Check for updates

The role of AI for MRI-analysis in multiple sclerosis—A brief overview

Jean-Pierre R. Falet[1,2,3]*[†], Steven Nobile[1][†], Aliya Szpindel[1], Berardino Barile[2,3], Amar Kumar[2,3], Joshua Durso-Finley[2,3], Tal Arbel[2,3] and Douglas L. Arnold[1]

[1]Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, QC, Canada, [2]Mila - Quebec AI Institute, Montreal, QC, Canada, [3]Department of Electrical and Computer Engineering, Centre for Intelligent Machines, McGill University, Montreal, QC, Canada

Magnetic resonance imaging (MRI) has played a crucial role in the diagnosis, monitoring and treatment optimization of multiple sclerosis (MS). It is an essential component of current diagnostic criteria for its ability to non-invasively visualize both lesional and non-lesional pathology. Nevertheless, modern day usage of MRI in the clinic is limited by lengthy protocols, error-prone procedures for identifying disease markers (e.g., lesions), and the limited predictive value of existing imaging biomarkers for key disability outcomes. Recent advances in artificial intelligence (AI) have underscored the potential for AI to not only improve, but also transform how MRI is being used in MS. In this short review, we explore the role of AI in MS applications that span the entire life-cycle of an MRI image, from data collection, to lesion segmentation, detection, and volumetry, and finally to downstream clinical and scientific tasks. We conclude with a discussion on promising future directions.

KEYWORDS
artificial intelligence, machine learning, magnetic resonance imaging, multiple sclerosis, precision medicine

1 Introduction

Multiple Sclerosis (MS) is a neuro-inflammatory disease of the central nervous system characterized by a wide spectrum of inflammatory and neurodegenerative changes (Compston and Coles, 2008), with clinical manifestations that vary greatly between individuals. Since the 1980s, magnetic resonance imaging (MRI) has been a cornerstone of MS diagnosis and management due to the ability to visualize demyelinating changes and axonal loss resulting from focal inflammation, using a combination of T2 and T1-weighted sequences (Hemond and Bakshi, 2018). The temporal evolution of lesions, which may initially enhance (Filippi et al., 2019), and subsequently expand, remain static, or decrease in size (Koopmans et al., 1989), can also be captured by MRI. A number of MRI biomarkers of MS diagnosis, prognosis, and treatment response, have also been described. These include T2-hyperintense white matter lesions, gadolinium-enhancing lesions, slowly enlarging lesions, paramagnetic rim lesions, cortical/deep gray matter lesions, and leptomeningeal enhancement (Filippi and Agosta, 2010; Filippi et al., 2020). Some of these biomarkers have been found to correlate strongly with key clinical outcomes. One example is the association between new/enlarging T2 lesions and clinical relapses (Rudick et al., 2006; Sormani et al., 2009; Sormani and Bruzzi, 2013).

Despite these advances, MRI-analysis continues to face problems that limit its potential (Maggi and Absinta, 2024). The longer acquisition times and higher field strengths required

to obtain measurements of many recently studied imaging biomarkers introduces new headaches for resource-limited settings. At many clinical sites, the evaluation of MRI continues to be done manually, which is a lengthy, error-prone, and highly variable procedure (Bozsik et al., 2022; Altay et al., 2013). A strongly predictive imaging biomarker of disability progression, especially progression which is independent of relapse activity (Müller et al., 2023), has yet to be found (Filippi et al., 2020). At the therapeutic level, the influx of disease modifying therapies has significantly improved the ability to suppress lesion formation and relapse risk (Amin and Hersh, 2023), but targeting disability progression remains a major challenge. The use of MRI in predicting disease course and facilitating treatment selection is still a work in progress.

The rapid pace of progress in artificial intelligence (AI) has led to new opportunities for MRI-analysis in MS. In contrast to classical statistical methods which focus on acquiring knowledge about a population given data sampled from the same distribution, the field of AI has developed machine learning (ML) methods that focus on learning predictive patterns from a dataset with the aim of making predictions (generalizing) on new data (Bzdok, 2017; Bzdok et al., 2018). Some of this work provides a different perspective on—and a new set of solutions to—the current limitations of MRI-analysis.

When using the MRI modality as part of an AI system, practitioners often prefer to use a set of hand-crafted, image-derived features, which are based on well established image markers (e.g., T2 lesion counts, brain volume). These are typically scalars derived from the voxel-level data, either manually, or through a semi-/fully-automated process. The values for these hand-crafted features, which are easy to interpret, can be stored in tabular form, and used to train a model for a specific task using a variety of ML methods. Alternatively, the raw voxel-level data can be provided directly as an input to ML models. Some types of ML, in particular deep learning (DL), which uses deep artificial neural networks (LeCun et al., 2015), can make use of the high information content in voxel-level data to *learn* (automatically, without explicit guidance from a human expert) abstract, lower-dimensional features of the image that might not be captured by traditional hand-crafted, image-derived features (e.g., the texture of the white matter in a certain brain region). A specific type of deep neural network called the convolutional neural network (CNN) (LeCun et al., 1989; Li et al., 2022) has significantly advanced digital image processing by automatically learning features from images, sometimes leading to superior performance in tasks like image classification and object detection. The theoretical benefits resulting from ML on raw images come at the cost of greater computational and dataset requirements (Berisha et al., 2021), and generally require more expertise in model training. Traditional, hand-crafted features therefore remain valuable, especially in scenarios with limited data or specific constraints (Lin et al., 2020; Zare et al., 2018; O'Mahony et al., 2019).

This review aims to introduce the reader to key areas in which AI is transforming MRI-analysis in MS (see Figure 1 for an overview). Given the vastness of the literature on this topic, this review is meant to provide a high-level overview of selected areas that are of interest to the MS community, showcasing published work on MS-specific applications. As such, this does not represent a comprehensive review of the literature. Where possible, we refer the reader to more in depth, dedicated reviews, in specific sections. First, we will explore how AI can be used for data collection (Section 2), before discussing the traditional tasks of lesion segmentation, detection, and volumetry (Section 3). Finally, we will discuss downstream scientific and clinical tasks (Sections 4, 5, and 6). We end with a discussion on promising future directions (Section 7).
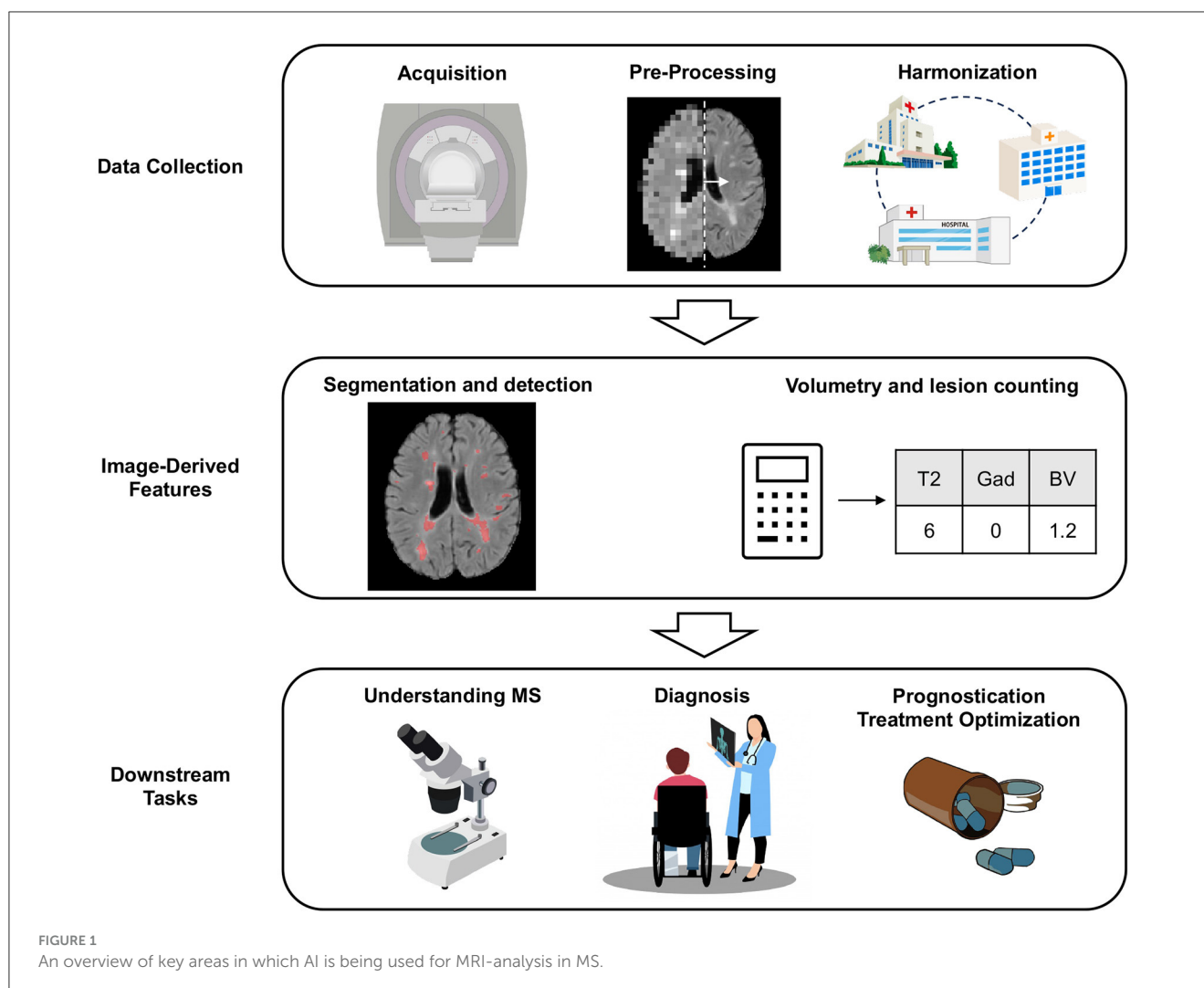
2 Acquisition, pre-processing, and harmonization

MRI has become essential for diagnosing MS and for monitoring it's evolution, primarily because of its higher sensitivity compared to clinical outcome measures of disease activity (McDonald et al., 1994). To reap the benefits of routine monitoring with MRI while minimizing the inconvenience for patients, caregivers, and resource utilization, many have turned to AI to improve the efficiency of MRI data collection. In this section, we will discuss three tasks pertaining to MRI collection: (1) acquiring the MRI images (acquisition), (2) processing the acquired images to improve their signal-to-noise ratio (pre-processing), and (3) transforming the pre-processed images from different scanners/sites to enable direct comparisons (harmonization).

Shortening the MRI acquisition time can be achieved by decreasing the number of sequences in the acquisition protocol, using generative models to synthesize the missing sequences. For example, Wei et al. (2019) showed that it is indeed possible to use a CNN to predict the FLAIR sequence from T1-weighted, T2-weighted, proton density, T1 spin-echo, and double inversion recovery (DIR) sequences. Others provided evidence to suggest that Generative Adversarial Networks [GANs, Goodfellow et al. (2014)] can synthesize DIR from the combination of T1 and T2/FLAIR (Finck et al., 2020, 2022), and T1 from T2-weighted FLAIR (Valencia et al., 2022). Although synthesis of gadolinium-enhanced T1-weighted sequences from low or non-contrast images is under-explored in MS, related work by Narayana et al. (2020) found that the presence of gadolinium-enhancing lesions can be predicted with moderate accuracy from non-contrast MRI.

Another strategy to speed data collection is to acquire lower resolution images, or images with a higher signal-to-noise ratio, and then use ML models in the post-processing phase to reconstruct higher-quality images. Various DL frameworks based on GANs and CNNs have been shown to produce higher-quality reconstructions that can improve lesion visualization and segmentation (Shaul et al., 2020; Zhao et al., 2019; Iwamura et al., 2023; Mani et al., 2021; Falvo et al., 2019). DL has also been used to optimize the more complex processing pipelines used for diffusion weighted imaging sequences (Golkov et al., 2016).

Finally, ML-based harmonization strategies can be used to address a frequently encountered problem in biomedical imaging research: small dataset sizes. Aggregating data from different data collection sites is complicated by the fact that each site may use different scanners and acquisition protocols, resulting in images that do not look alike. This is known to cause variability in tasks such as volume estimation (Clark et al., 2023; Bakshi et al.,

FIGURE 1
An overview of key areas in which AI is being used for MRI-analysis in MS.

2017). "Harmonization" is a solution to this problem that involves transforming the images so they all appear to come from the same distribution. Dewey et al. (2019) found benefits in the downstream task of brain volume estimation when images were first harmonized using DL. If direct visualization or comparisons between images from different datasets is not strictly necessary, one can also bypass the problem of harmonization by training models that are agnostic to the specific combination of sequences that is available for a particular patient (Havaei et al., 2016), or by searching for a set of hyperparameters that lead to comparable performance across a range of datasets (Gentile et al., 2023). It is worth noting that fake images can also be synthesized using DL to augment existing datasets. This is an open research problem and the magnitude of benefit probably depends on the context (Van Tulder and de Bruijne, 2015). Relatively little published research explores MRI generation specifically for MS datasets, but some authors have observed performance gains from augmentation with lesion-containing MRI images that are synthetically generated from the MRI images of healthy subjects (Salem et al., 2019; Basaran et al., 2022).

In summary, AI has shown promise in reducing the time taken to acquire and preprocess the MRI of MS patients, without significantly compromising the quality and utility of the MRI images. AI can also increase the ease with which data from different sources can be pooled together for further analysis, or for increasing the size of datasets which ML models use for training. Many of the methods that were reviewed in this section are at an early stage of development, and these tasks remain an active area of research.

3 Segmentation, lesion detection, and volumetry

Once a patient's MRI has been acquired and pre-processed, it is then ready to be used for clinical management and scientific research. Although the raw, voxel-level data can be fed directly as input to a ML model that is specifically trained for one of the downstream tasks described in Sections 4, 5, 6, there is often added value to taking an intermediate step consisting of identifying and quantifying established radiologic features in the images. These tasks include segmenting radiologic markers of MS, lesion detection, and the volumetric assessment of a variety of brain structures.

Current cross-sectional disease burden assessment typically consists of some variant on lesion volume, lesion count, and brain volume estimation. Monitoring of disease activity over time additionally calls for comparing volume estimates between time-points, and the detection of new or enlarging lesions. In most settings where radiologists and neurologists are responsible for performing these tasks, volume estimation is done qualitatively with high-level descriptors, while lesion detection is done using manual review of 2D slices. The process is lengthy, error-prone, and subject to significant inter- and intra-rater variability (Bozsik et al., 2022; Altay et al., 2013). For these reasons, there has been a growing appetite for at least partially automating these tasks using AI.

The segmentation of T2 lesions is one of the most well studied applications of ML in MS. The literature on automated MS lesion segmentation methods is vast, and methods range from classical ML to DL. We therefore refer the interested reader to several dedicated reviews for more details (García-Lorenzo et al., 2013; Danelakis et al., 2018; Spagnolo et al., 2023; Zeng et al., 2020; Doyle et al., 2018). There has been relatively less work on new (and/or enlarging) T2 lesion segmentation, but more emphasis has been placed on this task during recent challenges (Commowick et al., 2021). Beyond T2 hyper-intense lesions, DL has also been used to segment and detect imaging markers which are not currently integrated in most clinical settings. These include paramagnetic rim lesions (Barquero et al., 2020; Lou et al., 2021; Zhang et al., 2022), central vein sign on susceptibility-weighted images (Maggi et al., 2020), cortical lesions on 7T images (Rosa et al., 2022; La Rosa et al., 2020), gadolinium-enhancing lesions (Gaj et al., 2021; Karimaghaloo et al., 2010; Durso-Finley et al., 2020), and spinal cord lesions (Gros et al., 2019). The task of detecting lesions (including the detection of new lesions on follow-up images) has for the most part been studied in tandem with segmentation (Kamraoui et al., 2022; Salem et al., 2020; McKinley et al., 2020).

Although brain (parenchymal) volumetry has received less attention, DL has been used to segment the thalami of MS patients for the purpose of estimating its volume (Dwyer et al., 2021). DL methods have also been shown to perform well when compared to traditional methods for brain atrophy estimation (Zhan et al., 2023). Moreover, DL-based lesion-filling (or inpainting) has been shown to improve the performance of volumetric estimation methods that are usually sensitive to the presence of lesional tissue (Zhang et al., 2020; Clèrigues et al., 2023). Unfortunately, the large minimal detectable change in volume between clinically relevant intervals and the high inter-scanner variability still limit the utility of brain volume estimation in the clinic (Van Nederpelt et al., 2023). It is worth noting that a number of software packages for automated volumetric analysis and segmentation are available, and some already include DL methods (Billot et al., 2023).

Several challenges have been organized, in which groups compete for best performance on the same lesion segmentation task (either T2 lesion or new T2 lesion segmentation). These were hosted at the IEEE ISBI conference (Carass et al., 2017) and at MICCAI conferences (Styner et al., 2008; Commowick et al., 2018, 2021). In all cases, no model was found to be perfect, when evaluated on the basis of voxel-level segmentation metrics (under or over-segmentation) and lesion detection metrics (e.g., false positive rate), in comparison to the ground-truth segmentation obtained by human expert raters. Rather than indicative of a failure of ML for automatic segmentation, we argue that this finding should lead the community to rethink the way models are evaluated. In all challenges, performance was measured against the segmentation masks obtained from very few human experts, and on relatively small datasets of at most one hundred participants. Despite these challenge's best attempts to address the intra and inter-rater variability associated with the ground-truth lesion masks obtained from human experts (Bozsik et al., 2022; Altay et al., 2013), there remains no accepted consensus on what should constitute "ground truth". Where should one draw the lesion border, given that lesional tissue manifests as a continuous spectrum of intensity on MRI? How do we differentiate an enlarging lesion from confluent new lesions? How do we know if hyperintensities smaller than 3 mm [which are typically disregarded by expert raters (Filippi et al., 2019) to avoid false positive detections], are pathologically significant or not? Without answers to all these questions, finding that DL methods disagree with human experts is arguably insufficient to determine if they are truly inferior. To address this issue, some have proposed explicitly modeling the "label-style" that might be associated with a certain dataset or group of expert-raters (Nichyporuk et al., 2022). Others have avoided the use of ground-truth lesion masks altogether by framing lesion segmentation as an unsupervised anomaly detection task (Behrendt et al., 2023; Castellano et al., 2022; Luo et al., 2023; Pinaya et al., 2022). Training on soft-labels (as opposed to binary labels) (Gros et al., 2021; Lemay et al., 2022) and probabilistic lesion counting (Schroeter et al., 2022) are yet other possible solutions. In recognition of the importance of the problem of model evaluation in the case of image analysis, a large international consortium has recently published recommendations for model evaluation (Maier-Hein et al., 2024; Reinke et al., 2024). Still, more work has to be done to obtain answers to the problems specific to MS lesion segmentation.

To conclude, segmentation, lesion detection, and volumetry, are some of the oldest and most studied ML application in MS. In many cases, they reach performances that are acceptable for many clinical and research settings. More work is needed to determine how best to evaluate automated segmentation frameworks.

4 Improving our understanding of MS

With an increasing number of datasets containing MRI images of MS patients, and the plethora of open questions in MS research, one may ask: could AI help us uncover novel markers of MS diagnosis, evolution, and treatment response? For years, patients with MS have been categorized into a binary classification system consisting of relapsing-remitting and progressive clinical phenotypes (Lublin and Reingold, 1996). It was later found that significant overlap exists in disease evolution across these subtypes, prompting the introduction of subtype-agnostic evolution-focused terminology such as "relapse-associated worsening (RAW)" and "progression independent of relapse-activity (PIRA)" (Lublin et al., 2022). The current most accepted perspective is that individual differences in disease course can be traced back to different combinations of inflammatory, neurodegenerative, and

compensatory processes that lie along a continuous spectrum (Lassmann, 2019; Pitt et al., 2022; Vollmer et al., 2021).

This paradigm-shift, coupled with the fact that none of the existing MRI biomarkers have been particularly predictive of the key clinical outcome of disability progression (Filippi et al., 2020), has led researchers to search for alternative MRI-markers that could better explain the observed heterogeneity in disease evolution and treatment response. Notably, Eshaghi et al. (2021) and Pontillo et al. (2022) used an unsupervised ML algorithm called SuStaIn (Young, 2018) to identify disease subtypes characterized by distinct temporal progression patterns on MRI. Both groups found subtypes characterized by early cortical or deep gray matter atrophy, early signal changes in normal appearing white matter, and early T2 lesion accumulation. More work is needed to externally validate these subtypes and better understand their clinical correlates.

ML has also been used more directly to assist scientists in uncovering novel MRI markers. One strategy involves taking a pre-trained classifier (e.g., a model trained to predict MS diagnosis, or future disease activity) and producing "saliency-maps". These allow researchers to visualize the features that are thought to be "important" according to the classifier; for example, features associated with a diagnosis of MS, poorer prognosis, or specific phenotypes. By using heatmaps generated using layer-wise relevance propagation, Eitel et al. (2019) found that a CNN classifier pre-trained to predict MS diagnosis focused on T2-lesions and their location, along with non-lesional or gray matter areas that included the thalamus. Storelli et al. (2022) produced heatmaps from a CNN that was trained to predict EDSS-worsening, and identified differences in periventricular regions, white matter lesions and the corpus callosum, for EDSS-worsened patients. Zhang et al. (2021) interrogated different heatmap-generating techniques to better understand crucial brain regions that could help distinguish MS phenotypes, finding that the abnormalities associated with SPMS were more extensive compared to RRMS, the latter involving primarily the occipital region and, to a lesser extent, the frontal region. Finally, Kumar et al. (2022) proposed to identify candidate biomarkers of future new/enlarging T2 lesions in an RRMS population through a process called counterfactual image synthesis; specifically, by predicting how a patient's MRI would look like if they had a different future outcome (a counterfactual), and by taking the difference between the real (factual) and counterfactual images, markers that are predictive of future outcomes (in this case, lesion activity) can be revealed.

AI can therefore be useful to better understand disease evolution and heterogeneity. While exciting, this work remains largely at the level of methodological development, and more translational research will be needed.

5 Diagnosis

It is imperative that an MS diagnosis be confirmed rapidly, and accurately, to ensure that patients receive the best possible care. MS is currently diagnosed according to the 2017 McDonald criteria, which combines historical, MRI, and laboratory data (Thompson et al., 2018). While significant efforts have been made to accelerate MS diagnosis, the heterogeneity of the disease and

broad differential diagnosis still continues to put the clinician at risk of misdiagnoses, which can delay the initiation of an adequate treatment (Solomon et al., 2019; Brownlee and Solomon, 2021). Recent diagnostic criteria might provide increased sensitivity for the diagnosis, but at the cost of reduced specificity (Mescheriakova et al., 2018; Habek et al., 2018). In this section, we will discuss the use of AI for improving the accuracy and reliability of MS diagnosis. Note that there is some overlap with Section 3, since the detection of MS lesions on MRI is an important component of the diagnostic criteria (but not the only one). In the current section, the focus will be on the classification task of MS diagnosis, with the understanding that automated lesion segmentation and detection methods could be used upstream to provide image-derived features to an MS classifier.

Both classical ML and DL methods have been applied to the task of MS diagnosis, with MRI being the most common input modality for the classifier [we refer the reader to dedicated reviews on this topic for more details (Nabizadeh et al., 2022; Aslam et al., 2022; Shoeibi et al., 2021)]. Reported diagnostic sensitivity, and especially specificity, can be quite high [pooled sensitivity 92% (95%CI: 90%, 95%) and specificity 93% (95%CI: 90%, 96%), respectively, according to a recent meta-analysis (Nabizadeh et al., 2023)]. Even simple image-derived scalars such as the average of T1, T2*, and the total/myelin bound water content, have been found to be highly predictive (when used as input to train a supervised ML classifier) of an MS diagnosis (Neeb et al., 2019).

Differentiating MS from other diseases that can mimic it's presentation is also an important task in the clinic. Rocca et al. (2021) used a basic 3D-CNN with MRI as input to differentiate MS from neuromyelitis optica spectrum disorder (NMOSD), central nervous system vasculitis, and migraine, and found that the diagnostic accuracy exceeded that of human experts. Similarly, Kim et al. (2020) showed that MS could be differentiated from NMOSD using a 3D-CNN based on the ResNet architecture (He et al., 2016), as accurately as two neurologists. Huang et al. (2022) found that a transformer-based image classifier (Xu et al., 2021) could differentiate MS from NMOSD and myelin oligodendrocyte glycoprotein antibody disease as accurately as two neuroradiologists. MS could also be differentiated from hereditary diffuse leukodystrophy with spheroids using linear discriminant analysis (Mangeat et al., 2020), and from low grade tumors using MR-spectroscopy-derived features as input to a variety of ML models (Ekşi et al., 2021; Preul et al., 1996).

Overall, there is a growing amount of evidence supporting the use of AI in MS diagnosis.

6 Prognostication and treatment optimization

One of the main challenges for the clinician evaluating a patient with a new diagnosis of MS is to predict long-term prognosis (the evolution of the disease over time). The related task of treatment optimization (predicting which treatment will have the most beneficial effect) often depends on having an accurate prognosis. This begs the question: can AI do any better? Many early research efforts were focused on predicting the occurrence or timing of clinically-defined MS subtype transitions, using these as surrogate

markers of poor prognosis. However, as discussed in Section 4, there has been a tendency to de-emphasize these subtypes in the diagnosis and management of MS. Prognostication tasks that we will focus on in this section therefore involve the prediction of the evolution of specific manifestations of the disease, which include radiologic activity (new/enlarging T2 lesions), relapses, disability accumulation, and patient-reported outcomes.

Prognostication with respect to disability outcomes turns out to be a very challenging task, even for AI (Seccia et al., 2021). When predicting disability progression from hand-crafted, image-derived tabular features, Pellegrini et al. (2020) found that a variety of classical ML models could achieve only modest predictive performance (C-index ≤ 0.65). Nonetheless, predictive performance can vary greatly depending on what features are used as input, on the model, and on the optimization procedure. With regards to the input, Zhao et al. (2017) found that classical ML methods performed better when adding image-derived features from a 1-year follow-up MRI visit to the set of inputs, which otherwise consisted of data recorded at a baseline visit. The benefit of longitudinal follow-up was also highlighted in work that used SuStaIn (Young, 2018) for unsupervised temporal modeling of imaging trajectories. Specifically, Pontillo et al. (2022) were able to identify a "deep-gray-matter-first" subtype that was associated with long-term cognitive impairment, and Eshaghi et al. (2021) could identify a "lesion-led" subtype that was associated with both confirmed disability progression and relapse rate. Using long term clinical (non-imaging) follow-up data has also been shown to lead to a considerable performance boost when predicting progression (De Brouwer et al., 2021). All this evidence suggests that ML on longer-term MRI data represents a promising, though challenging, research direction.

With regards to the model type, Zhao et al. (2020) found that ensembles of gradient-boosted trees such as XGBoost and LightGBM performed better than alternative ML methods when predicting 5-year EDSS worsening from logitudinal data collected over 2 years, with an area under the curve (AUC) ranging from 0.79 to 0.83. Interestingly, their feature importance analysis [and that of others (Law et al., 2019)] suggests that clinical disability metrics (which includes the EDSS) might be more predictive than tabular image-derived features for this particular task.

It is possible that voxel-level MRI data, which has been understudied for the task of predicting clinical prognosis, could harbor more predictive features of prognosis than traditional image-derived features. In support of this hypothesis, Storelli et al. (2022) were able to train a CNN to predict 2-year EDSS and SDMT worsening with 75.0% sensitivity, and 87.5% specificity. It is also possible that non-trivial implementation details, such as the inclusion of a T2-lesion mask along with the raw MRI as input, could further boost performance (Tousignant et al., 2019). These studies hint at DL's potential to improve upon tabular, hand-crafted, image-derived features (e.g., T2 lesion volume). In an attempt to elucidate the relative contribution of voxel-level data to predicting disability progression Zhang et al. (2023) studied a dataset of 300 MS patients, with a very large feature set spanning numerous MRI sequences, laboratory data, demographic information, disability scores, and unstructured clinical notes. Imaging, tabular data, and notes were encoded and fused using various neural network architectures, and used for predicting EDSS milestones 3-years

later. While their best performing model made use of all three modalities (AUC 0.8380), a model trained without the MRI modality was only marginally worse (AUC 0.8078). Their study is limited by a small dataset size, with a comparatively large feature set, which could result in poor model optimization. More research is therefore needed to explore this important question, but this will require larger datasets, and additional methodological advances.

DL has also been used on radiologic markers of disease activity, which in certain cases are more sensitive to disease evolution than clinical measurements. A few studies have shown promising preliminary results in predicting the future appearance of new/enlarging T2 lesions from baseline MRI (Prabhakar et al., 2023; Durso-Finley et al., 2023, 2022). Tabular, hand-crafted image-derived features have also been used to classify a lesion as active or inactive (Peng et al., 2021). Similar to the task of predicting clinical prognosis (which focuses on predicting future disability-related outcomes), there remains the possibility that non-trivial methodological contributions may yield significant performance gains.

AI tools that aid in prognostication can be used for treatment optimization (for example, by favoring a more potent drug for a patient predicted to have highly active disease); however, it is also useful to consider the related task of estimating the "treatment effect" of a medication on the disease course. The most common treatment effect estimand that clinicians consider as part of treatment-related decisions is the *average* treatment effect, which typically is estimated using randomized clinical trials, and represents the average effect of a treatment on a population (compared to placebo or to a baseline drug). Some of the ML research cited in previous sections have presented results pertaining to treatment effect estimation. For example, the "lesion-led" subtype discovered by Eshaghi et al. (2021) appears to be specify a sub-group of individuals that experience a larger average treatment effect. Another line of work in causal ML aims to personalize treatment recommendations by predicting the treatment effect for a particular *individual* given their unique characteristics (Curth et al., 2024). For example, Durso-Finley et al. (2022) proposed a multi-headed CNN to predict the individual treatment effect of several treatments on new/enlarging T2-lesions, which used a person's MRI as input. Beyond treatment optimization, individual treatment effect estimation could also play a role in improving the statistical power of clinical trials by preferentially randomizing individuals who are predicted to benefit from an experimental therapy (Falet et al., 2022; Kanber et al., 2019).

In conclusion, although prognostication and treatment optimization remain challenging tasks, MRI-based ML research continues to improve upon previous baselines through diverse methodological innovations. Some models appear to identify subgroups of individuals that are more responsive to certain disease modifying therapies. These results are therefore paving the path toward precision medicine.

7 Discussion

In this review, we have presented several tasks where AI systems might already reliably outperform human experts in MS-specific applications. Indeed, a recent validation study by

Barnett et al. (2023) provided evidence supporting the use of AI tools for lesion detection and volumetric analyses, in both clinical settings and research studies. We also discussed tasks which are hardly feasible without recent advances in DL, such as MRI sequence synthesis and automated biomarker discovery.

As the performance of AI tools continues to improve, we will arguably see increasing interest in trustworthiness, because these AI systems are expected to take part in high-risk human decision-making. Trust in AI systems is built in numerous ways, one of which is by giving them the ability to explain the rationale behind a model's predictions, resulting in "explainable" AI systems (Došilović et al., 2018). Additionally, users should be aware of the level of confidence that a model has in a particular prediction, and how much this reflects the actual errors that a model might make. This line of work, often referred to as "uncertainty" estimation (and the related problem of calibration), allows users to know when to trust a model's predictions (Gawlikowski et al., 2023). In addition, to trust that a model will behave well in practice, there should be a good understanding of how it will generalize to new data, and whether or not it will be robust to distribution shifts (for example, if there is a change in acquisition protocol). The field of causal machine learning (Sanchez et al., 2022), which models the data generative process using causal models, promises improved out-of-distribution generalization, and represents an active field of research. MS researchers have begun to address all three of these topics, specifically explainable methods (see examples in Section 4), probabilistic modeling for uncertainty estimation (Nair et al., 2020; Durso-Finley et al., 2023), and structural causal models of MRI image generation (Reinhold et al., 2021), but more work is needed to truly enable trustworthy AI-assisted MRI analysis in MS.

Looking forward, it seems clear that highly capable AI systems based on large foundation models (Brown et al., 2020; Devlin et al., 2018; Touvron et al., 2023; Ramesh et al., 2021) will have a major impact on biomedical imaging research, including in MS. Certain chat-bots based on large language models (LLMs) can now arguably pass the Turing test (Jannai et al., 2023), and score higher than the average human on medical exams (Achiam, 2023). LLMs are increasingly being used in medical applications (Agbavor and Liang, 2022; Patel and Lam, 2023; Singhal et al., 2023; Jiang et al., 2023), and multi-modal inputs (which includes biomedical imaging) are becoming more common (Moor et al., 2023). Although foundation models remain understudied in MS applications, interesting future directions include using foundation models to improve generalization from small MS-specific datasets, through in-context learning (Dong et al., 2024), or fine-tuning. That said, in order to reap all the benefits of foundation models for MS-specific applications, several open problems need to be solved. These include sub-par reasoning capabilities (Rae et al., 2021; McKenzie et al., 2023; Arkoudas, 2023) which could be dangerous in high-stakes environments such as healthcare (Richens et al., 2020; Fraser et al., 2018), broader concerns regarding AI

safety (Bommasani et al., 2021; Anderljung et al., 2023; Urbina et al., 2022), and predictions that may be unacceptably skewed to the detriment of a particular group of people (Mehrabi et al., 2021). As more solutions to these problems are found, we can expect an increasing focus on large foundation models in the coming years, to help solve some of the most challenging tasks in MS MRI-analysis.

Author contributions

J-PF: Conceptualization, Visualization, Writing – original draft, Writing – review & editing. SN: Writing – original draft, Writing – review & editing. AS: Writing – original draft, Writing – review & editing, Visualization. BB: Writing – original draft, Writing – review & editing. AK: Writing – original draft, Writing – review & editing. JD-F: Writing – original draft, Writing – review & editing. TA: Supervision, Writing – original draft, Writing – review & editing. DA: Supervision, Writing – original draft, Writing – review & editing.

Funding

Conflict of interest

DA reports consulting fees from Biogen, Celgene, Frequency Therapeutics, Genentech, Merck, Novartis, Race to Erase MS, Roche, Sanofi–Aventis, Shionogi, and Xfacto Communications, grants from Immunotec and Novartis and an equity interest in NeuroRx.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., et al. (2023). Gpt-4 technical report. *arXiv preprint* arXiv:2303.08774.

Agbavor, F., and Liang, H. (2022). Predicting dementia from spontaneous speech using large language models. *PLOS Digital Health* 1:e0000168. doi: 10.1371/journal.pdig.0000168

Altay, E. E., Fisher, E., Jones, S. E., Hara-Cleaver, C., Lee, J. C., and Rudick, R. A. (2013). Reliability of classifying multiple sclerosis disease activity using magnetic resonance imaging in a multiple sclerosis clinic. *JAMA Neurol.* 70, 338–344. doi: 10.1001/2013.jamaneurol.211

Amin, M., and Hersh, C. M. (2023). Updates and advances in multiple sclerosis neurotherapeutics. *Neurodegener. Dis. Manag.* 13, 47–70. doi: 10.2217/nmt-2021-0058

Anderljung, M., Barnhart, J., Leung, J., Korinek, A., O'Keefe, C., Whittlestone, J., et al. (2023). Frontier ai regulation: Managing emerging risks to public safety. *arXiv* [preprint] arXiv:2307.03718. doi: 10.48550/arXiv.2307.03718

Arkoudas, K. (2023). GPT-4 can't reason. *arXiv* [preprint] arXiv:2308.03762. doi: 10.48550/arXiv.2308.03762

Aslam, N., Khan, I. U., Bashamakh, A., Alghool, F. A., Aboulnour, M., Alsuwayan, N. M., et al. (2022). Multiple sclerosis diagnosis using machine learning and deep learning: challenges and opportunities. *Sensors* 22:7856. doi: 10.3390/s22207856

Bakshi, R., Roy, S., Stern, W., Tummala, S., Yousuf, F., Zhu, A., et al. (2017). Volumetric analysis from a harmonized multisite brain MRI study of a single subject with multiple sclerosis. *Am. J. Neuroradiol.* 38:1501–1509. doi: 10.3174/ajnr.A5254

Barnett, M., Wang, D., Beadnall, H., Bischof, A., Brunacci, D., Butzkueven, H., et al. (2023). A real-world clinical validation for ai-based MRI monitoring in multiple sclerosis. *NPJ Digital Med.* 6:196. doi: 10.1038/s41746-023-00940-6

Barquero, G., La Rosa, F., Kebiri, H., Lu, P.-J., Rahmanzadeh, R., Weigel, M., et al. (2020). Rimnet: A deep 3D multimodal MRI architecture for paramagnetic RIM lesion assessment in multiple sclerosis. *NeuroImage: Clinical* 28:102412. doi: 10.1016/j.nicl.2020.102412

Basaran, B. D., Qiao, M., Matthews, P. M., and Bai, W. (2022). "Subject-specific lesion generation and pseudo-healthy synthesis for multiple sclerosis brain images," in *International Workshop on Simulation and Synthesis in Medical Imaging* (Cham: Springer), 1–11.

Behrendt, F., Bhattacharya, D., Krüger, J., Opfer, R., and Schlaefer, A. (2023). "Patched diffusion models for unsupervised anomaly detection in brain MRI," in *Proceedings of Machine Learning Research-Preprint* (Athens: IEEE), 1–14.

Berisha, V., Krantsevich, C., Hahn, P. R., Hahn, S., Dasarathy, G., Turaga, P., et al. (2021). Digital medicine and the curse of dimensionality. *NPJ Digital Med.* 4:153. doi: 10.1038/s41746-021-00521-5

Billot, B., Greve, D. N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., et al. (2023). Synthseg: Segmentation of brain MRI scans of any contrast and resolution without retraining. *Med. Image Anal.* 86:102789. doi: 10.1016/j.media.2023.102789

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., et al. (2021). On the opportunities and risks of foundation models. *arXiv* [preprint] arXiv:2108.07258. doi: 10.48550/arXiv.2108.07258

Bozsik, B., Tóth, E., Polyák, I., Kerekes, F., Szabó, N., Bencsik, K., et al. (2022). Reproducibility of lesion count in various subregions on MRI scans in multiple sclerosis. *Front. Neurol.* 13:843377. doi: 10.3389/fneur.2022.843377

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* 33, 1877–1901. doi: 10.48550/arXiv.2005.14165

Brownlee, W. J., and Solomon, A. J. (2021). Misdiagnosis of multiple sclerosis: time for action. *Multiple Sclerosis J.* 27:805–806. doi: 10.1177/13524585211005367

Bzdok, D. (2017). Classical statistics and statistical learning in imaging neuroscience. *Front. Neurosci.* 11:543. doi: 10.3389/fnins.2017.00543

Bzdok, D., Altman, N., and Krzywinski, M. (2018). Statistics versus machine learning. *Nat. Methods* 15:233. doi: 10.1038/nmeth.4642

Carass, A., Roy, S., Jog, A., Cuzzocreo, J. L., Magrath, E., Gherman, A., et al. (2017). Longitudinal multiple sclerosis lesion segmentation: resource and challenge. *Neuroimage* 148, 77–102. doi: 10.1016/j.dib.2017.04.004

Castellano, G., Placidi, G., Polsinelli, M., Tulipani, G., and Vessio, G. (2022). "Unsupervised brain MRI anomaly detection for multiple sclerosis classification," in *International Conference on Pattern Recognition* (Cham: Springer), 644–652.

Clark, K. A., O'Donnell, C. M., Elliott, M. A., Tauhid, S., Dewey, B. E., Chu, R., et al. (2023). Intersite brain MRI volumetric biases persist even in a harmonized multisubject study of multiple sclerosis. *J. Neuroimag.* 33, 941–952. doi: 10.1111/jon.13147

Clèrigues, A., Valverde, S., Salvi, J., Oliver, A., and Lladó, X. (2023). Minimizing the effect of white matter lesions on deep learning based tissue segmentation for brain volumetry. *Comp. Med. Imag. Graph.* 103:102157. doi: 10.1016/j.compmedimag.2022.102157

Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M. (2021). "Msseg-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure," in *MICCAI 2021-24th International Conference on Medical Image Computing and Computer Assisted Intervention* (Strasbourg: Hal Open Science), 126.

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

Compston, A., and Coles, A. (2008). Multiple sclerosis. *Lancet.* 372:1502–1517. doi: 10.1016/S0140-6736(08)61620-7

Curth, A., Peck, R. W., McKinney, E., Weatherall, J., and van der Schaar, M. (2024). Using machine learning to individualize treatment effect estimation: challenges and opportunities. *Clini. Pharmacol. Therapeut.* 115, 710–719. doi: 10.1002/cpt.3159

Danelakis, A., Theoharis, T., and Verganelakis, D. A. (2018). Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. *Comp. Med. Imag. Graphics* 70, 83–100. doi: 10.1016/j.compmedimag.2018.10.002

De Brouwer, E., Becker, T., Moreau, Y., Havrdova, E. K., Trojano, M., Eichau, S., et al. (2021). Longitudinal machine learning modeling of MS patient trajectories improves predictions of disability progression. *Comput. Methods Programs Biomed.* 208:106180. doi: 10.1016/j.cmpb.2021.106180

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* [preprint] arXiv:1810.04805. doi: 10.48550/arXiv.1810.04805

Dewey, B. E., Zhao, C., Reinhold, J. C., Carass, A., Fitzgerald, K. C., Sotirchos, E. S., et al. (2019). Deepharmony: a deep learning approach to contrast harmonization across scanner changes. *Magn. Reson. Imaging* 64, 160–170. doi: 10.1016/j.mri.2019.05.041

Dong, Q., Li, L., Dai, D., Zheng, C., Ma, J., Li, R., et al. (2024). A survey on in-context learning. *arXiv* [preprint] arXiv:2301.00234. doi: 10.18653/v1/2024.emnlp-main.64

Došilović, F. K., Brčić, M., and Hlupić, N. (2018). "Explainable artificial intelligence: a survey," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (Opatija: IEEE), 0210–0215.

Doyle, A., Elliott, C., Karimaghaloo, Z., Subbanna, N., Arnold, D. L., and Arbel, T. (2018). "Lesion detection, segmentation and prediction in multiple sclerosis clinical trials," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017* (Quebec City: Springer), 15–28.

Durso-Finley, J., Arnold, D. L., and Arbel, T. (2020). "Saliency based deep neural network for automatic detection of gadolinium-enhancing multiple sclerosis lesions in brain MRI," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019* (Shenzhen: Springer), 108–118.

Durso-Finley, J., Falet, J.-P., Mehta, R., Arnold, D. L., Pawlowski, N., and Arbel, T. (2023). "Improving image-based precision medicine with uncertainty-aware causal models," in *Medical Image Computing and Computer Assisted Intervention - MICCAI 2023* (Vancouver: MICCAI 2023),14224.

Durso-Finley, J., Falet, J.-P. R., Nichyporuk, B., Arnold, D. L., and Arbel, T. (2022). "Personalized prediction of future lesion activity and treatment effect in multiple sclerosis from baseline MRI," in *Proceedings of The 5th International Conference on Medical Imaging with Deep Learning* (New York: PMLR), 172.

Dwyer, M., Lyman, C., Ferrari, H., Bergsland, N., Fuchs, T. A., Jakimovski, D., et al. (2021). DeepGRAI (Deep Gray Rating via Artificial Intelligence): fast, feasible, and clinically relevant thalamic atrophy measurement on clinical quality T2-FLAIR MRI in multiple sclerosis. *NeuroImage Clin.* 30, 102652. doi: 10.1016/j.nicl.2021.102652

Eitel, F., Soehler, E., Bellmann-Strobl, J., Brandt, A. U., Ruprecht, K., Giess, R. M., et al. (2019). Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. *NeuroImage: Clinical* 24:102003. doi: 10.1016/j.nicl.2019.102003

Ekşi, Z., Özcan, E., Çakıroğlu, M., Öz, C., and Aralaşmak, A. (2021). Differentiation of multiple sclerosis lesions and low-grade brain tumors on MRS data: machine learning approaches. *Neurol Sci* 42, 389–3395. doi: 10.1007/s10072-020-04950-0

Eshaghi, A., Young, A. L., Wijeratne, P. A., Prados, F., Arnold, D. L., Narayanan, S., et al. (2021). Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat. Commun.* 12:2078. doi: 10.1038/s41467-021-22265-2

Falet, J.-P. R., Durso-Finley, J., Nichyporuk, B., Schroeter, J., Bovis, F., Sormani, M.-P., et al. (2022). Estimating individual treatment effect on disability progression in multiple sclerosis using deep learning. *Nat. Commun.* 26:1. doi: 10.1038/s41467-022-33269-x

Falvo, A., Comminiello, D., Scardapane, S., Scarpiniti, M., and Uncini, A. (2019). "A multimodal dense U-Net for accelerating multiple sclerosis MRI," in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)* (Pittsburgh, PA: IEEE), 1–6.

Filippi, M., and Agosta, F. (2010). Imaging biomarkers in multiple sclerosis. *J. Magnetic Res. Imag.* 31, 770–788. doi: 10.1002/jmri.22102

Filippi, M., Preziosa, P., Banwell, B. L., Barkhof, F., Ciccarelli, O., De Stefano, N., et al. (2019). Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. *Brain* 142, 1858–1875. doi: 10.1093/brain/awz144

Filippi, M., Preziosa, P., Langdon, D., Lassmann, H., Paul, F., Àlex Rovira, S.choonheim, M. M., et al. (2020). Identifying progression in multiple sclerosis: new perspectives. *Ann. Neurol.* 88, 438–452. doi: 10.1002/ana.25808

Finck, T., Li, H., Grundl, L., Eichinger, P., Bussas, M., Mühlau, M., et al. (2020). Deep-learning generated synthetic double inversion recovery images improve multiple sclerosis lesion detection. *Invest. Radiol.* 55, 318–323. doi: 10.1097/RLI.0000000000000640

Finck, T., Li, H., Schlaeger, S., Grundl, L., Sollmann, N., Bender, B., et al. (2022). Uncertainty-aware and lesion-specific image synthesis in multiple sclerosis magnetic resonance imaging: a multicentric validation study. *Front. Neurosci.* 16:889808. doi: 10.3389/fnins.2022.889808

Fraser, H., Coiera, E., and Wong, D. (2018). Safety of patient-facing digital symptom checkers. *Lancet* 392, 2263–2264. doi: 10.1016/S0140-6736(18)32819-8

Gaj, S., Ontaneda, D., and Nakamura, K. (2021). Automatic segmentation of gadolinium-enhancing lesions in multiple sclerosis using deep learning from clinical MRI. *PLoS ONE* 16:e0255939. doi: 10.1371/journal.pone.0255939

García-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L., and Collins, D. L. (2013). Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Med. Image Anal.* 17, 1–18. doi: 10.1016/j.media.2012.09.004

Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., et al. (2023). A survey of uncertainty in deep neural networks. *Artif. Intellig. Rev.* 56, 1513–1589. doi: 10.1007/s10462-023-10562-9

Gentile, G., Jenkinson, M., Griffanti, L., Luchetti, L., Leoncini, M., Inderyas, M., et al. (2023). Bianca-ms: An optimized tool for automated multiple sclerosis lesion segmentation. *Hum. Brain Mapp.* 44, 4893–4913. doi: 10.1002/hbm.26424

Golkov, V., Dosovitskiy, A., Sperl, J. I., Menzel, M. I., Czisch, M., Sämann, P., et al. (2016). Q-space deep learning: twelve-fold shorter and model-free diffusion MRI scans. *IEEE Trans. Med. Imaging* 35, 1344–1351. doi: 10.1109/TMI.2016.2551324

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. *arXiv* [preprint] arXiv:1406.2661. doi: 10.48550/arXiv.1406.2661

Gros, C., De Leener, B., Badji, A., Maranzano, J., Eden, D., Dupont, S. M., et al. (2019). Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *Neuroimage* 184, 901–915. doi: 10.1016/j.neuroimage.2018.09.081

Gros, C., Lemay, A., and Cohen-Adad, J. (2021). Softseg: Advantages of soft versus binary training for image segmentation. *Med. Image Anal.* 71:102038. doi: 10.1016/j.media.2021.102038

Habek, M., Pavičić, T., Ruška, B., Pavlović, I., Gabelić, T., Barun, B., et al. (2018). Establishing the diagnosis of multiple sclerosis in croatian patients with clinically isolated syndrome: 2010 versus 2017 mcdonald criteria. *Mult. Scler. Relat. Disord.* 25, 99–103. doi: 10.1016/j.msard.2018.07.035

Havaei, M., Guizard, N., Chapados, N., and Bengio, Y. (2016). "Hemis: Hetero-modal image segmentation," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference* (Athens: Springer), 469–477.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV: IEEE), 770–778.

Hemond, C. C., and Bakshi, R. (2018). Magnetic resonance imaging in multiple sclerosis. *Cold Spring Harb. Perspect. Med.* 8:5. doi: 10.1101/cshperspect.a028969

Huang, C., Chen, W., Liu, B., Yu, R., Chen, X., Tang, F., et al. (2022). Transformer-based deep-learning algorithm for discriminating demyelinating diseases of the central nervous system with neuroimaging. *Front. Immunol.* 13:897959. doi: 10.3389/fimmu.2022.897959

Iwamura, M., Ide, S., Sato, K., Kakuta, A., Tatsuo, S., Nozaki, A., et al. (2023). Thin-slice two-dimensional T2-weighted imaging with deep learning-based reconstruction: improved lesion detection in the brain of patients with multiple sclerosis. *Magnetic Res. Med. Sci.* 2022:0112. doi: 10.2463/mrms.mp.2022-0112

Jannai, D., Meron, A., Lenz, B., Levine, Y., and Shoham, Y. (2023). Human or not? A gamified approach to the turing test. *arXiv* [preprint] arXiv:2305.20010. doi: 10.48550/arXiv.2305.20010

Jiang, L. Y., Liu, X. C., Nejatian, N. P., Nasir-Moin, M., Wang, D., Abidin, A., et al. (2023). Health system-scale language models are all-purpose prediction engines. *Nature* 2023, 1–6. doi: 10.1038/s41586-023-06160-y

Kamraoui, R. A., Mansencal, B., Manjon, J. V., and Coupé, P. (2022). Longitudinal detection of new MS lesions using deep learning. *Front. Neuroimag.* 1:948235. doi: 10.3389/fnimg.2022.948235

Kanber, B., Nachev, P., Barkhof, F., Calvi, A., Cardoso, J., Cortese, R., et al. (2019). High-dimensional detection of imaging response to treatment in multiple sclerosis. *NPJ Digital Med.* 2:49. doi: 10.1038/s41746-019-0127-8

Karimaghaloo, Z., Shah, M., Francis, S. J., Arnold, D. L., Collins, D. L., and Arbel, T. (2010). "Detection of gad-enhancing lesions in multiple sclerosis using conditional random fields," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2010: 13th International Conference* (Beijing: Springer), 41–48.

Kim, H., Lee, Y., Kim, Y.-H., Lim, Y.-M., Lee, J. S., Woo, J., et al. (2020). Deep learning-based method to differentiate neuromyelitis optica spectrum disorder from multiple sclerosis. *Front. Neurol.* 11:599042. doi: 10.3389/fneur.2020.599042

Koopmans, R. A., Li, D. K., Oger, J. J., Mayo, J., and Paty, D. W. (1989). The lesion of multiple sclerosis: Imaging of acute and chronic stages. *Neurology* 39, 959–963. doi: 10.1212/WNL.39.7.959

Kumar, A., Hu, A., Nichyporuk, B., Falet, J.-P. R., Arnold, D. L., Tsaftaris, S., et al. (2022). "Counterfactual image synthesis for discovery of personalized predictive image markers," in *MICCAI Workshop on Medical Image Assisted Biomarkers' Discovery* (Springer), 113–124.

La Rosa, F., Abdulkadir, A., Fartaria, M. J., Rahmanzadeh, R., Lu, P.-J., Galbusera, R., et al. (2020). Multiple sclerosis cortical and wm lesion segmentation at 3t MRI: a deep learning method based on flair and mp2rage. *NeuroImage: Clinical* 27:102335. doi: 10.1016/j.nicl.2020.102335

Lassmann, H. (2019). Pathogenic mechanisms associated with different clinical courses of multiple sclerosis. *Front. Immunol.* 9:3116. doi: 10.3389/fimmu.2018.03116

Law, M. T., Traboulsee, A. L., Li, D. K., Carruthers, R. L., Freedman, M. S., Kolind, S. H., et al. (2019). Machine learning in secondary progressive multiple sclerosis: an improved predictive model for short-term disability progression. *Multiple Sclerosis J.-Exp. Transl. Clini.* 5:2055217319885983. doi: 10.1177/2055217319885983

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., et al. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1, 541–551. doi: 10.1162/neco.1989.1.4.541

Lemay, A., Gros, C., Naga Karthik, E., and Cohen-Adad, J. (2022). Label fusion and training methods for reliable representation of inter-rater uncertainty. *Mach. Learn. Biomed. Imag.* 1:1–27. doi: 10.59275/j.melba.2022-db5c

Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2022). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 6999–7019. doi: 10.1109/TNNLS.2021.3084827

Lin, W., Hasenstab, K., Cunha, G. M., and Schwartzman, A. (2020). Comparison of handcrafted features and convolutional neural networks for liver MR image adequacy assessment. *Sci. Rep.* 10:77264. doi: 10.1038/s41598-020-77264-y

Lou, C., Sati, P., Absinta, M., Clark, K., Dworkin, J. D., Valcarcel, A. M., et al. (2021). Fully automated detection of paramagnetic rims in multiple sclerosis lesions on 3t susceptibility-based MR imaging. *NeuroImage: Clinical* 32:102796. doi: 10.1016/j.nicl.2021.102796

Lublin, F. D., Häring, D. A., Ganjgahi, H., Ocampo, A., Hatami, F., Čuklina, J., et al. (2022). How patients with multiple sclerosis acquire disability. *Brain* 145, 3147–3161. doi: 10.1093/brain/awac016

Lublin, F. D., and Reingold, S. C. (1996). Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on clinical trials of new agents in multiple sclerosis. *Neurology* 46:907–911. doi: 10.1212/WNL.46.4.907

Luo, G., Xie, W., Gao, R., Zheng, T., Chen, L., and Sun, H. (2023). Unsupervised anomaly detection in brain MRI: Learning abstract distribution from massive healthy brains. *Comput. Biol. Med.* 154:106610. doi: 10.1016/j.compbiomed.2023.106610

Maggi, P., and Absinta, M. (2024). Emerging MRI biomarkers for the diagnosis of multiple sclerosis. *Multiple Sclerosis J.* 30, 1704–1713. doi: 10.1177/13524585241293579

Maggi, P., Fartaria, M. J., Jorge, J., Rosa, F. L., Absinta, M., Sati, P., et al. (2020). CVSNet: a machine learning approach for automated central vein sign assessment in multiple sclerosis. *NMR Biomed.* 33:4283. doi: 10.1002/nbm.4283

Maier-Hein, L., Reinke, A., Godau, P., Tizabi, M. D., Buettner, F., Christodoulou, E., et al. (2024). Metrics reloaded: recommendations for image analysis validation. *Nat. Methods* 21, 195–212. doi: 10.1038/s41592-023-02151-z

Mangeat, G., Ouellette, R., Wabartha, M., Leener, B. D., Plattén, M., Karrenbauer, V. D., et al. (2020). Machine learning and multiparametric brain MRI to differentiate hereditary diffuse leukodystrophy with spheroids from multiple sclerosis. *J. Neuroimag.* 30, 674–682. doi: 10.1111/jon.12725

Mani, A., Santini, T., Puppala, R., Dahl, M., Venkatesh, S., Walker, E., et al. (2021). Applying deep learning to accelerated clinical brain magnetic resonance imaging for multiple sclerosis. *Front. Neurol.* 12:685276. doi: 10.3389/fneur.2021.685276

McDonald, W. I., Miller, D. H., and Thompson, A. J. (1994). Are magnetic resonance findings predictive of clinical outcome in therapeutic trials in multiple sclerosis? the dilemma of interferon-beta. *Ann. Neurol.* 36, 14–18. doi: 10.1002/ana.410360106

McKenzie, I. R., Lyzhov, A., Parrish, A., Prabhu, A., Mueller, A., Kim, N., et al. (2023). Inverse scaling: when bigger isn't better. *Transact. Mach. Learn. Res*. Available online at: https://openreview.net/forum?id=DwgRm72GQF

McKinley, R., Wepfer, R., Grunder, L., Aschwanden, F., Fischer, T., Friedli, C., et al. (2020). Automatic detection of lesion load change in multiple sclerosis using convolutional neural networks with segmentation confidence. *NeuroImage: Clinical* 25:102104. doi: 10.1016/j.nicl.2019.102104

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Comput. Surv*. 54:6. doi: 10.1145/3457607

Mescheriakova, J. Y., Wong, Y. Y. M., Runia, T. F., Jafari, N., Samijn, J. P., de Beukelaar, J. W., et al. (2018). Application of the 2017 revised mcdonald criteria for multiple sclerosis to patients with a typical clinically isolated syndrome. *JAMA Neurol.* 75, 1392–1398. doi: 10.1001/jamaneurol.2018.2160

Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., et al. (2023). Foundation models for generalist medical artificial intelligence. *Nature* 616, 259–265. doi: 10.1038/s41586-023-05881-4

Müller, J., Cagol, A., Lorscheider, J., Tsagkas, C., Benkert, P., Yaldizli, Ö., et al. (2023). Harmonizing definitions for progression independent of relapse activity in multiple sclerosis: A systematic review. *JAMA Neurol.* 80, 1232–1245. doi: 10.1001/jamaneurol.2023.3331

Nabizadeh, F., Masrouri, S., Ramezannezhad, E., Ghaderi, A., Sharafi, A. M., Soraneh, S., et al. (2022). Artificial intelligence in the diagnosis of multiple sclerosis: a systematic review. *Mult. Scler. Relat. Disord.* 59:103673. doi: 10.1016/j.msard.2022.103673

Nabizadeh, F., Ramezannezhad, E., Kargar, A., Sharafi, A. M., and Ghaderi, A. (2023). Diagnostic performance of artificial intelligence in multiple sclerosis: a systematic review and meta-analysis. *Neurol. Sci.* 44, 499–517. doi: 10.1007/s10072-022-06460-7

Nair, T., Precup, D., Arnold, D. L., and Arbel, T. (2020). Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. *Med. Image Anal.* 59:101557. doi: 10.1016/j.media.2019.101557

Narayana, P. A., Coronado, I., Sujit, S. J., Wolinsky, J. S., Lublin, F. D., and Gabr, R. E. (2020). Deep learning for predicting enhancing lesions in multiple sclerosis from noncontrast MRI. *Radiology* 294, 398–404. doi: 10.1148/radiol2019191061

Neeb, H., Schenk, J., Neeb, H., and Schenk, J. (2019). Multivariate prediction of multiple sclerosis using robust quantitative MR-based image metrics. *Z. Med. Phys.* 29, 262–271. doi: 10.1016/j.zemedi.2018.10.004

Nichyporuk, B., Cardinell, J., Szeto, J., Mehta, R., Falet, J.-P., Arnold, D. L., et al. (2022). Rethinking generalization: The impact of annotation style on medical image segmentation. *Machine Learn. Biomed. Imag.* 1, 1–37. doi: 10.59275/j.melba.2022-2d93

O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., et al. (2019). Deep learning vs. traditional computer vision. *Adv. Intellig. Syst. Comput.* 943, 128–144. doi: 10.1007/978-3-030-17795-9_10

Patel, S. B., and Lam, K. (2023). ChatGPT: the future of discharge summaries? *Lancet Digital Health* 5, e107-e108. doi: 10.1016/S2589-7500(23)00021-3

Pellegrini, F., Copetti, M., Sormani, M. P., Bovis, F., de Moor, C., Debray, T. P., et al. (2020). Predicting disability progression in multiple sclerosis: Insights from advanced statistical modeling. *Multiple Sclerosis Journal* 26:1828–1836. doi: 10.1177/1352458519887343

Peng, Y., Zheng, Y., Tan, Z., Liu, J., Xiang, Y., Liu, H., et al. (2021). Prediction of unenhanced lesion evolution in multiple sclerosis using radiomics-based models: a machine learning approach. *Mult. Scler. Relat. Disord.* 53:102989. doi: 10.1016/j.msard.2021.102989

Pinaya, W. H., Tudosiu, P.-D., Gray, R., Rees, G., Nachev, P., Ourselin, S., et al. (2022). Unsupervised brain imaging 3D anomaly detection and segmentation with transformers. *Med. Image Anal.* 79:102475. doi: 10.1016/j.media.2022.102475

Pitt, D., Lo, C. H., Gauthier, S. A., Hickman, R. A., Longbrake, E., Airas, L. M., et al. (2022). Toward precision phenotyping of multiple sclerosis. *Neurol.-Neuroimmunol. Neuroinflammat.* 9:200025. doi: 10.1212/NXI.0000000000200025

Pontillo, G., Penna, S., Cocozza, S., Quarantelli, M., Gravina, M., Lanzillo, R., et al. (2022). Stratification of multiple sclerosis patients using unsupervised machine learning: a single-visit MRI-driven approach. *Eur. Radiol.* 32, 5382–5391. doi: 10.1007/s00330-022-08610-z

Prabhakar, C., Li, H. B., Paetzold, J. C., Loehr, T., Niu, C., Mühlau, M., et al. (2023). "Self-pruning graph neural network for predicting inflammatory disease activity in multiple sclerosis from brain MR images," in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2023*, eds. H. Greenspan, A. Madabhushi, P. Mousavi, S. Salcudean, J. Duncan, T. Syeda-Mahmood, et al. (Cham: Springer Nature Switzerland), 226–236.

Preul, M. C., Caramanos, Z., Collins, D. L., Villemure, J. G., Leblanc, R., Olivier, A., et al. (1996). Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. *Nature Medicine 1996 2:3* 2:323–325. doi: 10.1038/nm0396-323

Rae, J. W., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Song, F., et al. (2021). Scaling language models: methods, analysis & insights from training gopher. *arXiv* [preprint] arXiv:2112.11446. doi: 10.48550/arXiv.2112.11446

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., et al. (2021). "Zero-shot text-to-image generation," in *International Conference on Machine Learning* (New York: PMLR), 8821–8831.

Reinhold, J. C., Carass, A., and Prince, J. L. (2021). "A structural causal model for MR images of multiple sclerosis," in *Medical Image Computing and Computer Assisted Intervention-MICCAI 2021: 24th International Conference* (Strasbourg: Springer), 782–792.

Reinke, A., Tizabi, M. D., Baumgartner, M., Eisenmann, M., Heckmann-Nötzel, D., Kavur, A. E., et al. (2024). Understanding metric-related pitfalls in image analysis validation. *Nat. Methods* 21, 182–194. doi: 10.1038/s41592-023-02150-0

Richens, J. G., Lee, C. M., and Johri, S. (2020). Improving the accuracy of medical diagnosis with causal machine learning. *Nat. Commun.* 11:3923. doi: 10.1038/s41467-020-17419-7

Rocca, M. A., Anzalone, N., Storelli, L., Poggio, A. D., Cacciaguerra, L., Manfredi, A. A., et al. (2021). Deep learning on conventional magnetic resonance imaging improves the diagnosis of multiple sclerosis mimics. *Invest. Radiol.* 56, 252–260. doi: 10.1097/RLI.0000000000000735

Rosa, F. L., Beck, E. S., Maranzano, J., Todea, R. A., van Gelderen, P., de Zwart, J. A., et al. (2022). Multiple sclerosis cortical lesion detection with deep learning at ultra-high-field MRI. *NMR Biomed.* 35:e4730. doi: 10.1002/nbm.4730

Rudick, R. A., Lee, J.-C., Simon, J., and Fisher, E. (2006). Significance of T2 lesions in multiple sclerosis: a 13-year longitudinal study. *Ann. Neurol.* 60, 236–242. doi: 10.1002/ana.20883

Salem, M., Valverde, S., Cabezas, M., Pareto, D., Oliver, A., Salvi, J., et al. (2019). Multiple sclerosis lesion synthesis in MRI using an encoder-decoder U-Net. *IEEE Access* 7, 25171–25184. doi: 10.1109/ACCESS.2019.2900198

Salem, M., Valverde, S., Cabezas, M., Pareto, D., Oliver, A., Salvi, J., et al. (2020). A fully convolutional neural network for new T2-w lesion detection in multiple sclerosis. *NeuroImage: Clinical* 25:102149. doi: 10.1016/j.nicl.2019.102149

Sanchez, P., Voisey, J. P., Xia, T., Watson, H. I., ONeil, A. Q., and Tsaftaris, S. A. (2022). Causal machine learning for healthcare and precision medicine. *R. Soc. Open Sci.* 9:220638. doi: 10.1098/rsos.220638

Schroeter, J., Myers-Colet, C., Arnold, D. L., and Arbel, T. (2022). "Segmentation-consistent probabilistic lesion counting," in *International Conference on Medical Imaging with Deep Learning* (New York: PMLR), 1034–1056.

Seccia, R., Romano, S., Salvetti, M., Crisanti, A., Palagi, L., and Grassi, F. (2021). Machine learning use for prognostic purposes in multiple sclerosis. *Life* 11, 1–18. doi: 10.3390/life11020122

Shaul, R., David, I., Shitrit, O., and Raviv, T. R. (2020). Subsampled brain MRI reconstruction by generative adversarial neural networks. *Med. Image Anal.* 65:101747. doi: 10.1016/j.media.2020.101747

Shoeibi, A., Khodatars, M., Jafari, M., Moridian, P., Rezaei, M., Alizadehsani, R., et al. (2021). Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review. *Comput. Biol. Med.* 136:104697. doi: 10.1016/j.compbiomed.2021.104697

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., et al. (2023). Large language models encode clinical knowledge. *Nature* 620, 172–180. doi: 10.1038/s41586-023-06291-2

Solomon, A. J., Naismith, R. T., and Cross, A. H. (2019). Misdiagnosis of multiple sclerosis: Impact of the 2017 McDonald criteria on clinical practice. *Neurology* 92, 26–33. doi: 10.1212/WNL.0000000000006583

Sormani, M. P., Bonzano, L., Roccatagliata, L., Cutter, G. R., Mancardi, G. L., and Bruzzi, P. (2009). Magnetic resonance imaging as a potential surrogate for relapses in multiple sclerosis: a meta-analytic approach. *Ann. Neurol.* 65, 268–275. doi: 10.1002/ana.21606

Sormani, M. P., and Bruzzi, P. (2013). MRI lesions as a surrogate for relapses in multiple sclerosis: a meta-analysis of randomised trials. *Lancet Neurol.* 12, 669–676. doi: 10.1016/S1474-4422(13)70103-0

Spagnolo, F., Depeursinge, A., Schädelin, S., Akbulut, A., Müller, H., Barakovic, M., et al. (2023). How far MS lesion detection and segmentation are integrated into the clinical workflow? a systematic review. *NeuroImage: Clinical* 39:103491. doi: 10.1016/j.nicl.2023.103491

Storelli, L., Azzimonti, M., Gueye, M., Vizzino, C., Preziosa, P., Tedeschi, G., et al. (2022). A deep learning approach to predicting disease progression in multiple sclerosis using magnetic resonance imaging. *Invest. Radiol.* 57, 423–432. doi: 10.1097/RLI.0000000000000854

Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., et al. (2008). 3D segmentation in the clinic: a grand challenge II: MS lesion segmentation. *Midas J.* 2008, 1–6. doi: 10.54294/lmkqvm

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

Tousignant, A., Lemaître, P., Precup, D., Arnold, D. L., and Arbel, T. (2019). "Prediction of disease progression in multiple sclerosis patients using deep learning analysis of MRI data," in *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning*, eds. M. J. Cardoso, A. Feragen, B. Glocker, E. Konukoglu, I. Oguz, G. Unal, et al. (New York: PMLR), 483–492.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., et al. (2023). LLaMA: Open and efficient foundation language models. *arXiv* [preprint] arXiv:2302.13971. doi: 10.48550/arXiv.2302.13971

Urbina, F., Lentzos, F., Invernizzi, C., and Ekins, S. (2022). Dual use of artificial-intelligence-powered drug discovery. *Nat. Mach. Intellig.* 4, 189–191. doi: 10.1038/s42256-022-00465-9

Valencia, L., Clèrigues, A., Valverde, S., Salem, M., Oliver, A., Rovira, A., et al. (2022). Evaluating the use of synthetic T1-w images in new T2 lesion detection in multiple sclerosis. *Front. Neurosci.* 16:954662. doi: 10.3389/fnins.2022.954662

Van Nederpelt, D. R., Amiri, H., Brouwer, I., Noteboom, S., Mokkink, L. B., Barkhof, F., et al. (2023). Reliability of brain atrophy measurements in multiple sclerosis using MRI: an assessment of six freely available software packages for cross-sectional analyses. *Neuroradiology* 65, 1459–1472. doi: 10.1007/s00234-023-03189-8

Van Tulder, G., and de Bruijne, M. (2015). "Why does synthesized data improve multi-sequence classification?," in *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference* (Munich: Springer), 531–538.

Vollmer, T. L., Nair, K. V., Williams, I. M., and Alvarez, E. (2021). Multiple sclerosis phenotypes as a continuum the role of neurologic reserve. *Neurology* 11, 342–351. doi: 10.1212/CPJ.0000000000001045

Wei, W., Poirion, E., Bodini, B., Durrleman, S., Colliot, O., Stankoff, B., et al. (2019). Fluid-attenuated inversion recovery MRI synthesis from multisequence MRI using three-dimensional fully convolutional networks for multiple sclerosis. *J. Med. Imag.* 6, 014005–014005. doi: 10.1117/1.JMI.6.1.014005

Xu, W., Xu, Y., Chang, T., and Tu, Z. (2021). "Co-scale conv-attentional image transformers," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (Los Alamitos, CA: IEEE Computer Society), 9961–9970.

Young, A. L. (2018). Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nature Communications* 9, 1–16. doi: 10.1038/s41467-018-05892-0

Zare, M. R., Alebiosu, D. O., and Lee, S. L. (2018). "Comparison of handcrafted features and deep learning in classification of medical x-ray images," in *Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)* (Kota Kinabalu: IEEE), 1–5.

Zeng, C., Gu, L., Liu, Z., and Zhao, S. (2020). Review of deep learning approaches for the segmentation of multiple sclerosis lesions on brain MRI. *Front. Neuroinform.* 14:610967. doi: 10.3389/fninf.2020.610967

Zhan, G., Wang, D., Cabezas, M., Bai, L., Kyle, K., Ouyang, W., et al. (2023). Learning from pseudo-labels: deep networks improve consistency in longitudinal brain volume estimation. *Front. Neurosci.* 17:1196087. doi: 10.3389/fnins.2023.1196087

Zhang, H., Bakshi, R., Bagnato, F., and Oguz, I. (2020). "Robust multiple sclerosis lesion inpainting with edge prior," in *Machine Learning in Medical Imaging: 11th International Workshop, MLMI 2020, Held in Conjunction with MICCAI 2020* (Lima: Springer), 120–129.

Zhang, H., Nguyen, T. D., Zhang, J., Marcille, M., Spincemaille, P., Wang, Y., et al. (2022). QSMRim-Net: imbalance-aware learning for identification of chronic active multiple sclerosis lesions on quantitative susceptibility maps. *NeuroImage: Clinical* 34:102979. doi: 10.1016/j.nicl.2022.102979

Zhang, K., Lincoln, J. A., Jiang, X., Bernstam, E. V., and Shams, S. (2023). Predicting multiple sclerosis severity with multimodal deep neural networks. *BMC Med. Inform. Decis. Mak.* 23, 1–17. doi: 10.1186/s12911-023-02354-6

Zhang, Y., Hong, D., Mcclement, D., Oladosu, O., Pridham, G., and Slaney, G. (2021). Grad-cam helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. *J. Neurosci. Methods* 353:109098. doi: 10.1016/j.jneumeth.2021.109098

Zhao, C., Shao, M., Carass, A., Li, H., Dewey, B. E., Ellingsen, L. M., et al. (2019). Applications of a deep learning method for anti-aliasing and super-resolution in MRI. *Magn Reson Imaging* 64, 132–141. doi: 10.1016/j.mri.2019.05.038

Zhao, Y., Healy, B. C., Rotstein, D., Guttmann, C. R., Bakshi, R., Weiner, H. L., et al. (2017). Exploration of machine learning techniques in predicting multiple sclerosis disease course. *PLoS ONE* 12:e0174866. doi: 10.1371/journal.pone.0174866

Zhao, Y., Wang, T., Bove, R., Cree, B., Henry, R., Lokhande, H., et al. (2020). Ensemble learning predicts multiple sclerosis disease course in the summit study. *NPJ Digit. Med.* 3:135. doi: 10.1038/s41746-020-00338-8