Artificial intelligence applied to MRI data to tackle key challenges in multiple sclerosis

Sara Collorone[1], Llucia Coll[2], Marco Lorenzi[3], Xavier Lladó[4], Jaume Sastre-Garriga[2], Mar Tintoré[2], Xavier Montalban[2], Àlex Rovira[5], Deborah Pareto[5], Carmen Tur[1,2]

1 NMR Unit. Queen Square MS Centre, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, University College London, London, UK

2 Multiple Sclerosis Centre of Catalonia (Cemcat), Department of Neurology, Hospital Universitari Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain

3 Université Côte d'Azur, Inria Sophia Antipolis, Epione Research Project, France

4 Research Institute of Computer Vision and Robotics, University of Girona, Girona, Spain

5 Section of Neuroradiology, Department of Radiology (IDI), Vall d'Hebron University Hospital, Spain. Universitat Autònoma de Barcelona, Barcelona, Spain

Word count: Abstract: 139 words. Main manuscript: 2838 words. Number of references: 38.

Co-corresponding authors: Carmen Tur, email address: ctur@cem-cat.org; Sara Collorone, email address: s.collorone@ucl.ac.uk;

Abstract (150 words)

Artificial intelligence (AI) is the branch of science aiming at creating algorithms able to carry out tasks that typically require human intelligence. In medicine, there has been a tremendous increase in AI applications thanks to increasingly powerful computers and the emergence of big data repositories. Multiple sclerosis (MS) is a chronic autoimmune condition affecting the central nervous system with a complex pathogenesis, a challenging diagnostic process strongly relying on Magnetic Resonance Imaging (MRI), and a high and largely unexplained variability across patients. Therefore, AI applications in MS have the great potential of helping us better support the diagnosis, find markers for prognosis to eventually design more powerful randomised clinical trials and improve patient management in clinical practice, and eventually understand the mechanisms of the disease. This topical review aims to summarise the recent advances in AI applied to MRI data in MS to illustrate its achievements, limitations, and future directions.

INTRODUCTION

Artificial intelligence (AI) is the branch of science aiming at creating algorithms able to carry out tasks that typically require human intelligence.[1–3] AI methodologies used for medical research have mainly two objectives. On the one hand, AI can extract patterns from the data to understand their internal structure or classify them. For this purpose, data does not need to be *annotated*, that is, the output (e.g., the belonging to a given class) is unknown (unsupervised learning). On the other hand, AI can make predictions from the data available, which is *annotated.* That is, in such available data, not only the input data (e.g., set of predictors) but also the output data (e.g., disability milestone or death) are known and well characterised (supervised learning), allowing the *learning* of the *link* between input and output data. In both cases, AI usage goes through two phases: first, there is the creation of a model, with a series of iterations (i.e., the training) and tests until we obtain the desired output; then, there is a second phase characterised by the model application on new data (**Figure 1**).

The AI and machine learning (ML) approaches referred to here represent a set of mathematical algorithms able to learn from the data and make predictions on unseen information. A subset of ML relies on networks of multiple layers of artificial neurons and is defined as deep learning (DL) (**Table 1**). DL does not necessarily need pre-selected variables as input but can extract the variables necessary for the network task (**Figure 2**). For instance, when defining network's layers through convolution operators, DL allows to optimise spatial filters to automatically extract relevant image features to solve a given task.

Multiple sclerosis (MS) is a chronic inflammatory-demyelinating condition of the central nervous system (CNS). People with MS (pwMS) may live with a high physical and emotional burden with social and economic implications for both pwMS and healthcare systems. Disease diagnosis and monitoring strongly rely on routinely acquired brain and spinal cord magnetic resonance imaging (MRI). AI applications in MS can potentially help us better support the diagnosis, find markers for prognosis, facilitate accurate monitoring, and eventually

understand the mechanisms of the disease. Focusing on these main challenges, this review aims to summarise the recent advances in AI applied to MRI in MS, highlighting the key features of the most representative studies, as opposed to other, broader reviews which have been recently published.[4,5] This review also aims to illustrate its limitations and future directions.

SEARCH STRATEGY

We performed a search in PubMed based on the following criteria: (i) publication date: between 2013 and 2023; (ii) search terms: (multiple sclerosis or demyelination or demyelinating disease) AND (artificial intelligence or deep learning or machine learning) AND (MRI or neuroimaging or brain/cord imaging); (iii) language of publication: English; (iv) type of paper: original research. For the purpose of this narrative review, we have focused on four aspects: (i) diagnosis & differential diagnosis; (ii) prediction of clinical outcome; (iii) MRI lesion identification and segmentation, including detection of new and enlarging lesions; (iv) understanding of pathogenic mechanisms. Thus, after the first literature search, we manually selected the papers if they were included in one of these four categories. Papers not clearly included in any of these categories were not considered in the review. Thus, we did not include papers whose main focus was methodological or animal research. We also excluded review papers, editorials, and case reports. The PubMed search yielded 411 articles, published between 2013 and 2023, both included (**Figure 3**). After excluding those not meeting our inclusion criteria, we revised 185 papers for their inclusion in this narrative review (**Figure 3**). Most of these papers have been published between 2021 and 2023 (**Figure 4**).

MS DIAGNOSIS

The diagnosis of MS relies on integrating clinical, MRI, and laboratory findings to demonstrate disease dissemination in space and time and exclude alternative diagnoses, especially in the presence of the so-called red flags.[6] Indeed, the diagnosis of MS is not devoid of challenges: other conditions may mimic MS, clinically or radiologically. Additionally, there are special

populations, such as non-Caucasic individuals, those with comorbidities, paediatric patients, and older adults, where diagnostic criteria must be applied more carefully, and the differential diagnosis needs to be expanded.[6] In these circumstances, the use of AI algorithms to reach the correct diagnosis may be very useful.

In 2016, Eshaghi et al.[7] used a ML model to extract 157 features from T1-weighted (T1w) images to differentiate MS (N=25) from neuro-myelitis optica spectrum disorder (NMOSD) (N=30), and healthy controls (HCs) (N=35). The ML model achieved an accuracy of 74%, opening new possibilities for utilising ML for the differential diagnosis of MS. Rocca and colleagues compared the accuracy of a DL model from both T1w and T2-weighted (T2w) images based on convolutional neural networks (CNNs) to the one achieved by two experienced neuroradiologists to classify patients into different conditions: MS (N=70), NMOSD (N=91), vasculitis of the CNS (N=51), or migraine (N=56).[8] In all conditions, the DL model achieved a higher accuracy (around 90% or above) than human evaluators (**Table 2** for details). More recently, Seok et al.[9] also used a DL model based on CNN to differentiate between MS (N=86) and NMOSD (N=70) using 3D fluid attenuated inversion recovery (FLAIR) images and obtaining 76.1% accuracy. Other studies based on AI models have achieved similarly high accuracy differentiating MS from other conditions (**Table 2**).

All these studies have shown promising results, but there are still several limitations that need to be acknowledged, including relatively small datasets and heterogeneity of input data, discussed below.

PREDICTION OF MS EVOLUTION

MS is a highly heterogeneous condition with variable course, both in relation to the occurrence of future relapses and development of irreversible disability. Over the years, observational studies have identified MRI predictors of worse prognosis including a greater number of demyelinating lesions,[10] the presence of infratentorial,[11] cortical,[12] spinal cord,[13] or new lesions

in a short-term follow-up.[13] Importantly, all these risk factors are usually taken into account in the clinic to monitor patients and decide treatment strategies. However, the factors mentioned only account for a relatively small proportion of the variability in clinical outcomes. So, the potential for AI models to make a substantial contribution is clear.

Relapses

Only a few studies have used AI methodologies to predict the risk of future relapses at MS onset. Among the most relevant publications there is the one from Wottschel et al.,[14] which applied ML models to morphometric measures from MRI scans (both T1w and FLAIR/T2w images) of 400 patients at symptom onset to predict the occurrence of a second relapse over one year, achieving an accuracy of 71%. They showed that a greater whole brain T2 lesion load, a lower grey matter volume in the thalamus and the precuneus region, and a thinner cortex in the cuneus and inferior temporal gyrus at first attack were the best predictors of a second attack. Another study[15] carried out in in 364 patients with a first demyelinating attack which also used ML to predict a second clinical attack based on MRI scans (T1w after gadolinium administration and T2w images) achieved an accuracy of 70%. This accuracy was only achieved, though, when both MRI features, i.e., geometric features of individual lesions and whole-brain and region-of-interest-based volumes, and clinical & demographic characteristics were used.[15] Finally, a more recent study[16] used both a global and regional ML approach to predict the occurrence of a second clinical event over a three-year follow-up from brain MRI scans (T1w images) of 266 patients at disease onset. Specifically, input data were grey matter characteristics and white matter T1-hypointensities from 3D-T1-weighted images. Based on grey matter characteristics, the models achieved an accuracy of approximately 50%. After the inclusion of T1-hypointensities, the accuracy did not improve (see **Table 3** for more details). All these studies suggest that models based on input data from a single modality (e.g., T1w) have lower accuracy rates than those based on more complex MRI data or the combination of clinical and MRI-derived features.

Disability accumulation

Predicting disability accrual in pwMS has been a priority for the scientific community over the last years,[10] as this prediction has an impact on patient management. Studies have demonstrated that both high inflammatory lesion load and CNS atrophy observed in the MRI scan are associated with worse long-term prognosis.[10,17]

Among the AI studies which have focused on short- or mid-term predictions, Tousignant et al.[18] developed a DL model using multi-modal MRI data (T1w, T2w, T2-FLAIR, and post-contrast T1w images) of 465 pwMS from placebo arms of randomised clinical trials as input data to predict disease progression at one-year follow-up. When they only used data from the MR raw images (see **Table 3**), their model had a moderate performance with an Area Under the Curve (AUC) of 0.66 (± 0.055). However, when they supplemented the model with lesion masks from T2-weighted and contrast-enhanced T1-weighted sequences, the AUC improved to 0.701 (± 0.027), highlighting the importance of rich and informative input data for model performance. In 2020, Roca et al.[19] assessed the performance of a DL CNN model which used age, sex, and brain T2-FLAIR scans from 971 pwMS to predict EDSS scores at two-year follow-up. Their model showed moderate performance (mean EDSS score error: 1.7), which was worse for very low or high EDSS values. This possibly suggests the need for an adequate representation of the different output options when training the model to allow a correct learning process. Storelli et al.[20] also used a DL CNN to predict disability at two years in 48 pwMS, using both T1w and T2w images from a multi-centre training set of 325 pwMS. The output was a binary outcome for disability progression, based on EDSS and symbol digit modalities test (SDMT) scores. The CNN model showed a predictive accuracy that was high for EDSS (83.3%) and moderate for SMDT (67.7%) worsening, although the highest accuracy was achieved using both tests (85.7%).[20]

Finally, regarding the long-term predictions of disability using AI models, Zhao et al.[21] built a series of ML models to predict disability progression (binary output) at 5-year follow-up in 1693 pwMS. Input data were baseline and short-term follow-up, clinical and MRI data (T2w images). Of note, the best accuracy was achieved when the model included both MRI and clinical data from baseline and follow-up (around 75%) (**Table 3**), highlighting the need for both informative and rich input data to ensure model accuracy.

Segmentation of MRI lesions

The most widely used non-clinical tool for MS monitoring is the MRI. However, the assessment of the number of lesions and of the presence of new or enlarged lesions in follow-up MRI scans can be very time-consuming and requires expert neuroradiologists. Therefore, a great effort has been made over the years to develop strategies (software) to identify and segment lesions in T2-FLAIR images[22] and to identify the presence of new or enlarged lesions.[23,24]

In relation to the identification and segmentation of brain T2-FLAIR lesions, different studies have proposed DL algorithms that are emerging with increasing accuracy.[25] Yet the comparison across algorithms is challenging because they often use different datasets, not always publicly available. For the detection of new or enlarged lesions on follow-up MRI scans, also several algorithms have been proposed. In 2020, Salem et al. published the first longitudinal approach based on CNNs that dealt with lesion changes in brain MRI.[26] More recently, the same group has published an improved method also based on CNNs for new lesion segmentation.[27] Finally, regarding the identification of active lesions based on non-contrast MRI, a study on 1008 pwMS[28] was able to detect the presence and location (i.e., the MRI slice) of gadolinium-enhancing lesions using a CNN model based on pre-contrast T1w, T2w, and T2-FLAIR images, with high sensitivity and specificity of 78% and 73%, respectively.[28] After this study, other authors have also investigated DL algorithms to identify active lesions based on non-contrast MRI, showing high accuracy levels too.[29–32] Please see

Table 4 for more details on lesion segmentation studies carried out using AI-based algorithms. However, given the narrative, non-systematic nature of this review, a number of other relevant papers could not be included, despite their relevance.[33,34] Furthermore, it is to be highlighted that, although most lesion segmentation algorithms are based on DL, not all of them are, and still they may achieve high performances. For instance, a recent study presented a method called *Multiple Sclerosis Lesion Analysis at Seven Tesla* (MSLAST), based on connected component analysis, which was able to identify MS lesions from ultra-high-field MR images, with high sensitivity (71%).[35]

In sum, DL has shown a great potential for improving the performance of available automatic lesion segmentation tools. Additionally, apart from MRI lesions, new DL algorithms have also shown a great value in segmenting CNS tissues, mainly in the brain but also in the spinal cord. However, this topic was beyond the scope of this narrative review and should be covered by future reviews.

INVESTIGATION OF DISEASE MECHANISMS

The pathophysiological processes underlying disease progression in MS are not completely understood and are believed to be highly heterogeneous across people and stages of the disease. Conventional MRI techniques offer a range of metrics which are not specific to these processes and their pathological manifestations.[36] Consequently, these metrics have low to moderate correlations with clinical parameters.[37] In this context, AI may help us understand the pathogenetic mechanisms of MS.

AI-based focused on understanding disease mechanisms have mainly used two strategies, i.e., unsupervised ML models and DL-derived attention maps (**Table 5**). Among the studies using the first strategy, Eshaghi et al.[38] applied an unsupervised ML algorithm (SuStaIn[39]) on MRI scans (T1w, T2w, T2-FLAIR images) of 6322 pwMS to determine different longitudinal patterns of brain pathology. They identified three different patterns, i.e., cortex-led, with a more

neurodegenerative component, corresponding to those patients whose pathology began in the cortex; normal-appearing white matter-led, with a more chronic inflammatory component, corresponding to those patients whose pathology began in the normal-appearing white matter; and lesion-led, with a more acute inflammatory component, corresponding to pwMS whose pathology was initiated by visible lesions. The lesion-led subtype had the worse prognosis, with a faster progression of disability.[38] This model was later on applied by Pontillo et al.[40] to a different cohort of 425 pwMS also using T1w and T2-FLAIR scans. Although they found two patterns of disease instead of three, they were able to identify, as Eshaghi et al. did, a deep grey matter-led pattern associated with T2 lesion-related damage which implied a worse prognosis. These studies show the potential of AI algorithms to provide an accurate patient stratification that is both biologically reliable and prognostically meaningful.[38,40]

Regarding the studies that have used attention maps, which reflect those anatomical regions that the DL model deems more relevant to make a given DL-based prediction, Eitel et al.[41] found that posterior periventricular white matter regions were determinant for the diagnosis of MS. More recently, Coll et al.[42] found that the areas identified as most relevant to classify patients into more or less disabled (i.e., EDSS≥3.0 vs EDSS<3.0) were the frontotemporal cortex and cerebellum (**Figure 5**), suggesting that damage in these regions may be key to disability accrual (please see **Table 5** for more details).

Other AI-based strategies, such as supervised classification ML or DL models, have also been exploited in several studies to uncover the pathological processes underlying disability progression in MS. However, only in a few of those studies this was the primary aim. In this regard, a recent paper exploited the patterns of brain functional connectivity derived from resting state functional MRI to understand differences between relapsing and progressive forms of MS (**Table 5**).[43]

To conclude, although the studies that have used AI to understand pathogenetic mechanisms in MS are still relatively few, they certainly contribute to a greater characterisation of MS by expanding the concept of classical phenotypes. Nonetheless, the integration with new quantitative MRI techniques that can show damage in apparently normal tissues,[44] even at very early disease stages,[45] would be necessary.

LIMITATIONS IN AI-BASED RESEARCH IN MS

Although AI offers promising results in MS, there are still many limitations concerning AI-based research, which hamper its medical applications. Some of these limitations are intrinsically related to the ML/DL methodology, such as the so-called "black box" issue, i.e., the lack of transparency in the decision-making process of the AI model.[46] The consequences of this are twofold: from a methodological perspective, this issue may inadvertently lead to error propagation in data analysis, since it may be virtually impossible to evaluate erroneous paths, causing the same model to obtain different results on different datasets. From a practical perspective, the black-box issue may cause distrust among clinicians and healthcare providers, slowing down its implementation in the clinic. Secondly, the use of relatively small datasets is a clear limitation of most studies. In this context, though, the need for large datasets, which are difficult to obtain, can lead to a preference for data quantity over data quality, resulting in suboptimal data quality to train the models, which is also a problem in AI studies. Additionally, there is a large methodological variability among the different algorithms proposed to answer a similar research question, challenging their translation to clinical practice. In this regard, it is essential to conduct quantitative assessments of DL models and use large datasets with different MRI scanners and imaging protocols to assess their true potential as diagnostic and monitoring tools in the clinic. It is therefore fundamental to intensify future efforts towards collecting large-scale datasets to train DL models and improve their performance and robustness. Furthermore, there is still the unmet need for AI application guidelines to set standards for models' accuracy and data adequacy. Despite the presence of international frameworks for reporting clinical studies based on AI models (i.e., TRIPOD-AI for

diagnostic and prognostic prediction model studies[47]) there is still a large variability in its compliance among researchers. Finally, there are general limitations concerning the use of AI in healthcare that relate to ethical and equality issues that are beyond the scope of this review, but they will be part of the future scientific and public debate influencing future AI applications.

FUTURE DIRECTIONS

Future steps include the translation of AI-based research to clinical practice, as well as the development of new AI-based methodologies that help us tackle key challenges in MS beyond differential diagnosis and disease prediction, which are the areas where AI research applied to MS data has developed the most. For instance, AI applications may help design more efficient randomised clinical trials (RCTs), for instance selecting *a priori* those study participants with greatest likelihood to respond to treatment. This process, called predictive enrichment, was successfully explored by Falet and colleagues, who leveraged clinical and imaging data from six randomised clinical trials to predict treatment response.[48] Regarding those healthcare areas where AI-based research is most developed and closest to its use in clincial practice, there is radiology.[49] Hence, numerous AI-based devices for image analysis have received approval in both Europe and the USA, with a notable prevalence of those specifically associated with neurology, mainly stroke imaging.[50] For instance, several algorithms for an automatic identification of signs of acute ischaemic stroke in brain CT imaging, needed for the Alberta Stroke Program Early CT Score (ASPECTS), have already been commercialised.[50] In the MS field, a few algorithms for automatic lesion (and brain tissue) segmentation have also been commercialised. Nevertheless, the peer-reviewed scientific evidence supporting the efficacy of many of the commercialised AI-based products (in general, not only those related to neurology) is frequently absent or inaccessible.[50] This, in part, may account for the relatively limited adoption of these technologies in clinical practice, together with logistic and budgetary reasons. Therefore, future endeavours should focus on rigorous scientific validation to ensure a definitive integration of all these AI-based algorithms into clinical settings.

CONCLUSIONS

The use of AI in MS has made significant progresses in recent years. There is a growing recognition of the potential for AI to contribute to the diagnosis, monitoring, and prediction of MS. However, several challenges are still present in relation to AI-based research and are probably responsible for the significant gap between AI studies and their clinical utility. One of the challenges when using AI for diagnosis monitoring and prognosis in MS is the quality of the input data. Standardised MRI protocols and accurate and comprehensive data collection are essential for developing reliable AI models. Another challenge is ensuring the reproducibility of methods used in AI-based studies, particularly considering the use of different MRI scanners and imaging protocols across centres. The ability to reproduce and validate results is key to accepting and integrating AI into clinical practice. Additionally, we should adequately address the ethical issues derived from the implementation of AI to diagnose, monitor, and predict MS. Concerns about privacy, data security and potential biases in algorithmic decision-making need to be tackled to ensure ethical standards are met. Nonetheless, the high number of opportunities identified in relation to the use of AI in MS research and clinical practice will hopefully help address and overcome these challenges quickly. Furthermore, there is an unmet need for research to create AI prediction models capable of integrating longitudinal MRI data and exploring optimal methods for merging information from MR images and clinical data. Finally, in the future, AI may also enable better trial design and a deeper understanding of the mechanisms underlying irreversible disability accumulation, leading to more effective treatments and interventions.

DISCLOSURES

SC is supported by the Rosetrees Trust (A1332 and MS632) and was awarded a MAGNIMS-ECTRIMS fellowship in 2016.

LC reports no disclosures.

ML is partially supported by the French government labelled *PIA* program under its IDEX UCA JEDI project (ANR-15-IDEX-0001)

XL is currently being supported by the ICREA Academia Program. He has also received support from the DPI2020-114769RBI00 project funded by the Ministerio de Ciencia, Innovación y Universidades.

JSG serves as co-Editor for Europe for the Multiple Sclerosis Journal and as Editor-in-Chief of Revista de Neurología, receives research support from Fondo de Investigaciones Sanitarias (19/950 and 22/750) and in the last twelve months has served as a consultant/speaker for BMS, Roche, Sanofi, Janssen, and Merck.

MT has received compensation for consulting services, speaking honoraria and research support from Almirall, Bayer Schering Pharma, Biogen-Idec, Genzyme, Janssen, Merck-Serono, Novartis, Roche, Sanofi-Aventis, Viela Bio and Teva Pharmaceuticals. Data Safety Monitoring Board for Parexel and UCB Biopharma.

XM has received speaking honoraria and travel expenses for participation in scientific meetings, has been a steering committee member of clinical trials or participated in advisory boards of clinical trials in the past years with Abbvie, Actelion, Alexion, Biogen, Bristol-Myers Squibb/Celgene, EMD Serono, Genzyme, Hoffmann-La Roche, Immunic, Janssen Pharmaceuticals, Medday, Merck, Mylan, Nervgen, Novartis, Sandoz, Sanofi Genzyme, Teva Pharmaceutical, TG Therapeutics, Excemed, MSIF and NMSS.

AR serves/ed on scientific advisory boards for Novartis, Sanofi-Genzyme, Synthetic MR, TensorMedical, Roche, and Biogen, and has received speaker honoraria from Bayer, Sanofi-Genzyme, Merck-Serono, Teva Pharmaceutical Industries Ltd, Novartis, Roche, Bristol-Myers and Biogen, is CMO and co-founder of TensorMedical, and receives research support from

Fondo de Investigación en Salud (PI19/00950 and PI22/01589) from Instituto de Salud Carlos III, Spain.

DP has received a research grant from Biogen Idec, and grant support from Instituto Salud Carlos III co-funded by the European Union (PI18/00823, PI22/01709).

C. Tur is currently being funded by a Miguel Servet contract, awarded by the Instituto de Salud Carlos III (ISCIII), Ministerio de Ciencia e Innovación de España (CP2300117). She has also received a 2020 Junior Leader La Caixa Fellowship (fellowship code: LCF/BQ/PI20/11760008), awarded by “la Caixa” Foundation (ID 100010434), a 2021 Merck’s Award for the Investigation in MS, awarded by Fundación Merck Salud (Spain), a 2021 Research Grant (PI21/01860) awarded by the ISCIII, Ministerio de Ciencia e Innovación de España, and a FORTALECE research grant (FORT23/00034) also by the ISCIII, Ministerio de Ciencia e Innovación de España. In 2015, she received an ECTRIMS Post-doctoral Research Fellowship and has received funding from the UK MS Society. She is a member of the Editorial Board of Neurology Journal and Multiple Sclerosis Journal. She has also received honoraria from Roche and Novartis and is a steering committee member of the O’HAND trial and of the Consensus group on Follow-on DMTs.

FIGURE LEGENDS

Figure 1. Main objectives of machine and deep learning models

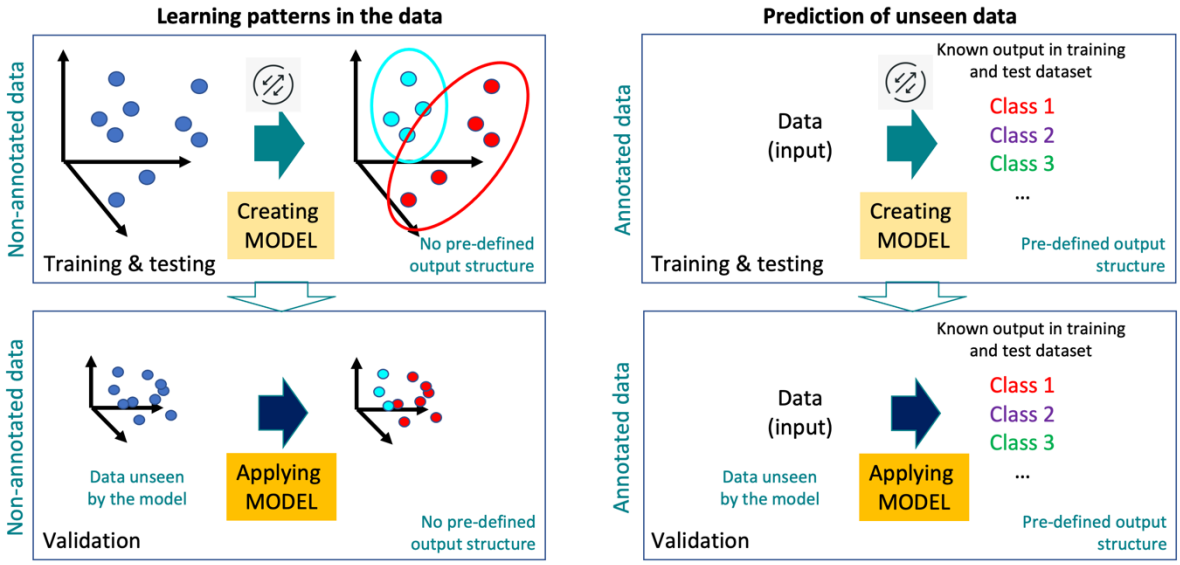

Figure 1 (legend). This figure illustrates the main aims of AI methods applied to data analysis: learning patterns from the data (left) and predicting unseen data (right). In both cases, there is always a first part consisting of building the model (i.e., training and testing), and a second part consisting of validating the model in new (unseen) data. Figure adapted from Tur and Collorone. Kranion, 2023. DOI: 10.24875/KRANION.M23000065.

Figure 2. Main types of AI models

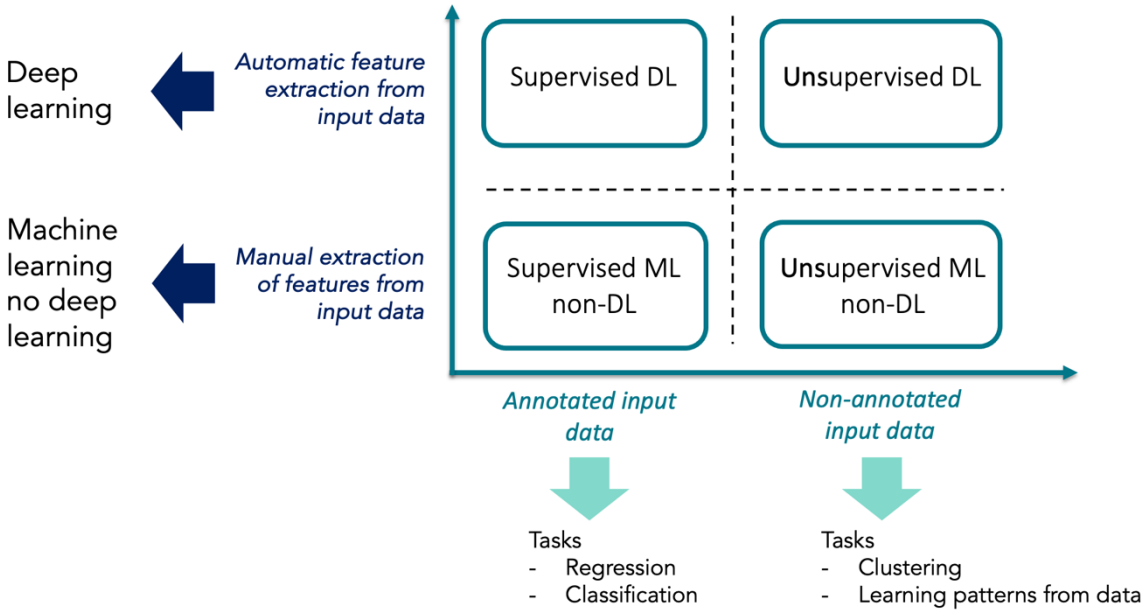

Figure 2 (legend). AI models can be divided into supervised and unsupervised learning models, depending on whether data are annotated (i.e., the output is known) or not (i.e., the output is unknown), respectively. AI models can also be divided into machine or deep learning models, depending on whether features (predictors) are already extracted or not, respectively. *Abbreviations*: DL: deep learning; ML: machine learning.

Figure 3. PRISMA chart describing article selection

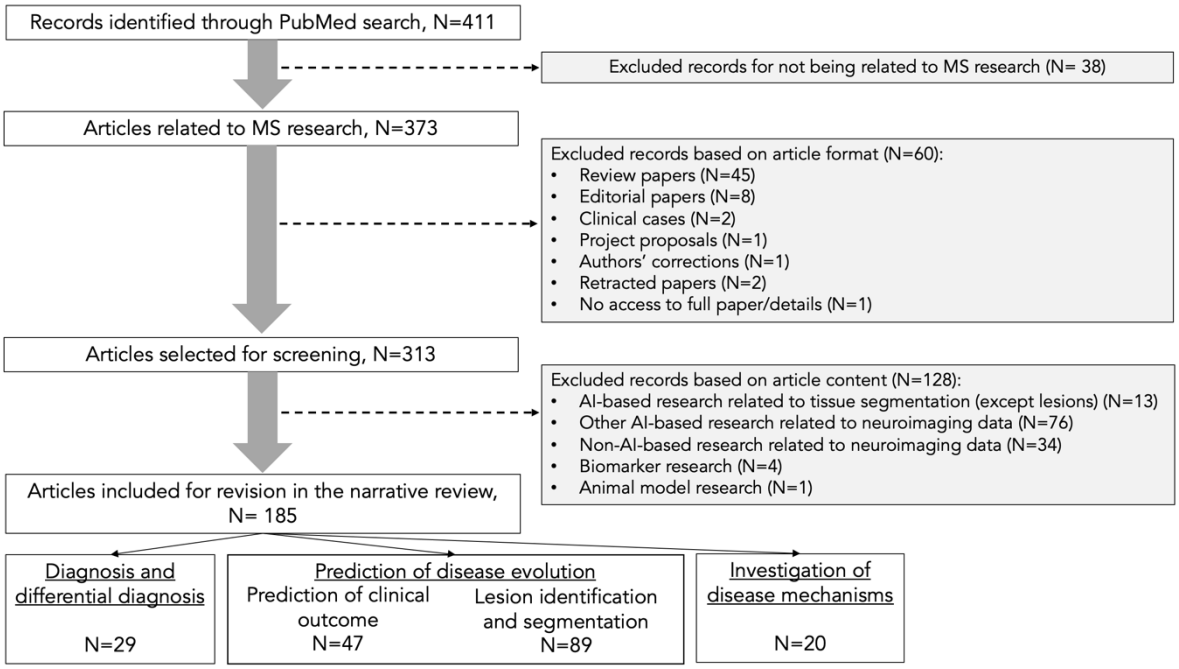

Figure 3 (legend). Although this is not a systematic review paper but a narrative review one, we have followed a fairly systematic approach to selecting the papers to be considered in our manuscript. After performing a PubMed search with the following terms: (multiple sclerosis or demyelination or demyelinating disease) AND (artificial intelligence or deep learning or machine learning) AND (MRI or neuroimaging or brain/cord imaging), 411 records were obtained. Of those, only 185 were considered for this review after excluding those not meeting our inclusion criteria.

Figure 4. Distribution of the research papers on AI applied to neuroimaging data in MS over time

Figure 4 (legend). This histogram shows the number of research articles (of those 185 selected) published per year, between 2013 and 2023. It is to be noted that most of the papers have been published in the last 3 years.

Figure 5. Attention maps to uncover the mechanisms of disability accumulation in MS

Figure 5 (legend). On the left (panels A and B), this figure shows an individual example of attention map (B) obtained after building a DL model which used structural imaging (T1-weighted and T2-FLAIR images [panel A]) to classify patients based on their disability level, i.e., below or above EDSS 3.0. On the right (panel C), this figure reflects the group-level results, for all the DL-derived classes, i.e., TP, TN, FP, FN. As can be observed, the most relevant brain areas for the DL-based prediction were the frontal and cerebellar cortices. See Coll et al. (ref.[42]), for more details. *Abbreviations*: DL: deep learning; EDSS: expanded disability status scale; FN: false negative; FP: false positive; TN: true negative; TP: true positive;

Table 1. Artificial Intelligence terminology

Terms	Definitions
Annotated data	Set of data where both the input variables (i.e., predictors in a given model) and the output variables (e.g., clinical outcome) are known and well characterised
Artificial Intelligence	Branch of science aiming at creating algorithms able to carry out tasks that typically require human intelligence
Attention Maps (or Salience Maps or Heat Maps)	Regions of the input data – typically of the brain MR images – that carry the most important information for the Deep Learning model
Bagging	Machine Learning technique aiming to improve the stability and accuracy of a model by combining predictions from multiple models trained on different subsets of the data
Bayesian Network	A graphical representation of probabilistic relationships among a set of variables, for instance diseases and symptoms
Convolutional Neural Networks	Deep Learning algorithm that can learn directly from the input data (images) their filters/characteristics and find patterns to recognise objects, classes, and categories
Deep Learning	Machine Learning using artificial neurons
Feature reduction methods	Selection (i.e., feature selection, LASSO, random forest, Relief) or dimensionality reduction (i.e., principal component analysis, independent component analysis, t-SNE) techniques improving the efficiency and performance of machine learning models, especially in the context of radiomics where datasets often contain a large number of features.
Gradient Boosting	Machine learning technique involving constructing an ensemble of models, typically decision trees, that are built sequentially. Each new model in the sequence is trained to correct the errors made by the previous models
Gradient Descent	An optimisation algorithm to find the coefficients that minimise the cost function
Input data	Set of variables or images (also called predictors) that are used to feed a statistical/AI model in order to predict an outcome, also called output data
K-Nearest Neighbours	Machine Learning classification algorithm based on similarity measures, usually distance functions, with known cases. It is a non-parametric learning algorithm, and it does not require training phase
Layer-wise Relevance Propagation	Deep Learning algorithm that allows to decompose the prediction computed over a sample (i.e., an image) down to relevance scores for the single input dimensions of the sample (i.e., pixels)
Linear Discriminant Analysis	Machine Learning classification algorithm limited for more than two-class classification problems
Logistic Regression	Machine Learning classification algorithm limited to two-class classification problems
Machine Learning	Set of mathematical algorithms able both to learn from the data and to predict
Multilayer Perceptron	Feedforward artificial neural network in which information travels in one direction—from the input layer through the hidden layers to the output layer—without forming cycle
Naïve Bayes Classifiers	Bayesian network model assuming that features are conditionally independent given the class label (naïve)
Non-annotated data	Set of data where the outcome variable (e.g., specific disease classification) is not known and must be investigated through AI methods
Output data	Variable in the dataset that contains the outcome that is being predicted by the (statistical/AI) model, for instance, a given classification based on disability scores
Principal Component Analysis	Machine Learning algorithm whose main objective is to transform a dataset of possibly correlated variables into a new set of uncorrelated variables, called principal components. These principal components are linear combinations of the original variables and capture the maximum variance present in the data
Probabilistic Neural Network	Deep Learning approach providing probabilistic outputs, making them suitable for applications where uncertainty in predictions needs to be quantified
Random Forest	Machine Learning classification algorithm consisting of many decisions trees
Support Vector Machine	Machine Learning supervised learning models with learning algorithms for classification and regression analysis

Table 1 (footnote). This table shows key definitions of the AI terms used in the included papers. *Abbreviations*; LASSO: least absolute shrinkage and selection operator; t-SNE: t-Distributed Stochastic Neighbour Embedding.

Table 2. Summary of selected studies focused on differential diagnosis

Reference	N	MRI protocol	AI method: algorithms	Model input	Model output	Model accuracy	Comment
Eshaghi et al., Neurology 2016 (ref.[7])	144 patients	1.5T and 3T scanners 3D-T1, FLAIR	Machine learning: random forest	157 features, including GM regional volumes, and cortical thicknesses and surface areas, extracted from 3D-T1 images	Disease group: MS vs NMOSD	74%	Variables reflecting deep grey matter volume were those responsible for the main differences between MS and NMOSD
Zurita et al., Neuroimage: Clin. 2018 (ref.[51])	157 subjects (107 MS; 50 HCs)	3T scanner T1 DTI Rs-fMRI	Machine learning: SVM	FA maps, DTI connectivity matrix, rsfMRI correlation matrix, normalized combination of DTI and rsfMRI matrices	Disease group: EDSS >1.5 MS vs EDSS ≤1.5 MS vs HCs	89%	The brain regions that contributed the most to the classification were: the right occipital, left frontal orbital, medial frontal cortices and lingual gyrus. Model performance was below 63% when comparing the two groups of patients with different levels of disability.
Eitel et al., Neuroimage: Clin. 2019 (ref.[41])	147 subjects (76 MS; 71 HCs)	3T scanner 3D-T1, 3D-FLAIR	Deep learning: CNN	FLAIR lesion masks and the entire FLAIR volume	Disease group: MS vs HC	87%	The authors used the LRP algorithm to assess the brain regions most relevant for CNN decisions, which were WM lesions and some GM regions such as thalamus
Mato-Abad et al., Eur J Neurol. 2019 (ref.[52])	34 patients (17 RIS; 17 CIS)	3T scanner DTI Structural sequence not specified	Machine learning: Bagging; Naive Bayes classifier Deep learning: Multilayer Perceptron	FA maps, MD maps, cortical thickness, cortical and subcortical grey matter volume	Disease group: RIS vs CIS	78%	The best models to predict the diagnosis of CIS and RIS used only three features: the left rostral middle frontal gyrus volume; FA in the right amygdala and right lingual gyrus
Rocca et al., Investigative Radiology 2021 (ref.[8])	268 patients	1.5T and 3T scanners 3D-T1, Axial T1 and T2-weighted images	Deep learning: CNN	T1-weighted and T2-weighted image features	Disease group (MS vs NMOSD vs migraine vs CNS vasculitis)	98.8%, 88.6%, 92.9%, 92.1%, for MS, NMOSD, migraine, or CNS vasculitis, respectively, vs rest	Human raters accuracy was 72.8% and 81.8% for MS; 4.4% for NMOSD, both; 53% and 64.8% for migraine; and 54.6% and 45.5 % for vasculitis
Seok et al., Sci Rep 2023 (ref.[9])	90 patients (86 MS, 70 NMOSD)	3T scanner 3D-T1, 3D-FLAIR, Axial T2-weighted images	Deep learning: CNN	5-channel 2D image obtained by concatenating selected five axial slices from 3D-FLAIR	Disease group: MS vs NMOSD	76.1%	White matter lesions were what the model focused on for classification.

Table 2 (footnote). This table shows a selection of studies focused on differential diagnosis. *Abbreviations:* AI: artificial intelligence; ANN: artificial neural network; CIS: clinically isolated syndrome; CM: chronic microangiopathy; CNN: convolutional neural networks; DTI: diffusion tensor imaging; EDSS: expanded disability status scale; FA: fractional anisotropy; FLAIR: fluid attenuated inversion recovery; FS: full spectra; LDA: linear discriminant analysis; LR: logistic regression; LRP: layer-wise relevance propagation; MD: mean diffusivity; MRS: magnetic resonance spectroscopy; MS: multiple sclerosis; NMOSD: neuromyelitis optica spectrum disorder; PCA: principal component analysis; PI: peak integration; PNN: probabilistic neural network; RIS: radiologically isolated syndrome; ROI: region of interest; rs-fMRI: resting state-functional MRI; SVM: support vector machine;

Table 3. Summary of selected studies focused on prediction of disease course: relapses and disability progression

Reference	N	Follow-up time	MRI protocol	AI method: algorithms	Model input	Model output	Model performance	Comment
Relapses								
Wottschel et al., Neuroimage: Clinical 2019 (ref.[14])	400 patients with a first demyelinating attack	1 year	Magnetic field not specified 3D-T1 FLAIR or PD/T2-weighted images	Machine learning: SVM	213 features extracted from structural imaging, including: whole-brain, lobar, and region-of-interest metrics	Second attack over 1-year follow-up	Accuracy: 70.8%	40 features selected as most relevant, including GM volumes and thicknesses, together with whole brain lesion load
Bendfeldt et al., Brain Imaging Behav 2019 (ref.[15])	364 patients with a first demyelinating attack	2 years	1.5T scanner T1-weighted after Gd administration and T2-weighted images	Machine learning: linear and non-linear SVM	Geometric features of individual lesions derived from MRI data (T2-, T1-, Gd-lesions); spatial information on 13 brain ROIs	Second attack over 2-year follow-up	Accuracy: 70.4%	The greatest accuracy was obtained when considering both MRI and clinical & demographic characteristics
Pareto et al., Neuroradiology 2022 (ref.[16])	266 patients with a first demyelinating attack	3 years	3T scanner 3D-T1 images	Machine learning: SVM, MKL	3D-T1 images: GM characteristics in 116 parcellations from the AAL atlas and maps of white matter hypointensities in the JHU atlas	Second attack over 3-year follow-up	Accuracy (GM + hypointensities): SVM: 51.44% MKL: 50.87%	Most relevant GM regions in the temporal, deep GM, and frontal lobe, followed by the cerebellum, parietal, and occipital lobe
Disability accumulation								
Short/mid-term								
Tousignant et al., Proceedings of Machine Learning Research 2019 (ref.[18])	465 patients with MS	1 year	Magnetic field not specified T1 and PD/T2-weighted images, FLAIR Post-contrast T1-weighted images	Deep learning: CNN	MR (raw) images at study baseline +/- lesion masks as an additional input	Confirmed disease progression at 1-year follow-up	Accuracy: 70.1%	Moderate performance, but clear improvement when lesion masks (on T2-weighted and Gd-T1 images) were added as inputs
Roca et al., Diagnostic and Interventional Imaging 2020 (ref.[19])	971 patients with MS	2 years	1.5T and 3T scanners 2D and 3D-FLAIR	Machine learning: random forest and manifold learning; Deep learning: CNN	Machine learning: 65 features, including lesion volumetric measures in white matter tracts and ventricle volumes, extracted from FLAIR	EDSS score at 2-year follow-up	Mean EDSS square error: - Random forest: 2.6 - CNN: 2.7 - Manifold learning: 3.2	Overall moderate performance. No clear superiority of DL over ML. Prediction of extreme disability scores more challenging

					images, and age Deep learning: FLAIR images and age			
Storelli et al., Investigative Radiology 2022 (ref.[20])	373 patients with MS	2 years	3T scanner 3D-T1 2D or 3D T2-weighted images	Deep learning: CNN	Coregistered T1-weighted and T2-weighted images in MNI space	Clinical worsening* on (i) EDSS, (ii) SDMT, (iii) EDSS or SDMT at 2-year follow-up	Accuracy for: - EDSS worsening: 83.3% - SDMT worsening: 67.7% - EDSS or SDMT worsening: 85.7%	High performance, and the EDSS+SDMT -based model had the highest accuracy, greater than (human) rater-based accuracy (70%)
Long-term								
Zhao et al., PloS ONE 2017 (ref.[21])	1693 patients with MS	5 years	Magnetic field not specified T2-weighted images, Other sequences not specified	Machine learning: SVM	35 features, including MRI (lesion load, BPF), clinical and demographic variables at baseline +/- 1-year follow-up data +/- 2-year follow-up data	Clinical worsening[$] on the EDSS at 5-year follow-up	Accuracy for: - Baseline-only data models: very low - Baseline + 1-year follow-up MRI data: 69% - Baseline + 2-year follow-up MRI data: 71-75%	Only when using baseline and longitudinal clinical and MRI data the ML model is better than a classical LR model

Table 3 (footnote). This table shows a selection of studies focused on predicting disease course. * Clinically worsening on the EDSS is defined as EDSS score increase ≥1.5 if baseline EDSS is 0, ≥1.0 if baseline EDSS is <6.0, and ≥0.5 if baseline EDSS is ≥6.0; clinical worsening on the SDMT is defined as a decrease ≥4 points in the follow-up SDMT (regardless of baseline SDMT score). $ Clinical worsening was defined as an increase in the EDSS of ≥1.5. *Abbreviations:* ALL: automated anatomical labelling; BPF: brain parenchymal fraction; CNN: convolutional neural networks; EDSS: expanded disability status scale; FLAIR: fluid attenuated inversion recovery; Gd: gadolinium; GM: grey matter; JHU: Johns Hopkins University; LR: logistic regression; MKL: multiple kernel learning; MRI: magnetic resonance imaging; MS: multiple sclerosis; ROI: region of interest; SDMT: symbol digit modalities test; SVM: support vector machine;

Table 4. Summary of selected studies focused on lesion segmentation

Reference	N	MRI protocol	AI method: algorithms	Model input	Model output	Model accuracy	Comment
Salem et al., Neuroimage Clin 2020 (ref.[26])	60 patients	3T scanner 1-year FU PD, Axial T2-weighted images, 3D-T1, Axial FLAIR	Deep learning: CNN	3D patches from the baseline and follow-up images of the four input modalities	New MS lesions	83%[a]	36 patients had new lesions at FU. True positive detection rate was 83.09% and false positive detection rate was 9.36%.
Narayana et al., Radiology 2020 (ref.[28])	1008 patients	1.5T and 3T scanners Pre- and post-contrast T1-weighted images, T2-weighted images, FLAIR	Deep learning: CNN	Automated T2-hyperintense lesion mask dilated by 3 voxels in each direction applied on pre-contrast T1-weighted images, T2-weighted images, and FLAIR	Enhancing lesions	75%	519 patients had enhancing lesions. Accuracy slice-wise: 82%, Sensitivity 78%, Specificity 73%
Salem et al., Front Neurosci 2022 (ref.[27])	32 patients	3T scanner 1-year FU 3D-FLAIR	Deep learning: CNN	3D patches from the baseline and follow-up images	New MS lesions	42%[a]	All patients had at least one new lesion at FU. Sensitivity: 53% Precision: 52%
Caba et al., Neuroimage 2023 (ref.[29])	4,924 patients	1.5T and 3T scanners 24-156-week FU T1-weighted images, T2-weighted images	Machine learning: Ensemble Classifier; Deep learning: CNN	32 radiomic features from raw images and inpainted patches concatenated together with T2-hyperintense lesion masks	Acute versus chronic MS lesions	75%	New and enlarging T2 lesions are detected through computing the difference between follow-up and baseline T2-hyperintense lesion masks
Khajetash et al., Biomed Phys Eng Express 2023 (ref.[30])	82 patients	1.5T scanner Pre- and post-contrast T1-weighted images, T2-weighted images, FLAIR	Machine learning: DT, KNN, LR, NB, RF, SVM	11 radiomics features for each lesion	Enhancing lesions	85%	Feature reduction methods (107 to 11): LASSO. Sensitivity 82%, Specificity 66%
Tavakoli et al., J Biomed Phys Eng 2023 (ref.[31])	82 patients	1.5T scanner Pre- and post-contrast T1-weighted images,	Machine learning: DT, GB, MLP, XGB	7 and 8 radiomics features for each lesion	Enhancing lesions	86%	Feature reduction methods (107 to 7 and 8): Boruta, Relief. Sensitivity 100%,

		T2-weighted images, FLAIR					Specificity 84%
Shekari et al., Mult Scler Relat Disord 2024 (ref.[32])	34 patients	1.5T scanner post-contrast T1-weighted, FLAIR, DWI	Machine learning: DT, KNN, LDA, LR, SVM	53 radiomics features for each lesion from DWI	Enhancing lesions	96%	Feature reduction methods (89 to 53): t-test, PCA, SBS, SFS, Relief. Sensitivity 91%, Specificity 100%

Table 4 (footnote). This table shows a selection of studies focused on lesion segmentation. [a] Accuracy computed as Dice similarity coefficient: 2 * TPs / (FNs + FPs + 2 * TPs). Where: TPs: number of voxels correctly predicted as lesions; FPs: number of voxels incorrectly predicted as lesions; and FNs: number of voxels incorrectly predicted as non-lesions. *Abbreviations:* CNN: convolutional neural networks; DT: decision tree; DWI: diffusion weighted imaging; FLAIR: fluid attenuated inversion recovery; FU: follow-up; GB: gradient boosting; KNN: K-Nearest Neighbours; LASSO: least absolute shrinkage and selection operator; LDA: linear discriminant analysis; LR: logistic regression; MLP: multi-layer perceptron; MS: multiple sclerosis; NB: naive bayes classifier; PCA: principal component analysis; PD: proton density; RF: random forest; SBS: sequential backward selection; SFS: sequential forward selection; SVM: support vector machine; XGB: extreme gradient boosting.

Table 5. Summary of selected studies focused on investigation of disease mechanisms

Reference	N	MRI protocol	AI method: algorithms	Model input	Model output	Model accuracy	Comment
Eitel et al., Neuroimage Clin 2019 (ref.[41])	147 subjects (76 MS; 71 HCs)	3T scanner 3D-T1, 3D-FLAIR	Deep learning: CNN Attention Maps	FLAIR images[a]	Disease group: MS vs HC	87%	Posterior periventricular white matter regions were determinant areas for MS diagnosis
Eshaghi et al., Nat Commun 2021 (ref.[38])	6322 patients with MS	1.5 and 3T scanners 2D or 3D-T1, FLAIR, T2-weighted images	Machine learning: SustaIn (ref.[39])	18 features from available images (regional volumes, lesion volumes, and T1/T2 ratio) of which 13 differed between training and validation datasets and were retained in the SustaIn model	Longitudinal patterns of brain pathology	NA	Three patterns identified: cortex-led, normal-appearing WM-led, and lesion-led- The lesion-led pattern was related to worse prognosis
Pontillo et al., Eur Radiol 2022 (ref.[40])	425 patients with MS	3T scanner 3D-T1, FLAIR images	Machine learning: SustaIn (ref.[39])	Regional and lesions volumes from T1-weighted and FLAIR-T2weighted images	Longitudinal patterns of brain pathology	NA	Two patterns identified: cortex-led, and DGM-lesion-led- The DGM-lesion-led pattern was related to worse prognosis
Coll et al., NeuroImage Clin 2023 (ref.[42])	268 patients with MS	1.5 and 3T scanners 3D-T1, FLAIR images	Deep learning: CNN Attention Maps	3D-T1 and FLAIR images	Disability (EDSS≥3.0 vs EDSS<3.0)	79%	Areas identified were the frontotemporal cortex and cerebellum
Yamin et al., Hum Brain Mapp 2023 (ref.[43])	100 patients with MS	3T scanner T2*-weighted single-shot EPI for RS fMRI; dual echo turbo spin echo (T2-weighted); 3D T1-weighted images	2-step Machine learning approach: Step 1: Unsupervised DS clustering; Step 2: SVM and LR with LASSO regularisation	RS FC metrics (derived from a low representation of the actual RS FC matrices)	Step 1: no outcome is provided (unsupervised learning) Step 2: clinical phenotypes (RRMS, PMS, HCs)	Accuracy for: -RRMS vs HCs: 72.51% -PMS vs HCs: 85.19% -RRMS vs PMS: 76.04%	RRMS and PMS (vs HCs): - Increased RS FC *within* basal ganglia subnetwork, especially between the bilateral thalami, - Decreased RS FC *within* the frontal, temporal, and occipital subnetworks PMS (vs RRMS): - Decreased RS FC *within* temporal, and occipital subnetworks - Decreased RS FC *between* several subnetworks

Table 5 (footnote). This table shows a selection of studies focused on investigating disease mechanisms. *Abbreviations:* CNN: convolutional neural networks; DGM: deep grey matter; DS: dominant set; EDSS: expanded disability status scale; EPI: echo-planar imaging; FC: functional connectivity; FLAIR: fluid attenuated inversion recovery; fMRI: functional MRI; HCs: healthy controls; LASSO: Least Absolute Shrinkage and Selection Operator; LR: logistic regression; MS: multiple sclerosis; PMS: progressive MS patients; RRMS: relapsing-remitting multiple sclerosis; RS: resting-state; WM: white matter.

REFERENCES

1. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. 1943. *Bull Math Biol* 1990; 52: 99–115; discussion 73-97.
2. Morris RGM. D.O. Hebb: The Organization of Behavior, Wiley: New York; 1949. *Brain Res Bull* 1999; 50: 437.
3. Turing AM. COMPUTING MACHINERY AND INTELLIGENCE. *Mind* 1950; 49: 433–460.
4. Bonacchi R, Filippi M, Rocca MA. Role of artificial intelligence in MS clinical practice. *NeuroImage Clin* 2022; 35: 103065.
5. Naji Y, Mahdaoui M, Klevor R, et al. Artificial Intelligence and Multiple Sclerosis: Up-to-Date Review. *Cureus*. Epub ahead of print 17 September 2023. DOI: 10.7759/cureus.45412.
6. Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol* 2018; 17: 162–173.
7. Eshaghi A, Wottschel V, Cortese R, et al. Gray matter MRI differentiates neuromyelitis optica from multiple sclerosis using random forest. *Neurology* 2016; 87: 2463–2470.
8. Rocca MA, Anzalone N, Storelli L, et al. Deep Learning on Conventional Magnetic Resonance Imaging Improves the Diagnosis of Multiple Sclerosis Mimics. *Invest Radiol* 2021; 56: 252–260.
9. Seok JM, Cho W, Chung YH, et al. Differentiation between multiple sclerosis and neuromyelitis optica spectrum disorder using a deep learning model. *Sci Rep* 2023; 13: 11625.
10. Tintore M, Rovira À, Río J, et al. Defining high, medium and low impact prognostic factors for developing multiple sclerosis. *Brain* 2015; 138: 1863–1874.
11. Chung KK, Altmann D, Barkhof F, et al. A 30-Year Clinical and Magnetic Resonance Imaging Observational Study of Multiple Sclerosis and Clinically Isolated Syndromes. *Ann Neurol* 2020; 87: 63–74.
12. Scalfari A, Romualdi C, Nicholas RS, et al. The cortical damage, early relapses, and

onset of the progressive phase in multiple sclerosis. *Neurology*; 90. Epub ahead of print 12 June 2018. DOI: 10.1212/WNL.0000000000005685.

13. Brownlee WJ, Altmann DR, Prados F, et al. Early imaging predictors of long-term outcomes in relapse-onset multiple sclerosis. *Brain* 2019; 142: 2276–2287.
14. Wottschel V, Chard DT, Enzinger C, et al. SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of multiple sclerosis. *NeuroImage Clin* 2019; 24: 102011.
15. Bendfeldt K, Taschler B, Gaetano L, et al. MRI-based prediction of conversion from clinically isolated syndrome to clinically definite multiple sclerosis using SVM and lesion geometry. *Brain Imaging Behav* 2019; 13: 1361–1374.
16. Pareto D, Garcia-Vidal A, Groppa S, et al. Prognosis of a second clinical event from baseline MRI in patients with a CIS: a multicenter study using a machine learning approach. *Neuroradiology* 2022; 64: 1383–1390.
17. Haider L, Chung K, Birch G, et al. Linear brain atrophy measures in multiple sclerosis and clinically isolated syndromes: a 30-year follow-up. *J Neurol Neurosurg Psychiatry* 2021; 92: 839–846.
18. Tousignant A, Lemaître P, Precup D, et al. Prediction of Disease Progression in Multiple Sclerosis Patients using Deep Learning Analysis of MRI Data. *Proceedings of Machine Learning Research* 2019; 102: 483–492.
19. Roca P, Attye A, Colas L, et al. Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI. *Diagn Interv Imaging* 2020; 101: 795–802.
20. Storelli L, Azzimonti M, Gueye M, et al. A Deep Learning Approach to Predicting Disease Progression in Multiple Sclerosis Using Magnetic Resonance Imaging. *Invest Radiol* 2022; 57: 423–432.
21. Zhao Y, Healy BC, Rotstein D, et al. Exploration of machine learning techniques in predicting multiple sclerosis disease course. *PLoS One* 2017; 12: e0174866.
22. Commowick O, Istace A, Kain M, et al. Objective Evaluation of Multiple Sclerosis Lesion

Segmentation using a Data Management and Processing Infrastructure. *Sci Rep* 2018; 8: 13650.

23. Rovira A, Corral JF, Auger C, et al. Assessment of automatic decision-support systems for detecting active T2 lesions in multiple sclerosis patients. *Mult Scler J* 2022; 28: 1209–1218.
24. Diaz-Hurtado M, Martínez-Heras E, Solana E, et al. Recent advances in the longitudinal segmentation of multiple sclerosis lesions on magnetic resonance imaging: a review. *Neuroradiology* 2022; 64: 2103–2117.
25. Zeng C, Gu L, Liu Z, et al. Review of Deep Learning Approaches for the Segmentation of Multiple Sclerosis Lesions on Brain MRI. *Front Neuroinform*; 14. Epub ahead of print 20 November 2020. DOI: 10.3389/fninf.2020.610967.
26. Salem M, Valverde S, Cabezas M, et al. A fully convolutional neural network for new T2-w lesion detection in multiple sclerosis. *NeuroImage Clin* 2020; 25: 102149.
27. Salem M, Ryan MA, Oliver A, et al. Improving the detection of new lesions in multiple sclerosis with a cascaded 3D fully convolutional neural network approach. *Front Neurosci*; 16. Epub ahead of print 24 November 2022. DOI: 10.3389/fnins.2022.1007619.
28. Narayana PA, Coronado I, Sujit SJ, et al. Deep Learning for Predicting Enhancing Lesions in Multiple Sclerosis from Noncontrast MRI. *Radiology* 2020; 294: 398–404.
29. Caba B, Cafaro A, Lombard A, et al. Single-timepoint low-dimensional characterization and classification of acute versus chronic multiple sclerosis lesions using machine learning. *Neuroimage* 2023; 265: 119787.
30. Khajetash B, Talebi A, Bagherpour Z, et al. Introducing radiomics model to predict active plaque in multiple sclerosis patients using magnetic resonance images. *Biomed Phys Eng Express* 2023; 9: 055004.
31. Tavakoli H, Pirzad Jahromi G, Sedaghat A. Investigating the Ability of Radiomics Features for Diagnosis of the Active Plaque of Multiple Sclerosis Patients. *J Biomed Phys Eng*; 13. Epub ahead of print 2023. DOI: 10.31661/jbpe.v0i0.2302-1597.

32. Shekari F, Vard A, Adibi I, et al. Investigating the feasibility of differentiating MS active lesions from inactive ones using texture analysis and machine learning methods in DWI images. *Mult Scler Relat Disord* 2024; 82: 105363.

33. Alijamaat A, NikravanShalmani A, Bayat P. Multiple sclerosis lesion segmentation from brain MRI using U-Net based on wavelet pooling. *Int J Comput Assist Radiol Surg* 2021; 16: 1459–1467.

34. Danelakis A, Theoharis T, Verganelakis DA. Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. *Comput Med Imaging Graph* 2018; 70: 83–100.

35. Fartaria MJ, Sati P, Todea A, et al. Automated Detection and Segmentation of Multiple Sclerosis Lesions Using Ultra–High-Field MP2RAGE. *Invest Radiol* 2019; 54: 356–364.

36. Filippi M, Brück W, Chard D, et al. Association between pathological and MRI findings in multiple sclerosis. *The Lancet Neurology* 2019; 18: 198–210.

37. Barkhof F. The clinico-radiological paradox in multiple sclerosis revisited. *Curr Opin Neurol* 2002; 15: 239–245.

38. Eshaghi A, Young AL, Wijeratne PA, et al. Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat Commun*; 12. Epub ahead of print December 2021. DOI: 10.1038/S41467-021-22265-2.

39. Young AL, Marinescu R V, Oxtoby NP, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat Commun* 2018; 9: 4273.

40. Pontillo G, Penna S, Cocozza S, et al. Stratification of multiple sclerosis patients using unsupervised machine learning: a single-visit MRI-driven approach. *Eur Radiol* 2022; 32: 5382–5391.

41. Eitel F, Soehler E, Bellmann-Strobl J, et al. Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. *NeuroImage Clin* 2019; 24: 102003.

42. Coll L, Pareto D, Carbonell-Mirabent P, et al. Deciphering multiple sclerosis disability

with deep learning attention maps on clinical MRI. *NeuroImage Clin* 2023; 103376.

43. Yamin MA, Valsasina P, Tessadori J, et al. Discovering functional connectivity features characterizing multiple sclerosis phenotypes using explainable artificial intelligence. *Hum Brain Mapp* 2023; 44: 2294–2306.
44. Cortese R, Collorone S, Ciccarelli O, et al. Advances in Brain Imaging in Multiple Sclerosis. *Ther Adv Neurol Disord* 2019; 12: 1756286419859722.
45. Collorone S, Prados F, Kanber B, et al. Brain microstructural and metabolic alterations detected in vivo at onset of the first demyelinating event. *Brain* 2021; 144: 1409–1421.
46. Castelvecchi D. Can we open the black box of AI? *Nature* 2016; 538: 20–23.
47. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 2021; 11: e048008.
48. Falet JPR, Durso-Finley J, Nichyporuk B, et al. Estimating individual treatment effect on disability progression in multiple sclerosis using deep learning. *Nat Commun*; 13. Epub ahead of print December 2022. DOI: 10.1038/S41467-022-33269-X.
49. Muehlematter UJ, Bluethgen C, Vokinger KN. FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks. *Lancet Digit Heal* 2023; 5: e618–e626.
50. van Leeuwen KG, Schalekamp S, Rutten MJCM, et al. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol* 2021; 31: 3797–3804.
51. Zurita M, Montalba C, Labbé T, et al. Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data. *NeuroImage Clin* 2018; 20: 724–730.
52. Mato-Abad V, Labiano-Fontcuberta A, Rodríguez-Yáñez S, et al. Classification of radiologically isolated syndrome and clinically isolated syndrome with machine-learning techniques. *Eur J Neurol* 2019; 26: 1000–1005.