 3 T (Castellaro et al., 2020; Suh et al., 2019). Single-center and multi-center retrospective studies imaging patients with clinically established diagnoses have demonstrated a significantly higher proportion of CVS-positive white matter lesions (%CVS +) in MS (mean pooled incidence: 79%, 95% CI: 68–87%) (Suh et al., 2019) as compared to other neurological disorders mimicking MS (mean pooled incidence: 38%, 95% CI: 18–63%) (Suh et al., 2019) such as cerebral small vessel disease (Campion et al., 2017), neuromyelitis optica spectrum disorder (NMOSD) (Cortese et al., 2018), inflammatory vasculopathies (Maggi et al., 2018), and migraine (Solomon et al., 2016). To distinguish MS from other neurological conditions, different CVS-based criteria have been proposed to date, some based on the percentage of perivenular lesions (from 35% to 60%) and others simply on the CVS lesion count (3-lesion or 6-lesion rule) (Maggi et al., 2018; Tallantyre et al., 2011; Mistry et al., 2016; Solomon et al., 2018). From a diagnostic perspective, retrospective studies have shown excellent diagnostic discrimination by applying the '40% rule' (Tallantyre et al., 2011) with sensitivity = 91% [95% CI, 82%-97%] and specificity = 96% [95% CI, 88%-100%]) (Castellaro et al., 2020). However, applying percentage-based criteria requires manual exclusion of lesions that are confluent or have multiple or eccentric veins, and performing the CVS evaluation on all the remaining lesions present in patients' brains, which is a time-consuming process difficult to accomplish in clinical practice.

Paramagnetic rim lesions (PRL) **-** Recent pathology studies have

Table 1
Summary of the methods proposed for the automated or semi-automated analysis of cortical lesions, the central vein sign, and paramagnetic rim lesions. The task is abbreviated as follows: segmentation (S), classification (C). If not specified, all sequences were 3D. Other abbreviations: k-nearest neighbors algorithm (K-NN), convolutional neural network (CNN), partial volume (PV).

Biomarker	Authors (year)	Method	Task	MRI sequences (magnetic field strength)	Dataset size (n. of sites)	Code available
Cortical lesions	Tardif,C. L., et al. (Tardif et al., 2010) (2010)	Laminar profile shape analysis	S	Quantitative high-resolution scan (3 T)	1 post mortem brain scan (1)	No
	Fartaria, M.J., et al. (Fartaria et al., 2016) (2016)	K-NN	S	FLAIR, MPRAGE, MP2RAGE, DIR (3 T)	39 MS patients (1)	No
	Fartaria, M.J., et al. (Fartaria et al., 2017) (2017)	K-NN with partial volume constraints	S	FLAIR, MPRAGE, MP2RAGE, DIR (3 T)	39 MS patients (1)	No
	Fartaria, M.J., et al. (Fartaria et al., 2019)(2019)	PV estimation and topological constraints	S	MP2RAGE (7 T)	25 MS patients (2)	No
	La Rosa, F., et al. (La Rosa et al., 2020) (2020)	CNN	S	FLAIR, MP2RAGE (3T)	90 MS patients (2)	Yes[b]
	La Rosa, F., et al. (La Rosa et al., 2020) (2020)	CNN	S and C	MP2RAGE, 2D T2*-w GRE, T2*w 3D-*EPI* (7T)	60 MS patients (1)	Yes[c]
	La Rosa, F., et al (2022) (La Rosa et al., 2022)	CNN	S and C	MP2RAGE, 2D T2*-w GRE, T2*w 3D-*EPI* (7T)	80 MS patients (2)	Yes[c]
Paramagnetic rim lesions	Barquero, G., et al. (Barquero et al., 2020) (2020)	CNN	C	FLAIR, T2*w 3D-*EPI* (3T)	124 MS patients (2)	No
	Lou, C., et al. (Lou et al., 2021) (2021)	Random forest classifier	C	FLAIR, T1-w, T2*w 3D-*EPI* (3T)	20 MS patients (1)	Yes[d]
	Zhang, H. et al. (2022) (Zhang et al., 2022)	Residual network and radiomic features	C	FLAIR, QSM (3T)	172 MS patients (1)	No
Central vein sign	Maggi, P., Fartaria, MJ, et al. (Maggi et al., 2020) (2020)	CNN	C	T2*w 3D-*EPI*, FLAIR (3T)	42 MS patients, 33 mimics, 5 others (3)	No
	Dworkin, J. D., et al. (Dworkin et al., 2018) (2018)	Probabilistic method	C	T2*w 3D-*EPI*, FLAIR (3T)	16 MS patients, 15 MS mimics (1)	Yes[e]

[b]https://github.com/FrancescoLR/MS-lesion-segmentation
[c]https://github.com/Medical-Image-Analysis-Laboratory/CLaiMS
[d]https://github.com/carolynlou/prlr
[e]https://github.com/jdwor/cvs

demonstrated that about 30% of chronic demyelinated lesions are pathologically characterized by perilesional accumulation of iron-laden microglia and macrophages, showing evidence of smouldering demyelination and axonal loss around an inactive hypocellular core (see Fig. 4) (Frischer et al., 2015; Luchetti et al., 2018). This type of MS lesion has been defined as "chronic active/smouldering lesions". Due to their peripheral paramagnetic iron rim, these lesions can be depicted using in-vivo susceptibility-based MRI techniques (T2*-weighted magnitude, phase images, and quantitative susceptibility mapping, QSM) at both 3 T and 7 T (Absinta et al., 2018; Absinta et al., 2016), and are therefore termed "paramagnetic rim lesions" (PRL).

Direct comparison among different MRI sequences and post-processing techniques for PRL detection is still limited. A recent study (Huang et al., 2022) has compared QSM and high-pass-filtered (HPF) phase imaging for identifying PRL. Of 2062 MS lesions detected in 80 patients, 9.1% were identified as PRL in both QSM and HPF phase, 9.8% were PRL only in HPF phase, and the rest were rim negative. QSM-identified PRL showed stronger association with clinical disability compared to those detected by HPF phase imaging.

Overall, in vivo studies have shown that about 50% of relapsing and about 60% of progressive MS patients have at least one PRL (Absinta et al., 2019; Maggi et al., 2020). Of clinical relevance, PRL accrual has been recently linked to a more aggressive disease course and disability accumulation at a younger age and/or shorter disease duration (Absinta et al., 2019). Reasons for such association directly rely on a few typical features of these lesions: PRL are destructive (Absinta et al., 2016; Kolb et al., 2021), they do not remyelinate (Absinta et al., 2016), and they can expand over time, (Absinta et al., 2019) demyelinating the surrounding tissue and injuring axons, as corroborated by the elevation of serum neurofilament light chain in patients with PRL who are not forming new white matter lesions (Maggi et al., 2021). The recent discovery that the paramagnetic rim can significantly shrink or disappear (Absinta et al., 2021; Dal-Bianco et al., 2021) holds promise regarding its potential use as an outcome measure in clinical trials designed to halt the chronic inflammation at the lesion edge. In addition to their prognostic role, PRL appear specific to MS, as they have been rarely detected in patients with other neurological conditions (52% of MS vs 7% of non-MS in a multicenter study of 438 individuals) (Maggi et al., 2020). PRL have the promise of becoming a clinically relevant biomarker to both improve MS diagnosis and monitor treatment efficacy over time.

Overall, there are not yet imaging guidelines for the visual detection of PRL which requires specific training and remains challenging and time-consuming. The development of ML-based approaches, described in the next section, may help alleviate these issues and facilitate PRL assessment.

Slowly evolving/expanding lesions (SELs) – A different computational approach, designed to detect in vivo longitudinal volumetric lesional changes not associated with gadolinium enhancement, identifies the so-called "slowly evolving/expanding lesions" or SELs. Linear and radial lesion expansion is computed as a function of the Jacobian determinant of the non-linear deformation field between baseline and follow up scans (linearity assessment requires a minimum of 3 scans) (Elliott et al., 2019). Advantages of this approach relate to the use of retrospective conventional T1-weighted and T2-weighted scans. re-analysis of the ORATORIO[a] clinical trial found reduced rate of T1-SELs expansion in progressive patients treated with ocrelizumab vs placebo (Elliott et al., 2019). A recent study showed that SELs are independent predictors of EDSS worsening after a median follow up of 9 years (Preziosa et al., 2022). The neuropathological correlate of SELs is currently not yet determined and preliminary data showed only modest correlation with PRL (Elliott et al., 2021).

Overall, CL, PRL, and CVS have the potential to considerably

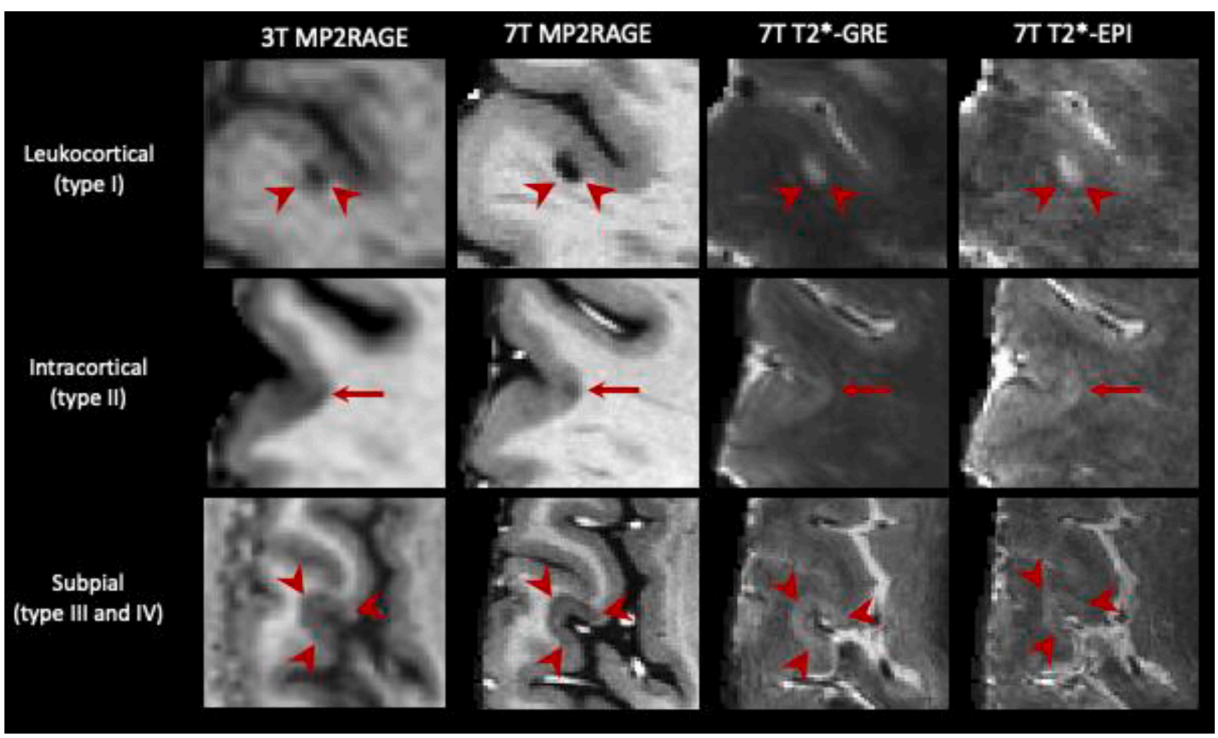

Fig. 1. Representative examples of the three main types of CL. From left to right: 3 T MP2RAGE (0.75 mm isometric), 7 T MP2RAGE (0.5 mm isometric), 7 T T2*-*EPI* (0.5 mm isometric) and 7 T T2*-GRE (0.5 mm isometric). CL, including leukocortical, intracortical, and subpial subtypes, are seen better at 7 T due to higher signal-to-noise ratios, allowing higher resolution scans, and increased susceptibility effects. The 7 T MP2RAGE image shown was obtained as the average of 4 acquisitions.

[a] A phase 3, randomized, parallel-group, double-blind, placebo-controlled trial.

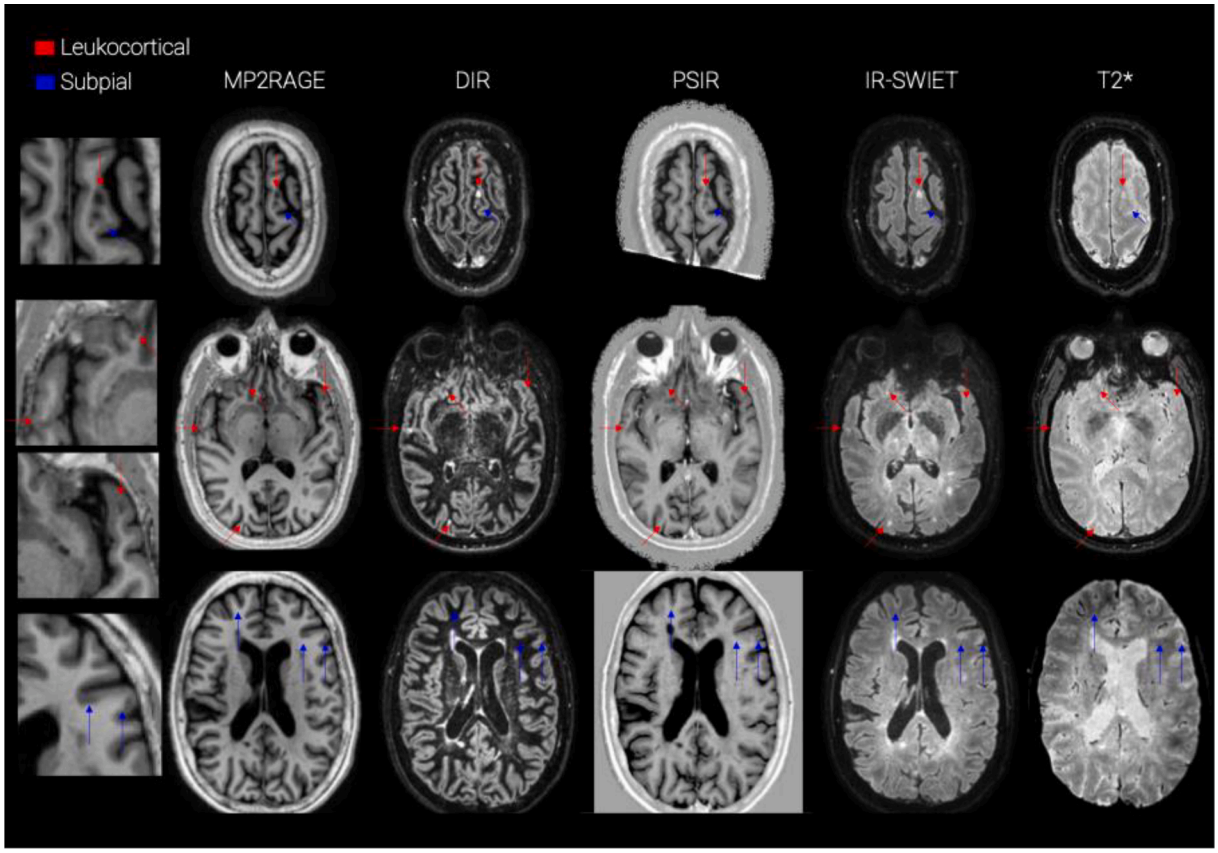

Fig. 2. examples of CL seen in different MRI contrasts at 3 T. From left to right: MP2RAGE, DIR, PSIR, IR-SWIET, T2*. Red arrows point to leukocortical lesions and blue arrows to subpial lesions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

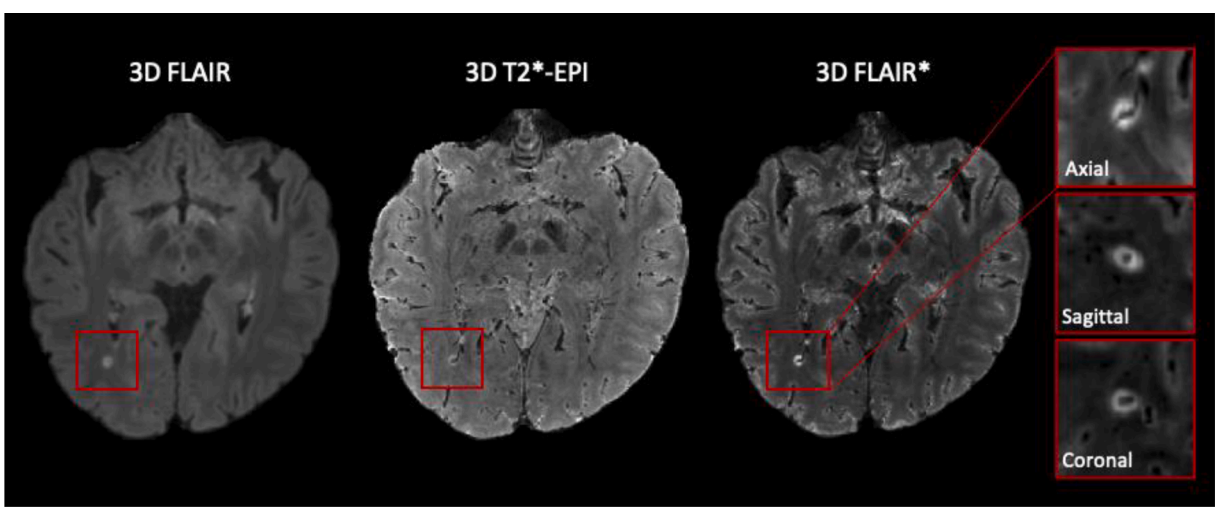

Fig. 3. A central vein running through a lesion visible in the three planes (zoomed-in boxes) in a 3D FLAIR* obtained combining FLAIR and T2*-*EPI* acquisitions at 3 T. Resampling was applied to the magnified images for visualization purposes. FLAIR, T2*-*EPI* and FLAIR* are the MRI contrasts that have been used by ML approaches for CVS detection. Refer to the supplementary material for additional examples of the CVS on different susceptibility-weighted imaging sequences.

improve the specificity of MS diagnosis (Junker et al., 2020; Maggi et al., 2018; Maggi et al., 2018). Moreover, studies have shown that CL, PRL, and SELs can be useful to assess prognosis (Calabrese et al., 2012; Absinta et al., 2016). Their manual assessment, however, particularly for CL, is both time-consuming and prone to inter-rater variability. As for conventional WML, some automated or semi-automated methods have been proposed to accelerate this task (Fartaria et al., 2017; Fartaria et al., 2019; La Rosa et al., 2020; La Rosa et al., 2020; Fartaria et al., 2016; Tardif et al., 2010; Barquero et al., 2020; Lou et al., 2021; Maggi et al., 2020; Dworkin et al., 2018). In the next section, we describe the challenges these approaches have been facing and how these differ from the segmentation of WML.

2.1. Added challenges for CL, PRL, and CVS assessment

Compared to conventional imaging biomarkers, the visual

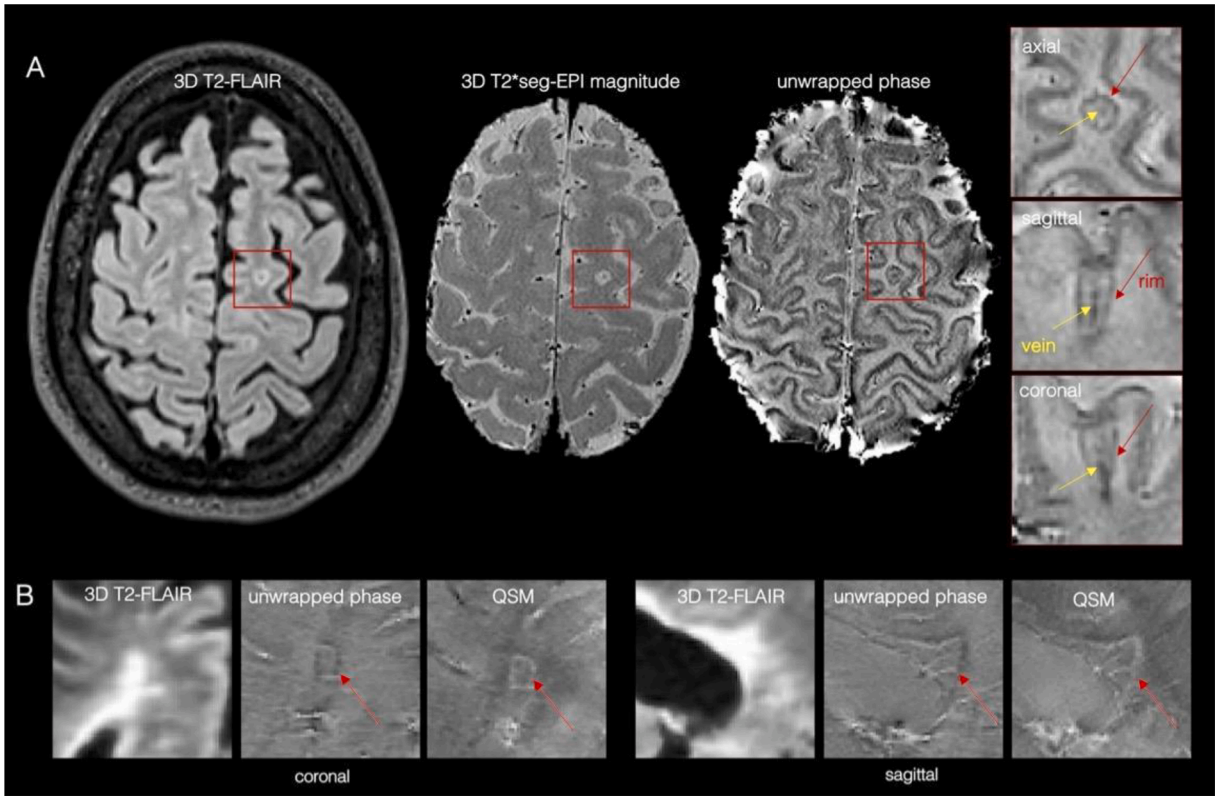

Fig. 4. (A) Representative paramagnetic rim lesion seen on a 3 T T2*-weighted seg-*EPI* magnitude and unwrapped filtered phase in the three orthogonal planes (zoomed-in red boxes, the rim is indicated with red arrows). The central vein (yellow arrows) is also clearly visible within the lesion. (B) Representative periventricular MS lesion with a paramagnetic rim. Paramagnetic rims are visible on both unwrapped phase and QSM-reconstructed images (white arrows). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

assessment of CL, PRL and CVS present some additional challenges.

Imaging and assessment guidelines- The first obstacle is represented by the lack of consensus guidelines for imaging protocols. Although efforts have been made to standardize the use of MRI in clinical practice for conventional biomarkers (Wattjes et al., 2021), guidelines are still in a preliminary stage for CL, PRL, and the CVS. CL were included in the MS diagnostic criteria in 2017 (Thompson et al., 2018), but, currently, there is no single gold standard sequence at 3 T for their detection in a clinical setting. PSIR, DIR, and MP2RAGE are all recommended by an international consensus (Filippi et al., 2019). However, these contrasts remain primarily acquired in research settings and are not yet widely used in clinical routine. Moreover, although 7 T MRI is increasingly used to detect CL, no guidelines have been presented yet to standardize their imaging sequences and their identification.

Regarding the CVS, in a 2016 consensus statement, the North American Imaging in MS Cooperative (NAIMS) proposed a standard radiological definition and suggested specific MRI acquisitions (Sati et al., 2016). Following these recommendations, recent studies have shown that high-resolution T2*w 3D-*EPI* or FLAIR* improve the detection of the CVS compared to clinical acquisitions (Castellaro et al., 2020; Suh et al., 2019). Nevertheless, a standardized clinical protocol for CVS detection is still missing. Among the three aforementioned biomarkers, PRL is probably the one at the earliest stages. Although recent studies support the feasibility of its assessment on clinical scans and its utility in improving the diagnosis and prognosis of MS (Maggi et al., 2020), there are currently no international guidelines for its definition nor a standardized MRI protocol for its analysis. Several different imaging modalities have been used for the PRL assessment, including phase 3D-*EPI*, susceptibility weighted imaging (SWI), QSM, and multi-echo T2* GRE at both 3 T and 7 T (Absinta et al., 2018; Absinta et al., 2016). However, there is a paucity of studies that have systematically compared the sensitivity of these acquisition techniques for PRL detection, especially when implemented at different field strengths.

These evolving or unclear criteria for CL, the CVS, and PRL, wide variety of imaging settings, and lack of clear guidelines for standardized protocols clearly jeopardize the development and wide use of these biomarkers and of targeted ML techniques.

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

2.2. *Machine learning specific challenges*

From a ML perspective, the automated segmentation or classification of CL, PRL, and the CVS faces new challenges as compared to their WML counterparts.

Limited datasets - An additional limitation, particularly for supervised DL-based approaches, is the scarcity and limited size of datasets in which these biomarkers were manually annotated. For their assessment, CL, CVS, and PRL all require advanced MRI sequences at high or ultra-

high magnetic field and experienced raters, and this makes it difficult to have large multi-site datasets. Although national MS registries exist in most countries, the data sharing of MRI in MS is still limited and often includes only conventional sequences (Vrenken et al., 2021). Moreover, the CVS or the rim-shape in PRL are visible only on a few slices per lesion, reducing, even more, the data available to train a supervised approach.

Inter-rater variability - The lack of standardization for both the definition and imaging of these biomarkers contributes to a modest inter-rater variability. Barquero et al. (Barquero et al., 2020) showed that, in a cohort of 124 MS patients, approximately 38% of PRL needed a consensus review from two raters classifying PRL independently (Cohen k of 0.73). Absinta et al. observed similar inter-rater agreement between three experts at 3 T (Fleiss coefficient of 0.71), with somewhat higher intra-rater reliability (Cohen k of 0.77) (Absinta et al., 2018). Similar values were reported at 7 T for the same set of patients, whereas the agreement between 3 T and 7 T annotations was substantial (Cohen k of 0.78). In a similar way, the inter-rater agreement was shown to be moderate for the segmentation of CL (Harrison et al., 2015; Nielsen et al., 2012; Geurts et al., 2011) and high, but not perfect, for the CVS (Cohen k of 0.9) (Maggi et al., 2020; Kau et al., 2013). Imaging quality and motion artifacts are other factors to consider as they can result in inconspicuity of all three biomarkers and, therefore, contribute to poor inter-rater agreement. Overall, the inter-rater variability represents an additional challenge for the development of automated approaches, as there might be large inconsistencies in the annotations of the training or testing set due to different raters performing the manual assessment.

3. Methods

Despite the recent discovery of the CVS and PRL and the above-mentioned challenges, a few groups have already attempted to support their analysis with automated or semi-automated ML methods. To these two novel biomarkers, we add also CL, which, although studied for several years, have only recently been assessed automatically. As there are no ML-based approaches to assess SELs yet, the prospect of analyzing this additional biomarker with ML is presented in the Discussion section. Overall, many fewer methods have been proposed for the assessment of CL, PRL, and the CVS compared to WML. In what follows, we briefly describe these state-of-the-art techniques by grouping them according to the biomarker they assess. A summary of the main characteristics for each method is presented in Table 1, and a scheme of the MRI sequences used to detect these three biomarkers at both 3 T and 7 T is shown in Fig. 5.

3.1. Cortical lesions

ML-based methods automatically segmenting CL have been explored with both 3 T and 7 T MRI. The first work (Tardif et al., 2010) present in the literature considered a postmortem MS brain imaged at 3 T with different sequences (T1, T2, and relative proton density) at high resolution (0.35 mm isotropic) (Tardif et al. (2012)). Tardif et al. (Tardif et al., 2010) proposed to first identify the cortical and white matter surfaces, then extract laminar profiles between the two tissues, and finally apply a k-means classifier to the profile intensity and shape features to parcellate the cortex and detect lesions. Although showing promising results on one postmortem MS brain, this method was never validated with larger cohorts nor in-vivo data. A few years later, Fartaria et al. (Fartaria et al., 2016) proposed the first automated method for the segmentation of both WM and cortical lesions. In their study, they analyzed a cohort of 39 early-stage MS patients and considered both conventional (FLAIR, MPRAGE) and advanced (MP2RAGE, DIR) MRI sequences at 3 T. In a nutshell, their method consisted of co-registering the different MRI contrasts, leveraging prior tissue probability maps from existing brain atlases of healthy subjects, and finally classifying each voxel either as being a lesion or healthy tissue with a k-nearest neighbor (k-NN) algorithm. Additionally, as post-processing, all lesions smaller than 3.6 μL were discarded, and a region-growing algorithm was applied to improve the lesion delineation. Results were promising, showing a CL detection rate of 62% when advanced imaging (FLAIR,

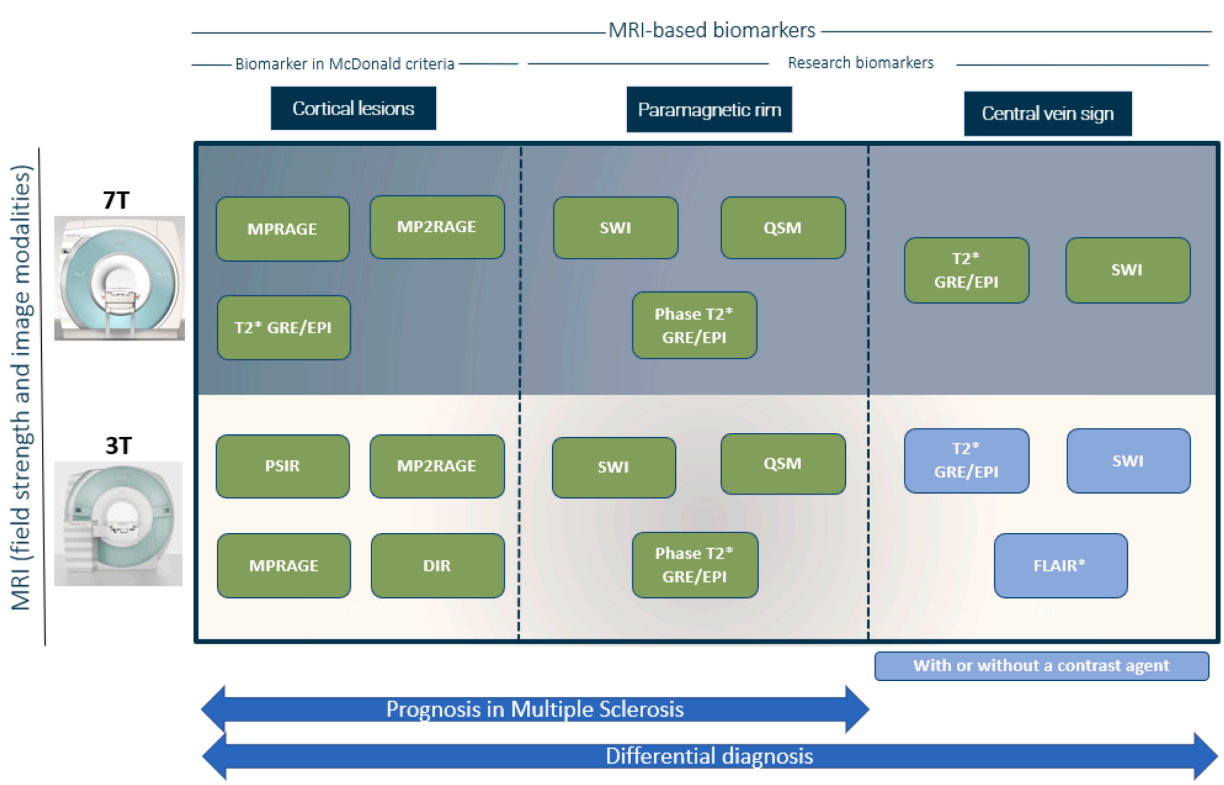

Fig. 5. Scheme showing the main MRI sequences used for detecting each biomarker at both 3 T and 7 T.

MP2RAGE, and DIR) was included. An extension of this segmentation framework with a Bayesian partial volume (PV) estimation method was presented by the same authors (Fartaria et al., 2017). They argued that CL, being generally small and located at the interface between WM and GM, suffer from strong PV effects. The addition of this PV model indeed improved the delineation of CL in terms of both total lesion volume estimation and dice coefficient (Fartaria et al., 2017).

The same research group also proposed a different segmentation method for WML and CL using only 7 T MP2RAGE images (called MSLAST: Multiple Sclerosis Lesion Analysis at Seven Tesla) (Fartaria et al., 2019). MSLAST computes tissue concentration maps with a PV algorithm and unifies them based on topological constraints. A connected-components analysis is then performed on gray matter and cerebrospinal fluid maps, and small components are classified as MS lesions. This method was evaluated with 25 MS patients' scans from two research centers and reached a 58% patient-wise CL detection rate (when 6 μL was considered as minimum lesion volume) with a false positive rate of 40%. Moreover, it showed promising scan-rescan repeatability within the same session, with a mean total lesion volume difference (WML and CL combined) of 0.29 mL (mean total lesion volume 5.52 mL), vs 0.13 mL for the manual segmentations. More recently, DL-based approaches have been presented as well (La Rosa et al., 2020; La Rosa et al., 2020). In the first study, La Rosa et al. proposed a framework for the automated segmentation of WML and CL at 3 T using FLAIR and MP2RAGE (La Rosa et al., 2020). Their method extracts 3D patches of 88x88x88 voxels from the two MRI contrasts and feeds them to a convolutional neural network (CNN). The CNN, inspired by the U-Net, has an encoder and decoder path, each one with three resolution levels. Evaluated on two datasets acquired in different centers, for a total of 90 MS patients, the framework showed competitive performance, with a CL detection rate of 76% and a false positive rate of 29%.

In a second study, the same group proposed a similar approach, this time tailored exclusively for the detection of CL using multi-contrast 7 T MRI (La Rosa et al., 2020). The contrasts considered were MP2RAGE, T2*-weighted GRE, and T2*-w 3D-*EPI*. A cohort of 60 patients was analyzed with a total of over 2000CL manually segmented by two experts. The CNN architecture proposed was similar to the one just described, but with a modified output. In addition to the CL segmentation, the CNN provided a classification into two types (leukocortical and intracortical/subpial lesions) and a separate branch with a simple tissue segmentation in WM/GM. CL were correctly classified into the two types by the network with an accuracy of 86%. Setting a minimum lesion size of 0.75 μL, it achieved a CL detection rate of 67% with, however, a quite high false positive rate of 42% (see Fig. 6). Importantly, about 24% of these false positives were retrospectively judged as CL or possible CL by an expert (La Rosa et al., 2020).

In a following publication (La Rosa et al., 2022), this method was further improved and evaluated on a multi-site dataset. Its main modifications included an added resolution level in the CNN architecture, a larger 3D patch input size of 96x96x96 voxels, and the use of the focal loss for training. Finally, a domain adaptation approach was applied to verify the performance on external datasets. On 20 MRI scans of patients imaged in a different center, this method achieved superior performance (CL detection rate of 71%) compared to MSLAST (48%) when setting a minimum lesion size of 6 μL.

3.2. The central vein sign

As of today, two automated ML methods for the classification of MS lesions as CVS+ (MS lesions showing the presence of the CVS) or CVS- (MS lesions without the CVS) have been proposed in the literature (Maggi et al., 2020; Dworkin et al., 2018). Both approaches were developed and evaluated only with 3 T MRI. Dworkin et al. (Dworkin et al., 2018) proposed a probabilistic method based on the Frangi vesselness filter (Frangi et al., 1496). They first perform an automated WML segmentation using T1 and FLAIR 3D MRI volumes acquired at 3 T

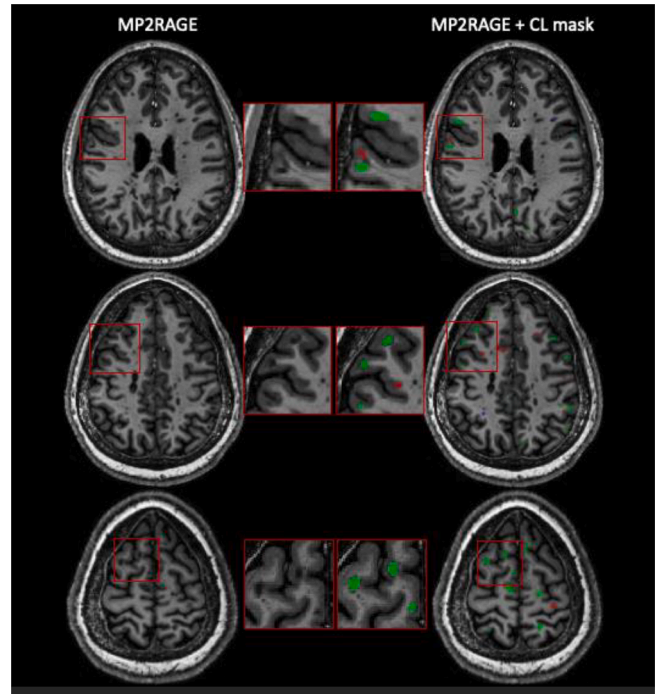

Fig. 6. Three representative axial slices from one MS patient showing the CL segmentation results of an automated CL segmentation method (La Rosa et al., 2020). 7 T MP2RAGE (left column) and CL mask (right column) showing true positives (green), false negatives (red), and false positives (blue) of the automated approach with respect to the CL manual segmentation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Valcarcel et al., 2018; Valcarcel et al., 2018) and obtain a map of the veins by applying the vesselness filter to a T2*w 3D *EPI* image. Confluent lesions are then separated, and lesion centers are detected by textural analysis (Dworkin et al., 2019). Periventricular lesions are discarded as suggested by consensus guidelines (Sati et al., 2016), and a permutation algorithm is applied to verify whether veins occur at the lesions' centers more often than would be expected due to random chance. Finally, to account for scan motion, the single lesion CVS + probabilities are weighted by the noise in their T2*-w 3D-*EPI* intensities and averaging across the total number of lesions, a patient-wise CVS value is obtained. This method was evaluated on a cohort of 31 adults, of whom 16 had MS. When considering a 40% cutoff rule, the method yielded a sensitivity of 0.94 and a specificity of 0.67 on a patient-wise classification level. The performance of the method on a lesion-wise level was not assessed. Although still far from experts' performance, this was a first attempt to automatize the CVS assessment and encouraged further improvements.

Maggi, Fartaria et al. (Maggi et al., 2020) introduced an optimized CNN for the automated CVS assessment, called CVSnet. CVSnet is inspired by the VGGnet (Simonyan and Zisserman, 2015) but composed of only three convolutional layers followed by rectified linear unit (ReLU) activations. Dropout was applied in each layer, and then two fully-connected layers of size 32 and 2, respectively, were added to provide the output. The authors selected 3D patches of size 21x21x21 voxels as input for the network, where each patch was centered on an MS lesion and FLAIR* was the only MRI contrast used. Moreover, an ensemble of 10 networks with the same architecture was trained and the probability outputs were averaged to provide the final prediction. This study considered a cohort of 80 patients imaged at three different sites, of whom 42 had MS, 35 an MS-mimic, and 5 an unknown diagnosis. On the test set, CVSnet reached a lesion-wise sensitivity, specificity, and accuracy of 0.83, 0.75, and 0.79, respectively. On a patient-wise level,

using a 50% cut-off, CVSnet achieved a sensitivity, specificity, and accuracy of 0.89, 0.92, and 0.90, respectively, outperforming the vesselness filter (Frangi et al., 1496) and approaching expert performance. However, as argued by the authors, these results are not directly comparable with those of Dworkin at al. (Dworkin et al., 2018), as the CVSnet considered different exclusion criteria to pre/select the lesions, and the initial lesion segmentation was performed manually.

3.3. Paramagnetic rim lesions

To our knowledge, only three methods have been proposed so far for the detection of rim-like features and classification of PRL (Barquero et al., 2020; Lou et al., 2021; Zhang et al., 2022). All three methods considered 3 T MRI sequences, whereas 7 T imaging has not yet been explored for the automated assessment of PRL. Barquero et al. (2020) introduced a DL-based approach (called RimNet) for the semi-automated classification of PRL, which considered 3D FLAIR and T2*w 3D-*EPI* and phase 3D-*EPI* images. RimNet's architecture is inspired by the VGGnet (Simonyan and Zisserman, 2015) and composed of two parallel CNN (one for either FLAIR or T2*w 3D-*EPI* image and one for the phase 3D-*EPI* image), where each CNN is made of three convolutional layers followed by a max-pooling operation. 3D patches of size 28x28x28 (centered around each MS lesion) are fed to each branch, and both high-level and low-level feature maps are concatenated. An automated lesion segmentation based on FLAIR and MPRAGE/MP2RAGE (La Rosa et al., 2020; La Rosa et al., 2019) was modified by an expert to split confluent lesions. The performance of RimNet was assessed on a cohort of 124 adults with MS who underwent 3 T MRI at two different sites with two scanners from the same vendor. Two experts annotated PRL independently and reached consensus in a joint session (462 PRL in total). The proposed multimodal approach based on FLAIR and phase 3D-*EPI* image achieves lesion-wise sensitivity and specificity of 0.70 and 0.95, respectively. When considering a previously identified clinical threshold of 4 PRL (Oladosu et al., 2021) for classifying patients as "chronic active" and "non-chronic active", RimNet reaches an accuracy of 0.90 and an F1-score of 0.84. These values are within 5% of the single experts' metrics, suggesting that RimNet could be a valuable tool in supporting the PRL analysis. The main drawback of RimNet, however, is that the method is not fully automated, as confluent lesions were split manually by an expert.

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

Inspired by these two previous works, Zhang et al. introduced QSM-RimNet (Zhang et al., 2022), a QSM-based approach that combines a two-branch feature extraction network and a synthetic minority oversampling technique. QSM-RimNet receives as input 3D patches of size 32x32x16 voxels where a masking out of non-lesional area is applied. One branch of the network employs residual blocks to extract convolutional features from QSM and FLAIR images, whereas the second consists in a fully-connected network that processes previously obtained radiomic features. Convolutional and radiomic features are concatenated and a minority oversampling network is used to alleviate the issue of class imbalance. Finally, a probability of being a PRL is assigned to each lesion. QSM-RimNet was evaluated with a stratified 5-folds cross-validation over 172 MS patients with a total of 177 PRL. Compared to RimNet and the automated approach of Lou et al., it outperformed both methods achieving a lesion-wise sensitivity and specificity of 0.68 and 0.99, respectively, although the differences were not statistically significant. Ablation studies showed that fusing convolutional and radiomic features improves the PRL identification (Zhang et al., 2022). Of note, QSM-Rimnet is not fully-automated as during training and evaluation it relies on manual corrections by experts of both PRL and confluent lesions. Similarly to RimNet, this strong limitation currently prevents its wider deployment and applicability.

Overall, two methods have tackled the PRL detection problem considering mainly the T2*-w 3D-*EPI* sequence and one method has focused on the QSM. Thus, none of the three frameworks has investigated the effect of differences in SWI and QSM processing on ML-based tools performance and this important aspect should be explored in future studies.

4. Discussion

The methods described in the present review tackle challenging and clinically relevant problems. Automated and reliable solutions for detecting, classifying, and segmenting CL, PRL, and CVS are needed to improve the standardization of these biomarkers and facilitate their assessment in clinical routine. As of today, however, these methods are still in an early stage and are slightly less sensitive than WML segmentation approaches.

Nevertheless, such tools would provide obvious advantages, either as stand-alone or adjunctive approaches as all three biomarkers are difficult and time-consuming to analyze using conventional radiological workflows. In these particular cases, manual reading is so involved that automated methods might actually boost the biomarkers' widespread adoption. First, they can substantially reduce analysis time, as compared to a manual rating. Maggi, Fartaria et al., for instance, showed that CVSNet was 600-fold faster on the test set compared to the manual assessment (4 s vs 40 min) when considering a 50% CVS + lesions criteria to distinguish MS from MS mimics (Maggi et al., 2020). A lower time gain, however, would be expected if CVS + lesion-count criteria, such as the 3-lesion and 6-lesion, were to be considered. Reduced analysis time can be predicted also for PRL and CL assessment. In La Rosa et al. (La Rosa et al., 2020), for instance, the automated CL segmentation of one subject is computed on average in 20 s. Although a direct comparison with the manual labeling was not reported, segmenting CL manually is known to be a much more time-consuming process. A second main advantage of automated ML methods is their ability to base their decision on 3D multi-contrast MRI analyzed simultaneously. This stands in contrast to expert reviews, which typically involve comparison of 2D slices across several contrast mechanisms in a variety of planes and are thus inherently limited in the amount of information that can be readily gleaned.

4.1. Common trends

Some common trends can be observed in most of the proposed pipelines. The large majority of the methods are supervised, relying on expert annotations. Regarding the DL-based approaches, they all used patch-based 3D CNN, exploiting the 3D intrinsic information, and often considered more than one MRI contrast simultaneously. In addition, a shared tendency consists of the use of relatively shallow architectures, with a limited number of trainable parameters, due to the lack of large datasets (La Rosa et al., 2020; La Rosa et al., 2020; Barquero et al., 2020;

Maggi et al., 2020). Combining this with extensive data augmentation techniques can help when datasets are small and unbalanced. Alternatively, other groups have tackled the problem of overfitting by proposing approaches based on classical ML techniques, such as k-NN (Fartaria et al., 2017; Fartaria et al., 2016) or random forest classifier (Commowick et al., 2018). In these studies, either intensity-based, radiomic, or probabilistic features are extracted and then fed to the respective classifier. Overall, their current performance is inferior compared to their DL-based counterparts.

In addition, some common pre-processing steps can also be identified. First, some methods use intensity normalization techniques, either based on entire 3D volumes (Lou et al., 2021; Dworkin et al., 2018; Fartaria et al., 2017; Fartaria et al., 2019; La Rosa et al., 2020; La Rosa et al., 2020) or on single patches (Barquero et al., 2020; Maggi et al., 2020). Second, the approaches using multiple MRI contrasts always register all images to the same space (Lou et al., 2021; Fartaria et al., 2017; Fartaria et al., 2019; La Rosa et al., 2020; La Rosa et al., 2020). Registration errors might affect the methods' performance. Finally, a shared pre-processing step in all approaches for the CVS or PRL assessment is the prior WML segmentation, obtained either manually (Maggi et al., 2020) or with an automated tool (Barquero et al., 2020; Lou et al., 2021; Dworkin et al., 2018). In both cases, this can be a source of error that negatively affects the subsequent biomarkers' classification accuracy.

4.2. Current limitations

Currently, a major limitation hinders the deployment of the above-described methods to the clinic: the methods proposed were trained and evaluated on small datasets acquired from one or at most two centers. Moreover, the MRI protocols used were often similar and not representative of the current diversity of images acquired in the clinics, including different processing, scans affected by noise and artifacts or protocols missing certain modalities. Therefore, the automated ML methods' robustness on larger datasets and different scanners, especially from multiple vendors, remains to be proven. This limitation is emphasized by the current lack of standardized acquisition protocols which increases the diversity of the MRI sequences considered for the same biomarkers. This also represents a major hurdle for potential regulatory approval of such methods. As regulatory approval is necessary for widespread adoption in the clinics, which is, in turn, the prerequisite for the availability of large datasets, this is currently a circular dependency issue.

In addition, the achieved performance levels of the automated ML methods are still inferior compared to the human experts. Considering the high inter-rater variability and the limited amount of data available, there is also a considerable risk of having methods that perform well on data annotated by a single expert and not as well with annotations from other raters. To mitigate this issue, several methods have already considered consensus annotations from two or more experts (La Rosa et al., 2020; Barquero et al., 2020; Maggi et al., 2020). Regarding CL, no automated method presented in the literature was compared, on the same dataset, with the experts' inter-rater variability, thus a proper evaluation is not possible. With respect to CVS, Maggi, Fartaria et al. (Maggi et al., 2020) compared the performance of CVSnet with the consensus of two experts. Following the "50% rule," CVSnet achieved on the testing set a classification accuracy of 79%, whereas the experts reached 100% accuracy in differentiating MS and mimic diseases. In a similar way, Barquero et al. (Barquero et al., 2020) compared RimNet's performance with those of two experts in classifying PRL. In a lesion-wise analysis, RimNet achieved a sensitivity of 71% and a negative predictive value of 96%, approaching the experts, who reached 78% and 98%, respectively.

Another main limitation is represented by the fact that some methods presented are not fully automated. CVSnet (Maggi et al., 2020), for instance, used manually annotated MS lesion masks in which lesions were excluded based on the NAIMS criteria (Sati et al., 2016), whereas in the pipeline proposed by Dworkin et al. (Dworkin et al., 2018), scans affected by noise were discarded following a manual rating. Similarly, RimNet (Barquero et al., 2020) exploits lesion masks where confluent lesions have been previously split into single units by an expert. In contrast, all methods described to date for CL segmentation or detection are fully automated (Fartaria et al., 2017; La Rosa et al., 2020; La Rosa et al., 2020; Fartaria et al., 2016). Another persistent issue in the automated analysis of the CVS and PRL is the presence of confluent lesions. Large, periventricular white matter lesions which include several single units pose additional challenges as the current methods classify each lesion singularly (Lou et al., 2021; Dworkin et al., 2018), and some of them extract 3D patches centered on the lesion of interest (Barquero et al., 2020; Maggi et al., 2020). In RimNet (Barquero et al., 2020), for instance, an expert manually split confluent lesions, whereas Lou et al. observed a consistent drop in performance in PRL classification in the presence of confluent lesions (Lou et al., 2021). Although methods to automatically split confluent lesions have been proposed (Dworkin et al., 2019; Zhang et al., 2021), further developments are needed in order to properly apply these in the presence of the CVS or PRL.

Finally, for every automated tool the regulatory environment remains a critical barrier, as up to date less than 90 AI/ML-based medical devices or algorithms have been approved by the US Food & Drugs Administration (FDA). This challenge, however, is not unique to the three biomarkers considered (Pinto et al., 2020) but shared also by automated approaches segmenting WML or estimating brain atrophy.

4.3. Future research avenues

Standardization of the biomarkers' assessment- The first two necessary steps toward the improvement of the above-referred approaches are the validation of the biomarkers' specific criteria and standardization of the relative MRI protocols. CL have been recently included in the MS diagnostic criteria (Thompson et al., 2018), however, a consensus on imaging and on their definition is still missing. In a similar way, PRL urgently need a consensus definition and standardized clinical protocols, whereas the initial criteria proposed for the CVS (Sati et al., 2016) need to be updated in light of the latest studies. This would clarify the automated methods' goals, which so far have been extremely dependent on specific expert labeling of each dataset or on the specific criteria adopted.

Standardization and extensive validation of the automated methods - Currently, it is difficult to compare the performance of automated ML methods considering different criteria (such as the minimum lesion size) and being evaluated on private datasets. In the future, the generalization of the proposed methods should be validated on large, multi-site datasets with standardized metrics. For this purpose, we urge research groups to organize grand challenges and release publicly available datasets with manual annotations of CL, PRL, and CVS. As already proved for several other tasks in medical imaging (Antonelli et al., 2021), including for WML segmentation (Carass et al., 2017; Commowick et al., 2018), such open data initiatives boost on the one hand the development of state-of-the-art methods, and on the other hand, help set benchmarks for a fair assessment. Only 5 of the 12 methods covered in this review are publicly available. In order to extend their usage and foster a culture of open science, research groups should make their code publicly available and possibly provide Docker (Docker, 2014)/Singularity (Kurtzer et al., 2017) images to facilitate their evaluation. Moreover, as successfully done for WML segmentation (Valverde et al., 2019), domain-adaptation techniques should also be explored in order to improve robustness of the automated ML methods to noise, artifacts, and different protocols. So far, all three biomarkers have been primarily studied at 3 T and 7 T, and therefore robust methods able to work with images acquired at both magnetic field strengths would be very valuable. Machine learning algorithms could exploit 7 T enhanced spatial resolution and tissue contrast by domain adaptation techniques

to improve their performance on 3 T imaging, which will continue to be the main tool for clinicians as well as for clinical research and trials for the foreseeable future. Although it would be highly desirable to have methods that work also at the most accessible field strength of 1.5 T, this seems currently unlikely as the sensitivity to these biomarkers is field-dependent.

Transfer learning - Considering the scarcity of large, annotated datasets, an additional strategy that should be explored consists of transfer learning. Sharing of neural network weights between research groups could foster interdisciplinary applicability of CNN trained on relatively large datasets towards different purposes, such as CL, PRL, and CVS, by fine-tuning the trained models in smaller datasets. Potential advantages would include a shorter training time and robust feature extraction across different MRI device manufacturers or different pulse sequence acquisition parameters (Valverde et al., 2021).

Longitudinal assessment - Another possible research direction is an expansion of the current methods to analyze longitudinal data. To the best of our knowledge, only one study has tackled the automated longitudinal assessment of CL at 3 T (Fartaria et al., 2019), whereas PRL evolution over time has not yet been assessed with automated approaches. CL are known to play a major role in disease progression (Mainero et al., 2015) and considerable changes in their volume were observed in longitudinal studies (Calabrese et al., 2008; Faizy et al., 2019). Of similar interest, PRL and slowly-evolving lesions (SELs) volume assessment over time is a plausible future clinical measure of treatment response (Absinta et al., 2021; Dal-Bianco et al., 2021; Elliott et al., 2019; Elliott et al., 2019). Therefore, automated longitudinal assessment of both CL and PRL could be of high relevance. Regarding SELs, longitudinal WML segmentation approaches (Lladó et al., 2012) could be adapted to track their evolution in a fully-automated way. This would facilitate their assessment as currently, following an automated cross-sectional WML segmentation, the lesion masks at each timepoint are manually reviewed (Elliott et al., 2019).

Joint assessment of multiple biomarkers- To date, all the methods proposed tackled the assessment of a single lesional biomarker, although in the case of CL some methods consider WML as well (Fartaria et al., 2019; La Rosa et al., 2020; Fartaria et al., 2016). Future work may aim at automatically analyzing multiple biomarkers in a unified framework (eg. with the same input images and algorithm) as this would be extremely useful for research purposes or in clinics. Moreover, ML-based algorithms have the potential to be useful also for prediction purposes. A few automated methods based either on MRI (Tousignant et al., 2021; Marzullo et al., 2019; Roca et al., 2020), optical coherence tomography (Montolío et al., 2021), or clinical information (Pinto et al., 2020) have already been presented to predict MS progression. Specifically to the biomarkers considered in the present review, Treaba et al. have proposed a ML approach for the regression of both CL and PRL, in the same cohort of patients, with disability progression (Treaba et al., 2021;3(3): fcab134.). In this prospective, longitudinal study, the authors analyzed brain scans of 100 MS patients using 7 T susceptibility-sensitive MRI in which CL and PRL were segmented manually. Although the study had some limitations, including the fact that the disability progression was assessed only by the EDSS and only one ML-based method (gradient boosting algorithm, XGBoost) was tested, it showed that 7 T MRI and the combination of different biomarkers are promising in predicting MS disability progression. Future studies should aim to combine the automated assessment of multiple biomarkers with clinical information and other relevant markers to predict clinical outcomes or treatment effect.

Explainable AI - As discussed in this paper, ML methods combined with specialized MRI sequences could play a fundamental role in supporting the diagnosis of, and prognostication in, MS. However, the complexity of DL algorithms hinders their interpretation, which has led some to consider these methods as "black boxes." The lack of an obvious connection between biology, pathophysiology, and features revealed by DL might diminish clinicians' confidence in these algorithms, again hindering the adoption of such tools in clinical research and healthcare. Explainable AI (XAI) methods are needed as to on one side provide uncertainty estimates regarding the output provided and on the other side transparency on the decisions taken by the DL-models. By explainability, we refer to a set of domain features such as pixels of an image or human-understandable high-level attributes that contribute to the output decision of the model and its internal working. To our knowledge, there are only two groups that have investigated XAI in MS. Eitel et al. (Eitel et al., 2019) explored explainability to reveal relevant voxel-wise locations that a trained CNN uses for distinguishing between a normal and MS brain MRI. They found that diagnostic success relied on the appearance of both lesions and non-lesional tissue (thalamus). Nair et al.Nair et al. (2020) studied the uncertainty of DL-based lesion segmentation to quantify the AI model reliability. Interestingly, their results showed that discarding lesions with high estimated uncertainty from the output segmentation would improve the performance of the model. These two pioneering approaches strengthen the idea that explainability and uncertainty measures can reliably provide new insights into how DL models for MS work and potentially improve them and increase their transparency.

Overall, we believe that developing explainable AI tools is crucial in the ML MS research roadmap and would have an impact at both methodological and clinical levels. First, explainable DL in MS would provide new insights into model decisions and help identify any bias. Second, the inclusion of uncertainty and explainability will help in increasing the confidence of clinicians considering their use, as well as improve the quality of decision making and ultimately the clinical impact. Finally, they may foster a better understanding of MS progression by generating biologically interpretable measures of inflammation and degeneration.

5. Conclusions

To summarize, automated or semi-automated ML-based approaches aiming to segment and classify CL, CVS, and PRL are still in an early stage. Nevertheless, these pioneering methods have the potential to provide standardized identification of the biomarkers and facilitate their large-scale assessment in clinical routines. Automated or semi-automated tools could considerably reduce the current amount of time and effort needed for a manual assessment. To date, however, some limitations still hinder a broader adoption of these tools. First, there is a general need for consensus criteria and standardized clinical protocols for all three aforementioned biomarkers. Further, a major barrier to the automated methods' deployment is their lack of validation on multi-center datasets acquired with different protocols. Future work should focus on improving the robustness of the automated methods, extending their framework with longitudinal data, and including interpretable measures into their decisions. Finally, we encourage research groups to organize grand challenges and release publicly available datasets. This would boost the development of new methods and provide benchmarks for a fair and standardized comparison that is currently lacking.

CRediT authorship contribution statement

Francesco La Rosa: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Maxence Wynen:** Methodology, Visualization, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Omar Al-Louzi:** Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Erin S Beck:** Methodology, Validation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Till Huelnhagen:** Resources, Writing – original draft, Writing – review & editing, Visualization. **Pietro Maggi:** Resources, Writing – original draft, Writing – review & editing, Visualization. **Jean-Philippe Thiran:** Investigation, Supervision, Project administration, Funding acquisition. **Tobias Kober:** Investigation, Writing – original draft, Writing – review & editing. **Russell T**

Shinohara: Validation, Investigation, Writing – original draft, Writing – review & editing. **Pascal Sati:** Validation, Investigation, Writing – original draft, Writing – review & editing. **Daniel S Reich:** Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing. **Cristina Granziera:** Validation, Resources, Writing – original draft. **Martina Absinta:** Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Meritxell Bach Cuadra:** Conceptualization, Methodology, Validation, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The University Hospital Basel (USB), as the employer of C.G., has received the following fees which were used exclusively for research support: (i) advisory board and consultancy fees from Actelion, Genzyme-Sanofi, Novartis, GeNeuro and Roche; (ii) speaker feesfrom Genzyme-Sanofi, Novartis, GeNeuro and Roche; (iii) research support from Siemens, GeNeuro, Roche. M.A. has received consultancy fees from GSK and Sanofi-Genzyme. P.M. has received support from Biogen and Cliniques universitaires Saint-Luc Fonds de Recherche Clinique. D.S.R. has received research support from Abata, Sanofi-Genzyme, and Vertex. The other authors have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

F.L.R. is supported by the Swiss National Foundation (SNF) Postdoc Mobility Fellowship (P500PB_206833), the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie project TRABIT (agreement No 765148) and the Novartis Foundation for Medical-Biological Research (#21A032). M.W. is supported by a Swiss government excellence scholarship (#2021.0087). O.A. is supported by a National Multiple Sclerosis Society (NMSS) - American Brain Foundation Clinician Scientist Development Award (FAN-1807–32163). E.S.B. is supported by a Career Transition Fellowship from the National Multiple Sclerosis Society. P.S., O.A., E.S.B., and D.S.R. are supported by the Intramural Research Program of the National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA. R.T.S. is partially supported by R01NS060910, R01MH112847, R01MH123550, U01NS116776, and R01NS112274 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. M.A. is supported by the Conrad N. Hilton Foundation (Marilyn Hilton Bridging Award for Physician-Scientists, grant #17313), the International Progressive MS alliance, the Roche Foundation for Independent Research, the Cariplo Foundation (grant #1677), and the FRRB Early Career Award (grant#1750327).

We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Ecole Polytechnique fedérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG). We thank Thomas Yu for proofreading the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2022.103205.

References

Absinta, M., Sati, P., Schindler, M., et al., 2016. Persistent 7-tesla phase rim predicts poor outcome in new multiple sclerosis patient lesions. J Clin Invest. 126 (7), 2597–2609. https://doi.org/10.1172/JCI86198.

Absinta, M., Sati, P., Fechner, A., Schindler, M.K., Nair, G., Reich, D.S., 2018. Identification of chronic active multiple sclerosis lesions on 3T MRI. Am J Neuroradiol. 39 (7), 1233–1238. https://doi.org/10.3174/ajnr.A5660.

Absinta, M., Sati, P., Masuzzo, F., et al., 2019. Association of chronic active multiple sclerosis lesions with disability in vivo. JAMA Neurol. 76 (12), 1474. https://doi.org/10.1001/jamaneurol.2019.2399.

Absinta, M., Maric, D., Gharagozloo, M., et al., 2021. A lymphocyte–microglia–astrocyte axis in chronic active multiple sclerosis. Nature 597 (7878), 709–714. https://doi.org/10.1038/s41586-021-03892-7.

Antonelli M, Reinke A, Bakas S, et al. The Medical Segmentation Decathlon. *ArXiv210605735 Cs Eess*. Published online June 10, 2021. Accessed November 8, 2021. http://arxiv.org/abs/2106.05735.

Barkhof, F., 2002. The clinico-radiological paradox in multiple sclerosis revisited. Curr. Opin. Neurol. 15 (3), 239–245. https://doi.org/10.1097/00019052-200206000-00003.

Barquero, G., La Rosa, F., Kebiri, H., et al., 2020. RimNet: A deep 3D multimodal MRI architecture for paramagnetic rim lesion assessment in multiple sclerosis. NeuroImage Clin. 28, 102412 https://doi.org/10.1016/j.nicl.2020.102412.

Beck ES, Maranzano J, Luciano NJ, et al. Cortical lesion hotspots and association of subpial lesions with disability in multiple sclerosis. *Mult Scler J*. Published online February 10, 2022:135245852110691. doi:10.1177/13524585211069167.

Beck, E.S., Gai, N., Filippini, S., Maranzano, J., Nair, G., Reich, D.S., 2020. Inversion recovery susceptibility weighted imaging with enhanced T2 Weighting at 3 T improves visualization of subpial cortical multiple sclerosis lesions. Invest. Radiol. 55 (11), 727–735. https://doi.org/10.1097/RLI.0000000000000698.

Calabrese, M., Filippi, M., Rovaris, M., et al., 2008. Morphology and evolution of cortical lesions in multiple sclerosis. A longitudinal MRI study. *NeuroImage*. 42 (4), 1324–1328. https://doi.org/10.1016/j.neuroimage.2008.06.028.

Calabrese, M., Filippi, M., Gallo, P., 2010. Cortical lesions in multiple sclerosis. Nat Rev Neurol. 6 (8), 438–444. https://doi.org/10.1038/nrneurol.2010.93.

Calabrese, M., Poretto, V., Favaretto, A., et al., 2012. Cortical lesion load associates with progression of disability in multiple sclerosis. Brain. 135 (10), 2952–2961. https://doi.org/10.1093/brain/aws246.

Calabrese, M., Romualdi, C., Poretto, V., et al., 2013. The changing clinical course of multiple sclerosis: A matter of gray matter: Cortical Pathology and Secondary Progressive MS. Ann. Neurol. 74 (1), 76–83. https://doi.org/10.1002/ana.23882.

Campion, T., Smith, R.J.P., Altmann, D.R., et al., 2017. FLAIR* to visualize veins in white matter lesions: A new tool for the diagnosis of multiple sclerosis? Eur. Radiol. 27 (10), 4257–4263. https://doi.org/10.1007/s00330-017-4822-z.

Carass, A., Roy, S., Jog, A., et al., 2017. Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. NeuroImage. 148, 77–102. https://doi.org/10.1016/j.neuroimage.2016.12.064.

Castellaro, M., Tamanti, A., Pisani, A.I., Pizzini, F.B., Crescenzo, F., Calabrese, M., 2020. The Use of the Central Vein Sign in the Diagnosis of Multiple Sclerosis: A Systematic Review and Meta-analysis. Diagnostics. 10 (12), 1025. https://doi.org/10.3390/diagnostics10121025.

Ciotti, J.R., Eby, N.S., Brier, M.R., et al., 2022. Central vein sign and other radiographic features distinguishing myelin oligodendrocyte glycoprotein antibody disease from multiple sclerosis and aquaporin-4 antibody-positive neuromyelitis optica. Mult Scler J. 28 (1), 49–60. https://doi.org/10.1177/13524585211007086.

Clarke, M.A., Pareto, D., Pessini-Ferreira, L., et al., 2020. Value of 3T Susceptibility-Weighted Imaging in the Diagnosis of Multiple Sclerosis. Am J Neuroradiol. 41 (6), 1001–1008. https://doi.org/10.3174/ajnr.A6547.

Commowick O, Cervenansky F, Cotton F, Dojat M. MSSEG-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure. In: ; 2021:1. Accessed October 28, 2021. https://hal.inria.fr/hal-03358968.

Commowick, O., Istace, A., Kain, M., et al., 2018. Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. Sci. Rep. 8 (1), 13650. https://doi.org/10.1038/s41598-018-31911-7.

Cortese, R., Magnollay, L., Tur, C., et al., 2018. Value of the central vein sign at 3T to differentiate MS from seropositive NMOSD. Neurology. 90 (14), e1183–e1190. https://doi.org/10.1212/WNL.0000000000005256.

Cortese, R., Collorone, S., Ciccarelli, O., Toosy, A.T., 2019. Advances in brain imaging in multiple sclerosis. Ther Adv Neurol Disord. 12 https://doi.org/10.1177/1756286419859722.

Dal-Bianco, A., Grabner, G., Kronnerwetter, C., et al., 2021. Long-term evolution of multiple sclerosis iron rim lesions in 7 T MRI. Brain. 144 (3), 833–847. https://doi.org/10.1093/brain/awaa436.

Merkel D. Docker: lightweight Linux containers for consistent development and deployment. *Linux J*. Published online March 1, 2014. Accessed March 15, 2022. https://dl.acm.org/doi/abs/10.5555/2600239.2600241.

Dworkin, J.D., Sati, P., Solomon, A., et al., 2018. Automated integration of multimodal MRI for the probabilistic detection of the central vein sign in white matter lesions. Am J Neuroradiol. 39 (10), 1806–1813. https://doi.org/10.3174/ajnr.A5765.

Dworkin, J.D., Linn, K.A., Oguz, I., et al., 2019. An automated statistical technique for counting distinct multiple sclerosis lesions. Am J Neuroradiol. 39 (4), 626–633. https://doi.org/10.3174/ajnr.A5556.

Eitel, F., Soehler, E., Bellmann-Strobl, J., et al., 2019. Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-

wise relevance propagation. NeuroImage Clin. 24, 102003 https://doi.org/10.1016/j.nicl.2019.102003.
Elliott, C., Wolinsky, J.S., Hauser, S.L., et al., 2019. Slowly expanding/evolving lesions as a magnetic resonance imaging marker of chronic active multiple sclerosis lesions. Mult Scler J. 25 (14), 1915–1925. https://doi.org/10.1177/1352458518814117.
Elliott, C., Belachew, S., Wolinsky, J.S., et al., 2019. Chronic white matter lesion activity predicts clinical progression in primary progressive multiple sclerosis. Brain. 142 (9), 2787–2799. https://doi.org/10.1093/brain/awz212.
Elliott, C., Belachew, S., Fisher, E., et al., 2021. MRI Characteristics of Chronic MS Lesions by Phase Rim Detection and/or Slowly Expanding Properties. Neurology. 96 (15 Supplement), 4101.
Faizy, T.D., Thaler, C., Ceyrowski, T., et al., 2017. Reliability of cortical lesion detection on double inversion recovery MRI applying the MAGNIMS-Criteria in multiple sclerosis patients within a 16-months period. Kassubek J, ed. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0172923.
Faizy, T.D., Broocks, G., Thaler, C., et al., 2019. Development of Cortical Lesion Volumes on Double Inversion Recovery MRI in Patients With Relapse-Onset Multiple Sclerosis. Front. Neurol. 10 https://doi.org/10.3389/fneur.2019.00133.
Fartaria, M.J., Bonnier, G., Roche, A., et al., 2016. Automated detection of white matter and cortical lesions in early stages of multiple sclerosis: Automated MS Lesion Segmentation. J. Magn. Reson. Imaging 43 (6), 1445–1454. https://doi.org/10.1002/jmri.25095.
Fartaria MJ, Roche A, Meuli R, Granziera C, Kober T, Bach Cuadra M. Segmentation of Cortical and Subcortical Multiple Sclerosis Lesions Based on Constrained Partial Volume Modeling. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*. Vol 10435. Lecture Notes in Computer Science. Springer International Publishing; 2017:142-149. doi:10.1007/978-3-319-66179-7_17.
Fartaria, M.J., Kober, T., Granziera, C., Bach, C.M., 2019. Longitudinal analysis of white matter and cortical lesions in multiple sclerosis. NeuroImage Clin. 23, 101938 https://doi.org/10.1016/j.nicl.2019.101938.
Fartaria, M.J., Sati, P., Todea, A., et al., 2019. Automated Detection and Segmentation of Multiple Sclerosis Lesions Using Ultra–High-Field MP2RAGE. Invest. Radiol. 54 (6), 356–364. https://doi.org/10.1097/RLI.0000000000000551.
Favaretto, A., Lazzarotto, A., Poggiali, D., et al., 2016. MRI-detectable cortical lesions in the cerebellum and their clinical relevance in multiple sclerosis. Mult Scler J. 22 (4), 494–501. https://doi.org/10.1177/1352458515594043.
Filippi, M., Preziosa, P., Banwell, B.L., et al., 2019. Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. Brain. 142 (7), 1858–1875. https://doi.org/10.1093/brain/awz144.
Frangi AF, Niessen WJ, Vincken KL, Viergever MA. Multiscale vessel enhancement filtering. In: *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98*. Vol 1496. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 1998:130-137. doi:10.1007/BFb0056195.
Frischer, J.M., Weigand, S.D., Guo, Y., et al., 2015. Clinical and pathological insights into the dynamic nature of the white matter multiple sclerosis plaque: dynamic nature of MS Plaque. Ann. Neurol. 78 (5), 710–721. https://doi.org/10.1002/ana.24497.
Gaitán, M.I., Yañez, P., Paday Formenti, M.E., et al., 2020. SWAN-Venule: An optimized MRI technique to detect the central vein sign in MS Plaques. Am J Neuroradiol. 41 (3), 456–460. https://doi.org/10.3174/ajnr.A6437.
Geurts JJG, Bo L, Pouwels PJW, Castelijns JA, Polman CH, Barkhof F. Cortical Lesions in Multiple Sclerosis: Combined Postmortem MR Imaging and Histopathology. Published online 2005:6.
Geurts, J.J.G., Roosendaal, S.D., Calabrese, M., et al., 2011. Consensus recommendations for MS cortical lesion scoring using double inversion recovery MRI. Neurology. 76 (5), 418–424. https://doi.org/10.1212/WNL.0b013e31820a0cc4.
Giorgio, A., Stromillo, M.L., Rossi, F., et al., 2011. Cortical lesions in radiologically isolated syndrome. Neurology. 77 (21), 1896–1899. https://doi.org/10.1212/WNL.0b013e318238ee9b.
Giorgio, A., Stromillo, M.L., Bartolozzi, M.L., et al., 2014. Relevance of hypointense brain MRI lesions for long-term worsening of clinical disability in relapsing multiple sclerosis. Mult Scler J. 20 (2), 214–219. https://doi.org/10.1177/1352458513494490.
Granziera, C., Wuerfel, J., Barkhof, F., et al., 2021. Quantitative magnetic resonance imaging towards clinical application in multiple sclerosis. Brain. 144 (5), 1296–1311. https://doi.org/10.1093/brain/awab029.
Haacke, E.M., Mittal, S., Wu, Z., Neelavalli, J., Cheng, Y.C.N., 2009. Susceptibility-weighted imaging: Technical aspects and clinical applications, Part 1. Am J Neuroradiol. 30 (1), 19–30. https://doi.org/10.3174/ajnr.A1400.
Hagens, M.H., Burggraaff, J., Kilsdonk, I.D., et al., 2019. Impact of 3 Tesla MRI on interobserver agreement in clinically isolated syndrome: A MAGNIMS multicentre study. Mult Scler J. 25 (3), 352–360. https://doi.org/10.1177/1352458517751647.
Harrison, D.M., Roy, S., Oh, J., et al., 2015. Association of cortical lesion burden on 7-T Magnetic Resonance imaging with cognition and disability in multiple sclerosis. JAMA Neurol. 72 (9), 1004. https://doi.org/10.1001/jamaneurol.2015.1241.
Hemond, C.C., Bakshi, R., 2018. Magnetic resonance imaging in multiple sclerosis. Cold Spring Harb Perspect Med. 8 (5), a028969 https://doi.org/10.1101/cshperspect.a028969.
Huang W, Sweeney EM, Kaunzner UW, Wang Y, Gauthier SA, Nguyen TD. Quantitative susceptibility mapping versus phase imaging to identify multiple sclerosis iron rim lesions with demyelination. *J Neuroimaging*. Published online March 9, 2022: jon.12987. doi:10.1111/jon.12987.
Ineichen, B.V., Beck, E.S., Piccirelli, M., Reich, D.S., 2021. New Prospects for ultra-high-field magnetic resonance imaging in multiple sclerosis. Invest. Radiol. 56 (11), 773–784. https://doi.org/10.1097/RLI.0000000000000804.
Junker, A., Wozniak, J., Voigt, D., et al., 2020. Extensive subpial cortical demyelination is specific to multiple sclerosis. Brain Pathol. 30 (3), 641–652. https://doi.org/10.1111/bpa.12813.
Kaisey, M., Solomon, A.J., Luu, M., Giesser, B.S., Sicotte, N.L., 2019. Incidence of multiple sclerosis misdiagnosis in referrals to two academic centers. Mult Scler Relat Disord. 30, 51–56. https://doi.org/10.1016/j.msard.2019.01.048.
Kau, T., Taschwer, M., Deutschmann, H., Schönfelder, M., Weber, J.R., Hausegger, K.A., 2013. The "central vein sign": is there a place for susceptibility weighted imaging in possible multiple sclerosis? Eur. Radiol. 23 (7), 1956–1962. https://doi.org/10.1007/s00330-013-2791-4.
Kaur, A., Kaur, L., Singh, A., 2021. State-of-the-Art Segmentation Techniques and Future Directions for Multiple Sclerosis Brain Lesions. Arch. Comput. Methods Eng. 28 (3), 951–977. https://doi.org/10.1007/s11831-020-09403-7.
Kidd, D., Barkhof, F., McConnell, R., Algra, P.R., Allen, I.V., Revesz, T., 1999. Cortical lesions in multiple sclerosis. Brain. 122 (1), 17–26. https://doi.org/10.1093/brain/122.1.17.
Kilsdonk, I.D., Jonkman, L.E., Klaver, R., et al., 2016. Increased cortical grey matter lesion detection in multiple sclerosis with 7 T MRI: a post-mortem verification study. Brain. 139 (5), 1472–1481. https://doi.org/10.1093/brain/aww037.
Kolb, H., Absinta, M., Beck, E.S., et al., 2021. 7T MRI Differentiates Remyelinated from Demyelinated Multiple Sclerosis Lesions. Ann. Neurol. 90 (4), 612–626. https://doi.org/10.1002/ana.26194.
Kurtzer, G.M., Sochat, V., Singularity, B.MW., 2017. Scientific containers for mobility of compute. Gursoy A, ed. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0177459.
La Rosa F, Fartaria MJ, Kober T, et al. Shallow vs Deep Learning Architectures for White Matter Lesion Segmentation in the Early Stages of Multiple Sclerosis. In: Crimi A, Bakas S, Kuijf H, Keyvan F, Reyes M, van Walsum T, eds. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Vol 11383. Lecture Notes in Computer Science. Springer International Publishing; 2019:142-151. doi:10.1007/978-3-030-11723-8_14.
La Rosa F, Beck ES, Abdulkadir A, et al. Automated Detection of Cortical Lesions in Multiple Sclerosis Patients with 7T MRI. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. Vol 12264. Lecture Notes in Computer Science. Springer International Publishing; 2020:584-593. doi:10.1007/978-3-030-59719-1_57.
La Rosa F, Abdulkadir A, Fartaria MJ, et al. Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE. *NeuroImage Clin*. Published online June 2020:102335. doi:10.1016/j.nicl.2020.102335.
La Rosa F, Beck ES, Maranzano J, et al. Multiple sclerosis cortical lesion detection with deep learning at ultra-high-field MRI. *NMR Biomed*. Published online March 31, 2022. doi:10.1002/nbm.4730.
Lladó, X., Oliver, A., Cabezas, M., et al., 2012. Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. Inf. Sci. 186 (1), 164–185. https://doi.org/10.1016/j.ins.2011.10.011.
Lou C, Sati P, Absinta M, et al. Fully Automated Detection of Paramagnetic Rims in Multiple Sclerosis Lesions on 3T Susceptibility-Based MR Imaging. *NeuroImage Clin*. Published online August 2021:102796. doi:10.1016/j.nicl.2021.102796.
Lu, P.J., Yoo, Y., Rahmanzadeh, R., et al., 2021. GAMER MRI: Gated-attention mechanism ranking of multi-contrast MRI in brain pathology. NeuroImage Clin. 29, 102522 https://doi.org/10.1016/j.nicl.2020.102522.
Luchetti, S., Fransen, N.L., van Eden, C.G., Ramaglia, V., Mason, M., Huitinga, I., 2018. Progressive multiple sclerosis patients show substantial lesion activity that correlates with clinical disease severity and sex: a retrospective autopsy cohort analysis. Acta Neuropathol (Berl). 135 (4), 511–528. https://doi.org/10.1007/s00401-018-1818-y.
Madsen, M.A.J., Wiggermann, V., Bramow, S., Christensen, J.R., Sellebjerg, F., Siebner, H.R., 2021. Imaging Cortical Multiple Sclerosis Lesions with Ultra-High Field MRI. Radiology and Imaging. https://doi.org/10.1101/2021.06.25.21259363.
Maggi, P., Absinta, M., Grammatico, M., et al., 2018. Central vein sign differentiates Multiple Sclerosis from central nervous system inflammatory vasculopathies: Central Vein Sign. Ann. Neurol. 83 (2), 283–294. https://doi.org/10.1002/ana.25146.
Maggi, P., Fartaria, M.J., Jorge, J., et al., 2020. CVSnet: A machine learning approach for automated central vein sign assessment in multiple sclerosis. NMR Biomed. 33 (5) https://doi.org/10.1002/nbm.4283.
Maggi, P., Sati, P., Nair, G., et al., 2020. Paramagnetic Rim Lesions are Specific to Multiple Sclerosis: An International Multicenter 3T MRI Study. Ann. Neurol. 88 (5), 1034–1042. https://doi.org/10.1002/ana.25877.
Maggi, P., Kuhle, J., Schädelin, S., et al., 2021. Chronic White Matter Inflammation and Serum Neurofilament Levels in Multiple Sclerosis. Neurology. 97 (6), e543–e553. https://doi.org/10.1212/WNL.0000000000012326.
Mainero, C., Louapre, C., Govindarajan, S.T., et al., 2015. A gradient in cortical pathology in multiple sclerosis by in vivo quantitative 7 T imaging. Brain. 138 (4), 932–945. https://doi.org/10.1093/brain/awv011.
Maranzano, J., Dadar, M., Rudko, D.A., et al., 2019. Comparison of Multiple Sclerosis Cortical Lesion Types Detected by Multicontrast 3T and 7T MRI. Am J Neuroradiol. 40 (7), 1162–1169. https://doi.org/10.3174/ajnr.A6099.
Martínez-Heras, E., Solana, E., Prados, F., et al., 2020. Characterization of multiple sclerosis lesions with distinct clinical correlates through quantitative diffusion MRI. NeuroImage Clin. 28, 102411 https://doi.org/10.1016/j.nicl.2020.102411.
Marzullo, A., Kocevar, G., Stamile, C., 2019. Prediction of multiple sclerosis patient disability from structural connectivity using convolutional neural networks. In, et al. 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE 2019, 2087–2090. https://doi.org/10.1109/EMBC.2019.8856845.
Mistry, N., Abdel-Fahim, R., Samaraweera, A., et al., 2016. Imaging central veins in brain lesions with 3-T T2*-weighted magnetic resonance imaging differentiates multiple

sclerosis from microangiopathic brain lesions. Mult Scler J. 22 (10), 1289–1296. https://doi.org/10.1177/1352458515616700.
Mittal, S., Wu, Z., Neelavalli, J., Haacke, E.M., 2009. Susceptibility-Weighted Imaging: Technical Aspects and Clinical Applications, Part 2. Am J Neuroradiol. 30 (2), 232–252. https://doi.org/10.3174/ajnr.A1461.
Montolío, A., Martín-Gallego, A., Cegoñino, J., et al., 2021. Machine learning in diagnosis and disability prediction of multiple sclerosis using optical coherence tomography. Comput. Biol. Med. 133, 104416 https://doi.org/10.1016/j.compbiomed.2021.104416.
Müller, J., La Rosa, F., Beaumont, J., et al., 2022. FLAWS: New Sensitive 3T MRI Contrasts for Cortical Lesion Detection in Multiple Sclerosis. Published online, Invest Radiol.
Nair, T., Precup, D., Arnold, D.L., Arbel, T., 2020. Exploring uncertainty measures in deep networks for Multiple sclerosis lesion detection and segmentation. Med. Image Anal. 59, 101557 https://doi.org/10.1016/j.media.2019.101557.
Nielsen, A.S., Kinkel, R.P., Tinelli, E., Benner, T., Cohen-Adad, J., Mainero, C., 2012. Focal cortical lesion detection in multiple sclerosis: 3 tesla DIR versus 7 tesla FLASH-T2*. J. Magn. Reson. Imaging 35 (3), 537–542. https://doi.org/10.1002/jmri.22847.
Nielsen, A.S., Kinkel, R.P., Madigan, N., Tinelli, E., Benner, T., Mainero, C., 2013. Contribution of cortical lesion subtypes at 7T MRI to physical and cognitive performance in MS. Neurology. 81 (7), 641–649. https://doi.org/10.1212/WNL.0b013e3182a08ce8.
Oladosu, O., Liu, W.Q., Pike, B.G., Koch, M., Metz, L.M., Zhang, Y., 2021. Advanced Analysis of Diffusion Tensor Imaging Along With Machine Learning Provides New Sensitive Measures of Tissue Pathology and Intra-Lesion Activity in Multiple Sclerosis. Front. Neurosci. 15, 634063 https://doi.org/10.3389/fnins.2021.634063.
Ontaneda, D., Sati, P., Raza, P., et al., 2021. Central vein sign: A diagnostic biomarker in multiple sclerosis (CAVS-MS) study protocol for a prospective multicenter trial. NeuroImage Clin. 32, 102834 https://doi.org/10.1016/j.nicl.2021.102834.
Pinto, M.F., Oliveira, H., Batista, S., et al., 2020. Prediction of disease progression and outcomes in multiple sclerosis with machine learning. Sci. Rep. 10 (1), 21038. https://doi.org/10.1038/s41598-020-78212-6.
Preziosa, P., Pagani, E., Meani, A., et al., 2022. Slowly Expanding Lesions Predict 9-Year Multiple Sclerosis Disease Progression. Neurol - Neuroimmunol Neuroinflammation. 9 (2), e1139 https://doi.org/10.1212/NXI.0000000000001139.
Reich DS, Lucchinetti CF, Calabresi PA. Multiple Sclerosis. Longo DL, ed. *N Engl J Med*. 2018;378(2):169-180. doi:10.1056/NEJMra1401483.
Roca, P., Attye, A., Colas, L., et al., 2020. Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI. Diagn Interv Imaging. 101 (12), 795–802. https://doi.org/10.1016/j.diii.2020.05.009.
Samaraweera, A.P.R., Clarke, M.A., Whitehead, A., et al., 2017. The Central Vein Sign in Multiple Sclerosis Lesions Is Present Irrespective of the T2* Sequence at 3 T: The Central Vein Sign in MS. J. Neuroimaging 27 (1), 114–121. https://doi.org/10.1111/jon.12367.
Sati, P., George, I.C., Shea, C.D., Gaitán, M.I., Reich, D.S., 2012. FLAIR*: A Combined MR Contrast Technique for Visualizing White Matter Lesions and Parenchymal Veins. Radiology 265 (3), 926–932. https://doi.org/10.1148/radiol.12120208.
Sati, P., Thomasson, D., Li, N., et al., 2014. Rapid, high-resolution, whole-brain, susceptibility-based MRI of multiple sclerosis. Mult Scler J. 20 (11), 1464–1470. https://doi.org/10.1177/1352458514525868.
Sati, P., Oh, J., Constable, R.T., et al., 2016. The central vein sign and its clinical evaluation for the diagnosis of multiple sclerosis: a consensus statement from the North American Imaging in Multiple Sclerosis Cooperative. Nat Rev Neurol. 12 (12), 714–722. https://doi.org/10.1038/nrneurol.2016.166.
Scalfari, A., Romualdi, C., Nicholas, R.S., et al., 2018. The cortical damage, early relapses, and onset of the progressive phase in multiple sclerosis. Neurology. 90 (24), e2107–e2118. https://doi.org/10.1212/WNL.0000000000005685.
Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv14091556 Cs*. Published online April 10, 2015. Accessed May 7, 2020. http://arxiv.org/abs/1409.1556.
Sinnecker, T., Dorr, J., Pfueller, C.F., et al., 2012. Distinct lesion morphology at 7-T MRI differentiates neuromyelitis optica from multiple sclerosis. Neurology. 79 (7), 708–714. https://doi.org/10.1212/WNL.0b013e3182648bc8.
Sinnecker, T., Clarke, M.A., Meier, D., et al., 2019. Evaluation of the Central Vein Sign as a Diagnostic Imaging Biomarker in Multiple Sclerosis. JAMA Neurol. 76 (12), 1446. https://doi.org/10.1001/jamaneurol.2019.2478.
Solomon, A.J., Schindler, M.K., Howard, D.B., et al., 2016. "Central vessel sign" on 3T FLAIR* for the differentiation of multiple sclerosis from migraine. Ann. Clin. Transl. Neurol. 3 (2), 82–87. https://doi.org/10.1002/acn3.273.
Solomon, A.J., Bourdette, D.N., Cross, A.H., et al., 2016. The contemporary spectrum of multiple sclerosis misdiagnosis: A multicenter study. Neurology. 87 (13), 1393–1399. https://doi.org/10.1212/WNL.0000000000003152.
Solomon, A.J., Watts, R., Ontaneda, D., Absinta, M., Sati, P., Reich, D.S., 2018. Diagnostic performance of central vein sign for multiple sclerosis with a simplified three-lesion algorithm. Mult Scler J. 24 (6), 750–757. https://doi.org/10.1177/1352458517726383.
Solomon, A.J., Naismith, R.T., Cross, A.H., 2019. Misdiagnosis of multiple sclerosis: Impact of the 2017 McDonald criteria on clinical practice. Neurology. 92 (1), 26–33. https://doi.org/10.1212/WNL.0000000000006583.
Sparacia, G., Agnello, F., Gambino, A., Sciortino, M., Midiri, M., 2018. Multiple sclerosis: High prevalence of the 'central vein' sign in white matter lesions on susceptibility-weighted images. Neuroradiol J. 31 (4), 356–361. https://doi.org/10.1177/1971400918763577.
Suh, C.H., Kim, S.J., Jung, S.C., Choi, C.G., Kim, H.S., 2019. The "Central Vein Sign" on T2*-weighted Images as a Diagnostic Tool in Multiple Sclerosis: A Systematic Review and Meta-analysis using Individual Patient Data. Sci. Rep. 9 (1), 18188. https://doi.org/10.1038/s41598-019-54583-3.
Tallantyre, E.C., Dixon, J.E., Donaldson, I., et al., 2011. Ultra-high-field imaging distinguishes MS lesions from asymptomatic white matter lesions. Neurology. 76 (6), 534–539. https://doi.org/10.1212/WNL.0b013e31820b7630.
Tardif, C.L., Bedell, B.J., Eskildsen, S.F., Collins, D.L., Pike, G.B., 2012. Quantitative Magnetic Resonance Imaging of Cortical Multiple Sclerosis Pathology. Mult Scler Int. 2012, 1–13. https://doi.org/10.1155/2012/742018.
Tardif CL, Collins DL, Eskildsen SF, Richardson JB, Pike GB. Segmentation of Cortical MS Lesions on MRI Using Automated Laminar Profile Shape Analysis. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. Vol 6363. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2010:181-188. doi:10.1007/978-3-642-15711-0_23.
Thaler, C., Kyselyova, A.A., Faizy, T.D., et al., 2021. Heterogeneity of multiple sclerosis lesions in fast diffusional kurtosis imaging. Bergsland N, ed. *PLOS ONE*. https://doi.org/10.1371/journal.pone.0245844.
Thompson, A.J., Banwell, B.L., Barkhof, F., et al., 2018. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 17 (2), 162–173. https://doi.org/10.1016/S1474-4422(17)30470-2.
Tousignant A, Lemaître P, Precup D, Arnold DL, Arbel T. Prediction of Disease Progression in Multiple Sclerosis Patients using Deep Learning Analysis of MRI Data. Accessed November 24, 2021. https://proceedings.mlr.press/v102/tousignant19a.html.
Tranfa, M., Tortora, M., Pontillo, G., et al., 2022. The central vein sign helps in differentiating multiple sclerosis from its mimickers: lessons from Fabry disease. Eur. Radiol. 32 (6), 3846–3854. https://doi.org/10.1007/s00330-021-08487-4.
Treaba, C.A., Granberg, T.E., Sormani, M.P., et al., 2019. Longitudinal Characterization of Cortical Lesion Development and Evolution in Multiple Sclerosis with 7.0-T MRI. Radiology 291 (3), 740–749. https://doi.org/10.1148/radiol.2019181719.
Treaba, C.A., Conti, A., Klawiter, E.C., et al., 2021;3(3):fcab134.. Cortical and phase rim lesions on 7 T MRI as markers of multiple sclerosis disease progression. Brain Commun. https://doi.org/10.1093/braincomms/fcab134.
Valcarcel, A.M., Linn, K.A., Khalid, F., et al., 2018. A dual modeling approach to automatic segmentation of cerebral T2 hyperintensities and T1 black holes in multiple sclerosis. NeuroImage Clin. 20, 1211–1221. https://doi.org/10.1016/j.nicl.2018.10.013.
Valcarcel, A.M., Linn, K.A., Vandekar, S.N., et al., 2018. MIMoSA: An Automated Method for Intermodal Segmentation Analysis of Multiple Sclerosis Brain Lesions: Method For Inter-Modal Segmentation Analysis. J. Neuroimaging 28 (4), 389–398. https://doi.org/10.1111/jon.12506.
Valverde, J.M., Imani, V., Abdollahzadeh, A., et al., 2021. Transfer Learning in Magnetic Resonance Brain Imaging: A Systematic Review. J Imaging. 7 (4), 66. https://doi.org/10.3390/jimaging7040066.
Valverde, S., Salem, M., Cabezas, M., et al., 2019. One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks. NeuroImage Clin. 21, 101638 https://doi.org/10.1016/j.nicl.2018.101638.
Vrenken H, Jenkinson M, Pham D, et al. Opportunities for Understanding MS Mechanisms and Progression With MRI Using Large-Scale Data Sharing and Artificial Intelligence. *Neurology*. Published online October 4, 2021:10.1212/WNL.0000000000012884. doi:10.1212/WNL.0000000000012884.
Wattjes, M.P., Ciccarelli, O., Reich, D.S., et al., 2021. 2021 MAGNIMS–CMSC–NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. Lancet Neurol. 20 (8), 653–670. https://doi.org/10.1016/S1474-4422(21)00095-8.
Ye, Z., George, A., Wu, A.T., et al., 2020. Deep learning with diffusion basis spectrum imaging for classification of multiple sclerosis lesions. Ann. Clin. Transl. Neurol. 7 (5), 695–706. https://doi.org/10.1002/acn3.51037.
Zeng, C., Gu, L., Liu, Z., Zhao, S., 2020. Review of Deep Learning Approaches for the Segmentation of Multiple Sclerosis Lesions on Brain MRI. Front Neuroinformatics. 14, 610967 https://doi.org/10.3389/fninf.2020.610967.
Zhang H, Zhang J, Yang G, Spincemaille P, Nguyen TD, Wang Y. Memory U-Net: Memorizing Where to Vote for Lesion Instance Segmentation. In: ; 2021. Accessed November 19, 2021. https://openreview.net/forum?id=JbWMYLN5Hba.
Zhang, H., Nguyen, T.D., Zhang, J., et al., 2022. QSMRim-Net: Imbalance-aware learning for identification of chronic active multiple sclerosis lesions on quantitative susceptibility maps. NeuroImage Clin. 34, 102979 https://doi.org/10.1016/j.nicl.2022.102979.