International Journal of Health Science (IJHS)

E-ISSN: 2827-9603
P-ISSN: 2827-9603

Research Article

Deep Learning as a Tool in Current MRI-Based Approaches for Predicting Clinical Disability Progression in Multiple Sclerosis

(A Systematic Review)

Shahifa Audy Rahima[1*], Mirza Wafiyudin Baehaqi[2], Putri Fortuna Sari[3], Dita Rahmania[4], Desie Yuliani[5], Ni Made Kurnia Jayanthi[6]

[1,5,6] Departement of Neurology, RSUD Wangaya, Denpasar, Indonesia
[2,3,4] Fakultas Kedokteran Universitas Jember, Indonesia
* Corresponding Author: e-mail: shahifarahima@gmail.com

Abstract: Multiple sclerosis (MS) is a chronic neurological disorder and one of the leading causes of long-term disability in young adults. Despite therapeutic advances, predicting disability progression remains challenging. Magnetic resonance imaging (MRI) plays a central role in disease monitoring; however, conventional prognostic models show limited predictive accuracy. Deep learning (DL) has emerged as a promising approach for extracting complex imaging patterns and improving outcome prediction. This systematic review, conducted in accordance with PRISMA guidelines, evaluated the potential of MRI-based deep learning models for predicting clinical disability progression in MS. A comprehensive literature search was performed in PubMed, Google Scholar, Cochrane Library, and IEEE Xplore using PICO criteria and relevant MeSH terms. Studies applying deep learning techniques to MRI data for disability progression prediction were included, and methodological quality was assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST). Of 226 identified records, seven studies met the inclusion criteria, with most demonstrating low risk of bias. Predictive performance ranged from moderate to strong, with area under the curve (AUC) values between 0.69 and 0.84. Regression models achieved root mean square error (RMSE) values as low as 1.33, while survival-based approaches reported concordance indices up to 0.72. Longitudinal deep learning models generally outperformed single timepoint approaches. Overall, MRI-based deep learning models show strong potential for predicting disability progression in MS and may support personalized disease monitoring and clinical decision-making.

Keywords: Deep Learning; Disability Progression; Magnetic Resonance Imaging (MRI); Multiple Sclerosis; Prediction Models

Received: October 14, 2025
Revised: November 12, 2025
Accepted: December 28, 2025
Published: March 02, 2026
Curr. Ver.: March 02, 2026

1. Introduction

Multiple sclerosis (MS) is a chronic immune-mediated disorder of the central nervous sys-tem characterized by inflammation, demyelination, and progressive neuroaxonal loss. It is a lead-ing cause of non-traumatic disability in young adults, with a highly variable clinical course (Reich et al., 2018; Thompson et al., 2018). Globally, MS affects approximately 2.8 million individuals, representing a substantial and grow-ing disease burden (MSIF, 2020). Despite advances in disease-modifying therapies, accurately predicting indi-vidual disability trajectories remains a major unmet clinical need (Havla et al., 2024).

Clinical disability is most commonly assessed using the Expanded Disability Status Scale (EDSS), the standard outcome measure in MS trials and longitudinal studies (Cinar & Yorgun, 2018). However, EDSS is limited by its ordinal and nonlinear structure, insensitivity to cognitive and upper limb dysfunc-tion, and inter-rater variability, with progression typically confirmed only retrospectively (Weinstock-Guttman et al., 2022). These limitations reduce its

value for early prognostic stratification and have motivated the develop-ment of predictive models for future disability risk.

MRI is central to MS diagnosis and monitoring, yet conventional prognostic models rely on handcrafted features such as lesion burden or global atrophy, which only partially explain clin-ical outcomes (Hemond & Bakshi, 2018). This weak correspondence between imaging findings and disability progression, the clinico-radiological paradox, highlights the need for more advanced analytical approaches (Mollison et al., 2017). Deep learning (DL) enables data-driven learning of hierarchical and nonlinear representations from high-dimensional MRI data and has shown strong performance in neuroimaging tasks, in-cluding lesion segmentation, disease classification, and outcome prediction (Litjens et al., 2017; Shen et al., 2017).

An increasing number of studies have applied DL to MRI for prognostic modeling in MS, targeting disability worsening, long-term outcomes, and progression phenotypes. However, such tasks are challenged by outcome heterogeneity, nonlinear clinical scales, and variable follow-up durations. A focused synthesis of MRI-based DL approaches for predicting disability progres-sion is therefore warranted.

This systematic review aims to evaluate MRI-based deep learning models for predicting clinical disability progression in MS, synthesizing evidence on model design, imaging strategies, outcome definitions, and predictive performance to inform future research and clinical translation.

2. Materials and Method

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines and was prospectively registered in the PROSPERO database under the registration number CRD42026127959. A comprehensive litera-ture search was systematically performed across four electronic databases, including PubMed (n = 10), Google Scholar (n = 207), Cochrane Library (n = 2), and IEEE Xplore (n = 7), covering studies published up to October 17, 2025. Search keywords were developed using relevant terms and synonyms and were combined using Boolean operators such as AND and OR to ensure comprehensive retrieval of eligible studies. All search results were imported into Rayyan.ai to facilitate organization and screening. Additional relevant articles were identified through manual searching of reference lists. As this review involved only the synthesis of data from previously published studies, ethical approval was not required.

The screening process was conducted independently by two reviewers using Rayyan.ai, which facilitated the identification and removal of duplicate records. Initial screening was per-formed based on titles and abstracts, followed by full-text assessment of potentially eligible arti-cles. Further screening ensured strict adherence to the predefined inclusion and exclusion criteria. Study eligibility was guided by the Population, Intervention, Comparison, Outcome, and Study design (PICOS) framework. Eligible studies included patients with clinically diagnosed multiple sclerosis and investigated deep learning based predictive models using magnetic resonance imag-ing data, with or without additional clinical variables. Model predictions were required to be compared against clinically assessed ground truth outcomes, with longitudinal clinical disability progression defined as the primary outcome of interest. Studies were excluded if they lacked longitudinal outcome assessment, did not employ MRI based deep learning methods, failed to report patient level evaluation, or were published in languages other than English.

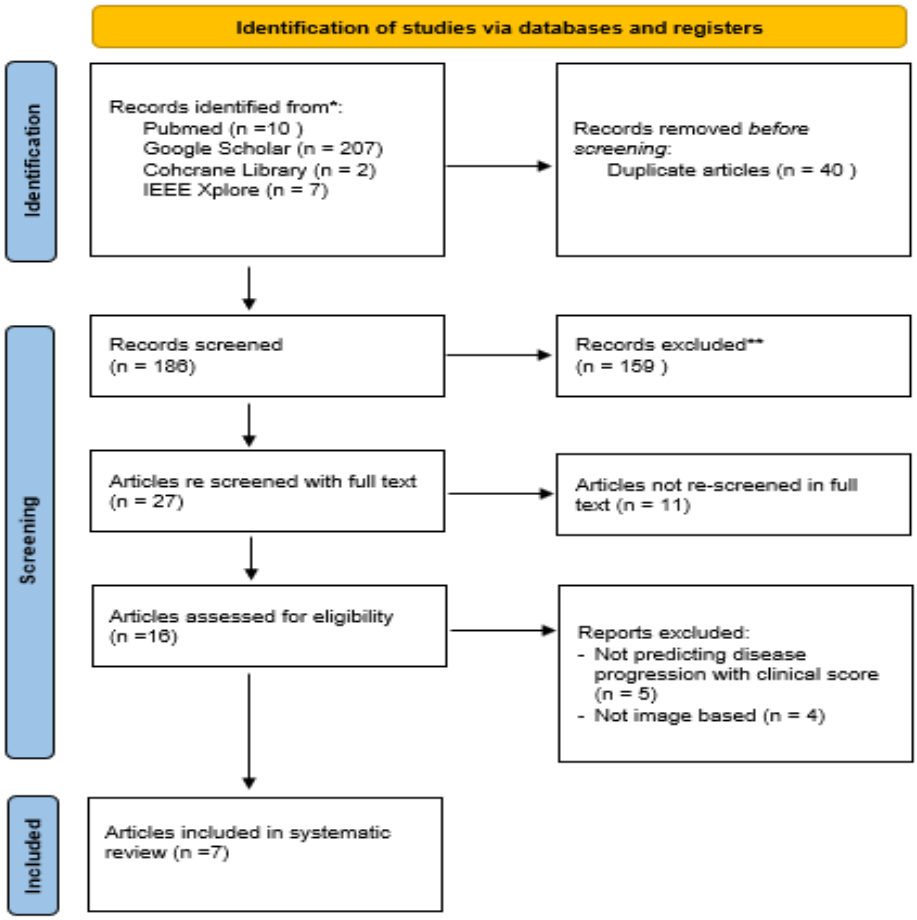
Data extraction was independently performed by two reviewers using a standardized ex-traction form to ensure accuracy and consistency. Extracted variables included authorship, year of publication, study design, sample size, population characteristics, type of deep learning model, magnetic resonance imaging input modalities, outcome definitions, and reported performance metrics. Methodological quality and risk of bias were assessed using the Prediction model Risk of Bias ASsessment Tool (PROBAST), which evaluates potential bias across four domains, namely participants, predictors, outcomes, and analysis (Wolff et al., 2019). Only studies that met the eligibility cri-teria and successfully passed critical appraisal were included in the final synthesis. The study se-lection and screening process was summarized and presented using a PRISMA flow diagram.

3. Results and Discussion

A total of 226 records were identified through systematic searches of electronic databases, including PubMed (n = 10), Google Scholar (n = 207), Cochrane Library (n = 2), and IEEE Xplore (n = 7). After removing 40 duplicate records, 186 unique articles remained and were screened based on their titles and abstracts. Of these, 159 records were excluded at this stage due to insufficient relevance to the review objectives. The remaining 27 articles were ad-vanced to full-text screening, of which 11 could not be retrieved for full-text assessment. Conse-quently, 16 full-text articles were evaluated for eligibility. During this stage, nine studies were excluded for not meeting the predefined inclusion criteria, primarily because they did not predict disease progression using clinical disability scores (n = 5) or did not employ image based approaches using magnetic resonance imaging data (n = 4). Ultimately, seven studies satisfied all eligibility requirements and were included in the final systematic review. The complete study selection process, encompassing identification, screening, eligibility assessment, and inclusion, is summarized in the PRISMA flow diagram (Figure 1). In the risk of bias assessment, five of the included studies were classified as low risk of bias, while two exhibited unclear risk of bias.

A summary of the data extraction of each included studies is presented in Table 1. A total of seven studies were included in the final systematic review, encompassing diverse cohorts, imaging modalities, deep learning architectures, and outcome definitions related to disability progression in multiple sclerosis (Coll et al., 2025; Kırkbir et al., 2025; J. D. Mayfield et al., 2024; J. Mayfield & Naqa, 2024; Roca et al., 2020; Storelli et al., 2022; Taloni et al., 2022). The included studies were published between 2020 and 2025 and represented data from multiple countries, including Italy, France, Spain, China, the United States, and Turkey. Sample sizes varied substantially, ranging from 181 to 1,000 patients, with several studies incorporating longitudinal follow-up periods extending up to five years or more. Most studies utilized brain magnetic resonance imaging as the primary input, while one study focused on longitudinal spinal cord MRI and another relied on MRI activity indicators de-rived from clinical records rather than raw imaging data. Disability progression was predominantly defined using Expanded Disability Status Scale based outcomes, either as categorical worsening, regression targets, or time to progression endpoints.

Regarding imaging inputs and feature representation, substantial methodological heterogeneity was observed across studies. Earlier works often incorporated lesion based or volumetric biomarkers, such as lesion load, or ventricular volume, whereas more recent studies increasingly favored end to end deep learning approaches that learned imaging representations directly from raw MRI data without explicit feature engineering. Several studies combined MRI derived features with baseline clinical variables, including age, baseline EDSS, cognitive scores, relapse his-tory, or demographic information, demonstrating improved predictive performance compared to imaging or clinical data alone. Architecturally, convolutional neural networks formed the back-bone of most models, with extensions incorporating recurrent neural networks, long short term memory units, transformers, or survival modeling frameworks to capture longitudinal and tem-poral disease dynamics. Across studies, predictive performance for disability progression was consistently moderate to strong, although direct comparison was limited by differences in out-come definitions and evaluation metrics. Classification based approaches reported area under the receiver operating characteristic curve values ranging from approximately 0.69 to 0.84, with the highest performance observed in models leveraging longitudinal imaging and transformer based architectures. Regression and survival based models achieved mean squared error values between 2.21 and 3.00, root mean square error values as low as 1.33, and concordance indices up to 0.72. Several studies demonstrated that deep learning models outperformed baseline clinical or demo-graphic predictors, highlighting the added value of MRI informed AI models for forecasting long term disability progression in multiple sclerosis.



Figur 1. *PRISMA flow diagram.*

Table 1. Qualitative summary of included studies.

No	Study	Data & N	MRI & Features	Additional Data	Model	Outcome	Follow-up	Metrics
1	Storelli et al., 2022 (Italy)	OFSEP cohort (MS challenge 2019); DS1=480, DS2=491; train=85, val=96, test=475	Brain MRI (T1w, T2w); CNN-learned lesion features; volumetrics (SIENAX)	Baseline EDSS, SDMT, demographics	3D CNN	EDSS-based disability worsening (± cognitive worsening)	2 years	Acc: 83.3% (EDSS), 67.7% (SDMT), 85.7% (combined); Human 70%; AUC=0.71
2	Roca et al., 2020 (France)	Multicenter MS imaging; Train+Val =971; Test=475	Brain MRI (3D FLAIR); lesion load, ventricular volume, spatial distribution	Age	CNN + Random Forest (aggregated)	EDSS regression (2-year)	2 years	MSE: 3.00 (challenge), 2.56 (RF), 2.21 (agg.); Age-only MSE=3.80
3	Coll et al., 2025 (Spain)	Prospective early MS; 259 (primary), 32 (external)	Brain MRI (T1w, T2-FLAIR); end-to-end DL	Age, relapse info	Efficient Net-based survival DL	PIRA (EDSS worsening independent of relapse)	Median 4.2 years	C-index=0.72; IBS=0.10; Acc=78% (int), 72% (ext); AUC year-1=0.694

No	Study	Data & N	MRI & Features	Additional Data	Model	Outcome	Follow-up	Metrics
4	Taloni et al., 2022 (Italy)	Two-center cohort; 181 (105+76)	Brain MRI (3D T1 MPRAGE); slice-based CNN	Clinical data (secondary)	ResNet50 (transfer learning)	EDSS progression	Mean 3.94 years (2–6)	Coronal AUC=0.72; Sagittal AUC=0.81; Axial AUC>0.69
5	Mayfield & El Naqa, 2024 (USA)	Two-site longitudinal MRI; 703 MS (609 ≥3 MRIs; 94 with 2 MRIs)	Sequential multisequence contrast-enhanced MRI (timepoints as video frames); end-to-end temporal imaging features	Baseline clinical variables	VOC-LSTM (MPS-, MERA-, TTN-LSTM); benchmarked vs VGG16-LSTM & ViViT	Binary EDSS classification (mild vs severe)	Longitudinal	Best: TTN-LSTM ROC-AUC=0.81; Acc=0.76; Prec=0.75; Rec=0.75; F1=0.75 (patient-level hold-out)
6	Mayfield et al., 2024 (USA)	Two-site MS cohort; 703 MS patients	Longitudinal spinal MRI (T1w, T2w, T1w+C); end-to-end CNN/Transformer (no lesion segmentation)	Baseline EDSS, demographics	VGG16-LSTM; ViViT	Future EDSS category (trinary) & exact EDSS	Longitudinal	EDSS: ViViT AUC=0.84; VGG16-LSTM AUC=0.74; Exact EDSS AUC=0.77 (ViViT); 0.73 (classification/ regression)
7	Kulabi et al., 2025 (Turkey)	Multicenter MS registry; 1,000 RRMS	Baseline MRI lesion indicators (categorical); candidate variables incl. gadolinium & spinal cord lesions	EDSS, KPS, relapse & treatment history	LSTM	EDSS at 5-year follow-up (regression)	5 years	RMSE (best model after feature selection)=1.33

Discussion

Deep learning (DL) has increasingly been adopted in neurological research as a powerful approach for modeling complex brain disorders, particularly in the analysis of neuroimaging data. Neurological diseases are characterized by high spatial and temporal heterogeneity, with pathological changes often distributed across multiple brain regions and evolving over time. Conventional neuroimaging analyses typically rely on predefined biomarkers, such as regional volumes, lesion burden, or summary statistics, which may oversimplify the underlying disease processes and limit prognostic accuracy at the individual level (Filippi et al., 2019; Mahajan & Ontaneda, 2017). DL-based models, by contrast, are capable of learning hierarchical and nonlinear representations directly from high-dimensional MRI data, enabling the integration of subtle spatial patterns, diffuse tissue alterations, and complex interactions that may not be captured by handcrafted features alone (Esteva et al., 2019). Within this broader neurological context, DL has shown growing potential for disease classification, outcome prediction, and progression modeling, motivating its application to chronic and heterogeneous conditions such as multiple sclerosis (MS), where accurate predic-tion of future disability progression remains a critical unmet need (Giovannoni et al., 2022).

Assessment of disease progression in multiple sclerosis has traditionally relied on clinical scales, most notably the Expanded Disability Status Scale (EDSS), which remains the cornerstone for monitoring disability accumulation and informing therapeutic decisions in both clinical practice and clinical trials (Cinar & Yorgun, 2018; Filippi et al., 2019) EDSS-based measures are routinely used to define dis-ease progression, guide treatment escalation, and evaluate therapeutic efficacy; however, they are inherently limited by nonlinearity, inter-rater variability, and an overemphasis on ambula-tory function, particularly at higher score ranges (Eshaghi et al., 2018; Mahajan & Ontaneda, 2017). Moreover, EDSS progression is typically assessed retrospectively, requiring longitudinal follow-up before clinical worsening can be confirmed, which constrains its utility for early prognostic stratification. Traditional MRI-

derived markers used alongside EDSS, such as lesion burden or global atrophy, similarly rely on predefined features and assume relatively simple relationships with future disability, which may inadequately reflect the complex and heterogeneous mechanisms underlying MS progression.[20,22] These limitations have prompted growing interest in data-driven approaches capable of integrating multidimensional MRI information to improve prediction of disability progression beyond conventional clinical and radiological assessment.

The use of deep learning (DL) for modeling disease progression in multiple sclerosis offers several advantages when compared with traditional clinical and MRI-based assessment strategies. MS-related disability progression reflects complex and spatially distributed pathological processes, including focal inflammatory lesions, diffuse microstructural damage, and neurodegeneration, which are only partially captured by conventional MRI metrics and clinical scales. DL-based approaches are capable of learning hierarchical and nonlinear representa-tions directly from high-dimensional MRI data, enabling the integration of subtle spatial patterns and tissue alterations that may be relevant to future disability accumulation but are dif-ficult to quantify using predefined biomarkers alone (Rocca et al., 2024). Furthermore, longitudinal DL frame-works provide a means to incorporate temporal disease evolution, which aligns with the progressive nature of MS and supports individualized risk stratification across different stages of the disease course (Coll et al., 2025). Together, these characteristics position DL-based MRI models as promis-ing tools to complement established clinical measures in the assessment and prediction of disability progression in MS.

Across the included studies, deep learning models demonstrated moderate to high predictive performance for disability progression in multiple sclerosis, with reported metrics consistently exceeding those of conventional baseline approaches but varying according to model architecture, MRI input strategy, and outcome formulation (Coll et al., 2025; Kırkbir et al., 2025; J. D. Mayfield et al., 2024; J. Mayfield & Naqa, 2024; Roca et al., 2020; Storelli et al., 2022; Taloni et al., 2022).

For baseline MRI-based approaches, Storelli et al. developed a three-dimensional con-volutional neural network trained on baseline brain MRI to predict EDSS worsening over a two-year period (Storelli et al., 2022). Using end-to-end learned features from T1-weighted and FLAIR images. The study reported a classification accuracy of 83.3% for EDSS worsening over a two-year follow-up, which increased to 85.7% when cognitive outcomes were jointly modeled, along-side an AUC of 0.71. Although this AUC value may appear modest in isolation, previous work has shown that EDSS-based progression prediction is intrinsically challenging due to the scale's nonlinearity, inter-rater variability, and limited sensitivity to subtle clinical change, particularly when predictions rely on a single baseline MRI timepoint. The observation that model performance exceeded that of human raters further supports the capacity of convolu-tional neural networks to capture prognostic information beyond conventional visual MRI interpretation, consistent with prior neuroimaging studies linking diffuse structural damage to later disability accumulation (Giovannoni et al., 2022).

Similarly, Taloni et al. demonstrated that slice-level CNN performance varied across anatomical orientations (Taloni et al., 2022). The model explored a complementary baseline MRI strategy by ap-plying a ResNet-based convolutional neural network to slice-level analysis of 3D T1-weighted MRI. Rather than treating the MRI volume as a unified input, the authors assessed predictive performance across different anatomical orientations. It achieved AUC values above 0.72 for coronal slices, reaching 0.81 for sagittal slices, while axial slices yielded lower performance at 0.69 over a mean follow-up of 3.94 years. This orientation-dependent variability suggests that disability-relevant information is unevenly distributed across MRI representations and may reflect the spatial organization of MS pathology. In particular, sagittal views may better capture periventricular and callosal regions known to be associated with functional impairment. Such variability suggests that disability-relevant information is not uniformly distributed across MRI representations and may reflect regional vulnerability patterns known to be associated with functional impairment in multiple sclerosis. Previous MRI studies have shown that damage in specific anatomical regions, including periventricular and deep gray matter structures, is more strongly associated with disability progression than global lesion burden alone (Cortese et al., 2021; Vaneckova et al., 2022).

Regression-based prediction of future EDSS values demonstrated a distinct perfor-mance profile compared with classification-based approaches, reflecting the intrinsic properties of the EDSS scale rather than limitations of the modeling strategy itself. In the

multicenter imaging challenge cohort analyzed by Roca et al., CNN-derived imaging features combined with ensemble learning achieved a mean squared error of 2.21 for two-year EDSS prediction on the test set, outperforming both a standalone random forest model (MSE 2.56) and an age-only ridge regression baseline (MSE 3.80) (Roca et al., 2020). This improvement highlights the capacity of deep learning–derived representations to enhance quantitative disability prediction beyond conventional machine learning and demographic benchmarks. A similar trend was observed in long-term prediction, where Kırkbir et al. reported root mean square error values of 1.46 for five-year EDSS prediction using LSTM-based models to sequential clinical data and categori-cal MRI activity indicators derived from imaging reports.[17] A baseline root mean square error is 1.46, which improved to 1.41 following hyperparameter optimization and further to 1.33 after feature selection These results indicate that temporal modeling and feature refinement can meaningfully reduce prediction error in continuous EDSS forecasting. Importantly, the remaining error should be interpreted in light of the ordinal, nonlinear, and unevenly distrib-uted nature of EDSS, where small numerical differences do not necessarily correspond to proportional clinical change. Consequently, regression-based modeling provides complemen-tary quantitative insight into disability trajectories, even as categorical and time-to-event for-mulations may offer greater robustness for specific clinical decision-making contexts.

In contrast, longitudinal and time-dependent deep learning models demonstrated higher and more stable discrimination metrics. Mayfield et al. investigated time-dependent deep learning models using longitudinal spinal cord MRI. The study compared recurrent ar-chitectures, including CNN–LSTM, with transformer-based models, specifically a Video Vi-sion Transformer, for predicting trinary EDSS categories and future disability status. [16] The ViViT model achieved an AUC of 0.84 for trinary EDSS classification, outperforming the CNN–LSTM model, which achieved an AUC of 0.74. When applied to exact EDSS predic-tion, performance declined across models, with AUC values of approximately 0.73. These findings reinforce the greater robustness of categorical outcome formulations relative to con-tinuous regression tasks and suggest that attention-based architectures may be better suited to capturing cumulative spinal cord damage and long-range temporal dependencies relevant to disability progression.

In a related but distinct investigation, Mayfield and El Naqa evaluated sequential multise-quence MRI using both classical and quantum-inspired deep learning architectures for disabil-ity forecasting (J. Mayfield & Naqa, 2024). The authors compared CNN–LSTM, ViViT, and Variational Quantum Cir-cuit–LSTM models. Across architectures, ROC–AUC values ranged from 0.73 to 0.77, with overall patient-level classification accuracy of approximately 75%. Notably, the quantum-inspired VQC–LSTM demonstrated predictive performance comparable to classical deep learning models while substantially reducing training time.

Survival-based modeling of progression independent of relapse activity further illustrated the clinical relevance of alternative outcome formulations, framing disability progression as a time-to-event outcome. Coll et al. reported a concordance index of 0.72 for PIRA prediction, with accuracy of 78% in internal validation and 72% in an external cohort, along-side an integrated Brier score of 0.10.18 Time-to-event modeling offers an advantage in capturing the cumulative and irreversible nature of disability progression, which is increasingly recognized as a key driver of long-term outcomes in multiple sclerosis (Eshaghi et al., 2018). Compared with AUC-based classification metrics, the concordance index provides a more clinically meaningful measure of prognostic ranking over time, supporting the potential utility of survival-based deep learning approaches in long-term risk stratification. The inclusion of external validation strengthens confidence in the model's generalizability and highlights the potential of survival-based deep learning approaches for capturing irreversible disability progression in multiple sclerosis.

Overall, predictive performance across the included studies ranged from classification accuracy of approximately 72% to 85%, AUC values between 0.69 and 0.84, root mean square error values between 1.33 and 1.46 for long-term EDSS prediction, mean squared error values of approximately 3.0 for baseline EDSS regression, and concordance indices up to 0.72. These variations reflect not only differences in model architecture but also fundamental constraints imposed by outcome definition, MRI input strategy, and follow-up duration. Baseline

convolutional models offer clinically accessible early risk stratification, whereas lon-gitudinal and survival-based frameworks provide enhanced prognostic insight when sequential imaging is available. Collectively, the findings suggest that the principal value of deep learning in multiple sclerosis lies in clinically meaningful risk stratification and longitudinal prognostic assessment rather than precise numerical forecasting of disability scores.

Despite the promising predictive performance observed across the included studies, several limitations should be considered when interpreting the applicability of deep learning based MRI models for predicting disability progression in multiple sclerosis. A major limitation relates to heterogeneity across datasets, including differences in cohort size, MRI acquisition protocols, follow up duration, and outcome definitions, which complicates direct comparison of predictive metrics and limits generalizability across clinical settings (J. D. Mayfield et al., 2024; Page et al., 2021; Storelli et al., 2022). Variabil-ity in scanner hardware, imaging sequences, and preprocessing pipelines may contribute to domain shift, potentially inflating performance in internally validated cohorts while reducing robustness when models are applied to external populations (Filippi et al., 2019; Mahajan & Ontaneda, 2017).

Another important limitation concerns the use of EDSS as the primary outcome measure. Although EDSS remains the clinical standard for assessing disability progression, it has well documented limitations, including nonlinearity, ordinal scaling, and limited sensitivity to subtle clinical change, particularly in early disease stages (Eshaghi et al., 2018). These characteristics likely con-tribute to the consistently weaker performance of regression based models compared with classification or time to event approaches observed across the included studies (Kırkbir et al., 2025; Roca et al., 2020).

Future research on deep learning based prediction of disability progression in multiple sclerosis should address both methodological development and clinical translation. From a technical perspective, greater emphasis is needed on standardized MRI preprocessing, harmonized outcome definitions, and robust multicenter validation to improve generalizability across scanners, institutions, and patient populations. Expanding datasets to include more diverse demographic and clinical profiles will be essential to reduce potential algorithmic bias and ensure consistent performance across different multiple sclerosis subgroups. In addition, integrating multimodal data, including longitudinal MRI, clinical history, cognitive assessments, and treatment exposure, may further enhance predictive accuracy and better reflect the multifactorial nature of disability progression. Improving model interpretability, calibration, and uncertainty estimation should also be prioritized to support meaningful clinical use.

Ethical and practical considerations will also play a critical role in the future deployment of deep learning models in multiple sclerosis care. Issues related to data privacy, informed consent, and transparency in algorithmic decision making must be carefully addressed, particularly given the sensitive and longitudinal nature of neuroimaging data. The development of clear guidelines for clinical implementation, including best practices for integrating predictive models into existing clinical workflows and electronic health record systems, will be necessary to prevent misuse or overreliance on automated outputs. Finally, prospective and longitudinal studies evaluating the real world impact of deep learning based predictions on treatment decisions, monitoring strategies, and long term patient outcomes are needed to establish the true clinical value of these approaches beyond technical performance metrics.

6. Conclusion

MRI-based deep learning models provide a robust framework for predicting disability progression in multiple sclerosis, consistently outperforming conventional imaging and clinical assessment strategies. By learning complex spatial and temporal patterns directly from high-dimensional MRI data, these models enable earlier and more precise risk stratification, particularly when longitudinal and time-to-event formulations are applied. Importantly, the ability of deep learning approaches to integrate multidimensional imaging information supports more refined clinical assessment and lays the foundation for personalized treatment planning, facilitating individualized therapeutic decisions aligned with each patient's predicted disease trajectory.

References

Cinar, B. P., & Yorgun, Y. G. (2018). What we learned from the history of multiple sclerosis measurement: Expanded Disability Status Scale. Archives of Neuropsychiatry, 55(Suppl. 1), S69.

Coll, L., Pareto, D., Aparicio-Serrano, F., Otero-Romero, S., Cobo-Calvo, A., Reinders, E., Alberich, M., Arévalo, M. J., Arrambide, G., & Auger, C. (2025). Deep learning to predict progression independent of relapse activity at a first demyelinating event. Brain Communications, 7(4), fcaf243. https://doi.org/10.1093/braincomms/fcaf243

Cortese, R., Giorgio, A., Severa, G., & De Stefano, N. (2021). MRI prognostic factors in multiple sclerosis, neuromyelitis optica spectrum disorder, and myelin oligodendrocyte antibody disease. Frontiers in Neurology, 12, 679881. https://doi.org/10.3389/fneur.2021.679881

Eshaghi, A., Prados, F., Brownlee, W. J., Altmann, D. R., Tur, C., Cardoso, M. J., De Angelis, F., Van De Pavert, S. H., Cawley, N., & De Stefano, N. (2018). Deep gray matter volume loss drives disability worsening in multiple sclerosis. Annals of Neurology, 83(2), 210–222. https://doi.org/10.1002/ana.25145

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

Filippi, M., Brück, W., Chard, D., Fazekas, F., Geurts, J. J. G., Enzinger, C., Hametner, S., Kuhlmann, T., Preziosa, P., & Rovira, À. (2019). Association between pathological and MRI findings in multiple sclerosis. The Lancet Neurology, 18(2), 198–210. https://doi.org/10.1016/S1474-4422(18)30451-4

Giovannoni, G., Popescu, V., Wuerfel, J., Hellwig, K., Iacobaeus, E., Jensen, M. B., García-Domínguez, J. M., Sousa, L., De Rossi, N., & Hupperts, R. (2022). Smouldering multiple sclerosis: The real MS. Therapeutic Advances in Neurological Disorders, 15, 17562864211066752. https://doi.org/10.1177/17562864211066752

Havla, J., Reeve, K., On, B. I., Mansmann, U., & Held, U. (2024). Prognostic models in multiple sclerosis: Progress and challenges in clinical integration. Neurological Research and Practice, 6(1).

Hemond, C. C., & Bakshi, R. (2018). Magnetic resonance imaging in multiple sclerosis. Cold Spring Harbor Perspectives in Medicine, 8(5), a028969. https://doi.org/10.1101/cshperspect.a028969

Kırkbir, İ. B., Kurt, B., Boz, C., Terzi, M., & Sarı, A. (2025). Predicting 5-year EDSS in multiple sclerosis with LSTM networks: A deep learning approach to disease progression. Journal of Clinical Neuroscience, 136, 111218. https://doi.org/10.1016/j.jocn.2025.111218

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Mahajan, K. R., & Ontaneda, D. (2017). The role of advanced magnetic resonance imaging techniques in multiple sclerosis clinical trials. Neurotherapeutics, 14(4), 905–923. https://doi.org/10.1007/s13311-017-0568-7

Mayfield, J. D., Murtagh, R., Ciotti, J., Robertson, D., & El Naqa, I. (2024). Time-dependent deep learning prediction of multiple sclerosis disability. Journal of Imaging Informatics in Medicine, 37(6), 3231–3249. https://doi.org/10.1007/s10278-024-01064-1

Mayfield, J., & El Naqa, I. (2024). Evaluation of VQC-LSTM for disability forecasting in multiple sclerosis using sequential multisequence MRI. Quantum Machine Intelligence, 6(2), 41. https://doi.org/10.1007/s42484-024-00162-1

Mollison, D., Sellar, R., Bastin, M., Chandran, S., Wardlaw, J., & Connick, P. (2017). The clinico-radiological paradox of cognitive function and MRI burden of white matter lesions in multiple sclerosis: A systematic review and meta-analysis. PLOS ONE, 12(5), e0177727. https://doi.org/10.1371/journal.pone.0177727

Multiple Sclerosis International Federation. (2020). Atlas of MS 2020: Mapping multiple sclerosis around the world. MSIF.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Reich, D. S., Lucchinetti, C. F., & Calabresi, P. A. (2018). Multiple sclerosis. New England Journal of Medicine, 378(2), 169–180. https://doi.org/10.1056/NEJMra1401483

Roca, P., Attye, A., Colas, L., Tucholka, A., Rubini, P., Cackowski, S., Ding, J., Budzik, J.-F., Renard, F., & Doyle, S. (2020). Artificial intelligence to predict clinical disability in patients with multiple sclerosis using FLAIR MRI. Diagnostic and Interventional Imaging, 101(12), 795–802. https://doi.org/10.1016/j.diii.2020.07.006

Rocca, M. A., Preziosa, P., Barkhof, F., Brownlee, W., Calabrese, M., De Stefano, N., Granziera, C., Ropele, S., Toosy, A. T., & Vidal-Jordana, À. (2024). Current and future role of MRI in the diagnosis and prognosis of multiple sclerosis. The Lancet Regional Health – Europe, 44, 100992. https://doi.org/10.1016/j.lanepe.2024.100992

Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. Annual Review of Biomedical Engineering, 19, 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

Storelli, L., Azzimonti, M., Gueye, M., Vizzino, C., Preziosa, P., Tedeschi, G., De Stefano, N., Pantano, P., Filippi, M., & Rocca, M. A. (2022). A deep learning approach to predicting disease progression in multiple sclerosis using magnetic resonance imaging. Investigative Radiology, 57(7), 423–432. https://doi.org/10.1097/RLI.0000000000000853

Taloni, A., Farrelly, F. A., Pontillo, G., Petsas, N., Gianni, C., Ruggieri, S., Petracca, M., Brunetti, A., Pozzilli, C., & Pantano, P. (2022). Evaluation of disability progression in multiple sclerosis via magnetic-resonance-based deep learning techniques. International Journal of Molecular Sciences, 23(18), 10651. https://doi.org/10.3390/ijms231810651

Thompson, A. J., Banwell, B. L., & Barkhof, F. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. The Lancet Neurology, 17(2), 162–173. https://doi.org/10.1016/S1474-4422(17)30470-2

Vaneckova, M., Piredda, G. F., Andelova, M., Krasensky, J., Uher, T., Srpova, B., Havrdova, E. K., Vodehnalova, K., Horakova, D., & Hilbert, T. (2022). Periventricular gradient of T1 tissue alterations in multiple sclerosis. NeuroImage: Clinical, 34, 103009. https://doi.org/10.1016/j.nicl.2022.103009

Weinstock-Guttman, B., Sormani, M. P., & Repovic, P. (2022). Predicting long-term disability in multiple sclerosis: A narrative review of current evidence and future directions. International Journal of MS Care, 24(4), 184–192.

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine, 170(1), 51–58. https://doi.org/10.7326/M18-1376