Turkish Journal of Neurology

Turk J Neurol 2026;32(4):257-266
Doi: 10.55697/tnd.2026.781

ORIGINAL ARTICLE

Machine learning-based risk profiling of fatigue in people with multiple sclerosis

Gizem Yağmur Yalçın[1], Meryem Koçaslan Toran[2], İsmail Toygar[3,4], Su Özgür[5], Tuncay Gündüz[6], Murat Kürtüncü[6]

[1]Department of Nursing, İstinye University, Faculty of Health Sciences, İstanbul, Türkiye
[2]Department of Midwifery, Üsküdar University, Faculty of Health Sciences, İstanbul, Türkiye
[3]Department of Gerontology, Muğla Sıtkı Koçman University, Faculty of Health Sciences, Muğla, Türkiye
[4]University College Cork, School of Nursing and Midwifery, Cork, Ireland
[5]Department of Biostatistics and Medical Informatics, Ege University, Faculty of Medicine, İzmir, Türkiye
[6]Department of Neurology, İstanbul University, İstanbul Faculty of Medicine, İstanbul, Türkiye

ABSTRACT

Objectives: This study aimed to evaluate fatigue-related clinical and sociodemographic risk profiles in people with multiple sclerosis (MS) using machine-learning algorithms and examined their relative contributions.

Patients and methods: This retrospective cross-sectional study included 932 people with MS enrolled between September 2021 and September 2023. Fatigue was assessed using the Fatigue Severity Scale. Cognitive processing speed was measured by the Symbol Digit Modalities Test, walking performance was assessed by the Timed 25-Foot Walk Test, depressive symptoms were measured by the Hamilton Depression Rating Scale, and disability was measured by the Expanded Disability Status Scale. Disease duration, MS type, and sociodemographic characteristics were recorded. The XGBoost, Random Forest, and LightGBM algorithms were evaluated.

Results: Fatigue was present in 56.1% of participants. The mean age of the patients was 42.9 ± 11.8 years (range, 18 to 77 years). Of the participants, 664 (71.2%) were female, and the remaining 268 (28.8%) were male. The three algorithms showed comparable, modest performance on the independent test set (AUC, 0.72-0.74; accuracy, 0.66-0.68; Matthews correlation coefficient [MCC], 0.30-0.35), with no single model superior across all metrics. Depressive symptoms, walking performance, and cognitive processing speed were among the leading contributors across the three models, although their relative rankings varied between algorithms. Participants with fatigue had significantly lower cognitive processing speed, poorer walking performance, and higher depression scores than those without fatigue ($p < 0.001$).

Conclusion: Tree-based machine-learning frameworks are useful for multivariable risk profiling in people with MS. Although predictive accuracy was modest and similar across algorithms, the models highlight the prominent roles of cognitive processing speed, physical mobility, and depression, supporting multidimensional fatigue risk assessment and facilitating earlier identification of individuals at increased risk of fatigue.

Keywords: Fatigue, machine learning, multiple sclerosis, risk profiles.

Multiple sclerosis (MS) is a chronic, immune-mediated neurological disease characterized by demyelination, inflammation, and neurodegeneration in the central nervous system.[1] The loss of myelin and axonal damage that occur during the disease can result in a wide range of symptoms, including motor, sensory, visual, and cognitive impairment, as well as pain, balance issues, depression, and fatigue.[2] Multiple sclerosis-related fatigue is a prevalent symptom, with prevalence rates ranging from 36.5 to 78% among patients. This symptom exerts a substantial

Correspondence: Meryem Koçaslan Toran, PhD. Üsküdar Üniversitesi, Sağlık Bilimleri Fakültesi, Ebelik Bölümü, 34662 Üsküdar, İstanbul, Türkiye.
E-mail: meryem.kocaslantoran@uskudar.edu.tr
Received: July 23, 2026 **Accepted:** October 21, 2026 **Published online:** November 16, 2026
Cite this article as: Yağmur Yalçın G, Koçaslan Toran M, Toygar İ, Özgür S, Gündüz T, Kürtüncü M. Machine learning-based risk profiling of fatigue in people with multiple sclerosis. Turk J Neurol 2026;32(4):257-266. doi: 10.55697/tnd.2026.781.

© Copyright 2026 by the Turkish Neurological Society. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0/

negative influence on various domains of function, including physical performance, cognitive function, social life, work capacity, and activities of daily living, and results in a decline in quality of life, markedly compromising patients' functional independence.[3]

Multiple sclerosis-related fatigue has been associated with a wide spectrum of factors, including high disability level, progressive disease course, long disease duration, sleep disturbances, depression, anxiety, cognitive impairment, low physical activity, smoking, and adverse socioeconomic conditions.[4] It is also noted that fatigue exerts a significant negative effect on walking performance, information processing speed, and quality of life. However, the heterogeneity of current findings regarding factors associated with MS-related fatigue necessitates a comprehensive evaluation of these relationships. Because fatigue management in MS typically involves neurologists, nurses, physiotherapists, and psychologists working together, approaches that identify individuals at higher risk of fatigue are particularly relevant to routine multidisciplinary care. Structured, multidisciplinary fatigue management programs improve fatigue-related outcomes in people with MS (PwMS).[5]

Machine learning approaches are increasingly used in the field of neurology, with applications in diagnostic assessment, treatment decisions, and risk estimation.[6] It is evident that classical statistics are predicated on a stringent set of assumptions, which can result in the failure to discern complex variable patterns. Conversely, machine learning models have the capacity to discern nonlinear or latent relationships within voluminous data sets.[7] The utilization of data-driven methodologies enables the prediction of outcomes with reduced human bias.[8]

Recent reviews have demonstrated that machine-learning approaches in MS research have predominantly focused on disease classification, subtype differentiation, lesion detection, and treatment-response estimation.[8,9] Despite the prevalence and disabling nature of fatigue as a symptom of MS, studies that employ machine-learning approaches to evaluate clinical and sociodemographic factors associated with fatigue remain limited, particularly in larger clinical samples.

Standardized clinical instruments are essential for mapping the multisystem manifestations of MS. In routine clinical practice and research, key dimensions such as physical disability, cognitive processing speed, depressive symptom severity, and ambulatory function are systematically assessed using well-established instruments, including the Expanded Disability Status Scale (EDSS), the Symbol Digit Modalities Test (SDMT), the Hamilton Depression Rating Scale (HDRS), and the Timed 25-Foot Walk Test (T25FW).[10,11] Previous studies have shown that impairments across these domains are associated with greater fatigue severity. Since fatigue is a multidimensional symptom that evolves throughout the disease course, evaluating these domains individually provides only a limited understanding of fatigue. Integrating routinely collected clinical measures using machine-learning approaches may provide a more comprehensive fatigue risk profile in PwMS, strengthen multidimensional risk assessment, and support earlier, individualized clinical assessment within multidisciplinary care.[12-14] Hence, this study aimed to evaluate the clinical and sociodemographic factors associated with fatigue in PwMS using machine-learning algorithms and to characterize multidimensional fatigue risk profiles.

PATIENTS AND METHODS

The retrospective cross-sectional study employed a machine-learning approach with the objective of evaluating clinical and sociodemographic factors associated with fatigue in PwMS. The study was conducted at the MS outpatient clinic of the İstanbul University, İstanbul Faculty of Medicine Hospital between September 2021 and September 2023. During the study period, 1,350 patients diagnosed with MS according to the McDonald criteria were assessed for eligibility. Of the 1,350 patients assessed for eligibility, 118 were excluded due to being in acute relapse, 78 due to cognitive impairment, 181 due to fatigue-related comorbidities, 36 due to being under 18 years of age, and five due to incomplete data. The final analytical sample consisted of 932 PwMS who met the inclusion criteria. In instances where multiple outpatient assessments were available for a given patient, solely the most recent comprehensive evaluation was considered in the analysis. The Human Research Ethics Committee of İstanbul University, İstanbul Faculty of Medicine waived the requirement for informed consent because of the retrospective study design and the use of de-identified routinely collected clinical data. The study protocol was approved

by the İstanbul University, İstanbul Faculty of Medicine Clinical Research Ethics Committee (Date: December 01, 2023, No: 2023/2160). The study was conducted in accordance with the principles of the Declaration of Helsinki.

The data were retrospectively extracted from the electronic health record (EHR) system using a standardized data extraction form developed by the researchers. The research team conducted the extraction process by means of a review of routinely recorded outpatient clinical assessments and neurological examination records. The sociodemographic and disease-related variables encompassed age, sex, marital status, body mass index, smoking, alcohol use, disease duration (years since diagnosis, as recorded in the EHR), and MS type. The clinical outcomes of interest were routinely recorded during outpatient follow-up visits. The clinical outcomes of interest were routinely recorded during outpatient follow-up visits and retrospectively extracted from the EHR system.

The study population comprised patients aged 18 years and over who had undergone evaluations, including FSS, EDSS, HDRS, SDMT, T25FW, and VAS-QoL, during routine outpatient follow-up visits. Patients who experienced an acute relapse during the assessment period were excluded from the study, as relapse-related inflammatory activity may transiently affect fatigue severity and functional status. Furthermore, patients with cognitive impairment, those with incomplete clinical records, or those with comorbid conditions that could directly contribute to fatigue, including hypothyroidism, diabetes mellitus, and iron deficiency anemia, were excluded from the study. The relevant clinical diagnoses and neurological assessments were confirmed by specialist physicians and documented in the EHR.

Measures

The Expanded Disability Status Scale is a widely used tool for the assessment of neurological disability in PwMS.[11] Scores within this system range from 0 to 10, with higher scores indicating greater neurological disability. All neurological evaluations were conducted by the same neurologist.

The T25FW is a performance-based assessment used for the evaluation of lower extremity function and walking speed in PwMS.[15] Participants were instructed to walk 25 feet as quickly and safely as possible. Two trials were performed, and the mean completion time (in seconds) was used for analysis.

The FSS is a 9-item instrument employed to evaluate the severity and functional impact of fatigue in daily life.[16] The validated Turkish version of the scale was used in this study.[17] Each item is evaluated using a 7-point Likert scale, with higher scores denoting increased fatigue severity. In the present study, the fatigue status of the participants was defined according to the mean FSS score. Participants with a mean FSS score of at least 4 were classified as having fatigue, whereas those with a score of less than 4 were classified as not having fatigue. The FSS score itself was not included among the independent variables in the machine-learning models and was used solely for the purpose of fatigue classification.

The HDRS was used to evaluate the severity of depressive symptoms.[18] The 17-item version of the scale was administered during routine clinical evaluations, with higher scores indicating greater depressive symptom severity.

The SDMT is a neuropsychological assessment used to evaluate information-processing speed, attention, and visuomotor coordination. During the test, participants were required to match symbols with corresponding numbers within a time limit of 90 sec. Scores on this scale were positively correlated with cognitive performance.

The VAS-QoL is a tool used to assess quality of life. The VAS-QoL was used during routine clinical follow-ups as a self-reported instrument to document patients' perceived quality of life on a visual scale ranging from 0 to 100.

Statistical analysis

Data were analyzed using IBM SPSS version 27.0 (IBM Corp., Armonk, NY, USA). The data were presented in the form of frequencies (n), percentages (%), mean and standard deviation, and median with interquartile range (IQR, 25th-75th percentile). The differences between participants experiencing fatigue and those not experiencing fatigue were evaluated using the chi-square test for categorical data. Since none of the continuous variables met the normality assumption (Shapiro-Wilk test, $p < 0.01$ in both groups), the Mann-Whitney U test was used for continuous variables.

Within the predictive modelling framework, machine-learning operations were deployed on cloud computing resources. A rigorous data preprocessing and validation protocol was implemented to ensure the robustness of

outcomes. The dataset was divided into two sets: a training set, constituting 80% of the dataset (n = 745), which was utilized for model development and hyperparameter tuning; and an independent testing set, constituting 20% of the dataset (n = 187) and reserved exclusively for final performance evaluation. The split was stratified by fatigue status with a fixed random seed. A five-fold cross-validation framework was used during the training phase to mitigate the risk of overfitting and ensure stable performance estimates.

The phenomenon of class imbalance was meticulously monitored. Given that the prevalence of fatigue (56.1%) exceeded that of nonfatigue (43.9%) by a margin of 12.2%, the target classes were deemed to be inherently balanced and did not necessitate the implementation of synthetic resampling techniques. The management of missing data was achieved through the implementation of a complete-case analysis strategy. Participants with incomplete records for any primary or secondary clinical variable were excluded prior to analysis to ensure maximum data consistency across frameworks. Hyperparameters of all three algorithms were optimized by grid search with five-fold stratified cross-validation on the training set, using the area under the ROC curve as the selection criterion (XGBoost: learning rate, maximum tree depth, subsample ratio, and number of trees; Random Forest: number of trees, maximum depth, and minimum samples per leaf; LightGBM: learning rate, number of leaves, number of trees, and minimum child samples). The consistency of clinical patterns and the contributions of variables were extracted using gain-based feature-importance calculations.

Grid sampling

The grid search methodology was employed to optimize the hyperparameters. The program systematically examined predefined parameter sets, terminating weak jobs promptly.[19]

XGBoost algorithm

XGBoost (Extreme Gradient Boosting) is a machine learning algorithm that builds a set of decision trees, each of which improves on the errors of the previous one.[20] The primary function of this approach is to regulate overfitting through L1/L2 regularization while also addressing issues such as missing data and imbalanced class sizes. The objective function comprises a loss term and a minor penalty for tree complexity [21]. The XGBoost model was trained iteratively to minimize prediction errors and optimize model performance. The sequential boosting mechanism is illustrated in Supplementary Material 1.

LightGBM (gradient boosting machine)

LightGBM, developed by Microsoft Research Asia (Microsoft Corp., Redmond, WA, USA) in 2017, was evaluated as a potential alternative boosting option.[22] The former is characterized by its superior processing speed and reduced memory requirements, which are advantageous when dealing with substantial data sets.[23] The process of parameter tuning encompassed the adjustment of several key variables, including the learning rate, the depth of the network, the number of leaves, and the sampling ratios.[24] LightGBM was evaluated on an equal footing with the other two algorithms.

Random forest

The Random Forest algorithm is a machine learning method that generates multiple decision trees, each constructed on a random subset of the data. These decision trees are then integrated to produce a consensus prediction. This approach, known as averaging, has been demonstrated to reduce variance and prevent overfitting.[25] In the present study, the variable-importance outputs of all three models were compared to identify predictors that ranked highly across algorithms.

Variable importance

The term variable importance is used to denote the relative contribution of each variable to the model output. Estimation can be facilitated by means of permutation-based methods or tree-based importance scores.[26]

Performance metrics

The performance of the model was evaluated using established evaluation measures, including the AUC, accuracy, precision, recall, F1 score, and MCC. Definitions of true positive, true negative, false positive, and false negative outcomes used in model evaluation are provided in Supplementary Material 2. Accuracy is a measure of the overall proportion of correct predictions; however, this measure may be misleading when class sizes differ. Precision is defined as the proportion of the predicted positives that are correct, while recall is the proportion of actual positive cases identified by the model. The F1 score integrates precision and recall into a unified metric, thereby providing a balanced assessment of performance when both false positives and false negatives are considered. The MCC was used as a summary index of binary

TABLE 1
Descriptive statistics according to fatigue groups (n = 932)

	Fatigue No (n = 409)					Fatigue Yes (n = 523)					Total (n = 932)					Test statistic		
	n	%	Mean ± SD	Median	IQR	n	%	Mean ± SD	Median	IQR	n	%	Mean ± SD	Median	IQR	χ^2	U	*p*
Age (year)			43.0 ± 12.0	43	35-51			42.8 ± 11.7	43	33-51			42.9 ± 11.8	43	34-51	0.794		0.373
Sex																	108766.5	0.657
Female	298	44.9				366	55.1				664	71.2						
Male	111	41.4				157	58.6				268	28.8						
Marital status																13.601		< 0.001
Married	242	39.5				371	60.5				613	65.8						
Single	167	52.4				152	47.6				319	34.2						
BMI (kg/m²)																2.585		0.460
Thin < 18.5	22	45.8				26	54.2				48	5.2						
Normal 18.5-24.99	221	43.2				291	56.8				512	54.9						
Overweight 25.0-29.99	115	41.8				160	58.2				275	29.5						
Obese ≥ 30	47	51.1				45	48.9				92	9.9						
Disease duration (year)			14.0 ± 8.6	13	7-19			13.7 ± 8.4	13	7-19			13.9 ± 8.5	13	7-19		108290.5	0.743
MS type																23.718		< 0.001
RRMS	371	47.3				413	52.7				784	84.1						
SPMS	27	25.2				80	74.8				107	11.5						
PPMS	11	26.8				30	73.2				41	4.4						
Smoking																0.486		0.486
No	283	43.1				374	56.9				657	70.5						
Yes	126	45.8				149	54.2				275	29.5						
Alcohol																0.538		0.463
No	368	43.4				479	56.6				847	90.9						
Yes	41	48.2				44	51.8				85	9.1						
EDSS			2.8 ± 1.7	2.5	1.5-4			2.8 ± 1.8	2	1.5-4			2.8 ± 1.8	2	1.5-4		108103.0	0.605
T25FW (s)			8.5 ± 12.4	6	5-7.5			14.9 ± 28.9	7	6-10.5			12.1 ± 23.3	6.5	5.5-9		68771.0	< 0.001
VAS-QoL			67.8 ± 26.4	75	50-90			56.6 ± 27.2	60	40-80			61.6 ± 27.4	65	50-80		133783.0	< 0.001
SDMT			37.5 ± 16.6	39	26-49			29.5 ± 15.0	30	19-39			33.0 ± 16.2	33	22-45		137795.0	< 0.001
HDRS			5.0 ± 4.7	4	2-7			8.4 ± 6.0	7	4-12			6.9 ± 5.7	6	2-10		68529.0	< 0.001

SD, standard deviation; IQR, interquartile range; χ^2, Chi-square test statistics; U, Mann Whitney U test statistics; $p < 0.05$ Significance level; BMI, body mass index; MS, multiple sclerosis; RRMS, relapsing-remitting multiple sclerosis; SPMS, secondary progressive multiple sclerosis; PPMS, primary progressive multiple sclerosis; EDSS, The Expanded Disability Status Scale; T25FW, Timed 25-Foot Walk Test; VAS-QoL, Visual Analog Scale-quality of life; SDMT, Symbol Digit Modalities Test; HDRS, Hamilton Depression Rating Scale, Disease duration is given in years.

classification quality, as it is capable of accounting for all elements of the confusion matrix and provides a balanced measure even in cases of uneven class distributions.[27]

Statistical and clinical evaluation

The application of distribution-based methods, incorporating chi-square tests and Mann-Whitney U tests, was undertaken to facilitate the evaluation of disparities between participants experiencing fatigue and those not experiencing it. As statistical significance may be influenced by sample size, the clinical relevance of the variables was also taken into consideration during the process of variable selection. The clinical and sociodemographic variables included in the machine-learning analysis were determined based on previous MS fatigue literature and multidisciplinary expert opinion involving a neurologist, an MS nurse, and a physiotherapist.

RESULTS

In this study, 56.1% of participants reported fatigue. The mean age of the patients was 42.9 ± 11.8 years (range, 18 to 77 years). Of the participants, 664 (71.2%) were female, and the remaining 268 (28.8%) were male. The sociodemographic and disease-related characteristics of the participants are shown in Table 1. There were statistically significant differences in marital status, type of MS, and scores on the SDMT, HDRS, and T25FW between those with and without fatigue.

The variable importance scores identified by the predictive modeling are presented in Table 2. The HDRS and T25FW were among the highest-ranking variables in all three models (HDRS = 0.134, 0.144, and 0.207; T25FW = 0.121, 0.138, and 0.209 for XGBoost, Random Forest, and LightGBM, respectively). The SDMT also ranked highly, third in Random Forest (0.144, equal to HDRS) and

TABLE 2
Variable importance scores across machine learning models

XGBoost	Importance	Random forest	Importance	LightGBM	Importance
HDRS	0.134	HDRS	0.144	T25FW	0.209
T25FW	0.121	SDMT	0.144	HDRS	0.207
Marital status	0.090	T25FW	0.138	SDMT	0.153
SDMT	0.084	Age	0.121	Age	0.098
VAS-QoL	0.077	Disease duration	0.111	Disease duration	0.092
EDSS	0.071	VAS-QoL	0.105	VAS-QoL	0.069
Age	0.070	EDSS	0.090	EDSS	0.054
MS type	0.068	BMI	0.047	BMI	0.046
BMI	0.067	Marital status	0.026	Marital status	0.044
Disease duration	0.063	Sex	0.023	Alcohol	0.016
Smoking	0.056	Smoking	0.022	Sex	0.008
Sex	0.055	MS type	0.019	Smoking	0.003
Alcohol	0.047	Alcohol	0.011	MS type	0.001

HDRS, Hamilton Depression Rating Scale; T25FW: Timed 25-Foot Walk Test; SDMT: Symbol Digit Modalities Test; VAS-QoL: Visual Analog Scale-Quality of Life; EDSS: The Expanded Disability Status Scale; MS, multiple sclerosis; BMI, body mass index. Gain-based importance normalised to sum to 1 within each model.

TABLE 3
Performance metrics of the evaluated machine learning models

	XGBoost	RF	LightGBM
Accuracy	0.658	0.684	0.674
AUC	0.739	0.733	0.722
Precision	0.681	0.702	0.700
Recall	0.733	0.762	0.733
F1 score	0.706	0.731	0.716
Matthews correlation coefficient	0.299	0.353	0.334

AUC, area under the curve; RF, random forest. Metrics computed on the independent test set (n = 187).

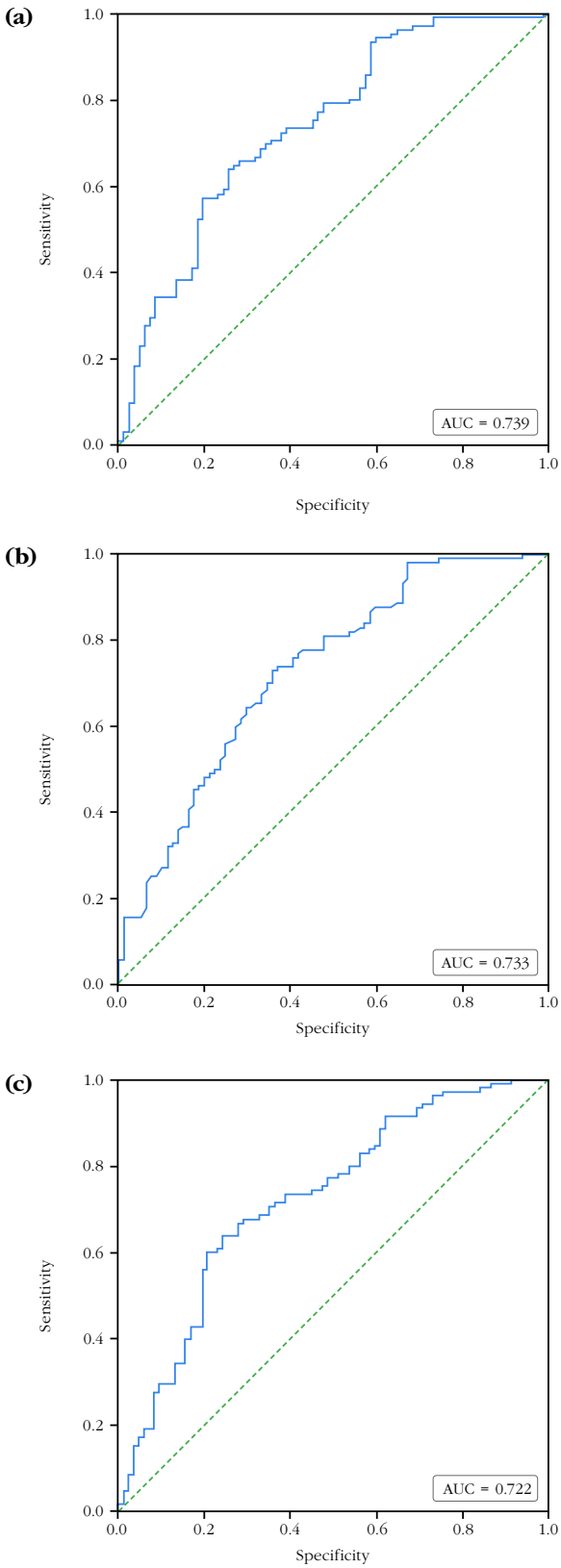

Figure 1. Receiver operating characteristic curves of the three models on the independent test set (n = 187). **(a)** XGBoost; **(b)** Random forest; **(c)** LightGBM.
AUC, area under the curve.

LightGBM (0.153) and fourth in XGBoost (0.084), where marital status (0.090) ranked third. The ordering of the remaining variables differed between algorithms; VAS-QoL, age, and EDSS occupied the middle ranks, whereas MS type, disease duration, body mass index, sex, smoking, and alcohol use contributed little in any model.

The performance metrics of the three models on the independent test set are presented in Table 3, and the corresponding ROC curves are illustrated in Figure 1. XGBoost achieved an AUC of 0.739, accuracy of 0.658, and MCC of 0.299; Random Forest achieved an AUC of 0.733, accuracy of 0.684, and MCC of 0.353; and LightGBM achieved an AUC of 0.722, accuracy of 0.674, and MCC of 0.334. No model was superior across all metrics.

DISCUSSION

The present study used machine-learning algorithms to evaluate the association between clinical and sociodemographic factors and fatigue in PwMS. The three evaluated models showed comparable, modest performance (AUC, 0.72-0.74), and no single algorithm was superior across all metrics. The substantial overlap among their higher-ranking contributors, together with differences in the ranking of other variables, indicates the associations that are robust to the choice of algorithm. Overall, the findings suggest that routinely collected clinical measures can be integrated to characterize multidimensional fatigue risk profiles, highlighting the contributions of cognitive performance, walking function, depressive symptoms, and disease-related characteristics to fatigue.

Among all the variables that were evaluated, the SDMT was among the leading contributors in all three machine-learning models (third in Random Forest and LightGBM, fourth in XGBoost). Its prominence across different algorithms suggests that information-processing speed may be an important component of fatigue-related clinical profiles in MS. Participants experiencing fatigue demonstrated lower SDMT performance, which is consistent with the evidence that slower processing speed is common in MS.[28] The SDMT is regarded as a sensitive measure of information-processing speed and may reflect subtle cognitive dysfunction that is not always apparent during routine neurological examination.[12] As indicated by previous findings, diminished performance

in the SDMT is associated with relapse activity, cognitive impairment, and the overall disease burden in MS.[29] However, the exact relationship between information-processing speed and fatigue remains inconsistent across previous literature.[30] These findings suggest that cognitive processing speed may represent an important component of comprehensive fatigue assessment and should be considered when evaluating fatigue risk in PwMS.

The T25FW also demonstrated a comparatively elevated contribution within the models, thereby corroborating prior findings that impaired walking performance and reduced physical functioning may be associated with the severity of fatigue in MS.[31] It is important to note that walking difficulties may result in an increased physical effort during daily activities, which can contribute to a perceived state of fatigue. Walking tests are also considered useful indicators of overall physical capacity in clinical fatigue evaluation.[32,33] These findings support the inclusion of objective walking performance measures as part of a comprehensive clinical assessment of fatigue in PwMS. Similarly, higher HDRS scores among participants with fatigue are consistent with the well-established association between depressive symptoms and fatigue in PwMS.[14] Frequently, depression and fatigue coexist in MS, and these conditions may interact bidirectionally, exerting a negative effect on both psychological well-being and functional status.[34] These findings reinforce the importance of including psychological status as part of a comprehensive clinical assessment of fatigue in PwMS.

The model output was also influenced by MS type and disease duration, although the relative contributions of these factors were lower than those of SDMT, T25FW, and HDRS. In accordance with preceding research, progressive MS phenotypes and extended disease duration were associated with a heightened fatigue burden, indicative of cumulative neurological impairment.[35] In the final models, the variables of EDSS, age, and marital status were retained; however, it was found that their relative contributions to variance in fatigue-related classification were limited.

In addition, it is important to recognize that all three models exhibited relatively modest overall classification performance. Consequently, these algorithms should not be interpreted as a definitive clinical prediction tool for individual diagnostic decision-making. Instead, its clinical utility lies in pattern identification and multivariable risk profiling. Fatigue in MS is inherently multifactorial, reflecting the interaction of neurological, psychological, and systemic factors. The finding that moderate classification performance can be achieved using routinely collected clinical measures suggests that machine-learning approaches may support multivariable fatigue risk profiling in PwMS, although individual-level prediction remains challenging. Compared with conventional statistical methods based on linear assumptions, tree-based algorithms such as XGBoost offer advantages in capturing complex nonlinear relationships within heterogeneous clinical data. A previous study applied an XGBoost model with SHAP-based variable-importance ranking to routinely recorded clinical data in a Turkish MS cohort and reported that the leading contributors to five-year disability differed across subgroups defined by symptoms at disease onset.[36] In the present study, HDRS and T25FW ranked highest in all three algorithms and SDMT ranked among the top four, whereas the ranking of the remaining variables differed. Therefore, comparing several tree-based algorithms helps to distinguish associations that are robust to the modelling approach from those that are dependent on algorithm.

These findings have practical relevance for clinical care in MS. Since fatigue assessment and follow-up are routinely carried out within multidisciplinary MS teams, a risk profile based on routinely collected measures could support structured fatigue screening in outpatient neurology practice, extending assessment beyond self-reported tiredness to the interrelated cognitive, physical, and psychological domains identified here, and help direct at-risk individuals toward earlier evaluation, referral, and targeted intervention.

The present study had several limitations. First, the retrospective cross-sectional design limits the causal interpretation of the observed associations. Second, the data were collected from a single-center outpatient clinic, which may limit the generalizability of the findings. Third, variables such as sleep quality, medication use, and detailed neuroimaging findings were not available in the EHR system and therefore could not be included in the analysis. Fourth, although the leading contributors overlapped substantially across models, their relative rankings and those of the remaining variables differed between algorithms; therefore, importance scores should be interpreted as indicative rather than

definitive. It is suggested that future multicenter studies incorporating longitudinal follow-up and a broader range of biological and behavioral variables may contribute to a more comprehensive understanding of fatigue-related clinical profiles in MS.

In conclusion, this study demonstrated the utility of tree-based machine-learning frameworks for multivariable risk profiling of fatigue in PwMS. The three algorithms achieved modest yet informative and comparable performance (AUC, 0.72-0.74), with depressive symptoms, walking performance, and cognitive processing speed emerging as leading contributors across models, with rankings that varied between algorithms. Rather than serving as a definitive tool for individual prediction, these findings highlight the interconnected cognitive, physical, and psychological dimensions of fatigue and support its multidimensional assessment. Such profiling may inform structured fatigue screening within multidisciplinary MS care and help identify PwMS who could benefit from earlier, targeted evaluation.

Data Sharing Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions: G.Y.Y.: Conceptualization, methodology, data curation, investigation, writing-original draft, writing-review & editing; M.K.T.: Conceptualization, methodology, data curation, validation, visualization, writing-original draft, writing-review & editing; İ.T.: Investigation, data curation, participant recruitment, writing-review & editing, supervision; S.Ö.: Investigation, data curation, participant recruitment, formal data analysis, writing-review & editing; T.G.: Investigation, resources, writing-review & editing; M.K.: Supervision, methodology, resources, validation, visualization, writing-review & editing.

Conflict of Interest: The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding: The authors received no financial support for the research and/or authorship of this article.

AI Disclosure: The authors declare that artificial intelligence (AI) tools were not used, or were used solely for language editing, and had no role in data analysis, interpretation, or the formulation of conclusions. All scientific content, data interpretation, and conclusions are the sole responsibility of the authors. The authors further confirm that AI tools were not used to generate, fabricate, or 'hallucinate' references, and that all references have been carefully verified for accuracy.

REFERENCES

1. Lorenzut S, Negro ID, Pauletto G, Verriello L, Spadea L, Salati C, et al. Exploring the pathophysiology, diagnosis, and treatment options of multiple sclerosis. J Integr Neurosci 2025;24:25081. doi: 10.31083/JIN25081.
2. Murayama R, Cai Y, Nakamura H, Hashimoto K. Demyelination in psychiatric and neurological disorders: Mechanisms, clinical impact, and novel therapeutic strategies. Neurosci Biobehav Rev 2025;174:106209. doi: 10.1016/j.neubiorev.2025.106209.
3. Oliva Ramirez A, Keenan A, Kalau O, Worthington E, Cohen L, Singh S. Prevalence and burden of multiple sclerosis-related fatigue: A systematic literature review. BMC Neurol 2021;21:468. doi: 10.1186/s12883-021-02396-1.
4. Balshi A, Dempsey JP, Manning N, Kriemler M, Kumbar S, Leuenberger G, et al. Clinical characteristics associated with multiple sclerosis fatigue severity. Mult Scler Relat Disord 2025;100:106528. doi: 10.1016/j.msard.2025.106528.
5. Qomi M, Rakhshan M, Ebrahimi Monfared M, Khademian Z. The effect of distance nurse-led fatigue management on fatigue, sleep quality, and self-efficacy in patients with multiple sclerosis: A quasi-experimental study. BMC Neurol 2023;23:71. doi: 10.1186/s12883-023-03115-8.
6. Chen JH, Asch SM. Machine learning and prediction in medicine - beyond the peak of inflated expectations. N Engl J Med 2017;376:2507-9. doi: 10.1056/NEJMp1702071.
7. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods 2018;15:233-4. doi: 10.1038/nmeth.4642.
8. Ponce de Leon-Sanchez ER, Dominguez-Ramirez OA, Herrera-Navarro AM, Rodriguez-Resendiz J, Paredes-Orta C, Mendiola-Santibañez JD. A deep learning approach for predicting multiple sclerosis. Micromachines (Basel) 2023;14:749. doi: 10.3390/mi14040749.
9. Pilehvari S, Morgan Y, Peng W. An analytical review on the use of artificial intelligence and machine learning in diagnosis, prediction, and risk factor analysis of multiple sclerosis. Mult Scler Relat Disord 2024;89:105761. doi: 10.1016/j.msard.2024.105761.
10. Armutlu K, Keser I, Korkmaz N, Akbiyik DI, Sümbüloğlu V, Güney Z, et al. Psychometric study of Turkish version of Fatigue Impact Scale in multiple sclerosis patients. J Neurol Sci 2007;255:64-8. doi: 10.1016/j.jns.2007.01.073.
11. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An Expanded Disability Status Scale (EDSS). Neurology 1983;33:1444-52. doi: 10.1212/wnl.33.11.1444.
12. Benedict RHB, Amato MP, DeLuca J, Geurts JJG. Cognitive impairment in multiple sclerosis: Clinical management, MRI, and therapeutic avenues. Lancet Neurol 2020;19:860-71. doi: 10.1016/S1474-4422(20)30277-5.
13. Oreja-Guevara C, Ayuso Blanco T, Brieva Ruiz L, Hernández Pérez MÁ, Meca-Lallana V, Ramió-Torrentà L. Cognitive dysfunctions and assessments in multiple sclerosis. Front Neurol 2019;10:581. doi: 10.3389/fneur.2019.00581.

14. AlSaeed S, Aljouee T, Alkhawajah NM, Alarieh R, AlGarni H, Aljarallah S, et al. Fatigue, depression, and anxiety among ambulating multiple sclerosis patients. Front Immunol 2022;13:844461. doi: 10.3389/fimmu.2022.844461.
15. Tiftikçioğlu Bİ. Multiple Sclerosis Functional Composite (MSFC): Scoring instructions. Noro Psikiyatr Ars 2018;55:S46-8. doi: 10.29399/npa.23330.
16. Tedeschi G, Dinacci D, Lavorgna L, Prinster A, Savettieri G, Quattrone A, et al. Correlation between fatigue and brain atrophy and lesion load in multiple sclerosis patients independent of disability. J Neurol Sci 2007;263:15-9. doi: 10.1016/j.jns.2007.07.004.
17. Akdemir A, Türkçapar MH, Orsel SD, Demirergi N, Dag I, Ozbay MH. Reliability and validity of the Turkish version of the Hamilton Depression Rating Scale. Compr Psychiatry 2001;42:161-5. doi: 10.1053/comp.2001.19756.
18. Shmueli A. The Visual Analog rating Scale of health-related quality of life: An examination of end-digit preferences. Health Qual Life Outcomes 2005;3:71. doi: 10.1186/1477-7525-3-71.
19. Microsoft. Hyperparameter tuning a model (v2). Available from: https://learn.microsoft.com/en-us/azure/machine-learning/how-to-tune-hyperparameters?view=azureml-api-2 [Accessed: January 15, 2026]
20. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16); 2016 Aug 13-17; San Francisco, CA, USA. New York: Association for Computing Machinery (ACM); 2016. p. 785-94. doi: 10.1145/2939672.2939785.
21. Inoue T, Ichikawa D, Ueno T, Cheong M, Inoue T, Whetstone WD, et al. XGBoost, a machine learning method, predicts neurological recovery in patients with cervical spinal cord injury. Neurotrauma Rep 2020;1:8-16. doi: 10.1089/neur.2020.0009.
22. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems (NeurIPS 2017); 2017 Dec 4-9; Long Beach, CA, USA. Red Hook (NY): Curran Associates, Inc.; 2017. p. 3146-54.
23. Liang W, Luo S, Zhao G, Wu H. Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms. Mathematics 2020;8:765. doi: 10.3390/math8050765.
24. Özgür S, Orman M. Application of deep learning technique in next generation sequence experiments. J Big Data 2023;10:42. doi: 10.1186/s40537-023-00726-1.
25. Özgür S, Taner S, Bozcuk GG, Ekberli G. Exploring the predictive role of inflammatory markers in neuropathic bladder-related kidney damage with machine learning. J Pediatr Res 2024;11:1-10. doi: 10.4274/jpr.galenos.2024.08624.
26. Genuer R, Poggi JM, Tuleau-Malot C. Variable selection using random forests. Pattern Recognit Lett 2010;31:2225-36. doi: 10.1016/j.patrec.2010.03.014.
27. Hicks SA, Strümke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep 2022;12:5979. doi: 10.1038/s41598-022-09954-8.
28. Loy BD, Taylor RL, Fling BW, Horak FB. Relationship between perceived fatigue and performance fatigability in people with multiple sclerosis: A systematic review and meta-analysis. J Psychosom Res 2017;100:1-7. doi: 10.1016/j.jpsychores.2017.06.017.
29. Benedict RH, Morrow S, Rodgers J, Hojnacki D, Bucello MA, Zivadinov R, et al. Characterizing cognitive function during relapse in multiple sclerosis. Mult Scler 2014;20:1745-52. doi: 10.1177/1352458514533229.
30. Wattjes MP, Ciccarelli O, Reich DS, Banwell B, de Stefano N, Enzinger C, et al. 2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. Lancet Neurol 2021;20:653-70. doi: 10.1016/S1474-4422(21)00095-8.
31. Vister E, Tijsma ME, Hoang PD, Lord SR. Fatigue, physical activity, quality of life, and fall risk in people with multiple sclerosis. Int J MS Care 2017;19:91-8. doi: 10.7224/1537-2073.2015-077.
32. Burschka JM, Keune PM, Menge U, Hofstadt-van Oy U, Oschmann P, Hoos O. An exploration of impaired walking dynamics and fatigue in multiple sclerosis. BMC Neurol 2012;12:161. doi: 10.1186/1471-2377-12-161.
33. Drebinger D, Rasche L, Kroneberg D, Althoff P, Bellmann-Strobl J, Weygandt M, et al. Association between fatigue and motor exertion in patients with multiple sclerosis-a prospective study. Front Neurol 2020;11:208. doi: 10.3389/fneur.2020.00208.
34. Malhi GS, Mann JJ. Depression. Lancet 2018;392:2299-312. doi: 10.1016/S0140-6736(18)31948-2.
35. Healy BC, Zurawski J, Chitnis T, Weiner HL, Glanz BI. Patient-reported outcomes associated with transition to secondary progressive multiple sclerosis. Qual Life Res 2022;31:1799-805. doi: 10.1007/s11136-021-03034-6.
36. Buçan Kırkbir İ, Kurt B, Boz C, Terzi M. Determination of important predictors for the fifth year Expanded Disability Status Scale scores of patients with multiple sclerosis using machine learning. Turk J Neurol 2024;30:157-66. doi: 10.55697/tnd.2024.107.