



OPEN

Multiple sclerosis diagnosis and phenotype identification by multivariate classification of in vivo frontal cortex metabolite profiles

Kelley M. Swanberg^{1,2,5}✉, Abhinav V. Kurada^{1,5}, Hetty Prinsen² & Christoph Juchem^{1,2,3,4}

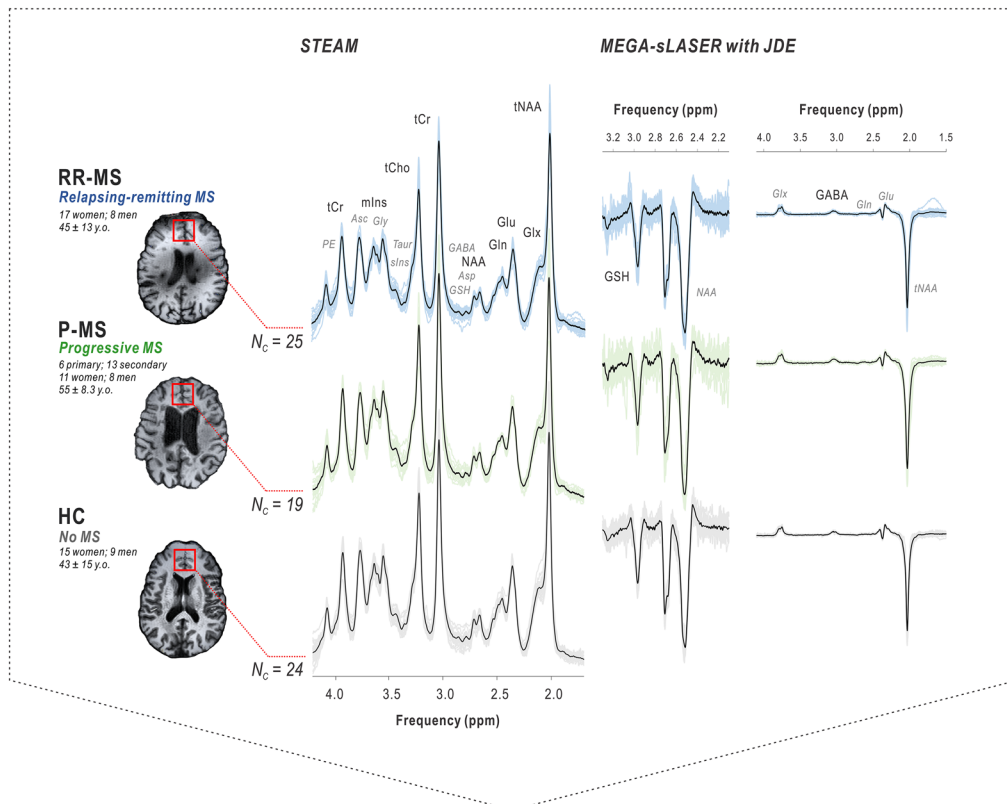
Multiple sclerosis (MS) is a heterogeneous autoimmune disease for which diagnosis continues to rely on subjective clinical judgment over a battery of tests. Proton magnetic resonance spectroscopy (¹H MRS) enables the noninvasive in vivo detection of multiple small-molecule metabolites and is therefore in principle a promising means of gathering information sufficient for multiple sclerosis diagnosis and subtype classification. Here we show that supervised classification using ¹H-MRS-visible normal-appearing frontal cortex small-molecule metabolites alone can indeed differentiate individuals with progressive MS from control (held-out validation sensitivity 79% and specificity 68%), as well as between relapsing and progressive MS phenotypes (held-out validation sensitivity 84% and specificity 74%). Post hoc assessment demonstrated the disproportionate contributions of glutamate and glutamine to identifying MS status and phenotype, respectively. Our finding establishes ¹H MRS as a viable means of characterizing progressive multiple sclerosis disease status and paves the way for continued refinement of this method as an auxiliary or mainstay of multiple sclerosis diagnostics.

Multiple sclerosis is an inflammatory neurodegenerative condition that damages both white and grey matter in the central nervous system. Heterogeneity in the clinical presentation of multiple sclerosis can complicate its diagnosis, typically achieved by a combination of symptomatic report, neurological assessment, magnetic resonance imaging, and occasionally lumbar puncture¹. While recent revisions to the McDonald diagnostic criteria² as applied to magnetic resonance imaging, particularly *T*₁-weighted sequences that can demonstrate local abnormalities in blood–brain barrier permeability to injectable gadolinium contrast indicative of inflammatory lesion activity and *T*₂-weighted FLAIR that can indicate lesions of some age, have improved diagnostic accuracy for new multiple sclerosis cases, specificity remains low³.

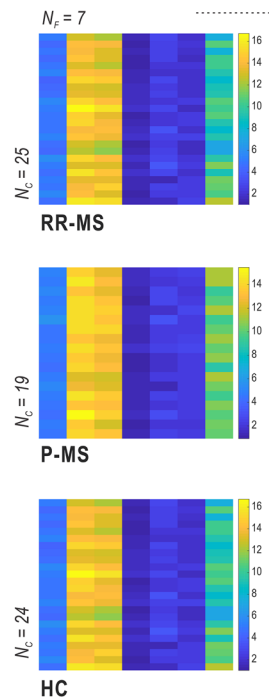
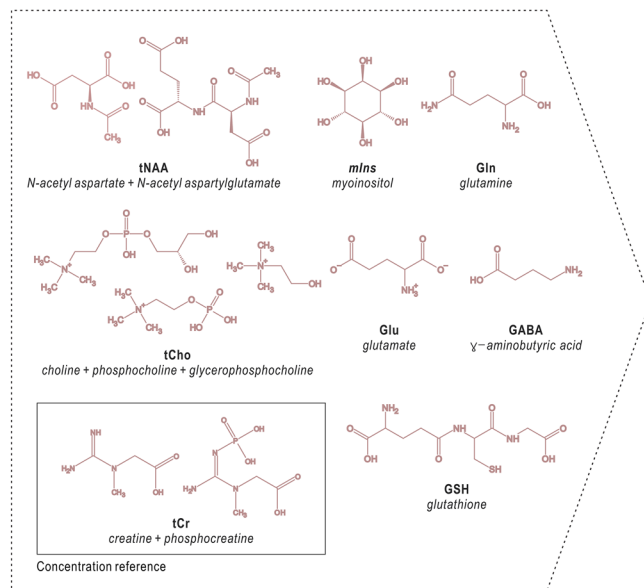
Compounding both the difficulty and importance of accurate multiple sclerosis diagnosis is the existence of diverse disease courses with disparate responsiveness to currently available disease-modifying therapies. The majority of individuals with multiple sclerosis exhibit the relapsing–remitting phenotype, marked largely by intermittent immunological flares called relapses, for which the arsenal of modern pharmacotherapies, including monoclonal antibodies, inhibitors of immune cell proliferation or migration, and other immunosuppressive drugs, has demonstrated relative efficacy. Up to a third or more of these individuals, however, will transition to the secondary progressive phenotype, marked rather by steady neurodegeneration exhibited by cortical atrophy and functional decline largely resistant to currently available treatments¹. About 15% of patients manifest a progressive phenotype from the outset with minimal overt relapse, called primary progressive multiple sclerosis. Uncertainty surrounds not only initial identification but also phenotypic classification of multiple sclerosis, especially during the transition from relapsing to progressive manifestation, for which a mean duration of diagnostic uncertainty as high as three years has previously been calculated⁴.

¹Department of Biomedical Engineering, Columbia University Fu Foundation School of Engineering and Applied Science, 351 Engineering Terrace, 1210 Amsterdam Avenue, Mail Code: 8904, New York, NY 10027, USA. ²Department of Radiology and Biomedical Imaging, Yale University School of Medicine, New Haven, CT, USA. ³Department of Radiology, Columbia University College of Physicians and Surgeons, New York, NY, USA. ⁴Department of Neurology, Yale University School of Medicine, New Haven, CT, USA. ⁵These authors contributed equally: Kelley M. Swanberg and Abhinav V. Kurada. ✉email: k.swanberg@columbia.edu

1. Data Collection



2. Feature Extraction



3. Supervised Learning

◀Figure 1. Frontal cortex metabolite profiling by proton magnetic resonance spectroscopy in individuals with relapsing–remitting, progressive, and no multiple sclerosis. Seven metabolites were measured in a single 27-cc cubic voxel in the prefrontal cortex of individuals with relapsing–remitting ($N = 25$), progressive ($N = 19$), or no ($N = 24$) multiple sclerosis. Individual spectra shown in color; group averaged spectra shown in black. Signals from five compounds, including total N-acetyl aspartate (tNAA: N-acetyl aspartate plus N-acetyl aspartylglutamate), total choline (tCho: choline, phosphocholine, and glycerophosphocholine), myoinositol (mIns), glutamate (Glu), and glutamine (Gln), were obtained via macromolecule-suppressed STimulated Echo Acquisition Mode (STEAM; echo time T_E 10 ms, repetition time T_R 3 s, mixing time T_M 50 ms, inversion time T_I 300 ms); glutathione (GSH) and GABA were estimated by J -difference editing (JDE) on a semi-Localization by Adiabatic SElective Refocusing sequence (MEGA-sLASER; T_E 72 ms, T_R 3 s). Total creatine (tCr) served as a concentration reference for these seven metabolite features, which were calculated in millimolar (mM) assuming 10 mM tCr; additional metabolites listed in italics, including aspartate (Asp), ascorbate (Asc), phosphorylethanolamine (PE), glycine (Gly), taurine (Taur), scylloinositol (sIns), GSH and GABA from STEAM, and glutamate, glutamine, and/or NAA in J -difference-edited spectra, were employed in spectral data quantification for one or more experiments but not included in the machine-learning feature set. RR-MS relapsing–remitting multiple sclerosis, P-MS progressive multiple sclerosis, HC healthy control, ppm parts per million, N_C number of cases, N_F number of features.

Despite continued shortcomings of current imaging-supported diagnostic pipelines in identifying multiple sclerosis, magnetic resonance techniques in general are attractive as potential diagnostic tools. Magnetic resonance is noninvasive and safe, facilitating its repeated use not only for initial identification of disease state but also for continued monitoring of treatment and transition between phenotypes. In part for these reasons, ‘advanced’ magnetic resonance techniques, including magnetic transfer imaging^{5,6}, diffusion tensor imaging⁷, and proton magnetic resonance spectroscopy⁸ have been explored for their potential use in multiple sclerosis prognosis, disease progression, or diagnosis⁹.

Among these, proton magnetic resonance spectroscopy (¹H MRS) is unique in its ability to simultaneously query concentrations of several small-molecule metabolites in one or more regions of interest. This is particularly advantageous for diagnosing a disease like multiple sclerosis, associated to some degree of reproducibility with abnormalities in multiple ¹H-MRS-visible metabolites, including N-acetyl aspartate, choline, myoinositol, glutamate, glutathione, GABA, and possibly common ¹H-MRS quantification reference creatine⁸. It is also potentially advantageous for differentiation among different multiple sclerosis phenotypes, especially given that progressive subtypes may predominantly exhibit cortical as opposed to active white-matter lesioning¹⁰, the former largely invisible to conventional clinical imaging^{11–13}. While some evidence exists that relapsing and progressive multiple sclerosis may express differential metabolic signatures in white-matter N-acetyl aspartate and perhaps creatine as well as grey-matter N-acetyl aspartate and inositols, direct phenotypic comparisons by proton spectroscopy remain sparse⁸.

Given previously published findings of $80 \pm 13\%$ (average \pm standard deviation over comparisons reported) sensitivity and $63 \pm 13\%$ specificity for MS diagnostics combining imaging and CSF analysis¹⁴, the 80% thresholds previously recommended for each index to qualify a measure as a novel Alzheimer’s disease diagnostic biomarker¹⁵ might be reasonably applied to the improvement of multiple sclerosis diagnosis as well. One major shortcoming of proton magnetic resonance spectroscopy as currently employed is the typically low sensitivity and specificity of any single metabolite to disease effects, and currently published investigations demonstrate multiple sclerosis to be no exception. Reductions in N-acetyl aspartate, the metabolite demonstrating the most extensively reported abnormality in multiple sclerosis central nervous tissue, are neither reliably reproduced nor specific to this disease⁸. Similarly, a promising finding of diffusely localized cortical and subcortical choline concentration as a 100% sensitive and 90% specific biomarker for relapsing–remitting multiple sclerosis in a cohort of 20 individuals¹⁶ has yet to be replicated among a larger sample.

Despite the failure of nearly thirty years of proton magnetic resonance spectroscopy research to identify a single-molecule diagnostic biomarker for multiple sclerosis, it is plausible that the rich multivariate data sets provided by this method enable the accurate classification of individuals by multiple sclerosis state and/or phenotype when assessed using measures that more fully address their inherent complexity. Recursive methods for classifying data via simultaneous consideration of multiple variables are growing in feasibility and therefore popularity for uncovering patterns that may be too subtle and/or complex for traditional hypothesis testing, typically of one dependent variable at a time or combination of two at most, for instance, as ratios. Two of the most widely employed methods for classifying small data sets from multiple sclerosis patients have been support vector machines (SVM), used on a variety of data types to separate multiple sclerosis patients from control^{17–26}, each other^{19,27,28}, future non-converters with clinically isolated syndrome²⁹, and individuals with other neurological disorders³⁰, as well as random forest algorithms, used to separate patients from control^{18,22,25,26,31–34} and individuals with neuromyelitis optica³¹. Additional techniques used to classify multiple sclerosis state or subtype on the basis of non-MRS data sets have included neural networks^{18,25,26,35–39}, K-nearest neighbors^{17,20,25,27,37} (KNN), decision trees^{17,18,26,40}, logistic regression^{17,27}, Naïve Bayes²⁵, and least squares²⁷ or maximum likelihood estimation⁴¹. A range of classifiers has additionally been employed, also on non-MRS data, to characterize or predict disease conversion⁴², symptom severity^{43–51}, or treatment effect^{52–55}, and especially to automatically segment MRI-visible multiple sclerosis lesions⁵⁶.

With a nested optimization pipeline designed with care to the considerations inherent in potentially overfitting flexible classifiers to the small data sets typical of ¹H MRS studies, here we apply a number of supervised classification algorithms to explore the feasibility of diagnosing multiple sclerosis from proton magnetic resonance spectroscopy measurements of prefrontal cortex metabolite concentrations alone (Fig. 1); details regarding the

Model hyperparameters optimized			
Model	Scaling transformation ^a	Model-specific hyperparameters	
QDA	min-max (0 to 1) unit normalization standardization quantile transformation maximum value normalization quadratic features	N/A	
KNN		neighbors	2-15
		p	1, 2
		γ	0.01, 0.025, 0.05, 0.075, 0.91, 0.1, 0.11, 0.125, 0.15, 0.2, 0.25, 0.5, 1
		c	0.1, 0.2, 0.3, 0.4, 0.5:0.05:1.5, 2, 3, 4, 5, 10
SVM		kernel	Poly (2-6), RBF, linear
Final hyperparameters by question and model type			
Model	Question	Scaling transformation	Model-specific hyperparameters
QDA	Control vs. MS	unit normalization	N/A
	Control vs. RR-MS	unit normalization	N/A
	Control vs. P-MS	quantile transformation	N/A
	RR vs. P-MS	quantile transformation	N/A
KNN	Control vs. MS	unit normalization	5 neighbors; p=2
	Control vs. RR-MS	unit normalization	7 neighbors; p=2
	Control vs. P-MS	standardization	7 neighbors; p=2
	RR vs. P-MS	min-max (0 to 1)	5 neighbors; p=2
SVM	Control vs. MS	unit normalization	$\gamma=(\sigma_x^2 N_f)^{-1}$; c=1; RBF kernel
	Control vs. RR-MS	unit normalization	$\gamma=(\sigma_x^2 N_f)^{-1}$; c=1; RBF kernel
	Control vs. P-MS	min-max (0 to 1)	$\gamma=1.1$; c=0.7; RBF kernel
	RR vs. P-MS	quantile transformation	$\gamma=(\sigma_x^2 N_f)^{-1}$; c=1; RBF kernel

Table 1. Model hyperparameters optimized against average training cross-validation accuracy. ^aScaling optimized over default model-specific hyperparameters. σ_x^2 feature variance, N_f number of features, *RBF* radial basis function, *QDA* quadratic discriminant analysis, *KNN* K-nearest neighbors, *SVM* support vector machines.

methodological validation of data collection procedures⁵⁷; the acquisition, processing, and univariate analysis⁵⁸; and proof-of-concept classification analysis⁵⁹ of this data set have been reported previously. Due to the low degrees of freedom offered by a small number of cases, we split our diagnostic pipeline into two independent problems: First, deciding whether an individual has multiple sclerosis; second, deciding whether individuals with multiple sclerosis exhibit a relapsing–remitting or progressive phenotype. Binary classifications between controls and each phenotype separately were further developed to inform the interpretation of potential differences in model accuracies.

Analysis pipelines thus constituted four types of binary classification. The first established among all available participants the absence or presence of multiple sclerosis. The second decided among all participants with multiple sclerosis whether they exhibited the relapsing or progressive phenotype. Two additional comparisons between control and each multiple sclerosis subtype were also performed to enhance the interpretability of relative model accuracies. Support vector machines (SVM), K-nearest neighbors (KNN), and quadratic discriminant analysis (QDA) were selected for detailed report among all approaches employed for testing based on preliminary held-out validation accuracies in the relapsing versus progressive question. Additional classification algorithms tested but not refined or reported included random forests, gradient-boosted decision trees, extremely randomized trees, and logistic regression.

Dimensionality reduction on the already small feature set of seven metabolites (glutamate, glutamine, GABA, glutathione, total choline, total N-acetyl aspartate, and myoinositol) was not performed to avoid reduction in model interpretability. Data were split into training and validation sets by removing one data set at a time for use as a held-out validation set, yielding a total of $N_1 \times 2$ runs for data with smaller group size N_1 . Leave-one-out cross-validation (LOOCV) statistics were calculated on each training set for aggregate use in feature selection

and optimization of training hyperparameters (Table 1). This bootstrapping strategy, in which LOOCV accuracies were aggregated across multiple training sets, each itself missing one held-out validation set, struck a balance between the low statistical confidence offered by the small validation set N_s of a single-loop approach and the computational intensity of separately optimizing features and hyperparameters against individual training sets for each held-out validation case (i.e., a true test case) in a fully nested approach. Notably, the small range of parameters optimized did not render hyperparameter overfitting a concern for this particular application. Under-sampling proceeded similarly for cross-validation as for held-out validation in that each of the training cases was removed once for use as a cross-validation set for that training run, with a total of $N_2 \times 2$ cross-validations for data with the smaller training group size N_2 to ensure class balancing in the population of cross-validation sets used for each training run. Class sizes were then balanced after removal of the aforementioned held-out validation and cross-validation sets using the Synthetic Minority Over-Sampling Technique (SMOTE) such that data points were interpolated in the smaller class along line segments joining each case to a predefined number of its nearest neighbors until the class equaled the size of the larger⁶⁰.

Each training set was scaled using a matrix calculated over that training set only, after which the same scaling was transferred to the held-out validation case. This enabled each separate model to minimize bias to held-out validation cases while also allowing for the calculation of population validation set statistics via the inclusion of all participants in both training and held-out validation sets. Optimized hyperparameters and model performance were averaged across all runs for each algorithm (Fig. 2).

Results

Classifiers separating progressive from relapsing multiple sclerosis exhibited increased performance relative to those separating multiple sclerosis from control. Average held-out validation accuracies were higher for separations of relapsing versus progressive (QDA 79%, KNN 74%, and SVM 68%) than for multiple sclerosis versus control (QDA 60%, KNN 63%, SVM 58%). Accuracy averaged across all three classifiers was significantly higher for models separating relapsing from progressive multiple sclerosis than for all multiple sclerosis from control ($73.7 \pm 5.5\%$ versus $60.3 \pm 2.5\%$, $t(2) = 3.82$, two-tailed $p < 0.05$).

The same trend was also seen in higher held-out validation accuracies for progressive versus control classifications (QDA 68%, KNN 74%, and SVM 74%) than for relapsing–remitting versus the same control (QDA 52%, KNN 58%, and SVM 52%), yielding a mean difference that also exceeded the threshold for statistical significance despite a small sample size of 3 classifiers ($72.0 \pm 3.5\%$ versus $54.0 \pm 3.5\%$, $t(2) = 6.30$, two-tailed $p < 0.01$) (Fig. 3, 4; Table 2).

Quadratic discriminant analysis and K-nearest neighbors but not support vector machines yielded 80+% sensitivity or specificity for at least one classification question. Quadratic discriminant analysis yielded the highest classification accuracies for identification of progressive multiple sclerosis as distinct from relapsing–remitting (held-out validation accuracy 79%; sensitivity 84%, specificity 74%; area under the receiver operating characteristic or AUROC 0.86) as well as from control (held-out validation accuracy 68%; sensitivity 58%, specificity 79%; AUROC 0.83). K-nearest neighbors exhibited slightly lower performance than QDA for its top-performing applications, distinguishing progressive from relapsing multiple sclerosis (held-out validation accuracy 74%; sensitivity 68%, specificity 78%; AUROC 0.78) and from control (held-out validation accuracy 74%; sensitivity 95%, specificity 53%; AUROC 0.72). By contrast, support vector machines failed to reach either 80% sensitivity or specificity for any question, though overall accuracy for the classification of progressive multiple sclerosis relative to relapsing–remitting (held-out validation accuracy 68%; sensitivity 68%, specificity 68%; AUROC 0.68) and control (held-out validation accuracy 74%; sensitivity 79%, specificity 68%; AUROC 0.77) still exceeded those of the other two classifications (Table 2).

Metabolite feature importance differed more by classification question than by algorithm. Total choline ranked among the most important three features for all progressive versus relapsing models but none of the control versus multiple sclerosis models. Total choline was also the most important feature for distinguishing progressive from control in QDA, while it did not feature among the top three for any model distinguishing relapsing from control. Similarly, glutamine and glutathione ranked among the three most important features for two of the three progressive versus relapsing models but only one (glutamine) or none (glutathione) of those distinguishing multiple sclerosis from control. Additionally, GABA ranked among the three most important features for two of three models distinguishing progressive MS from control but none distinguishing relapsing–remitting MS from control.

Inversely, two metabolites (myoinositol and glutamate) ranked among the three most important retained features for all three control versus multiple sclerosis models but only one (myoinositol) or none (glutamate) of the progressive versus relapsing models. Both myoinositol and glutamate were each also ranked among the top three features for at least five of six (myoinositol) or all six (glutamate) models distinguishing each multiple sclerosis subtype from control (Table 3).

Discussion

We have presented evidence that individuals with progressive multiple sclerosis can be accurately distinguished from those with relapsing–remitting and no multiple sclerosis on the sole basis of frontal-cortex metabolite concentrations, including glutamate, glutamine, glutathione, GABA, *N*-acetyl aspartate, choline, and myoinositol, measured by proton magnetic resonance spectroscopy at 7 T. By contrast, the same data did not support accurate separation from control of relapsing–remitting multiple sclerosis or of both multiple sclerosis subtypes pooled as a single classification. This finding underlines the potential of proton magnetic resonance spectroscopy

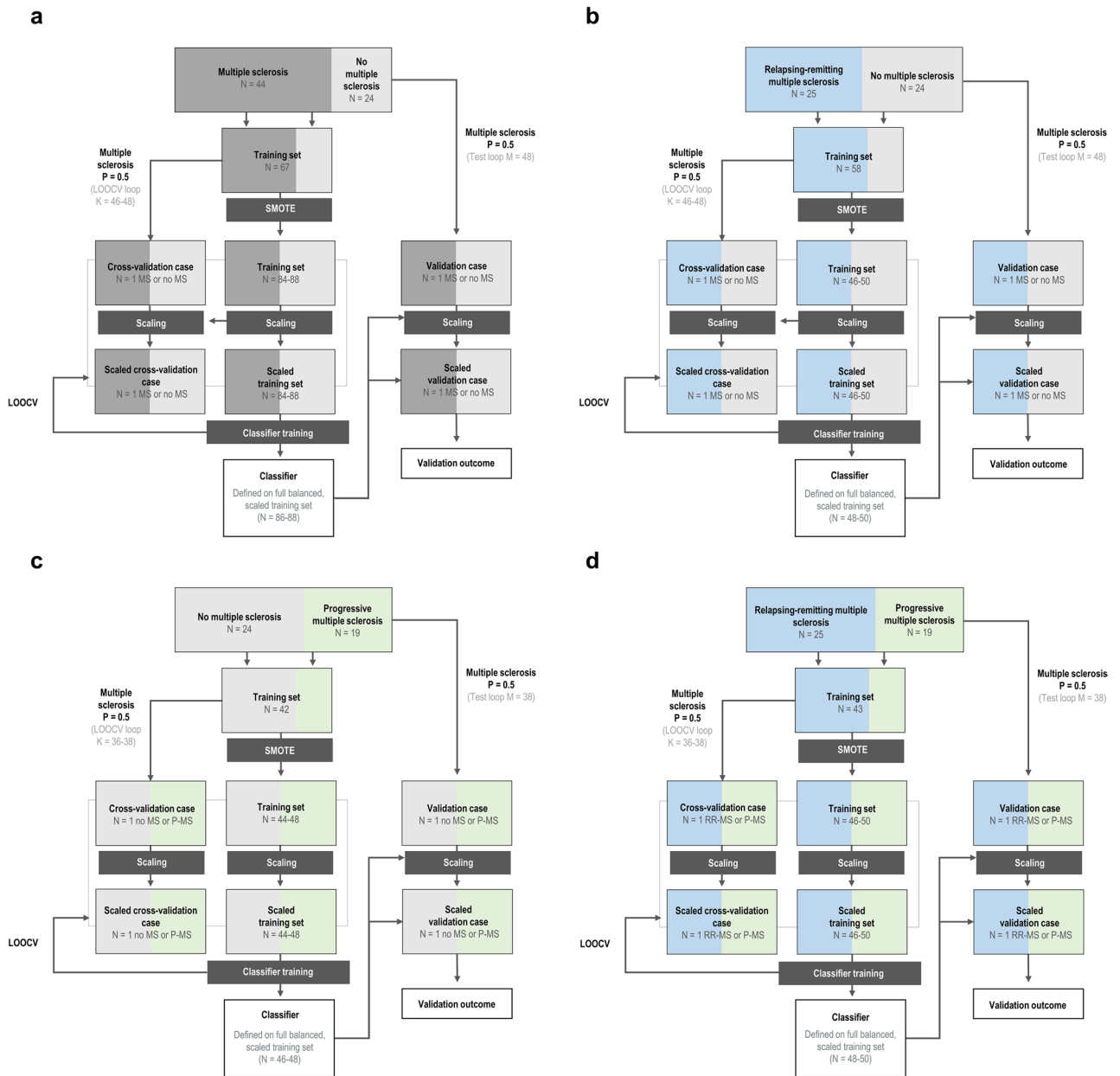


Figure 2. Training, optimization, and validation pipeline for multivariate classifiers of multiple sclerosis state and type by frontal cortex metabolite profiles. Classification pipelines constituted two binary decisions: The first (A) established the presence of multiple sclerosis, while the second (D) characterized the phenotype. Models distinguishing from control relapsing–remitting multiple sclerosis (B) and progressive multiple sclerosis (C) were also implemented for context in interpreting performance differences between (A) and (D). One data set at a time was held out for use as the validation set, yielding a total of $N_1 \times 2$ runs for data with smaller group size N_1 . Undersampling proceeded similarly for an inner cross-validation loop used for hyperparameter optimization and feature selection, in that each of the training cases was removed once for use as a cross-validation set for that training run, with a total of $N_2 \times 2$ cross-validations for data with the smaller training group size N_2 to ensure class balancing in the population of cross-validation sets used for each training run. Class sizes were then balanced using the Synthetic Minority Over-Sampling Technique (SMOTE). Each training set was scaled using a matrix calculated over that training set only, after which the same scaling was transferred to the held-out validation case. Dimensionality reduction was not performed in order to avoid degrading model interpretability. Classification algorithms tested included support vector machines, K-nearest neighbors, quadratic discriminant analysis, random forests, gradient-boosted decision trees, extremely randomized trees, and logistic regression; only the first three were selected for reporting based on performance separating progressive and relapsing–remitting multiple sclerosis. LOOCV leave-one-out cross-validation.

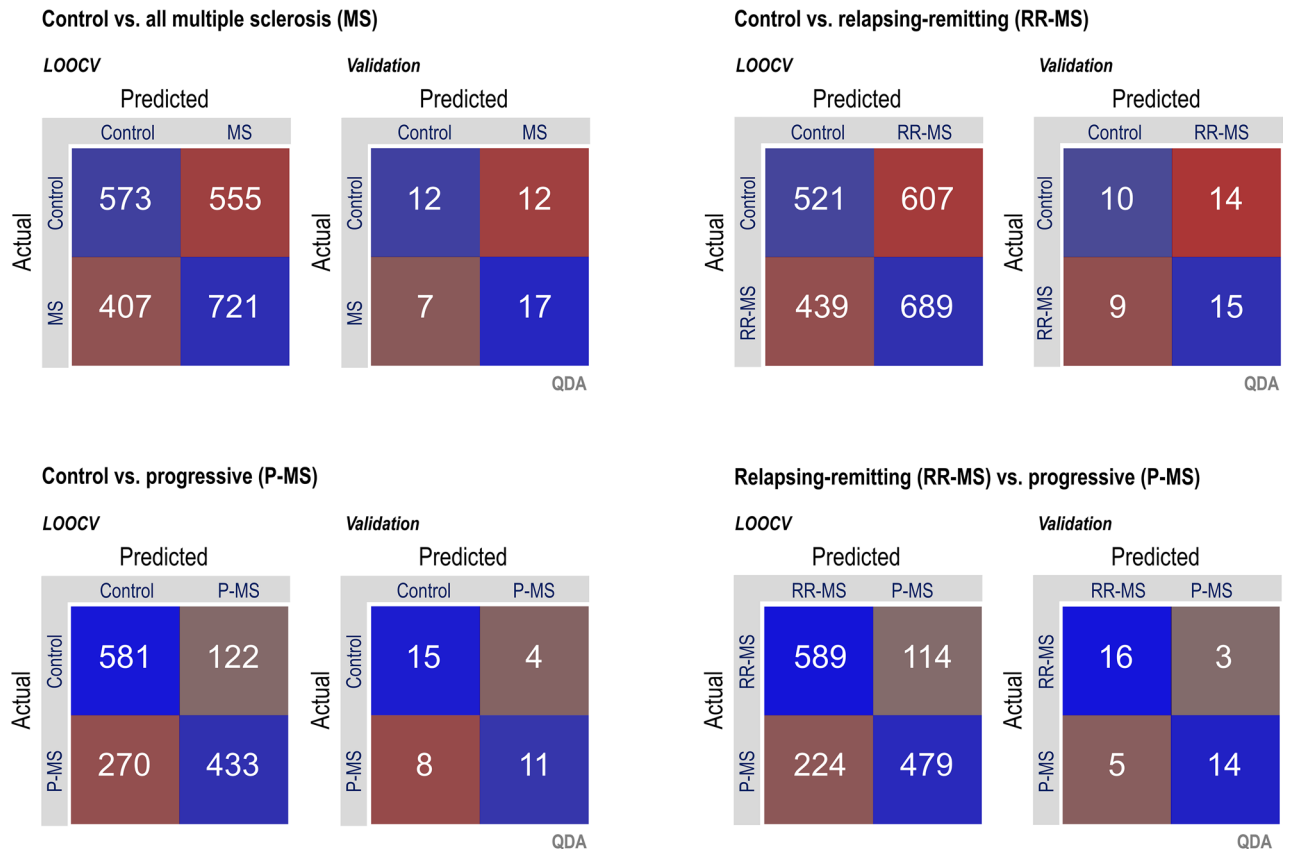


Figure 3. Confusion matrices for top-performing pipeline in each binary classification. Quadratic discriminant analysis (QDA) proved to be the top-performing classification algorithm by area under the receiver operating characteristic for all four questions, with higher validation sensitivity and specificity for classification of progressive patients than relapsing–remitting or all multiple sclerosis patients taken together, despite a smaller group size (progressive $N = 19$ versus relapsing–remitting $N = 25$ or control $N = 24$). LOOCV leave-one-out cross-validation.

as an auxiliary diagnostic tool for progressive multiple sclerosis. Additionally, it highlights the as-yet untapped potential of multivariate approaches to investigating the in vivo metabolic signatures of multiple sclerosis and its various phenotypes, especially under experimental conditions that enable the separation of visually overlapping but physiologically distinct spectral signals like glutamate and glutamine.

Proton magnetic resonance spectroscopy is an experimental technique of low sensitivity that has historically led to findings of metabolic abnormality in multiple sclerosis that are subtle and inconsistent at best⁸. Notably, application of data derived from this method alone led to satisfactory differentiation between progressive multiple sclerosis and control, and between progressive and relapsing–remitting multiple sclerosis, in a cohort that exhibited unequivocally significant univariate differences between the former two groups for only two metabolites (GABA and glutamate), and between the latter for only one (GABA)⁵⁸. Our classification pipelines distinguished patients from control with a clarity exceeding the subtle pattern of results yielded by univariate statistical techniques, demonstrating the additional information offered by multivariate approaches to multiple-sclerosis-related abnormalities in small molecule concentrations measurable by proton magnetic resonance spectroscopy. The additional diagnostic utility of simultaneously considering multiple metabolic signals is perhaps especially true when assessing data from tissue voxels that appear grossly normal on T_1 -weighted MR scans without contrast, and which would not be expected to demonstrate the more substantial abnormalities, including but not limited to decreased N-acetyl aspartate, increased choline, and differences in macromolecule and lipid contribution to the spectral baseline, previously reported for tissue that is already visibly lesioned⁸.

Composite imaging and CSF-based multiple sclerosis diagnostic strategies have reported low average sensitivities and specificities across the literature¹⁴. At least one investigation using the 2001 McDonald criteria⁶¹ including MR imaging of a cohort with suspected multiple sclerosis has reported higher than 80% diagnostic sensitivity and specificity upon follow-up after several years⁶², but the majority of imaging-based diagnostics tested have not exceeded these thresholds⁶³. Our finding, in our top-performing hyperparameter-optimized classification scheme, of 84% sensitivity and 74% specificity of differentiating progressive from relapsing–remitting multiple sclerosis on the basis of select metabolite concentrations in normal-appearing prefrontal cortex alone also does not exceed but approaches this standard. Importantly, it does so despite very small training samples ($N < 25$ per group without synthetic oversampling) using minimally transformed, biologically interpretable features in a pipeline controlled for overfitting.

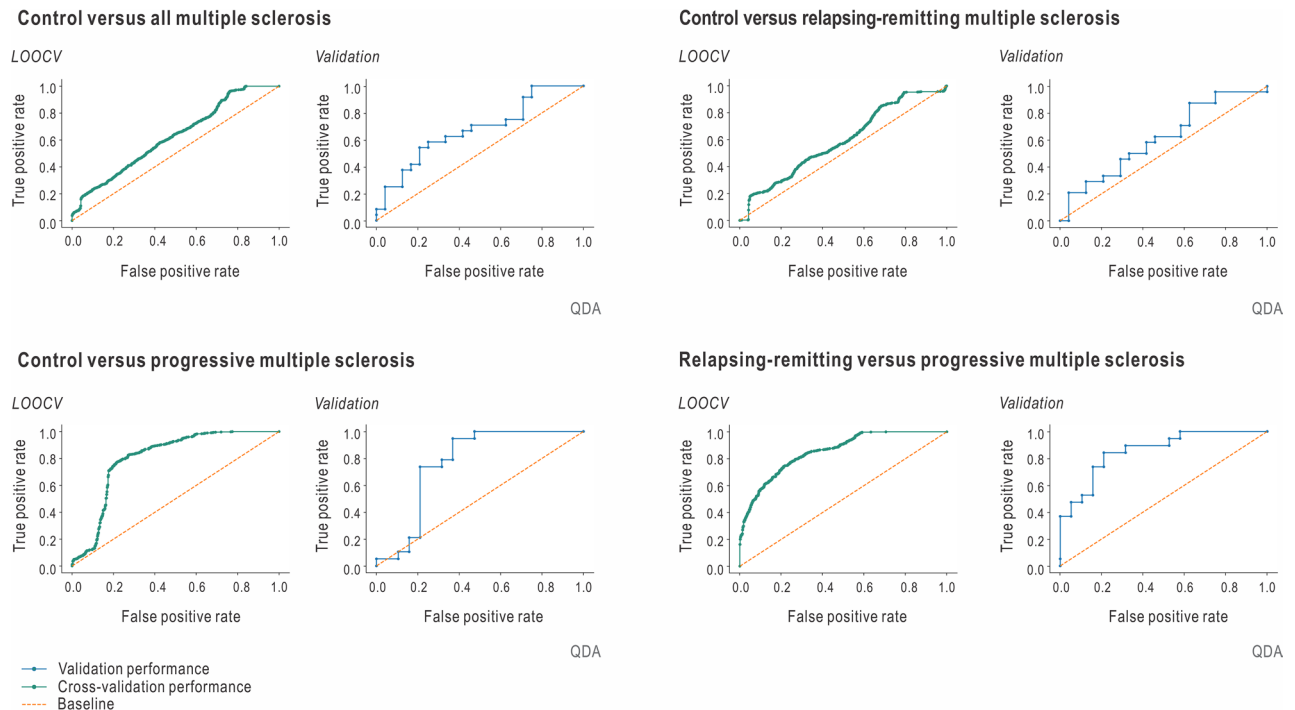


Figure 4. Receiver operating characteristic (ROC) curves for classification cross-validation and validation performance by question. Quadratic discriminant analysis (QDA) proved to be the top-performing classification algorithm by area under the receiver operating characteristic for all four questions. While it is visually clear that metabolite-based classifications between controls and multiple sclerosis patients, or between controls and relapsing–remitting multiple sclerosis patients, exhibited performance only marginally better than chance, ROC curves exhibited substantially increased convexity for questions involving the classification of progressive multiple sclerosis patients specifically, also reflected in higher sensitivity and specificities as shown in Fig. 3. Multiple sclerosis groups defined as positives in all plots; for the relapsing–remitting versus progressive question a positive is defined as relapsing–remitting multiple sclerosis. LOOCV: leave-one-out cross-validation.

That reported accuracy decreases with sample size, suggestive of the dangerous potential for overfitting in models produced by iterative learning methods, has been previously shown in a literature review of machine-learning application to multiple neurological disorders⁶⁴. The same review reported that only one-fifth of the literature examined employed held-out validation sets on which classification parameters were not optimized. The present study does not feature among them per se, as we did optimize a limited range of model hyperparameters and perform feature selection based on cross-validation accuracies pooled over the entire data set, though all training was conducted on cases kept entirely separate from the held-out validations. Overfitting, however, was not a concern, given the low number of variables optimized (only one for the consistently top-performing model QDA) and features (de)selected. In addition, all three of the final models tuned on the relapsing–remitting versus progressive multiple sclerosis classification data set were subsequently applied on a similar but distinct 7-Tesla ¹H-MRS data set to separate individuals with either posttraumatic stress disorder or major depression from healthy controls, yielding comparable (>70% sensitivity and specificity) held-out validation performance using only tailored feature selection without any further hyperparameter tuning^{65,66}. While a detailed presentation of these analyses lies beyond the scope of this manuscript and can be found in the referenced sources, comparable model performance on this completely held-out independent test set further supports the notion that the present findings are not simple artifacts of overfitting.

Classification accuracies for distinguishing either relapsing–remitting or general multiple sclerosis from control on the basis of the metabolites measured in this study were only a few percentage points above chance (50%), while at least one classifier of progressive versus relapsing–remitting multiple sclerosis nearly achieved the 80% threshold of sensitivity and specificity previously recommended for a diagnostic biomarker¹⁵. In addition to demonstrating that the pipeline employed here does not inevitably lead to overfitting for data sets that lack clear signals, this pattern of results underscores a phenotype-specific heterogeneity in multiple sclerosis as viewed through the lens of the metabolites measured here. This suggestion dovetails with previous reports of discrepant single-metabolite normal-appearing brain tissue abnormalities in relapsing–remitting and progressive multiple sclerosis patients^{67–69} to further emphasize the necessity for targeted study of multiple sclerosis progression as opposed to, or in parallel with, relapse, rather than multiple sclerosis as a reified monolith. The same argument may in principle be extended to our own conflation of secondary and primary progression into a single category of progressive multiple sclerosis, which is an admitted limitation of the present study borne of limited sample size.

Our results also suggest consistently divergent contributions by certain metabolites to the separation of relapsing–remitting or progressive multiple sclerosis subtypes from control versus from each other. It was demonstrated

Model	Control versus all multiple sclerosis				Control versus relapsing				Control versus progressive				Relapsing versus progressive							
	Acc.	Sn/Sp	R/P	F ₁	AUC	Acc.	Sn/Sp	R/P	F ₁	AUC	Acc.	Sn/Sp	R/P	F ₁	AUC	Acc.	Sn/Sp	R/P	F ₁	AUC
QDA																				
CV	57±5	64/51	64/57	60	0.62	54±5	61/46	61/53	57	0.59	72±4	62/83	62/78	69	0.80	76±4	84/68	84/72	78	0.85
Validation	60	71/50	71/59	64	0.68	52	63/42	52/63	57	0.61	68	58/79	58/73	65	0.83	79	84/74	84/76	80	0.86
KNN																				
CV	58±6	59/57	59/58	58	0.58	57±3	55/60	55/58	56	0.61	72±3	89/54	89/66	76	0.72	72±4	63/81	63/77	69	0.77
Validation	63	67/58	67/62	64	0.60	58	54/63	54/59	56	0.60	74	95/53	95/67	78	0.72	74	68/78	68/76	72	0.78
SVM^a																				
CV	58±4	57/58	57/58	57	0.36	54±2	49/58	49/54	51	0.49	71±4	76/66	76/69	72	0.74	68±3	68/69	68/68	68	0.70
Validation	58	58/58	58/58	58	0.41	52	46/58	46/52	49	0.52	74	79/68	79/71	75	0.77	68	68/68	68/68	68	0.68

Table 2. Binary classifier performance by feature-selected algorithm and participant groups. All statistics reported as means ± standard deviations. Models with held-out validation set AUC > 0.75 are in bold. ^aNo features eliminated for relapsing versus progressive. CV cross-validation, Sn sensitivity, Sp specificity, R recall, P precision, Acc. Accuracy, AUC area under receiver operating characteristic (ROC) or AUROC, QDA quadratic discriminant analysis, KNN K-nearest neighbors, SVM support vector machines.

Model	Control versus all multiple sclerosis		Control versus relapsing		Control versus progressive		Relapsing versus progressive	
	Δ Value	Metabolites (PI mean \pm S.D.)	Δ Value	Metabolites (PI mean \pm S.D.)	Δ Value	Metabolites (PI mean \pm S.D.)	Δ Value	Metabolites (PI mean \pm S.D.)
QDA								
Sn/Sp	+6/+8	mIns (0.13 \pm 0.08)	+7/-9	mIns (0.17 \pm 0.04)	+15/+10	tCho (0.21 \pm 0.06)	+7/+12	tCho (0.24 \pm 0.03)
Acc.	+7%	GABA (0.10 \pm 0.06)	-1% ^a	Glu (0.15 \pm 0.03)	+13%	Glu (0.20 \pm 0.06)	+9%	Glu (0.23 \pm 0.04)
AUC	+0.09	Glu (0.09 \pm 0.08)	+0.06	tNAA (0.15 \pm 0.03)	+0.16	GABA (0.18 \pm 0.07)	+0.14	GABA (0.22 \pm 0.03)
		tCho (0.04 \pm 0.06)		GABA (0.13 \pm 0.03)		tNAA (0.16 \pm 0.07)		GSH (0.13 \pm 0.03)
						Gln (0.15 \pm 0.05)		tNAA (0.08 \pm 0.03)
KNN								
Sn/Sp	+4/+16	mIns (0.29 \pm 0.05)	-12/+18	mIns (0.18 \pm 0.03)	+3/+2	Glu (0.12 \pm 0.06)	+2/+2	mIns (0.12 \pm 0.05)
Acc.	+10%	Glu (0.22 \pm 0.04)	+3%	Glu (0.08 \pm 0.03)	+2%	mIns (0.10 \pm 0.06)	+2%	tCho (0.09 \pm 0.04)
AUC	+0.09	Gln (0.16 \pm 0.05)	+0.12	tNAA (0.06 \pm 0.03)	-0.01	Gln (0.08 \pm 0.06)	+0.02	GSH (0.06 \pm 0.03)
				GABA (0.01 \pm 0.02)				tNAA (0.06 \pm 0.03)
								GABA (0.05 \pm 0.04)
								Glu (0.03 \pm 0.03)
SVM								
Sn/Sp	+7/0	Glu (0.02 \pm 0.02)	+14/-3	Glu (0.05 \pm 0.02)	+1/0	Glu (0.09 \pm 0.06)	--	tCho (0.10 \pm 0.03)
Acc.	+4%	mIns (0.02 \pm 0.03)	+5%	mIns (0.05 \pm 0.03)	+0.5%	mIns (0.07 \pm 0.06)		GSH (0.10 \pm 0.02)
AUC	-0.05	GABA (0.007 \pm 0.007)	-0.005	Gln (0.005 \pm 0.007)	+0.03	GABA (0.05 \pm 0.05)		Gln (0.10 \pm 0.03)
		GSH (0.006 \pm 0.006)				tCho (0.04 \pm 0.05)		mIns (0.06 \pm 0.03)
		tCho (<0)				Gln (0.03 \pm 0.06)		GABA (0.05 \pm 0.02)
								Glu (0.04 \pm 0.02)
								tNAA (0.03 \pm 0.02)

Table 3. Effect of metabolite selection on binary classifier leave-one-out cross-validation (LOOCV) performance. Grey indicates no feature selection implemented due to resultant reductions in model cross validation accuracy. ^aFeature elimination performed in this algorithm despite no substantial improvement in CV accuracy due to appreciable increase in AUROC. *PI* permutation importance, *S.D.* standard deviation, *Sn* sensitivity, *Sp* specificity, *Acc.* Accuracy, *AUC* area under receiver operating characteristic (ROC) or AUROC, *QDA* quadratic discriminant analysis, *KNN* K-nearest neighbors, *SVM* support vector machines, *mIns* myoinositol, *Glu* glutamate, *tCho* total choline, *Gln* glutamine, *tNAA* total *N*-acetyl aspartate, *GSH* glutathione.

in particular that myoinositol and glutamate were consistently ranked among the most important metabolites for distinguishing from control relapsing–remitting and progressive multiple sclerosis both separately and as a single category of disease. By contrast, total choline, glutathione, and glutamine ranked among the top three features for at least two of the three classifiers separating progressive from relapsing–remitting multiple sclerosis, but never (total choline or glutathione) or only once (glutamine) in any of the three models separating any multiple sclerosis patient from control. It is crucial not to over-interpret feature importances within any single model, especially those deriving from models with low baseline accuracies against which this parameter is calculated. These patterns were consistently observed, however, across multiple model questions and types, including those with accuracies exceeding 70%.

The validity of reporting separate concentrations of glutamate and glutamine derived from ¹H MRS experiments at field strengths lower than 7 Tesla is controversial, given previously demonstrated spectral overlap at these fields even under the clean spectral conditions offered by simulations⁷⁰. Our results suggest that at least when considered in concert with other compounds measurable by ¹H MRS, these two metabolites may weight differently in characterizing the commonalities versus the differences between the in vivo cortical metabolic signatures of relapsing–remitting and progressive multiple sclerosis. These findings therefore re-emphasize the crucial need for spectroscopy research on the role of glutamate in multiple sclerosis pathophysiology to be conducted under conditions that enable its unequivocal separation from glutamine, especially considering a literature that continues to demonstrate a relative absence of such investigations⁸.

It has been argued that both relapsing and progressive multiple sclerosis phenotypes reflect serial, parallel, or otherwise interrelated mechanisms in both inflammation and neurodegeneration⁷¹ and that diffusely amplified manifestation of the latter may be especially relevant for the progressive state, regardless of whether primary or secondary⁷². One possible implication of our finding that prefrontal 7 Tesla ¹H MRS-visible metabolite concentrations were sufficient to identify progressive but not unspecified or relapsing–remitting multiple sclerosis relative to control may therefore be that diffuse inflammatory and neurodegenerative changes were more visible to our methods than focal inflammatory ones in the normal-appearing cortical voxels under study. To the extent that such tissue changes may not track perfectly with strictly “relapsing–remitting” or “progressive” labels, multivariate assessment of ¹H MRS-visible metabolite signatures may thus yield even higher accuracies with respect to finer subgroups based on more detailed views of clinical presentation, e.g. involving additional qualifiers regarding activity and progression as per the 2013 revisions to multiple sclerosis clinical course definitions⁷³. Moreover, such ¹H-MRS biomarkers might reflect arguably more physiologically and clinically meaningful subgroups of shared disease mechanism, prognosis, and appropriate treatment thereof (e.g., progressive multiple sclerosis patients

who show some inflammatory disease activity and expected response to certain disease-modifying therapies versus those without overt relapse who may not) than the categories employed here.

One factor that additionally bears mention for future extensions of these ^1H MRS-based methods toward any claim of direct diagnostic utility is the consideration of positive controls representing a clinically realistic range of differential diagnoses, which the present study, employing only healthy participants as a negative control and patients with another phenotype of multiple sclerosis as a limited positive control, of course lacks. Low specificity to multiple sclerosis against other neurodegenerative conditions is one limitation of, for example, serum or plasma neurofilament light chain (NfL) as a diagnostic as opposed to prognostic or treatment effect biomarker⁷⁴, even as parallel developments in machine learning and networking technologies are otherwise enhancing the feasibility of automated blood-based testing in routine point-of-care use⁷⁵.

On the sole basis of in vivo proton magnetic resonance spectroscopic metabolite concentrations derived from normal-appearing frontal cortex voxels at 7 Tesla, progressive multiple sclerosis could be distinguished from relapsing–remitting multiple sclerosis or control with near to or greater than 70% sensitivity and specificity. By contrast, relapsing–remitting multiple sclerosis, or both multiple sclerosis subtypes pooled as a single classification, could not be accurately distinguished from control using the same approach. While further steps toward clinical translation will demand larger-scale replication of these findings involving more clinically diverse control groups, our results demonstrate the potential of multivariate statistical classifiers as applied to proton magnetic resonance spectroscopy as an auxiliary diagnostic tool for progressive multiple sclerosis as well as of multivariate approaches to researching in vivo metabolic signatures of diverse multiple sclerosis phenotypes, especially under ultra-high-field conditions enabling the separation of key spectral signals like glutamate and glutamine.

Methods

Participants and acquisition. Metabolite spectra were obtained as previously reported⁵⁸ from a single 27-cm³ cubic voxel in the prefrontal cortex using a 7 T head-only scanner (Varian Medical Systems, Inc., Palo Alto, CA, USA) at the Yale University Magnetic Resonance Research Center (MRRC) with actively shielded gradients and zero-through third-order shims. Spin handling and signal reception were achieved via a custom-built eight-channel transceiving radiofrequency head coil as previously described in detail for a similar protocol⁵⁷. Briefly, short echo time STimulated Echo Acquisition Mode (STEAM; echo time T_E 10 ms, repetition time T_R 3 s, mixing time T_M 50 ms, inversion time T_I 300 ms) and semi-Localization by Adiabatic SElective Refocusing (sLASER; T_E 72 ms, T_R 3 s) with J -difference editing (JDE) for GSH (on 4.56 ppm to isolate 2.95-ppm $^7\text{CH}_2$) and GABA (on 1.89 ppm to isolate 3.01-ppm $^4\text{CH}_2$) sequences were custom written for VnmrJ software. Water suppression by the CHEMical Shift Selective (CHESS) module in all sequences, outer-volume suppression by cuboid saturation bands in STEAM, and inversion recovery for macromolecule suppression in STEAM and JDE-GABA sLASER were additionally employed. B_0 shim coefficients through third order were calculated on gradient-echo images (T_E 3.8, 4.0, 4.3, 5.3, 6.8 ms) in B0DETOX⁷⁶, and B_1 phase shimming was achieved by in-house software IMAGO, written in MATLAB (v. 2013b, MathWorks, Natick, MA, USA).

STEAM and at least one J -difference acquisition were completed for 68 of 72 original study participants⁵⁸, leaving groups for progressive multiple sclerosis ($N = 19$; 6 primary, 13 secondary; 11 women, 8 men; $55 \pm \text{S.D. } 8.3$ years old; time since first MS diagnosis 20 ± 13 years), relapsing–remitting multiple sclerosis ($N = 25$; 17 women, 8 men; 45 ± 13 years old; time since first MS diagnosis 9 ± 7 years) and healthy controls with no multiple sclerosis ($N = 24$; 15 women, 9 men; 43 ± 15 years old). All participants provided informed consent to study enrollment prior to scanning, and all experimental procedures were conducted in accordance with the Declaration of Helsinki and as approved by Yale School of Medicine Human Investigation Committee protocol #1107008743 and Columbia University Institutional Review Board protocol AAAQ9641.

Spectral processing and quantification. Spectral processing was performed in ^1H MRS analysis free-ware INSPECTOR^{77,78} as described in detail previously^{58,79}. Briefly, signals from individual receive channels were corrected for eddy currents using water-unsuppressed references⁸⁰, phase- and frequency-aligned, and averaged with weighting by receive channel sensitivities⁸¹. Summed metabolite spectra were zero-order phased but not truncated, zero-filled, or line-broadened before direct quantification or alignment between summed editing conditions for difference spectrum calculation, as applicable.

Spectral quantification was achieved by linear combination modeling with basis spectra density-matrix simulated in SpinWizard⁸² using previously published chemical shift and J -coupling values^{83,84}. Basis sets included ascorbate, aspartate, choline, creatine, GABA, glycerophosphocholine, glutathione, glutamate, glutamine, glycine, myoinositol, N -acetyl aspartate, N -acetyl aspartylglutamate, phosphocholine, phosphocreatine, phosphorylethanolamine, scylloinositol, and taurine. GSH J -difference spectral basis sets included glutathione and N -acetyl aspartate, while GABA J -difference spectra employed GABA, total N -acetyl aspartate, glutamate, and glutamine (Fig. 1). Spectral quantification was performed in LCModel⁸⁵, and precision was assessed for each basis function fit using Cramér-Rao lower bounds⁸⁶. Absolute quantification estimates of total N -acetyl aspartate (N -acetyl aspartate plus N -acetyl aspartylglutamate), total choline (choline, phosphocholine, and glycerophosphocholine), glutamate, glutamine, myoinositol, glutathione, and GABA were achieved by normalizing by a 10 mM voxel concentration of total creatine (creatine and phosphocreatine). Total N -acetyl aspartate, total choline, myoinositol, and glutamate concentrations were corrected using significant linear regression betas on subject age, calculated from the control cohort as detailed previously⁵⁸. These seven metabolite concentrations were used as input features for each classification pipeline. All participants in this analysis had full STEAM and GSH data sets; two subjects (one control and one progressive MS patient) were missing GABA concentrations that were interpolated as the average of their respective classes prior to analysis.

Pipeline implementation. All classification pipelines were developed in Python 3.7 and implemented as per functions provided in Scikit-learn⁸⁷.

Classification performance assessment. Cross-validation accuracies were averaged over all runs for each classification pipeline. Additionally, receiver operating characteristics were approximated for each algorithm by plotting sensitivities and specificities for each of $N(N - 1)$ cross-validation runs. Sensitivity (true positive classifications over all real positives) and specificity (true negative classifications over all real negatives) were calculated over all classifications. Model error was additionally reported as precision, recall, and the composite thereof F_1 , or $2(\text{precision} \times \text{recall})/(\text{precision} + \text{recall})$, where precision is the proportion of positive classifications that were true and recall is the proportion of positive cases that were identified. In each challenge versus control, multiple sclerosis was defined as the positive classification; in the relapsing versus progressive question, relapsing multiple sclerosis was defined as positive for the calculation of the aforementioned statistics.

Metabolite feature selection. Feature selection was performed on hyperparameter-optimized models and consisted of recursive feature elimination according to permutation importance. Permutation importance is proportional to the reduction in model accuracy when the association between feature values and classifications is broken by randomization⁸⁸. Starting from hyperparameter-optimized models retaining all seven metabolites, the feature with the lowest permutation importance was removed and LOOCV performance recalculated over the whole data set. Feature elimination was continued as long as the resultant model exhibited superior performance (higher average LOOCV accuracy and/or area under the receiver operating characteristic) relative to the original.

Group statistics. All group statistics are reported as means \pm standard deviations unless otherwise noted. Model accuracies were compared across classification questions using a Student's *t*-test⁸⁹ scripted in MATLAB (v. 2018a, MathWorks, Natick, MA, USA) with two-tailed significance testing and α at 0.05.

Data availability

The extracted feature sets analyzed during the current study are available from the corresponding author on reasonable request.

Received: 3 October 2021; Accepted: 29 July 2022

Published online: 16 August 2022

References

- Thompson, A. J., Baranzini, S. E., Geurts, J., Hemmer, B. & Ciccarelli, O. Multiple sclerosis. *Lancet* **391**, 1622–1636 (2018).
- Thompson, A. J. *et al.* Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* **17**, 162–173 (2018).
- Gobbin, F. *et al.* Sensitivity and specificity of 2017 McDonald criteria for multiple sclerosis in patients with clinically isolated syndrome. *Mult. Scler. J.* **24**, 531–532 (2018).
- Sand, I. K., Krieger, S., Farrell, C. & Miller, A. E. Diagnostic uncertainty during the transition to secondary progressive multiple sclerosis. *Mult. Scler. J.* **20**, 1654–1657 (2014).
- Santos, A. C. *et al.* Magnetization transfer can predict clinical evolution in patients with multiple sclerosis. *J. Neurol.* **249**, 662–668 (2002).
- Khaleeli, Z. *et al.* Magnetization transfer ratio in gray matter a potential surrogate marker for progression in early primary progressive multiple sclerosis. *Arch Neurol-Chicago* **65**, 1454–1459 (2008).
- Fox, R. J. Picturing multiple sclerosis: Conventional and diffusion tensor imaging. *Semin. Neurol.* **28**, 453–466 (2008).
- Swanberg, K. M., Landheer, K., Pitt, D. & Juchem, C. Quantifying the metabolic signature of multiple sclerosis by in vivo proton magnetic resonance spectroscopy: Current challenges and future outlook in the translation from proton signal to diagnostic biomarker. *Front. Neurol.* **10**, 1173. <https://doi.org/10.3389/fneur.2019.01173> (2019).
- Fox, R. J., Beall, E., Bhattacharyya, P., Chen, J. T. & Sakaie, K. Advanced MRI in multiple sclerosis: Current status and future challenges. *Neurol. Clin.* **29**, 357 (2011).
- Kutzelnigg, A. *et al.* Cortical demyelination and diffuse white matter injury in multiple sclerosis. *Brain* **128**, 2705–2712 (2005).
- Calabrese, M., Filippi, M. & Gallo, P. Cortical lesions in multiple sclerosis. *Nat. Rev. Neurol.* **6**, 438–444 (2010).
- Newcombe, J. *et al.* Histopathology of multiple-sclerosis lesions detected by magnetic-resonance-imaging in unfixed postmortem central-nervous-system tissue. *Brain* **114**, 1013–1023 (1991).
- Bo, L., Vedeler, C. A., Nyland, H. L., Trapp, B. D. & Mork, S. J. Subpial demyelination in the cerebral cortex of multiple sclerosis patients. *J. Neuropathol. Exp. Neurol.* **62**, 723–732 (2003).
- Schaffler, N. *et al.* Accuracy of diagnostic tests in multiple sclerosis—A systematic review. *Acta Neurol. Scand.* **124**, 151–164 (2011).
- The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group. Consensus report of the working group on molecular and biochemical markers of Alzheimer's Disease. *Neurobiol. Aging* **19**, 285–285 (1998).
- Inglese, M. *et al.* Diffusely elevated cerebral choline and creatine in relapsing-remitting multiple sclerosis. *Magnet. Reson. Med.* **50**, 190–195 (2003).
- Bang, S. *et al.* Establishment and evaluation of prediction model for multiple disease classification based on gut microbial data. *Sci. Rep. UK* **9**, 1–9 (2019).
- Goyal, M. *et al.* Computational intelligence technique for prediction of multiple sclerosis based on serum cytokines. *Front. Neurol.* **10**, 781 (2019).
- Kocevar, G. *et al.* Graph theory-based brain connectivity for automatic classification of multiple sclerosis clinical courses. *Front. Neurosci.* **10**, 478. <https://doi.org/10.3389/fnins.2016.00478> (2016).
- Torabi, A., Daliri, M. R. & Sabzposhan, S. H. Diagnosis of multiple sclerosis from EEG signals using nonlinear methods. *Aust. Phys. Eng. Sci. Med.* **40**, 785–797. <https://doi.org/10.1007/s13246-017-0584-9> (2017).
- Reitz, S. C. *et al.* Multi-parametric quantitative MRI of normal appearing white matter in multiple sclerosis, and the effect of disease activity on T2. *Brain Imaging Behav.* **11**, 744–753. <https://doi.org/10.1007/s11682-016-9550-5> (2017).
- Kontschieder, P. *et al.* Quantifying progression of multiple sclerosis via classification of depth videos. *Med. Image Comput. Comput. Assist. Interv.* **17**, 429–437. https://doi.org/10.1007/978-3-319-10470-6_54 (2014).

23. Weygandt, M. *et al.* MRI pattern recognition in multiple sclerosis normal-appearing brain areas. *PLoS ONE* **6**, e21138. <https://doi.org/10.1371/journal.pone.0021138> (2011).
24. Zurita, M. *et al.* Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data. *Neuroimage Clin.* **20**, 724–730. <https://doi.org/10.1016/j.nicl.2018.09.002> (2018).
25. Sacca, V. *et al.* Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state fMRI connectivity data. *Brain Imaging Behav.* **13**, 1103–1114. <https://doi.org/10.1007/s11682-018-9926-9> (2019).
26. PerezDelPalomar, A. *et al.* Swept source optical coherence tomography to early detect multiple sclerosis disease: The use of machine learning techniques. *PLoS ONE* **14**, e0216410. <https://doi.org/10.1371/journal.pone.0216410> (2019).
27. Fiorini, S. *et al.* A machine learning pipeline for multiple sclerosis course detection from clinical scales and patient reported outcomes. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **4443–4446**, 2015. <https://doi.org/10.1109/EMBC.2015.7319381> (2015).
28. Taschler, B. *et al.* Spatial modeling of multiple sclerosis for disease subtype prediction. *Med. Image Comput. Comput. Assist. Interv.* **17**, 797–804. https://doi.org/10.1007/978-3-319-10470-6_99 (2014).
29. Wotischel, V. *et al.* Predicting outcome in clinically isolated syndrome using machine learning. *Neuroimage-Clin.* **7**, 281–287 (2015).
30. Guo, P., Zhang, Q., Zhu, Z., Huang, Z. & Li, K. Mining gene expression data of multiple sclerosis. *PLoS ONE* **9**, e100052. <https://doi.org/10.1371/journal.pone.0100052> (2014).
31. Eshaghi, A. *et al.* Gray matter MRI differentiates neuromyelitis optica from multiple sclerosis using random forest. *Neurology* **87**, 2463–2470 (2016).
32. Richiardi, J. *et al.* Classifying minimally disabled multiple sclerosis patients from resting state functional connectivity. *Neuroimage* **62**, 2021–2033. <https://doi.org/10.1016/j.neuroimage.2012.05.078> (2012).
33. Yoo, Y. *et al.* Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls. *Neuroimage Clin.* **17**, 169–178. <https://doi.org/10.1016/j.nicl.2017.10.015> (2018).
34. Andersen, S. L. *et al.* Metabolome-based signature of disease pathology in MS. *Mult. Scler. Relat. Disord.* **31**, 12–21. <https://doi.org/10.1016/j.msard.2019.03.006> (2019).
35. Eitel, F. *et al.* Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. *Neuroimage Clin.* **24**, 102003. <https://doi.org/10.1016/j.nicl.2019.102003> (2019).
36. Lotsch, J. *et al.* Machine-learned data structures of lipid marker serum concentrations in multiple sclerosis patients differ from those in healthy subjects. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms18061217> (2017).
37. Alaqtash, M. *et al.* Automatic classification of pathological gait patterns using ground reaction forces and machine learning algorithms. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **453–457**, 2011. <https://doi.org/10.1109/IEMBS.2011.6090063> (2011).
38. Zhang, J., Tong, L., Wang, L. & Li, N. Texture analysis of multiple sclerosis: A comparative study. *Magn. Reson. Imaging* **26**, 1160–1166. <https://doi.org/10.1016/j.mri.2008.01.016> (2008).
39. Ahmadi, A., Davoudi, S. & Daliri, M. R. Computer aided diagnosis system for multiple sclerosis disease based on phase to amplitude coupling in covert visual attention. *Comput. Methods Prog. Biomed.* **169**, 9–18 (2019).
40. Ohanian, D. *et al.* Identifying key symptoms differentiating myalgic encephalomyelitis and chronic fatigue syndrome from multiple sclerosis. *Neurology (E-Cronicon)* **4**, 41–45 (2016).
41. Ostmeier, J. *et al.* Statistical classifiers for diagnosing disease from immune repertoires: A case study using multiple sclerosis. *BMC Bioinform.* **18**, 401. <https://doi.org/10.1186/s12859-017-1814-6> (2017).
42. Corvol, J. C. *et al.* Abrogation of T cell quiescence characterizes patients at high risk for multiple sclerosis after the initial neurological event. *Proc. Natl. Acad. Sci. USA* **105**, 11839–11844 (2008).
43. Flauzino, T. *et al.* Disability in multiple sclerosis is associated with age and inflammatory, metabolic and oxidative/nitrosative stress biomarkers: Results of multivariate and machine learning procedures. *Metab. Brain Dis.* **34**, 1401–1413 (2019).
44. Jackson, K. C. *et al.* Genetic model of MS severity predicts future accumulation of disability. *Ann. Hum. Genet.* **84**, 1–10 (2019).
45. Mesaros, S. *et al.* Diffusion tensor MRI tractography and cognitive impairment in multiple sclerosis. *Neurology* **78**, 969–975 (2012).
46. Zhong, J. D. *et al.* Combined structural and functional patterns discriminating upper limb motor disability in multiple sclerosis using multivariate approaches. *Brain Imaging Behav.* **11**, 754–768 (2017).
47. Crimi, A. *et al.* Predictive value of imaging markers at multiple sclerosis disease onset based on gadolinium- and USPIO-enhanced MRI and machine learning. *PLoS ONE* **9**, e93024 (2014).
48. Bendfeldt, K. *et al.* Multivariate pattern classification of gray matter pathology in multiple sclerosis. *Neuroimage* **60**, 400–408 (2012).
49. Kiiski, H. *et al.* Machine learning EEG to predict cognitive functioning and processing speed over a 2-year period in multiple sclerosis patients and controls. *Brain Topogr.* **31**, 346–363. <https://doi.org/10.1007/s10548-018-0620-4> (2018).
50. Tacchella, A. *et al.* Collaboration between a human group and artificial intelligence can improve prediction of multiple sclerosis course: A proof-of-principle study. *F1000Res* **6**, 2172. <https://doi.org/10.12688/f1000research.13114.2> (2017).
51. Brosch, T., Yoo, Y., Li, D. K., Traboulsee, A. & Tam, R. Modeling the variability in brain morphology and lesion distribution in multiple sclerosis by deep learning. *Med. Image Comput. Comput. Assist. Interv.* **17**, 462–469. https://doi.org/10.1007/978-3-319-10470-6_58 (2014).
52. Fagone, P. *et al.* Identification of CD4(+) T cell biomarkers for predicting the response of patients with relapsing-remitting multiple sclerosis to natalizumab treatment. *Mol. Med. Rep.* **20**, 678–684 (2019).
53. Lyu, T., Lock, E. F. & Eberly, L. E. Discriminating sample groups with multi-way data. *Biostatistics* **18**, 434–450 (2017).
54. Ghalwash, M. F., Ramljak, D. & Obradovic, Z. Patient-specific early classification of multivariate observations. *Int. J. Data Min. Bioinform.* **11**, 392–411 (2015).
55. Baranzini, S. E. *et al.* Prognostic biomarkers of IFN β therapy in multiple sclerosis patients. *Mult. Scler. J.* **21**, 894–904 (2015).
56. Llado, X. *et al.* Segmentation of multiple sclerosis lesions in brain MRI: A review of automated approaches. *Inform. Sci.* **186**, 164–185 (2012).
57. Prinsen, H., de Graaf, R. A., Mason, G. F., Pelletier, D. & Juchem, C. Reproducibility measurement of glutathione, GABA, and glutamate: Towards in vivo neurochemical profiling of multiple sclerosis with MR spectroscopy at 7T. *J. Magn. Reson. Imaging* **45**, 187–198. <https://doi.org/10.1002/jmri.25356> (2017).
58. Swanberg, K. M. *et al.* In vivo evidence of differential frontal cortex metabolic abnormalities in progressive and relapsing-remitting multiple sclerosis. *NMR Biomed.* **11**, e4590 (2021).
59. Kurada, A. V., Swanberg, K. M., Prinsen, H. & Juchem, C. Diagnosis of multiple sclerosis subtype through machine learning analysis of frontal cortex metabolite profiles. *Proc. Int. Soc. Magn. Reson. Med.* **2019**, 4871 (2019).
60. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357. <https://doi.org/10.1613/jair.953> (2002).
61. McDonald, W. I. *et al.* Recommended diagnostic criteria for multiple sclerosis: Guidelines from the International Panel on the Diagnosis of Multiple Sclerosis. *Ann. Neurol.* **50**, 121–127 (2001).
62. Dalton, C. M. *et al.* Application of the new McDonald criteria to patients with clinically isolated syndromes suggestive of multiple sclerosis. *Ann. Neurol.* **52**, 47–53 (2002).
63. Whiting, P. *et al.* Accuracy of magnetic resonance imaging for the diagnosis of multiple sclerosis: Systematic review. *BMJ-Br. Med. J.* **332**, 875–878 (2006).

64. Sakai, K. & Yamada, K. Machine learning studies on major brain diseases: 5-year trends of 2014–2018. *Jpn. J. Radiol.* **37**, 34–72 (2019).
65. Swanberg, K. M. *et al.* Abnormal glutamate metabolism in prefrontal cortex of post-traumatic stress disorder linked to comorbidity with major depression. *Proc. Int. Soc. Magn. Reson. Med.* **2022**, 3344 (2022).
66. Swanberg, K. M. *Optimization of Sensitivity to Disease-Associated Cortical Metabolic Abnormality by Evidence-Based Quantification of In Vivo Proton Magnetic Resonance Spectroscopy Data from 3 Tesla and 7 Tesla*. Ph.D. Thesis, Columbia University. <https://doi.org/10.7916/2nv4-q759> (2022).
67. Bagory, M. *et al.* Implementation of an absolute brain 1H-MRS quantification method to assess different tissue alterations in multiple sclerosis. *IEEE Trans. Biomed. Eng.* **59**, 2687–2694. <https://doi.org/10.1109/TBME.2011.2161609> (2012).
68. Pelletier, D. *et al.* 3-D echo planar (1)HMRS imaging in MS: Metabolite comparison from supratentorial vs central brain. *Magn. Reson. Imaging* **20**, 599–606. [https://doi.org/10.1016/s0730-725x\(02\)00533-7](https://doi.org/10.1016/s0730-725x(02)00533-7) (2002).
69. Sarchielli, P. *et al.* Localized (1)H magnetic resonance spectroscopy in mainly cortical gray matter of patients with multiple sclerosis. *J. Neurol.* **249**, 902–910. <https://doi.org/10.1007/s00415-002-0758-5> (2002).
70. Tkac, I., Oz, G., Adriany, G., Ugurbil, K. & Gruetter, R. In vivo 1H NMR spectroscopy of the human brain at high magnetic fields: Metabolite quantification at 4T vs. 7T. *Magn. Reson. Med.* **62**, 868–879. <https://doi.org/10.1002/mrm.22086> (2009).
71. Lassmann, H. What drives disease in multiple sclerosis: Inflammation or neurodegeneration? *Clin. Exp. Neuroimmunol.* **1**, 2–11 (2010).
72. Lassmann, H. Pathogenic mechanisms associated with different clinical courses of multiple sclerosis. *Front. Immunol.* **9**, 3116. <https://doi.org/10.3389/fimmu.2018.03116> (2018).
73. Lublin, F. D. *et al.* Defining the clinical course of multiple sclerosis: The 2013 revisions. *Neurology* **83**, 278–286. <https://doi.org/10.1212/WNL.0000000000000560> (2014).
74. Ferreira-Atuesta, C., Reyes, S., Giovanonni, G. & Gnanapavan, S. The evolution of neurofilament light chain in multiple sclerosis. *Front. Neurosci.* **15**, 642384. <https://doi.org/10.3389/fnins.2021.642384> (2021).
75. Gasparin, A. T. *et al.* Hilab system, a new point-of-care hematology analyzer supported by the internet of things and artificial intelligence. *Sci. Rep.* **12**, 10409. <https://doi.org/10.1038/s41598-022-13913-8> (2022).
76. Juchem, C., Umesh Rudrapatna, S., Nixon, T. W. & de Graaf, R. A. Dynamic multi-coil technique (DYNAMITE) shimming for echo-planar imaging of the human brain at 7 Tesla. *Neuroimage* **105**, 462–472. <https://doi.org/10.1016/j.neuroimage.2014.11.011> (2015).
77. Gajdošík, M., Landheer, K., Swanberg, K. M. & Juchem, C. INSPECTOR: Free software for magnetic resonance spectroscopy data inspection, processing, simulation and analysis. *Sci. Rep.* **11**, 2094. <https://doi.org/10.1038/s41598-021-81193-9> (2021).
78. Columbia Technology Ventures. *INSPECTOR*. http://innovation.columbia.edu/technologies/cu17130_inspector-magnetic-resonance-spectroscopy-software-for-optimized-data-extraction/licensing. (2019).
79. Swanberg, K. M., Prinsen, H., Coman, D., de Graaf, R. A. & Juchem, C. Quantification of glutathione transverse relaxation time T2 using echo time extension with variable refocusing selectivity and symmetry in the human brain at 7 Tesla. *J. Magn. Reson.* **290**, 1–11. <https://doi.org/10.1016/j.jmr.2018.02.017> (2018).
80. Klose, U. In vivo proton spectroscopy in presence of eddy currents. *Magn. Reson. Med.* **14**, 26–30. <https://doi.org/10.1002/mrm.1910140104> (1990).
81. Wright, S. M. & Wald, L. L. Theory and application of array coils in MR spectroscopy. *NMR Biomed.* **10**, 394–410. [https://doi.org/10.1002/\(sici\)1099-1492\(199712\)10:8%3c394::aid-nbm494%3e3.0.co;2-0](https://doi.org/10.1002/(sici)1099-1492(199712)10:8%3c394::aid-nbm494%3e3.0.co;2-0) (1997).
82. de Graaf, R. A., Chowdhury, G. M. & Behar, K. L. Quantification of high-resolution (1)H NMR spectra from rat brain extracts. *Anal. Chem.* **83**, 216–224 (2011).
83. Govindaraju, V., Young, K. & Maudsley, A. A. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed.* **13**, 129–153 (2000).
84. Govind, V., Young, K. & Maudsley, A. A. Corrigendum: Proton NMR chemical shifts and coupling constants for brain metabolites. Govindaraju V, Young K, Maudsley AA, *NMR Biomed.* 2000; 13: 129–153. *NMR Biomed.* **28**(923–924), 2015. <https://doi.org/10.1002/nbm.3336> (2015).
85. Provencher, S. W. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magn. Reson. Med.* **30**, 672–679. <https://doi.org/10.1002/mrm.1910300604> (1993).
86. Cavassila, S., Deval, S., Huegen, C., Van Ormondt, D. & Graveron-Demilly, D. Cramer-Rao bound expressions for parametric estimation of overlapping peaks: Influence of prior knowledge. *J. Magn. Reson.* **143**, 311–320 (2000).
87. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
88. Scikit-learn. *4.2. Permutation Feature Importance*. https://scikit-learn.org/stable/modules/permutation_importance.html.
89. Glantz, S. A. *Primer of Biostatistics* 7th edn. (McGraw-Hill Medical, 2012).

Acknowledgements

Special thanks are due to Ms. Yvette Strong of the Yale Center for Clinical Investigation (YCCI) as well as the physician and nursing staff at the Yale-New Haven Hospital Interventional Immunology Center for assiduous participant recruitment efforts. Acknowledgement of software support in spectral analysis is due to Robin A. de Graaf, Ph.D. at Yale University, and experimental planning for data set collection due to Christina Azevedo, M.D., and Daniel Pelletier, M.D., at Yale University and the University of Southern California. This research has received support from the National Multiple Sclerosis Society (NMSS) grant “In Vivo Metabolomics of Oxidative Stress with 7 Tesla Magnetic Resonance Spectroscopy” (RG 5319); it has been further made possible by CTSA Grant Number UL1 TR000142 from the National Center for Advancing Translational Science (NCATS), components of the National Institutes of Health (NIH), and NIH roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of the NIH.

Author contributions

K.M.S. and A.V.K. contributed equally to the manuscript. K.M.S. engaged in experimental data acquisition and processing, statistical analysis, and manuscript writing; A.V.K. engaged in statistical analysis and manuscript editing; H.P. engaged in experimental data acquisition and processing and manuscript editing; C.J. engaged in experimental data acquisition and manuscript editing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022