ARTICLE

ASCPT

Predicting disease activity in patients with multiple sclerosis: An explainable machine-learning approach in the Mavenclad trials

Sreetama Basu[1] | **Alain Munafo[1]** | **Ali-Frederic Ben-Amor[2]** | **Sanjeev Roy[2]** | **Pascal Girard[1]** | **Nadia Terranova[1]** (ORCID)

[1]Merck Institute for Pharmacometrics, Merck Serono S.A. (an affiliate of Merck KGaA, Darmstadt, Germany), Lausanne, Switzerland

[2]Ares Trading SA (an affiliate of Merck KGaA, Darmstadt, Germany), Eysins, Switzerland

Correspondence
Nadia Terranova, Merck Institute for Pharmacometrics, Merck Serono S.A., EPFL Innovation Park–Building I, CH-1015 Lausanne, Switzerland.
Email: nadia.terranova@merckgroup.com

Abstract

Multiple sclerosis (MS) is among the most common autoimmune disabling neurological conditions of young adults and affects more than 2.3 million people worldwide. Predicting future disease activity in patients with MS based on their pathophysiology and current treatment is pivotal to orientate future treatment. In this respect, we used machine learning to predict disease activity status in patients with MS and identify the most predictive covariates of this activity. The analysis is conducted on a pooled population of 1935 patients enrolled in three cladribine tablets clinical trials with different outcomes: relapsing–remitting MS (from CLARITY and CLARITY-Extension trials) and patients experiencing a first demyelinating event (from the ORACLE-MS trial). We applied gradient-boosting (from XgBoost library) and Shapley Additive Explanations (SHAP) methods to identify patients′ covariates that predict disease activity 3 and 6 months before their clinical observation, including patient baseline characteristics, longitudinal magnetic resonance imaging readouts, and neurological and laboratory measures. The most predictive covariates for early identification of disease activity in patients were found to be treatment duration, higher number of new combined unique active lesion count, higher number of new T1 hypointense black holes, and higher age-related MS severity score. The outcome of this analysis improves our understanding of the mechanism of onset of disease activity in patients with MS by allowing their early identification in clinical settings and prompting preventive measures, therapeutic interventions, or more frequent patient monitoring.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
Classical pharmacometric methods are computationally intensive and have a limited ability to exploit high-dimensional covariate data. Hence, incorporating machine-learning (ML) approaches that are better suited for big data in

clinical analysis has the potential to improve prediction in model-informed drug development.

WHAT QUESTION DID THIS STUDY ADDRESS?

This study presents a predictive modeling approach for identifying patients with MS who will have an onset of disease activity in 3 and 6 months along with identification of clinical covariates that drive the model prediction using explainable ML methods.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

This study illustrates how incorporation of interpretable ML methods such as Shapley Additive Explanations (SHAP) with complex, nonlinear "black box" ML models, such as XgBoost into Drug Discovery and Development questions can lead to efficient exploration of high-dimensional patient covariates and assess their contribution to composite clinical end points in MS. The interpretability methods such as SHAP make the decision process of ML models transparent, increasing trust in the models, whereas ensemble methods such as XgBoost enable improved prediction by capturing nonlinear effects and covariate interactions in a data set with a high number of diverse covariates.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

The study demonstrates a framework to use complex "black box" ML models and explainability methods to analyze high-dimensional data and present clinical insights and interpretable covariate importance to further enrich the currently available domain knowledge of MS disease progression and drug response mechanisms in clinical pharmacology.

INTRODUCTION

Multiple sclerosis (MS) is an inflammatory neurological condition that is the leading disabling disease in young adults[1,2] with compounded socioeconomic impact.[3] The pathology and clinical presentation of MS have been widely studied.[1,4,5] Cladribine tablets (Mavenclad®) represent an attractive treatment owing to its benefit–risk balance and convenience of use as an oral disease-modifying drug (DMD) for use in MS treatment.[6,7] The recommended dose is 3.5 mg/kg, consisting of two annual courses, each comprising two treatment weeks 1 month apart and starting at the beginning of two consecutive years. Reviews of cladribine tablets for relapsing MS treatment can be found in Deeks[8] and Rammohan et al.[9]

Cladribine functions by preferentially reducing lymphocytes, key immune cells underlying MS pathogenesis.[10] The clinical pharmacology of cladribine tablets have been investigated in detail including the pharmacokinetics, pharmacodynamics, and exposure–response relationships.[11] Network meta-analysis of 10,825 articles covering 44 studies assessing 12 disease-modifying treatments (DMTs) showed that cladribine tablets are comparatively effective and a safe alternative to other DMTs in both active relapsing–remitting MS (RRMS) and high disease activity populations[12] with respect to multiple clinical

end points such as annualized relapse rate, confirmed disease progression, and no evidence of disease activity.

The efficacy and safety of cladribine tablets in the treatment of patients with MS in different stages of the evolution of the disease have been studied by integrating data from multiple phase III clinical trials.[13,14] The CLARITY (NCT00213135) study in patients with RRMS showed that annualized relapse rates and worsening sustained disability were reduced in patients treated with cladribine tablets compared with patients on placebo.[15,16] The efficacy observed in CLARITY was maintained without further active treatment during CLARITY-Extension (NCT00641537).[14,17] ORACLE-MS (NCT00725985) clinical data showed the efficacy of cladribine in delaying the conversion of patients with a first clinically isolated event to clinically definite MS (CDMS).[18] A significant treatment effect of cladribine tablets was observed in patients who later converted to CDMS and were switched to a different DMD.[19]

The pooled data set from the three trials results in a richly informative population with a long observation period of more than 6 years, with heterogeneous baseline and time-varying covariates that can be used to explore the relationship of multiple efficacy end points with patient (baseline) characteristic and treatment effects. Incorporation of machine-learning (ML) methods to efficiently handle the analysis of high-dimensional data for model-informed

TABLE 1 Input covariates for the P3-T-24 and P3-T-12 models

Patient characteristics + baselines	Age, sex, race, dose (number of weeks of treatment), weight, age of onset of disease, time since first attack, lymphocytes_baseline, EDSS_baseline		
Neurological assessment	Global Age-Related Multiple Sclerosis Severity Score, KFSS1–Bowel and Bladder Functions, KFSS1–Brain Stem Functions, KFSS1–Cerebellar Functions, KFSS1–Cerebral or Mental Functions, KFSS1–Pyramidal Functions, KFSS1–Sensory Functions, KFSS1–Visual or Optic Functions		
MRI assessment	Total number of T1 Gd+ lesions, total T1 hypointense (black holes), total number of T2/flair lesions, T1 Gd+ (volume in mm3), T1 hypointense lesions (volume in mm3), T2 lesions (volume in mm3), combined unique lesion count, new T1 hypointense (black holes)		
Laboratory	Biochemistry: alanine aminotransferase, albumin, alkaline phosphatase, aspartate aminotransferase, bilirubin, blood urea nitrogen, calcium, creatine kinase, creatinine, sodium, potassium, urate, serum protein	Hematology: basophils, basophils/leukocytes, eosinophils, eosinophils/leukocytes, erythrocytes, hematocrit, hemoglobin, leukocytes, lymphocytes, lymphocytes/leukocytes, monocytes, monocytes/leukocytes, neutrophils, neutrophils/leukocytes, platelets	Urinalysis: urine pH, glucose

Note: The laboratory covariates are not collected in routine clinical practice. Hence, the input to P4-T-24 and P4-T-12 models have the same set of input covariates as P3 models except for the laboratory covariates.

Abbreviations: EDSS, Expanded Disability Status Scale; Gd+, gadolinium enhancing; KFSS, Kurtzke Functional Systems Scores; MRI, magnetic resonance imaging; P3-T-12, phase III 12 weeks; P3-T-24, phase III 24 weeks; P4-T-12, phase IV 12 weeks; P4-T-12, phase IV 24 weeks.

drug development has been widely supported and evidenced in the literature,[20-23] including in predicting the MS disease course in patients and the conversion to secondary progressive MS.[24-27] This multivariate data set presents an opportunity to characterize patients who will have disease activity in the future with data-driven ML modeling. Earlier studies have assessed the relation of various baseline and time-varying patient characteristics for overall clinical outcome through time-to-event modeling.[28-31] The focus of our current analysis is on the early identification of patients who will experience the onset of disease activity within a 6-month period and explore which covariates contribute to their early identification.

In this article, we present our findings by applying interpretable ML methods such as Shapley Additive Explanations (SHAP) in combination with XGBoost, an ensemble tree-boosting method, a state-of-the-art approach for many ML challenges.[32-34] These methods generate insights into the contribution of the input covariates into model predictions, thus rendering the decision process of the so-called "black-box" models transparent and informative for clinical patient management. To this end, we studied the top predictive covariates in our models affecting the future disease activity status of patients with MS.

METHODS

In this study, combining the patient populations from ORACLE-MS, CLARITY, and CLARITY-Extension, a disease activity event, for a patient while on cladribine treatment or placebo or observational follow-up, is defined as meeting any of the following five criteria:

C1: one qualified relapse and at least one new T1 gadolinium enhancing (Gd+) lesion during the previous 48 weeks.

C2: one qualified relapse and at least two new or enlarging T2 lesions during the previous 48 weeks.

C3: two or more qualified relapses in absence of any magnetic resonance imaging (MRI) finding during the previous 48 weeks.

C4: 3-month sustained Expanded Disability Status Scale (EDSS) progression.

C5: Required switching to an alternative DMT.

Sustained EDSS progression was defined as an increase in the EDSS score of ≥1 point if baseline EDSS was between ≥1.0 and ≤4.5, ≥1.5 points if baseline EDSS was 0, or ≥0.5 if baseline EDSS was ≥5.0 during a period of at least 3 months.

Although the first four criteria measure various dimensions of disease activity, the last criterion, C5, is a marker of treatment persistence and can be driven by lack of efficacy or a tolerability issue. Because of the complex and multidimensional aspect of disease progression in MS, there is a strong interest in studying composite end points that combine several clinical end points to address better treatment planning and MS patient monitoring. To note, our analysis involves all patients from the three cladribine trials, including those in the placebo arm not receiving cladribine.

Materials

In our analysis, we adopted a supervised ML framework which algorithmically learns the complex relationship

between the high-dimensional input covariate set with the dependent or output covariate. We trained and validated four models that make predictions of disease activity for patients approximately 3 months/12 weeks (T-12) and 6 months/24 weeks (T-24) in advance, as this matches the frequency of patient clinical visits and sampling schedules of various covariates in the trials. Our phase III T-12 (P3-T-12) and T-24 (P3-T-24) models are based on a set of 57 independent covariates, including various patient baseline and time-varying covariates such as patient characteristics, neurological and MRI-based assessments, laboratory measurements (refer to Table 1). The phase IV T-12 (P4-T-12) and T-24 (P4-T-24) models in contrast to the P3 models are based on 25 independent covariates, excluding laboratory covariates, to mimic the scenario of routine clinical practice or phase IV trials where laboratory covariates are not usually collected.

This covariate list includes those that were collected and can be matched across the combined population of 1935 individuals in the three trials. Covariates missing for greater than 20% of the patients are not included (e.g., lymphocyte subset CD4, CD8). This is because such covariates increase the dimension of the feature space without contributing to the predictive power of the model. These can further adversely affect generalizability of model performance. Last observation carried forward imputation was adopted to have a fixed number of input variables for the XGBoost algorithm. The input to the T-12 models are baseline covariates and time-varying covariates from 12 to 20 weeks before the first observation of disease activity for the patient or from the time of their last available observation record. Similarly, input to the T-24 models are baseline covariates and time-varying covariates from 21 to 30 weeks before the first observation of disease activity or from the time of their last available observation record. The representation of the cladribine dose to the model was chosen as the number of weeks of treatment received instead of cumulative dose because the former had fewer confounders such as patient body weight.

The dependent variable is a binary indicator of whether a patient will meet any of the five disease activity criteria in 3 or 6 months. The covariates used in the computation of the dependent variable are explicitly removed from the input set of model covariates to ensure that the retrospective analysis mimics how the information could be used prospectively, where the covariates for clinical determination of disease activity in patients in the future are not available. These covariates are qualified relapse count, new T1 Gd+ lesions count, new and enlarging T2 lesions count, EDSS, and whether the patient required switching to another DMT.

ML methods

The overview of our analysis framework is presented in Figure 1. First, the available patient data is split in an 80%–20% ratio using a stratified random sampling strategy for model training and testing, respectively. During the training process, GridSearch cross-validation[35] was used to select optimal model hyperparameters. We used 10 times repeated 10-fold cross-validation to optimize the model hyperparameters (fraction of covariates to subsample at each split = 0.8, tree depth = 6, minimum weight of leaf nodes for regularization = 15, learning rates = 0.05, and number of estimators = 150), which generalizes best the model performance on unseen data. In cross-validation, the model is trained on a fraction of the training data, and its performance is estimated on the unseen "out of sample" (test) data, and this process is repeated by randomly shuffling the data splits to obtain an averaged performance estimate over several runs. In this way, the training of the classifier and the evaluation of its predictive performance are based on statistically unrelated training and test sets. Therefore, cross-validation provides a more generalizable estimation of the performance of a model on new unseen data.[35] Finally, the trained model is applied on the completely left out test data set to get an unbiased estimate of model performance on future unseen data. This training and evaluation protocol results in a good balance of preventing overfitting and ensuring generalizability of model performance on future unseen data, also termed the *bias-variance trade-off in ML*.

We employed the gradient-boosting, tree-based ensemble ML algorithm XGBoost that has been empirically demonstrated to be highly effective in a variety of problems.[32,33] Final model predictions are obtained by aggregating the predictions of individual decision trees by weighted voting. XGBoost can efficiently handle multicollinearity between covariates as a result of considering each individual covariate independently for the splits during the decision tree building. In fact, because of the ability of ensembles to handle correlated covariates and their interaction well, there is no requirement for covariate selection in the final model. The models have been built with the Python (Version 3.7) interface of the "xgb" library (Version 0.90).

Difficulties in interpreting complex ML models such as XGBoost and their predictions limit the practical applicability of and confidence in ML. In clinical settings, it is critical to understand not only the global aspects of covariate dynamics at a population level but also to understand it specifically for each individual patient. The goal of SHAP is to explain the model prediction for each patient by computing the contribution of each covariate to the prediction using cooperative game theory. The global
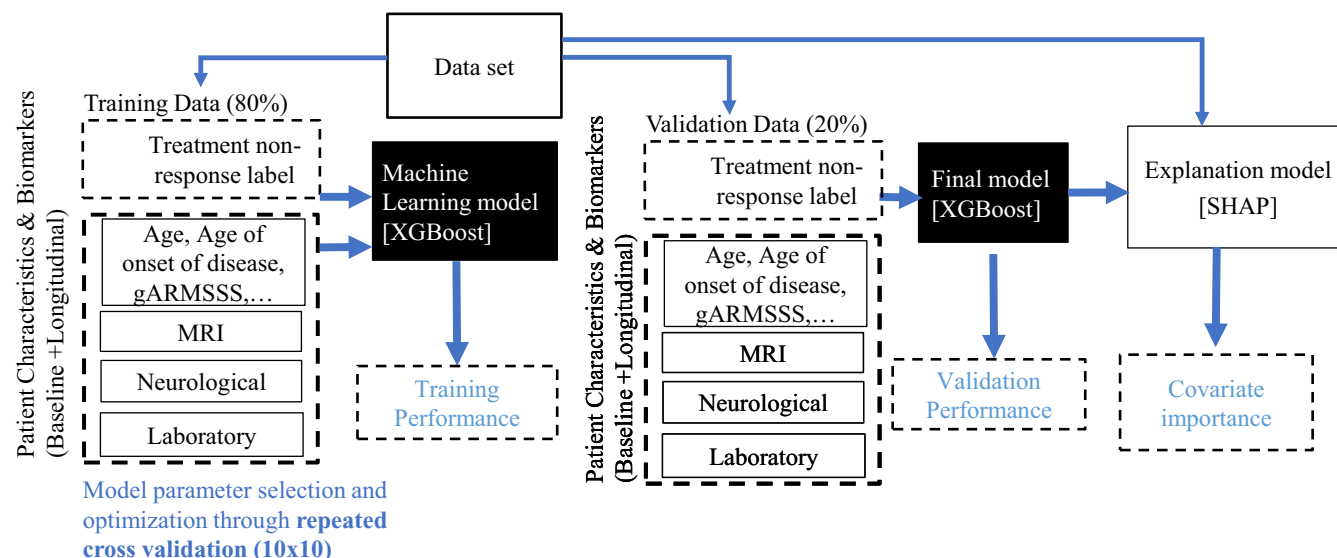
FIGURE 1 Overview of our analysis framework. The available data is split into a 80–20 fraction using stratified random sampling as training and testing data. The training data are used to select optimal XGBoost model parameters using repeated cross-validation, and the final model performance is estimated on the completely unseen test data. In the final step, an explainable machine-learning model SHAP is used to study the covariate contribution to the model predictions and assess covariate importance. gARMSSS, Global Age-Related Multiple Sclerosis Severity Score; MRI, magnetic resonance imaging; SHAP, Shapley Additive Explanations
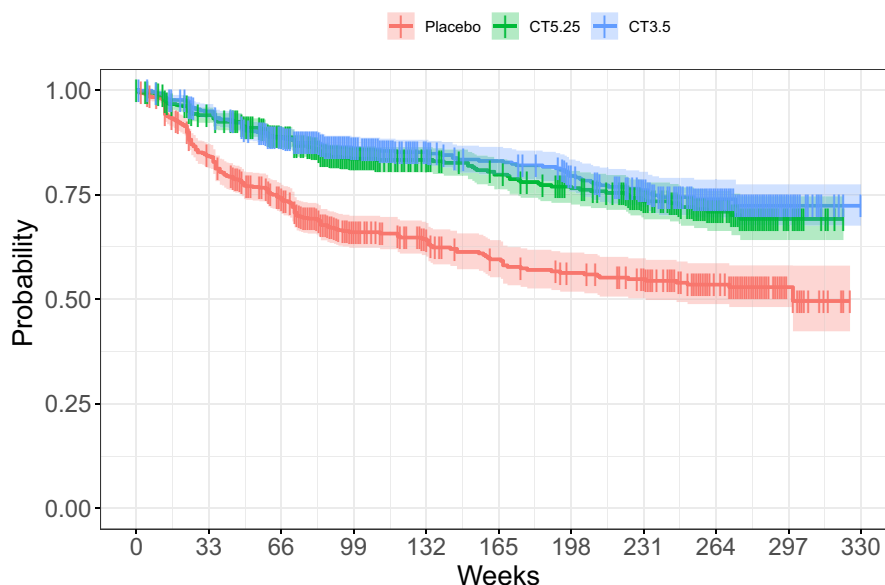

FIGURE 2 Kaplan–Meier survival curves for disease activity in patients in the combined trial population from ORACLE-MS, CLARITY, and CLARITY-Extension. The survival curves are stratified by the treatment arm assignment at the start of the observation period for these three-armed trials. We see that the disease activity free survival probability in the placebo arm (red) drops lower compared with the two treated arms (CT3.5 in blue and CT5.25 in green), showing that there is higher prevalence of disease activity in the placebo population. Vertical bars represent the time of censoring. CT3.5, cumulative cladribine dose of 3.5 mg/kg over 96 weeks; CT5.25, cumulative cladribine dose of 5.25 mg/kg over 96 weeks

ranking of important covariates is obtained by ordering the covariates by their mean absolute SHAP values for all patients (exemplified in Figure 3 and Figure 4). The results were obtained with the Python SHAP library (Version 0.35).

RESULTS

First, after combining the patient populations from the three trials, we examined the proportions of patients with disease activity in the three-treatment arms: placebo,

CT3.5 (cumulative cladribine dose of 3.5 mg/kg over 96 weeks), and CT5.25 (cumulative cladribine dose of 5.25 mg/kg over 96 weeks). Looking at Kaplan–Meier curves of time to first disease activity (Figure 2), there is a higher prevalence of disease activity among patients in the placebo arm compared with the two cladribine treatment arms. These results are in accordance with observations in studies demonstrating the efficacy of cladribine versus placebo in reducing clinical relapses, disability progression, and MRI-assessed disease activity and some aspects of health-related quality of life.[8,16,17,19] Of our combined population of 1935 patients with more than 6 years of observation, approximately 25% (497 patients) met one or more of the five criteria for disease activity. Second, most patients who had disease activity had it early during the observation period, with 80% of observed events in the first 96 weeks of our more than 6 years of observation.

In our analysis of disease activity, we treated all five criteria in the definition equally and assumed no temporal ordering among them. Hence, a patient may meet simultaneously or sequentially multiple criteria during the observation period to qualify as having disease activity, but the time of disease activity is defined as the one corresponding to the first occurrence of any criteria. To quantify the informativeness of each of these criteria, we performed a sensitivity analysis by dropping out one criterion at a time and calculating what percentage of our total number of patients with disease activity fail to be detected with the remaining four criteria.

Table 2 shows the contribution of each of the five criteria toward qualifying disease activity in patients. Dropping C1 criteria does not result in the loss of identification of any patients with disease activity, and in fact the remaining criteria are sufficient to identify this set of patients. This could be explained by the overlap of the relapse count and MRI lesion observations in criteria C2 and C3. On the other hand, dropping the C4 criterion, which is 3-month sustained EDSS progression, results in missing the detection of approximately 42% of the patients with disease activity. This implies that C1 criteria do not uniquely identify disease activity in patients and that the C4 criterion is extremely important as patients meeting the disease activity criteria by C4 do not seem to exhibit the remaining four criteria at any timepoint. Simultaneously, it highlights the

need to include the remaining criteria in our composite analysis objective given the multidimensional aspect of MS, which requires MRI, relapses, and DMT switch observations for qualifying disease activity.

Next, we used a supervised multivariate ML method to train and validate models to predict which patients will have disease activity 3 and 6 months in advance of their clinical determination matching the frequency of patient visits according to the design of the clinical trials and current clinical practice. Specifically, we used XGBoost classifiers and several metrics suitable for studying model performance in cases of unbalanced data (Methods, Table 3, and Appendix S1: Table S1). Overall, we achieved comparable model performance for both 6-month outcome prediction models P3-T-24 and P4-T-24 on test data at balanced accuracies of 80% (Table 3). Sensitivity and specificity are, respectively, 84% and 76% and 81% and 78%, whereas the areas under the receiver operating characteristic curve (ROC) for both models are also 80%. In case of the 3-month outcome prediction models P3-T-12 and P4-T-12, P3-T-12 maintains a similar level of performance with balanced accuracy and area under ROC of 80%, but for P4-T-12, the performance reduces to around 75% for these metrics (refer to Appendix S1: Table S1). Balanced accuracy is a metric in binary classification problems to deal with imbalanced data sets and is defined as the average of sensitivity and specificity. Although intuitively it is expected that the 3-month model should perform better than the 6-month model, given we are closer to the time of event, there are other factors at play such as the frequency of covariate assessment that can impact the performance of the models.

The top 20 most predictive covariates and their relationship to the predicted output by the models are presented in Figure 3 (P3-T-24 and P4-T-24) and in Appendix S1: Figure S1 (P3-T-12 and P4-T-12) in descending order for the models, sorted by their mean absolute SHAP values. This represents a global view at the population level of the important covariates. Covariate importance is understood here as the contribution of covariates in the model to discriminate between the patients who in the future will have disease activity and those who do not. It is interesting to note that there is a strong overlap of the top predictive covariates for all the models

No C1: one qualified relapse and one new T1 Gd+ in 48 weeks	0%
No C2: one qualified relapse and two NE T2 in 48 weeks	3.6%
No C3: two qualified relapses in 48 weeks	6.6%
No C4: 3-month sustained EDSS progression	42.3%
No C5: switching DMT	16.9%

TABLE 2 Percentage of patients with disease activity not detected because of dropping criterion X

Abbreviations: DMT, disease-modifying treatment; EDSS, Expanded Disability Status Scale; Gd+, gadolinium enhancing; NE, new and enlarging.

TABLE 3 Performance estimation of models P3-T-24 and P4-T-24

		P3-T-24		P4-T-24	
		Train (*n* = 1356)	Test (*n* = 340)	Train (*n* = 1356)	Test (*n* = 340)
Specificity	TN/(TN + FP)	0.76	0.76	0.77	0.78
Sensitivity	TP/(TP + FN)	0.81	0.84	0.78	0.81
Balanced accuracy	(Sensitivity + Specificity)/2	0.79	0.80	0.78	0.8
AUC-ROC	Area under curve of ROC	0.79	0.80	0.78	0.8

Note: The table lists the model performance on training and test data with several metrics.

Abbreviations: FP, false positive; FN, false negative; P3-T-24, phase III 24 weeks; P4-T-12, phase IV 24 weeks; ROC, receiver operating characteristic curve; TP, true positive; TN, true negative.
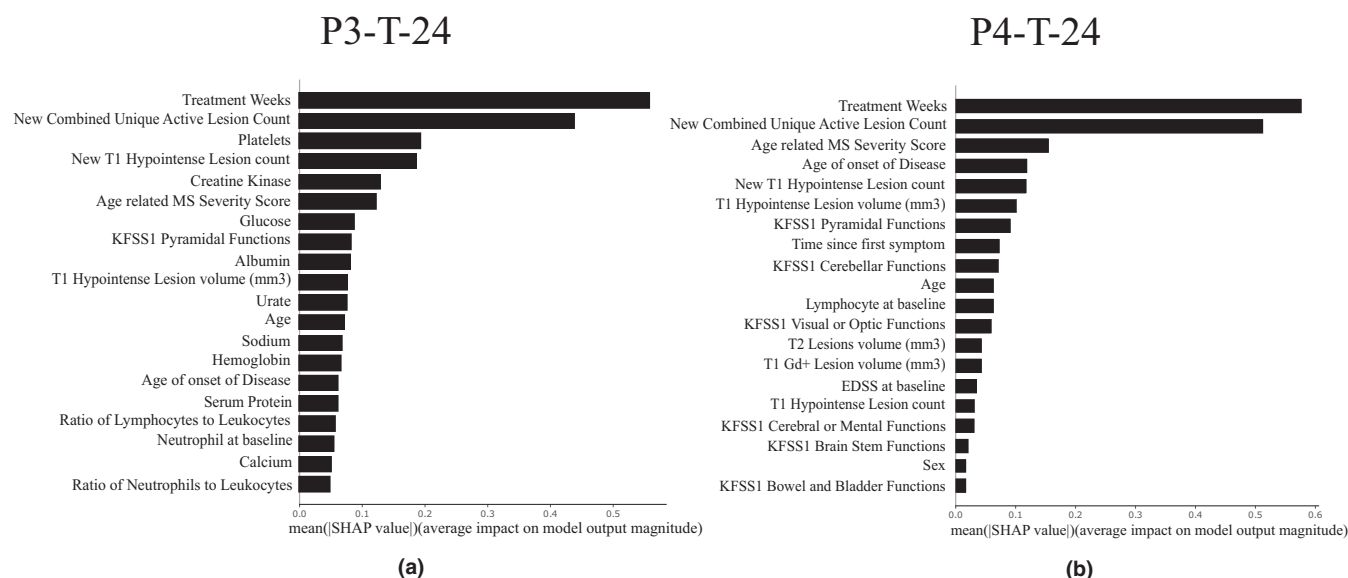
FIGURE 3 Covariates predictive of disease activity in patients 6 months in advance. The list of top predictive covariates sorted in decreasing order by their absolute mean SHAP values from the (a) P3-T-24 and (b) P4-T-24 models. We see that for both models there is a strong overlap of top predictive covariates, including the numbers of weeks of cladribine treatment received, the magnetic resonance imaging measures of new combined unique lesion count and new T1 hypointense lesion count, and other clinically well understood disability measures such as age-related multiple sclerosis severity score. In absence of laboratory covariates in the P4 model, other well-known predictive and prognostic covariates become more important such as T1 hypointense lesion volume, age of onset of disease, and time since first symptom. EDSS, Expanded Disability Status Scale; Gd+, gadolinium enhancing; KFSS, Kurtzke Functional Systems Scores; P3-T-24, phase III 24 weeks; P4-T-12, phase IV 24 weeks; SHAP, Shapley Additive Explanations

including the number of weeks of cladribine treatment received, the new combined unique active (CUA) lesions count (of T1 Gd+ lesions and T2 lesions, without double counting), and the new T1 hypointense lesions count as well as disability measures such as the Age-Related MS Severity Score (ARMSS).[36] Figure 4 shows the global relationship of the top six predictive covariates with the P3-T-24 model output. The model-derived probability for a patient to have disease activity in the future decreases with the increasing number of weeks of treatment up to 4 weeks, and there is no substantial reduction in model prediction for more than 4 weeks of treatment. Similarly, increasing counts of new CUA lesions and new T1

hypointense lesions raise the models' predicted probability of the patient having disease activity in the future. ARMSS, a clinically well-understood representation measure of disability[36] ranging from 0 to 10, also shows that a more advanced disease stage relative to a patient's age cohort (i.e., increasing ARMSS values) increases the model-predicted probability of the patient to have disease activity. The other top predictive factors are laboratory variables such as platelets, creatine kinase, albumin, urate, and so on and prognostic factors such as time since first symptoms and age of onset of disease as well as neurological measures such as Kurtzke Functional Systems Scores Pyramidal Functions.
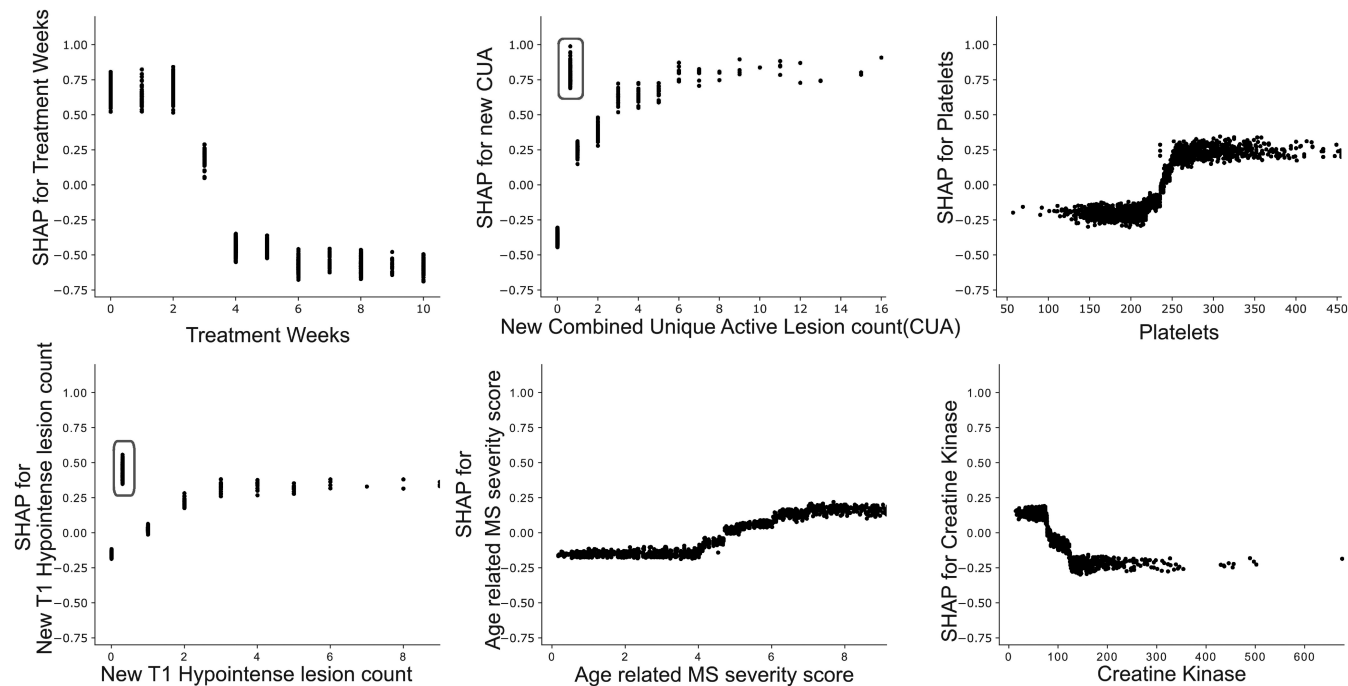
FIGURE 4 Dependency plots show the global relationship between top predictive covariates and the output variable for the P3-T-24 model. More positive SHAP values push model output toward a more confident prediction of disease activity in patients. For example, with the increasing number of weeks of cladribine treatment received, from 0 (placebo) up to 4 weeks, there is a decrease in model output toward a prediction of no disease activity for the patients. The missing values are imputed to population means only for visualization of these dependency plots and are highlighted with gray circles, noticeably for the new CUA lesion count and the new hypointense lesion counts. It is observed that patients who had a missing value for CUA lesion are most at risk for future disease activity events. It shows that missingness for CUA is not at random and in fact informative and related to the event of interest. CUA, combined unique active; MS, multiple sclerosis; SHAP, Shapley Additive Explanations

DISCUSSION

In this study, we developed multivariate ML models to identify predictors of future disease activity in patients with MS 3 and 6 months in advance of clinical determination. It is important to note our observations of disease activity in patients are both right censored as a result of patient dropout as well as interval censored because of missing covariates such as MRI lesions counts, EDSS, or the time of qualified relapses, which are unavailable in the interval between the end of the CLARITY trial and enrollment in the CLARITY-Extension trial. The dropout of patients from the CLARITY and CLARITY-Extension trials (approximately 90% of enrolled patients completed all three arms) were previously investigated, and the efficacy results with the treatment switching were found to be not substantially biased by informative dropout.[37]

The ML models achieve comparable levels of performance with about 80% balanced accuracy on both the training and testing data sets, indicating no obvious model overfitting. XGBoost deals with a large number of input covariates while efficiently handling correlated covariates and missing values without an impact on performance. Given the imbalance in the data where disease activity was

observed in only about 25% of the patient population, in addition to using a boosting algorithm, we also used majority class undersampling as well as a cost-sensitive objective function. For example, mispredicting that a patient will have disease activity incurs a 10% additional penalty compared with mispredicting that a patient will not have disease activity, resulting in higher sensitivity than specificity of the models. Such ML models in conjunction with explainability methods (SHAP) allow us to gain insights on how the most predictive covariates can discriminate between patients who will have disease activity and those who will not.

Among the top predictive covariates, the number of weeks of cladribine treatment received has the largest impact. It is not surprising given the demonstrated effectiveness of the drug in patients experiencing a first demyelinating event and patients with RRMS with four weekly treatments for a sustained effect observed over 4 years.[13,16,18,19] We found that the other top predictive covariates for early detection of disease activity in patients are higher count of new CUA lesions, higher count of new T1 hypointense (black hole) lesions, and clinical disability measures such as higher ARMSSS in all our models. These results are explained by the clinical understanding of the

onset of action of cladribine, which shows a reduction in CUA lesion count in treated patients.[38] In the P4 models, which excluded the laboratory covariates not routinely collected in clinical practice, we observed other well-known markers such as age of onset of disease, time since first symptom, volumes of T2 lesions, and T1 hypointense lesions become more predictive[39,40] in the absence of the laboratory covariates that have not been studied in the MS clinical literature. The role of laboratory covariates such as urate, a predictor of disease activity in our P3-T-12 model (Appendix S1: Figure S1), in MS disease has been previously investigated,[41-45] but its precise understanding is currently lacking. Further investigations should focus on the assessment of the prognostic (treatment independent) versus predictive (treatment dependent) nature of these factors. It is interesting to note the strong overlap of predictive covariates among all our models, which gives further evidence to their robustness.

Such a predictive model will aid in treatment planning, for example, in deciding the frequency of patient visits for at-risk patients, or prompt other clinical interventions. Although our analysis population is different from the label population for cladribine in several countries where it has been approved, this high-dimensional data set enables us to present a feasibility study on how data-driven ML models could help in the future as a clinical decision support tool by presenting an interpretable prediction about how clinically available covariates drive the probability of future disease status in patients with MS. In this regard, it is also interesting to note the updates to clinically relevant covariates attributed to the evolving understanding of MS. For example, recent trials include newer neurological measures of disability progression such as a timed 25-m walk and the nine hole-peg test that have been found to be robust and stable compared with traditional clinical trial end points, such as 3-month EDSS progression.[46]

Prior work has investigated various clinical end points in patient populations with MS using ML techniques to access covariate importance.[24] However, these studies used the in-built covariate importance of various tree-based ensemble methods that have been known in the literature to carry various biases.[47] Our work differs in two key aspects. First, the analysis objective that serves as the dependent output variable for our ML models encompasses multiple clinical end points, which is important as demonstrated by the obtained results to have a more complete picture of disease activation and/or progression. Second, we use the state-of-the-art explainable ML technique SHAP,[48-50] which overcomes the known limitations of the in-built covariate importance assessment methods of the ensemble decision tree methods.

Our findings further highlight the importance of quality MRI evidence in the management of MS disease progression and patient monitoring as highlighted in the MRI in MS guidelines for MS.[51] Although such quality and detailed information regarding counts of various new lesion types are difficult to collect in routine clinical practice and are usually only available in research or during drug development, they are important predictive covariates for MS disease activity in patients. The current model treats the task of predicting the future probability of disease activity as a binary classification. Dynamic prediction models such as recurrent neural networks taking as input longitudinal covariates and predicting an updated probability of the risk of disease activity could be explored in future work.

CONCLUSION

In summary, we conducted an analysis on integrated data from multiple clinical trials at various stages of MS (patients experiencing a first demyelinating event and patients with RRMS) to investigate the predictive covariates of onset of disease activity. Disease activity, as defined by the five criteria, was found to be more frequent among the placebo population, and 3-month sustained EDSS progression was the most informative among the five criteria. The novelty of our work lies in the use of explainable ML methods for training multivariate predictive models combining patient baseline characteristics as well as longitudinal MRI readouts and neurological and laboratory measures to identify patients 3 and 6 months before their clinical observation of disease activity. We trained and validated multivariate nonlinear models with the XgBoost algorithm at 80% balanced accuracy and area under ROC, with different subsets of input covariates available during phase III and phase IV trials to predict patient outcomes of disease activity 3 and 6 months in advance. Predicted disease activity probability in patients was found to decrease with cladribine treatment duration increase, the most important predictor. Additional top predictive covariates from explainable ML models were the count of new CUA lesions, new T1 hypointense lesions, and ARMSS score. Such multivariate predictive models to identify patients early who in the future will have disease activity can improve the available knowledge of the underlying mechanism of disease activity in patients with MS to enable better patient monitoring and treatment planning.

AUTHOR CONTRIBUTIONS

N.T., A.M., and A.-F.B.-A. designed the research. N.T., S.R., P.G., and S.B. performed the research. S.B. analyzed the data. S.B., A.M., A.-F.B.-A., S.R., P.G., and N.T. wrote the manuscript.

ORCID

Nadia Terranova 🄳 https://orcid. org/0000-0002-0033-3695

REFERENCES

1. Lassmann H. Multiple sclerosis pathology. *Cold Spring Harb Perspect Med.* 2018;8:a028936. doi:10.1101/cshperspect.a028936
2. Thompson AJ, Baranzini SE, Geurts J, Hemmer B, Ciccarelli O. Multiple sclerosis. *The Lancet.* 2018;391:1622-1636. doi:10.1016/S0140-6736(18)30481-1
3. Leddy S, Dobson R. Multiple sclerosis. *Medicine.* 2020;48(9):588-594. doi:10.1016/j.mpmed.2020.06.008
4. Oh J, Vidal-Jordana A, Montalban X. Multiple sclerosis: clinical aspects. *Curr Opin Neurol.* 2018;31:752-759. doi:10.1097/WCO.0000000000000622
5. Dobson R, Giovannoni G. Multiple sclerosis – a review. *Eur J Neurol.* 2019;26:27-40. doi:10.1111/ene.13819
6. Torkildsen O, Myhr KM, Bø L. Disease-modifying treatments for multiple sclerosis – a review of approved medications. *Eur J Neurol.* 2016;23:18-27. doi:10.1111/ene.12883
7. Tintore M, Vidal-Jordana A, Sastre-Garriga J. Treatment of multiple sclerosis – success from bench to bedside. *Nat Rev Neurol.* 2019;15:53-58. doi:10.1038/s41582-018-0082-z
8. Deeks ED. Cladribine tablets: a review in relapsing MS. *CNS Drugs.* 2018;32:785-796. doi:10.1007/s40263-018-0562-0
9. Rammohan K, Coyle PK, Sylvester E, et al. The development of cladribine tablets for the treatment of multiple sclerosis: a comprehensive review. *Drugs.* 2020;80:1901-1928. doi:10.1007/s40265-020-01422-9
10. Stuve O, Soelberg Soerensen P, Leist T, et al. Effects of cladribine tablets on lymphocyte subsets in patients with multiple sclerosis: an extended analysis of surface markers. *Ther Adv Neurol Disord.* 2019;12:175628641985498.

11. Hermann R, Karlsson MO, Novakovic AM, Terranova N, Fluck M, Munafo A. The clinical pharmacology of cladribine tablets for the treatment of relapsing multiple sclerosis. *Clin Pharmacokinet.* 2018;58:283-297. doi:10.1007/s40262-018-0695-9
12. Siddiqui MK, Khurana IS, Budhia S, Hettle R, Harty G, Wong SL. Systematic literature review and network meta-analysis of cladribine tablets versus alternative disease-modifying treatments for relapsing–remitting multiple sclerosis. *Curr Med Res Opin.* 2018;34:1361-1371. doi:10.1080/03007995.2017.1407303
13. Cook S, et al. Cladribine tablets in the treatment of patients with multiple sclerosis (MS): an integrated analysis of safety from the MS clinical development program. *Sinapse. Conf. 4th Int. porto Congr. Mult. sclerosis. Port.* 2017.
14. Cook S, Leist T, Comi G, et al. Safety of cladribine tablets in the treatment of patients with multiple sclerosis: an integrated analysis. *Mult Scler Relat Disord.* 2019;29:157-167. doi:10.1016/j.msard.2018.11.021
15. Cook S, Vermersch P, Comi G, et al. Safety and tolerability of cladribine tablets in multiple sclerosis: the CLARITY (CLAdRIbine Tablets treating multiple sclerosis orallY) study. *Mult Scler J.* 2011;17:578-593. doi:10.1177/1352458510391344
16. Rammohan K, Giovannoni G, Comi G, et al. Cladribine tablets for relapsing-remitting multiple sclerosis: efficacy across patient subgroups from the phase III CLARITY study. *Mult Scler Relat Disord.* 2012;1:49-54. doi:10.1016/j.msard.2011.08.006
17. Giovannoni G, Soelberg Sorensen P, Cook S, et al. Safety and efficacy of cladribine tablets in patients with relapsing–remitting multiple sclerosis: results from the randomized extension trial of the CLARITY study. *Mult Scler J.* 2018;24:1594-1604. doi:10.1177/1352458517727603
18. Leist TP, Comi G, Cree BAC, et al. Effect of oral cladribine on time to conversion to clinically definite multiple sclerosis in patients with a first demyelinating event (ORACLE MS): a phase 3 randomised trial. *Lancet Neurol.* 2014;13:257-267. doi:10.1016/S1474-4422(14)70005-5
19. Comi G Freedman MS, Cree BAC, et al. Cladribine tablets in the ORACLE-MS study open-label maintenance period: analysis of efficacy in patients after conversion to clinically definite multiple sclerosis (CDMS). *Multiple Sclerosis. Conference: 32nd Congress of the European Committee for Treatment and Research in Multiple Sclerosis, ECTRIMS.* 2016.
20. Terranova N, Venkatakrishnan K, Benincosa LJ. Application of machine learning in translational medicine: current status and future opportunities. *AAPS J.* 2021;23:1-10.
21. Hutchinson L, Steiert B, Soubret A, et al. Models and machines: how deep learning will take clinical pharmacology to the next level. *CPT Pharma Syst Pharmacol.* 2019;8:131-134.
22. Chan P, Zhou X, Wang N, Liu Q, Bruno R, Jin JY. Application of machine learning for tumor growth inhibition – overall survival modeling platform. *CPT Pharma Syst Pharmacol.* 2021;10:59-66.
23. Park DJ et al. Development of machine learning model for diagnostic disease prediction based on laboratory tests. *Sci Rep.* 2021;11:1-11.
24. Zhao Y, Wang T, Bove R, et al. Ensemble learning predicts multiple sclerosis disease course in the SUMMIT study. *NPJ Digit. Med.* 2020;3(1):135. doi:10.1038/s41746-020-00338-8
25. Law MTK, Traboulsee AL, Li DKB, et al. Machine learning in secondary progressive multiple sclerosis: an improved

predictive model for short-term disability progression. *Mult Scler J – Exp Transl Clin*. 2019;5:205521731988598.

26. Cavaliere C, Vilades E, Alonso-Rodríguez M, et al. Computer-aided diagnosis of multiple sclerosis using a support vector machine and optical coherence tomography features. *Sensors (Switzerland)*. 2019;19(23):5323.

27. Seccia R, Gammelli D, Dominici F, et al. Considering patient clinical history impacts performance of machine learning models in predicting course of multiple sclerosis. *PLoS One*. 2020;15:1-18.

28. O'Riordan JI, Thompson AJ, Kingsley DP, et al. The prognostic value of brain MRI in clinically isolated syndromes of the CNS. A 10-year follow-up. *Brain*. 1998;121:495-503. doi:10.1093/brain/121.3.495

29. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*. 2014;83:278-286. doi:10.1212/WNL.0000000000000560

30. Tedeholm H, Skoog B, Lisovskaja V, Runmarker B, Nerman O, Andersen O. The outcome spectrum of multiple sclerosis: disability, mortality, and a cluster of predictors from onset. *J Neurol*. 2015;262:1148-1163. doi:10.1007/s00415-015-7674-y

31. Gasperini C, Prosperini L, Tintoré M, et al. Unraveling treatment response in multiple sclerosis: a clinical and MRI challenge. *Neurology*. 2019;92:180-192. doi:10.1212/WNL.0000000000006810

32. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. *San Fr. Calif*. 2016.

33. Nielsen D. Tree boosting with XGBoost why does XGBoost win "Every" machine learning competition? *Tree Boost. With XGBoost – Why Does XGBoost Win 'Every' Mach. Learn. Compet*. 2016.

34. Sundrani S, Lu J. Computing the hazard ratios associated with explanatory variables using machine learning models of survival data. *JCO Clin Cancer Informatics*. 2021;364–378:364-378. doi:10.1200/cci.20.00172

35. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Int Jt Conf Artif Intell*. 1995;14(2):1137-1145.

36. Manouchehrinia A, Westerlind H, Kingwell E, et al. Age related multiple sclerosis severity score: disability ranked by age. *Mult Scler*. 2017;8:1938-1946. doi:10.1177/1352458517690618

37. Gorrod HB, Latimer NR, Damian D, Hettle R, Harty GT, Wong SL. Impact of nonrandomized dropout on treatment switching adjustment in the relapsing–remitting multiple sclerosis CLARITY trial and the CLARITY extension study. *Value Heal*. 2019;22(7):772-776. doi:10.1016/j.jval.2018.11.015

38. Comi G, Cook SD, Giovannoni G, et al. MRI outcomes with cladribine tablets for multiple sclerosis in the CLARITY study. *J Neurol*. 2013;260(4):1136-1146.

39. Fisniku LK, Chard DT, Jackson JS, et al. Gray matter atrophy is related to long-term disability in multiple sclerosis. *Ann Neurol*. 2008;64:247-254. doi:10.1002/ana.21423

40. Cierny D, Lehotsky J, Hanysova S, et al. The age at onset in multiple sclerosis is associated with patient's prognosis. *Bratislava Med J*. 2017;118:374-377. doi:10.4149/BLL_2017_071

41. Peng F, Zhang B, Zhong X, et al. Serum uric acid levels of patients with multiple sclerosis and other neurological diseases. *Mult Scler*. 2008;14:188-196. doi:10.1177/1352458507082143

42. Guerrero AL, Gutiérrez F, Iglesias F, et al. Serum uric acid levels in multiple sclerosis patients inversely correlate with disability. *Neurol Sci*. 2011;32:347-350. doi:10.1007/s10072-011-0488-5

43. Ljubisavljevic S, Stojanovic I, Vojinovic S, et al. Association of serum bilirubin and uric acid levels changes during neuroinflammation in patients with initial and relapsed demyelination attacks. *Metab Brain Dis*. 2013;28:629-638. doi:10.1007/s11011-013-9409-z

44. Simental-Mendía E, Simental-Mendía LE, Guerrero-Romero F. Serum uric acid concentrations are directly associated with the presence of benign multiple sclerosis. *Neurol Sci*. 2017;38:1665-1669. doi:10.1007/s10072-017-3043-1

45. Atya HB, Ali SA, Hegazy MI, El Sharkawi FZ. Urinary urea, uric acid and hippuric acid as potential biomarkers in multiple sclerosis patients. *Indian J Clin Biochem*. 2018;33:163-170. doi:10.1007/s12291-017-0661-6

46. Kalincik T, Sormani MP, Tur C. Has the time come to revisit our standard measures of disability progression in multiple sclerosis? *Neurology*. 2021;96:12-13. doi:10.1212/WNL.0000000000011120

47. Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinform*. 2007;8(1):25. doi:10.1186/1471-2105-8-25

48. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inform Proc Sys*. 2017;30.

49. Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *J Comput Aided Mol Des*. 2020;34:1013-1026. doi:10.1007/s10822-020-00314-0

50. Lundberg SM, Erion G, Chen H, et al. Explainable AI for trees: from local explanations to global understanding. *arXiv preprint*. 2019;arXiv:1905.04610.

51. Filippi M, Rocca MA, Ciccarelli O, et al. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. *Lancet Neurol*. 2016;15:292-303. doi:10.1016/S1474-4422(15)00393-2

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.