Received: 10 February 2022 | Revised: 4 March 2022 | Accepted: 4 March 2022

DOI: 10.1111/jon.12991

JOURNAL OF NEUROIMAGING
THE OFFICIAL JOURNAL OF THE AMERICAN SOCIETY OF NEUROIMAGING
WILEY

CLINICAL INVESTIGATIVE STUDY

Classification of multiple sclerosis patients based on structural disconnection: A robust feature selection approach

Simona Schiavi[1,2] | Alberto Azzari[2] | Antonella Mensi[2] | Nicole Graziano[3] | Alessandro Daducci[2] | Manuele Bicego[2] | Matilde Inglese[1,3,4,5] | Maria Petracca[3,6]

[1]Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOGMI), University of Genoa, Genoa, Italy

[2]Department of Computer Science, University of Verona, Verona, Italy

[3]Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, New York, USA

[4]Department of Radiology, Icahn School of Medicine at Mount Sinai, New York, New York, USA

[5]IRCCS Ospedale Policlinico San Martino, Genoa, Italy

[6]Department of Human Neurosciences, Sapienza University of Rome, Rome, Italy

Correspondence
Prof. Matilde Inglese, Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOGMI), University of Genoa, Largo Paolo Daneo 3, 16132 Genoa, Italy.
Email: m.inglese@unige.it

Funding information
This study was supported in part by grants from NMSS (RG 5120A3/1) and Teva Neuroscience (CNS-2014-221).

Abstract

Background and Purpose: Although structural disconnection represents the hallmark of multiple sclerosis (MS) pathophysiology, classification attempts based on structural connectivity have achieved low accuracy levels. Here, we set out to fill this gap, exploring the performance of supervised classifiers on features derived from microstructure informed tractography and selected applying a novel robust approach.

Methods: Using microstructure informed tractography with diffusion MRI data, we created quantitative connectomes of 55 MS patients and 24 healthy controls. We then used a robust approach—based on two classical methods of feature selection— to select relevant features from three network representations (whole connectivity matrices, node strength, and local efficiency). Classification accuracy of the selected features was tested with five different classifiers, while their meaningfulness was tested via correlation with clinical scales. As a comparison, the same classifiers were run on features selected with the standard procedure in network analysis (thresholding).

Results: Our procedure identified 11 features for the whole net, five for local efficiency, and seven for node strength. For all classifiers, the accuracy was in the range 64.5%-91.1%, with features extracted from the whole net reaching the maximum, and overcoming results obtained with the standard procedure in all cases. Correlations with clinical scales were identified across functional domains, from motor and cognitive abilities to fatigue and depression.

Conclusion: Applying a robust feature selection procedure to quantitative structural connectomes, we were able to classify MS patients with excellent accuracy, while providing information on the white matter connections and gray matter regions more affected by MS pathology.

KEYWORDS
classification, machine learning, microstructure informed tractography, multiple sclerosis, quantitative structural connectivity, robust feature selection

INTRODUCTION

Multiple sclerosis (MS) is a chronic disease of the central nervous system characterized by inflammation and neurodegeneration.[1–3] Both disease diagnosis and in vivo investigation of disease pathophysiology rely on the application of standard and advanced MRI techniques.[4] Over the last years, following the developments witnessed in the artificial intelligence field, classification of MS patients via imaging-based machine learning approaches has been pursued in order to aid MS diagnosis and/or offer new insights in MS pathophysiology.[5–8] Primary information source for individual classification in MS appears to be white matter (WM) lesions.[9] Indeed, WM lesions represent the imaging hallmark of MS and the basis of the so-called disconnection syndrome, caused by trans-axonal degeneration of fibers connecting secluded gray matter (GM) regions.[10] Being a pathophysiologically relevant and consistent MS feature,[11,12] brain structural connectivity seems an ideal candidate for classification tasks.

So far, however, the only study that has attempted a structural connectivity-based classification of MS patients has reached a classification accuracy lower than 80%.[8] This might be due to several reasons. Among these, the intrinsic limitation of fiber (streamlines) tracking in presence of WM lesions and microstructural damage,[13] and the fact that the selection of relevant features to be fed to the classification model is based on empirically set thresholds.[14,15]

Here, in order to overcome these limitations, we based our classification on structural metrics derived from quantitative connectomes,[16] whose ability to characterize brain connectivity beyond impairment caused by focal lesions has been already exploited in MS.[17] To further improve the classification accuracy, we applied a robust feature selection (FS) procedure before testing different machine learning classifiers. Finally, to confirm the relevance of the selected features, not only in terms of discrimination between MS patients and healthy controls but also in terms of clinical meaningfulness, we explored their relationships with patients' clinical status.

METHODS

Study population

Fifty-five MS patients[1] (36 female [F], age 50.45 ± 11.28 years, disease duration 15.5 ± 11.6 years, 22 primary progressive, 20 secondary progressive, and 13 relapsing-remitting, median Expanded Disability Status Scale [EDSS] 4 range: 0-6.5) and 24 healthy controls (HC) (11 F, age 50.3 ± 8.5 years) were prospectively enrolled. Clinical examination, performed within 1 week from the MRI scan, included scales assessing neurological dysfunction (EDSS), motor performance (manual dexterity-9 Hole Peg Test and ambulation-25 Foot Walk Test), cognitive performance (attention-Symbol Digit Modalities Test, visuospatial memory-Brief Visuospatial Memory Test, verbal memory-California Verbal Learning Test, verbal fluency-Controlled Oral Word Association Test, executive functions-card sorting test), fatigue (Modified Fatigue Impact Scale), and depression (Beck Depression Inventory). Table 1 contains a summary of clinical and demographical data. Written informed consent was obtained from all participants before the beginning of the study procedures, according to the Declaration of Helsinki. The protocol was approved by the Institutional Review Board of the Icahn School of Medicine at Mount Sinai.

TABLE 1 Demographic and clinical data

	Healthy controls	MS patients
Age (years)	50.3 ± 8.5	50.45 ± 11.28
Sex	11 female, 13 male	36 female, 19 male
Phenotype	–	22 primary progressive 20 secondary progressive 13 relapsing-remitting
Disease duration (years)	–	15.5 ± 11.6
EDSS median (range)	–	4 (0-6.5)
9HPT (seconds)	–	28.13 ± 7.84
25FWT (seconds)	–	8.44 ± 10.22
SDMT	–	48.6 ± 12.21
BVMT	–	17.96 ± 8.87
CVLT	–	52.65 ± 13.47
COWAT	–	38.45 ± 13.40
CARD SORTING	–	9.46 ± 4.74
MFIS	–	37.11 ± 20.68
BDI	–	9.16 ± 7.17

Note: Summary of demographic and clinical data for the subjects involved in this study. All measures are reported as mean ± standard deviation unless otherwise indicated.
Abbreviations: 25FWT, 25-foot walk test; 9HPT, 9-hole peg test; BDI, Beck depression inventory; BVMT, brief visuospatial memory test; COWAT, controlled oral word association test; CVLT, California verbal learning test; EDSS, Expanded Disability Status Scale; MFIS, modified fatigue impact scale; MS, multiple sclerosis; SDMT, symbol digit modalities test; std, standard deviation.

MRI acquisition

All subjects underwent MRI on a Siemens Skyra 3T scanner (Siemens, Erlangen, Germany) with a 32-channels head coil. The MRI protocol included the following sequences: axial T2-weighted 3-dimensional (repetition time [TR]: 8000 ms, echo time [TE]: 95 ms, spatial resolution 0.5 × 0.5 × 3.0 mm^3); sagittal T1-weighted 3-dimensional magnetization-prepared rapid gradient echo (TR/TE: 3000/2.47 ms, inversion time: 1000 ms, spatial resolution 0.8 × 0.8 × 0.8 mm3; generalized autocalibrating partially parallel acquisitions with acceleration factor $R = 2$); twice-refocused spin echo echo-planar imaging sequence for diffusion MRI with b values of 1000 and 2000 seconds/mm2 and 30

TABLE 2 Features selected for each representation (whole network, local efficiency, and node strength)

Whole net		
Features	**IFS**	**LASSO**
L.superiorparietal → R.supramarginal	1.000	1.000
R.thalamus → R.paracentral	0.987	0.934
L.middletemporal → R.superiorparietal	1.000	0.914
L.supramarginal → R.postcentral	0.987	0.909
L.paracentral → R.caudalmiddlefrontal	1.000	0.757
R.caudate → R.lateraloccipital	0.975	0.701
L.lateraloccipital → L.thalamus	1.000	0.678
R.bankssts → R.precuneus	0.937	0.653
L.superiorparietal → L.caudate	1.000	0.605
R.thalamus → R.precuneus	0.987	0.590
L.rostralmiddlefrontal → R.superiorfrontal	1.000	0.524
Local efficiency		
Features	**IFS**	**LASSO**
R.thalamus	1.000	1.000
R.caudate	1.000	1.000
R.cuneus	1.000	0.623
R.pericalcarine	0.924	1.000
R.superiorfrontal	0.557	1.000
Node strength		
Features	**IFS**	**LASSO**
L.parsopercularis	1.000	1.000
L.superiorparietal	1.000	0.661
L.thalamus	1.000	0.532
R.thalamus	1.000	1.000
R.caudate	1.000	0.716
R.putamen	1.000	1.000
R.superiorfrontal	0.532	0.997

Note: For each feature, we report how many times it has been selected among the different leave one out runs and repetitions for the two different feature-selection schemes: infinite feature selection (IFS) and least absolute shrinkage and selection operator (LASSO). In the features name, L indicates left and R right.

directions each (repeated for both phase acquisitions), in addition to six $b = 0$ images (TR/TE: 4700/100 ms, flip angle 80°, spatial resolution $1.8 \times 1.8 \times 2\ mm^3$).

MRI processing

All images were processed as described in Schiavi et al[17] according to the pipeline illustrated in Figure 1. Briefly, after segmenting T2-hyperintense and T1-hypointense lesions,[18] we used the relative masks to fill the T1-weighted images with FMRIB software library (FSL) (https://fsl.fmrib.ox.ac.uk) and processed them with FreeSurfer version 6.0 (http://surfer.nmr.mgh.harvard.edu), to obtain an cortical parcellation comprising 85 regions of interest.[19] The obtained parcellations were visually inspected by a neurologist with more than 5 years of experience in neuroimaging, but no additional manual editing was performed on them. After state-of-the-art processing steps on diffusion MR images to build the whole brain probabilistic tractogram,[20] we ran the convex optimization modeling for microstructure informed tractography (COMMIT)[16,21] to assign the intra-axonal signal fraction to each entry of the connectivity matrix and derive a more quantitative structural connectome. As done in previous works investigating differences in structural connectivity between HC and MS patients,[17,22,23] for each node of the whole brain connectomes, we extracted the node's strength (NS; which is the sum of weights of links connected to the node) and local efficiency (LE; which is the average inverse shortest path length in the neighborhood of the node, and is related to the clustering coefficient) using Brain Connectivity Toolbox.[24] We then ran the FS approaches described below on the entire set of connections (Whole Net [WN] ($85 \times 85 - 85$)/2 features), and on the set of NS and LE (85 features each) separately.

Feature selection

The goal of the FS step was to identify the set of most important features within each network representation (WN, NS, LE) for the discrimination between healthy subjects and MS patients. To achieve this, we devised a robust procedure, based on a leave one out (LOO) cross-validation scheme[25] already applied in other scenarios when estimating feature stability.[26] Briefly, given a population of N subjects, we applied N times a FS method, each time on a set in which a single element has been removed (ie, in the first round we applied the FS method to all subjects except the first, in the second round to all subjects except the second, and so on). Good features are those that are selected as important for discrimination in most of the rounds (ie, features which are extracted even if we have a small perturbation of the dataset). Given the nonidentical sets of features extracted in the different perturbations of the original dataset, only features that were selected in more than 50% of the rounds were considered as discriminative within our robust framework. To further increase the robustness of the results, we applied two different FS schemes:

1. Least absolute shrinkage and selection operator (LASSO): an effective and widely applied method for FS,[27] based on a penalized regression model. To estimate the parameter of the model in each LOO run and increase the robustness against random fluctuations, we repeated 10 times a 10-fold cross-validation, keeping the value that minimized the mean square error on the left-out fold (averaged over the 10 runs).

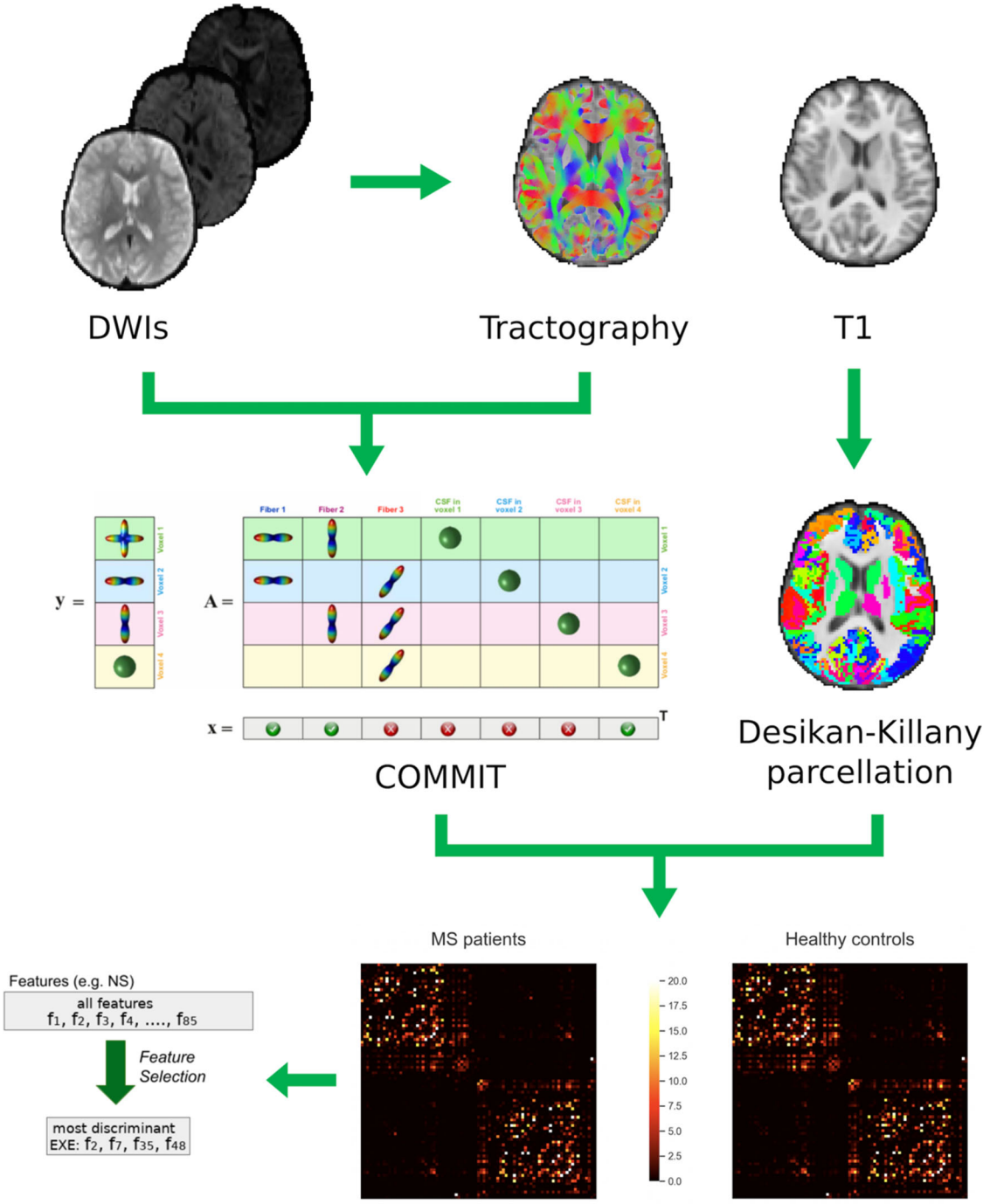

FIGURE 1 Main steps of processing pipeline. From diffusion-weighted imaging (DWI), we built the whole brain probabilistic tractography and then ran the convex optimization modeling for microstructure informed tractography (COMMIT). In parallel, we parcellated the T1 images with the Desikan–Killiany atlas. Combining the cortical parcellation with COMMIT weights, we built the quantitative connectomes used for the feature selection procedure. MS, multiple sclerosis; NS, node strength

2. Infinite feature selection (IFS): a very recent graph-based feature ranking approach[28] that exploits properties of power series matrices and Markov chains derived from a graph representation of the FS problem. This method has shown to be very competitive and effective in many problems: for example, IFS has been successfully employed for cancer classification[28,29] that, similarly to our case, typically involves few subjects and many features: diagnosis of amyotrophic lateral sclerosis[30] and exploring gender differences in cortical morphological networks,[31] just to cite a few.

Only features selected by both methods were considered as robust for our results.

TABLE 3 Classification accuracies before and after the application of the proposed feature selection procedure

Representation	Max	Average	NN	KNN	SVM-LIN	SVM-RBF	RF
WN	0.8101	0.7443	0.6456	0.6709	0.7975	0.8101	0.7975
WN + FS	0.9114	0.8658	0.8481	0.8354	0.8608	0.9114	0.8734
LE	0.8354	0.7949	0.7468	0.7848	0.8354	0.7848	0.8228
LE + FS	0.8481	0.8329	0.8228	0.8481	0.8228	0.8481	0.8228
NS	0.7975	0.7139	0.6582	0.7342	0.6962	0.6835	0.7975
NS + FS	0.8608	0.8278	0.8354	0.7722	0.8608	0.8481	0.8228

Note: For each representation: whole net (WN), local efficiency (LE), and node strength (NS), we report the accuracy results before and after the application of the feature selection (FS) procedure for each classifier (NN, KNN, SVM-LIN, SVM-RBF, and RF) as well as the maximum (Max) and the average scores among them (Max and Average).
Abbreviations: KNN, Kth nearest neighbor; NN, nearest neighbor; RF, random forest; SVM-LIN, Support Vector Machine linear kernel; SVM-RBF, Support Vector Machine radial basis function kernel.

TABLE 4 Classification accuracies obtained with the standard feature selection approach

Representation	Max	Average	NN	KNN	SVM-LIN	SVM-RBF	RF
WN + StandardFS	0.7722	0.7038	0.6582	0.7089	0.7089	0.6709	0.7722
StandardFS + LE	0.8354	0.7989	0.7884	0.7884	0.8101	0.7722	0.8354
StandardFS + NS	0.8228	0.7519	0.6582	0.7342	0.8228	0.7468	0.7975

Note: Classification accuracies of the different classifiers (NN, KNN, SVM-LIN, SVM-RBF, and RF) and maximum (Max) and average obtained on whole net (WN), local efficiency (LE), and node strength (NS) preprocessed with the standard thresholding procedure (StandardFS).
Abbreviations: KNN, Kth nearest neighbor; NN, nearest neighbor; RF, random forest; SVM-LIN, Support Vector Machine linear kernel; SVM-RBF, Support Vector Machine radial basis function kernel.

Classification

The classification analysis aimed at quantitatively measuring the discrimination capability of a given set of features. To compute the classification accuracy, we applied the LOO cross-validation protocol to five different classifiers, ranging from the simple nearest neighbor technique up to the more complex support vector machines and random forests (nearest neighbor and *K*th nearest neighbor with Euclidian distance, support vector machine with linear kernel, support vector machine with radial basis function kernel, and random forest with 100 trees). The parameters of the classifiers have been automatically set via cross-validation on the training set.

In order to have a comparison with standard analyses, we also computed the accuracies of the different classifiers when applying the "Standard FS" procedure, which represents standard preprocessing done in this field.[14] This scheme consists of postprocessing usually applied to the whole connectome (ie, on the WN representation), and it retains only those features (ie, connections) that are present in at least 50% of the subjects involved in the study.[14,15] Even if these results are not directly comparable with those obtained with the proposed approach (in the previous case, the LE/NS have been computed before applying the FS), with this comparison we can get an idea on how the proposed approach compares with established preprocessing pipelines (which implies filtering with Standard FS and computing quantities such LE/NS from the filtered net). To aid in this comparison, we also computed the No Information Rate (or Null Information Rate [NIR]). The NIR is the accuracy obtained with a "silly" classifier that assigns every object to the class that is most probable a priori, that is, to the class that is the most frequent in the training set. The NIR is typically used to have an idea of the minimum significant accuracy below which a classifier is considered useless.

Relationships between extracted features and clinical scores

To confirm the meaningfulness of the selected features, we tested their correlation with clinical scales. Bivariate correlations were run between clinical scores and selected features for each network representation (WN, NS, LE). Significance level was set at $p < .002$, that is, Bonferroni corrected for multiple comparisons (0.05/23 as the number of features tested).

RESULTS

Extracted features

We applied the proposed FS pipeline to the three representations (WN, NS, LE). Obtained features (ie, those features that were selected by both LASSO and IFS in at least 50% of the runs) are listed in Table 2 and graphically represented in Figure 2. For each representation and for each FS scheme, we reported the feature together with its selection percentage (ie, how many times such feature has been selected among

TABLE 5 Correlations with clinical scores

		Whole net											Node efficiency					Node strength						
		L.SPG R.SM	R.Th R.PC	L.MT R.SP	L.SM R.PoC	L.PC R.CMF	R.Ca R.LO	L.LO L.Th	R.BT R.PCu	L.SPG L.Ca	R.Th R.PCu	L.RMF R.SF	R_Th	R_Ca	R_Cu	R_PCAL	R_SFG	L_POP	L_SPG	L_Th	R_Th	R_Ca	R_Pu	R_SFG
EDSS	Spearman rho	−.524***	−.275	−.212	−.354*	.064	−.109	−.385*	−.112	−.226	−.346*	−.305*	−.012	−.150	.391*	.357*	.018	.175	.129	−.211	−.353*	−.198	−.405*	.039
	p	<.001	.051	.135	.011	.656	.447	.005	.434	.110	.013	.030	.935	.295	.005	.010	.898	.220	.367	.137	.010	.163	.003	.786
9HPT	Pearson *r*	−.369**	−.352*	−.129	−.190	−.021	−.009	−.303*	−.200	−.053	−.345*	−.316*	−.387**	−.329*	.172	.142	.007	−.056	.04	−.236	−.578***	−.403**	−.471***	.054
	p	.010	.014	.381	.195	.885	.952	.036	.172	.719	.016	.029	.007	.022	.242	.335	.964	.705	.789	.107	<.001	.005	.001	.717
25FWT	Pearson *r*	−.071	−.182	−.096	−.072	−.057	−.087	−.163	−.123	−.186	−.233	−.246	−.265	−.271	.014	.002	−.025	−.039	−.199	−.279	−.303*	−.244	−.290*	.218
	p	.627	.210	.512	.622	.698	.554	.263	.399	.202	.107	.088	.066	.06	.926	.99	.865	.792	.170	.052	.034	.091	.043	.133
SDMT	Pearson *r*	.213	.239	.141	.216	.159	.047	.425**	.215	.409**	.287*	.347*	.321*	.445***	−.259	−.226	.058	.237	.128	.380**	.515***	.460***	.363**	.263
	p	.133	.091	.325	.128	.266	.745	.002	.129	.003	.041	.013	.021	.001	.066	.111	.684	.094	.372	.006	<.001	.001	.009	.062
BVMT	Pearson *r*	.291*	.249	.224	.397**	.067	.144	.446§	.116	.380**	.208	.199	.305*	.415**	−.044	−.339*	.151	.232	−.047	.360**	.416**	.515***	.432***	.318*
	p	.036	.075	.111	.004	.637	.307	.001	.414	.006	.139	.157	.028	.002	.756	.014	.286	.098	.742	.009	.002	<.001	.001	.021
CVLT	Pearson *r*	.271	.278*	−.041	.209	−.003	.054	.350*	.204	.331*	.144	.09	.144	.171	−.167	−.090	.181	.104	−.055	.259	.280*	.304*	.226	0
	p	.052	.046	.771	.137	.984	.703	.011	.148	.017	.307	.527	.309	.225	.238	.525	.199	.464	.699	.064	.044	.029	.107	.999
COWAT	Pearson *r*	−.055	.422**	−.054	.053	.117	−.017	.224	−.030	.157	.106	.074	.494***	.436***	−.110	−.178	.242	.116	−.067	.224	.418**	.427***	.310*	.265
	p	.695	.002	.704	.704	.403	.902	.107	.829	.260	.452	.599	<.001	.001	.431	.202	.081	.408	.633	.106	.002	.001	.024	.055
CARD	Pearson *r*	.007	−.156	.354*	.033	.105	−.029	−.054	−.066	.544***	.016	.046	.207	.354*	−.076	−.076	.149	−.030	.145	.241	.229	.196	.157	.079
SORTING	*p*	.962	.271	.010	.815	.460	.837	.705	.641	<.001	.912	.748	.141	.010	.59	.593	.293	.831	.304	.085	.103	.164	.267	.579
MFIS	Pearson *r*	−.268	−.118	−.269	−.115	−.188	−.079	−.348*	−.150	−.164	−.038	.103	−.158	−.217	−.125	−.041	−.033	.071	−.033	−.314*	−.288*	−.359**	−.512***	.138
	p	.057	.410	.056	.420	.186	.581	.012	.293	.250	.790	.471	.268	.126	.383	.773	.817	.623	.816	.025	.040	.010	<.001	.336
BDI	Pearson *r*	−.260	−.036	−.285*	−.155	−.171	−.005	−.271	−.091	−.392**	−.194	−.063	−.059	−.275	.065	.016	.099	−.081	−.137	−.285*	−.235	−.272	−.402**	.113
	p	.066	.801	.042	.277	.23	.973	.054	.525	.004	.172	.658	.682	.051	.651	.909	.488	.571	.339	.043	.097	.054	.003	.429

*Correlation is significant at the 0.05 level (two-tailed).

**Correlation is significant at the 0.01 level (two-tailed).

***Correlations surviving Bonferroni correction ($p < .002$, two-tailed).

Abbreviation: 25FWT, 25-foot walk test; 9HPT, 9-hole peg test; BDI, Beck depression inventory; BT, bankssts; BVMT, brief visuospatial memory test; Ca, caudate; CMF, caudalmiddlefrontal; COWAT, controlled oral word association test; Cu, cuneus; CVLT, California verbal learning test; EDSS, Expanded Disability Status Scale; L, left; LO, lateraloccipital; MFIS, modified fatigue impact scale; MT, middletemporal; PC, paracentral; PCAL, pericalcarine cortex; PCu, precuneus; PoC, postcentral; POP, parsopercularis; Pu, putamen; R, right; RMF, rostralmiddlefrontal; SDMT, symbol digit modalities test; SF, superiorfrontal; SFG, superiorfrontal; SM, supramarginal; SPG, superiorparietal; Th, thalamus.

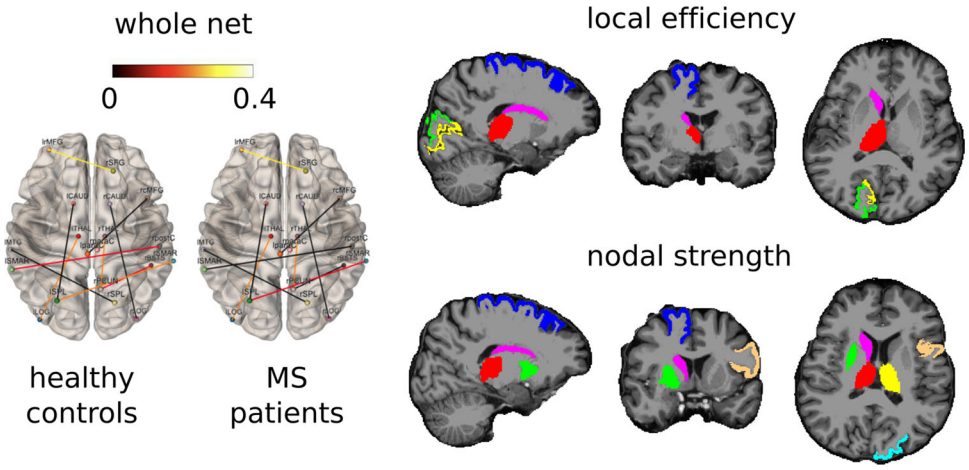

FIGURE 2 Selected features for whole net (left), local efficiency (top right), and node strength (bottom right). For whole net, we graphically show the selected connection and their mean strengths in healthy controls and multiple sclerosis (MS) patients. For local efficiency and node strength, we show the selected nodes color-coded as follows: right superior frontal gyrus in blue; right cuneus in green; right pericalcarine in yellow; right thalamus in red; right caudate in magenta; left superior parietal in orange; left parsopercularis in light blue; left thalamus in yellow

the different LOO runs and repetitions—eg, 1.00 indicates that the feature has always been selected).

Classification results

In Table 3, we report the accuracy results for each classifier on WN, NS, and LE before and after applying the proposed FS methodology. In the former case, each subject is represented with all features derived from the quantitative connectomes; in the latter, each subject is represented only with the features surviving FS. Average and maximum accuracy computed across all classifiers for a given representation are also shown.

In Table 4, we report the accuracies of the different classifiers applied to WN processed with Standard FS as well as NS/LE representations derived from the network preprocessed with Standard FS.

The NIR was 0.6962.

Clinical impact of selected features

Correlations between selected features and clinical scores are shown in Table 5.

DISCUSSION

In this study, we introduced a robust FS procedure that allowed us to improve the discrimination of MS patients from HC based on structural connectivity while identifying clinically meaningful features. We devised a robust procedure that combines the results of two classical methods of FS and tested their classification accuracy with five different classifiers. Finally, the different classifiers were also used to investigate the accuracy obtained with the standard thresholding procedure.

The investigation of structural connectivity through diffusion tractography and the extraction of meaningful properties to distinguish patients from HC is a well-established procedure.[20] However, this methodology selects features in the whole network via a thresholding procedure solely based on empirical assumptions, with local and global graph-based metrics being extracted from the thresholded network.[14,15]

On the contrary, our approach was based on the selection of connections and nodes' local properties that were consistently extracted by two different FS procedures (LASSO and IFS) in more than 50% of the runs. Indeed, in the majority of cases LASSO and IFS agreement in the selection of individual features reached 70%. This approach offers two main advantages: (i) it reduces the dimensionality of the space of features (in our case from 3570 connections—deriving from 85 × 85 cells of the connectivity matrix minus 85 diagonal entries divided by 2—to 11 connections and from 85 nodes' local properties to 5 or 7 [for node strength and local efficiency, respectively]); (ii) it offers insights on which WM tracts/GM nodes are more affected by the disease and, likely, represents the substrates of the observed clinical disability.

We achieved a higher accuracy when employing the proposed FS methodology than when applying the standard FS approach, thus confirming the importance of reducing the dimensionality to improve classification. Indeed, with our approach the mean accuracy was higher than 82% for all tested classifiers, which, compared to the NIR of 69.62%, confirms the potentiality of the proposed method. Here, a critical observation is that our analysis is limited due to the small sample size. Indeed, these accuracies might become even higher by employing a dataset with a larger number of both MS and HC. Nevertheless, the newly proposed procedure outperformed classification accuracies obtained with the standard FS procedure in all the investigated cases.

Perhaps our most relevant finding is the fact that the largest accuracy increments as well as the highest overall accuracy were obtained by using the FS procedure on the whole net. This suggests that the

procedure of FS proposed here might be used as an alternative of the classical network analysis procedure, which usually reduces the dimensionality of the space of features by deriving global metrics like efficiency and strength. Indeed, analyzing the outcomes of the standard FS procedure, an accuracy improvement over classification attempted with the entire set of features was obtained only for NS.

The major potential of our approach is that it can directly select connections more affected by the disease instead of combining and merging the information of all the connections ending in the same node. In particular, the best performance was obtained by selecting only 11 connections. This model reached a classification accuracy of 91%, a percentage considerably higher than the ones previously reported by classification works based on structural/functional connectivity analysis.[7,8]

Regarding the biological meaningfulness of the extracted features, those with higher selection percentage at both LASSO and IFS were more consistently associated with clinical scores, confirming that our approach was able to identify relevant indicators of the patients' clinical status. Among the selected features, NS of deep GM (thalamus, putamen, caudate) showed the strongest correlations with a variety of clinical disability measures (eg, manual dexterity, fatigue, cognition—measured in terms of processing speed, visual memory, and verbal fluency). This finding is in line with the known relevance of deep GM damage for clinical disability in MS,[32–34] and once again underscores how disconnection of this relay from cortical areas and wider brain circuits, here expressed by the reduction in NS, is a fundamental contributor to the clinical expression of disability.[11] Of note, while the highest discriminative accuracy was reached with the set of features derived from whole net connectivity, in agreement with the prominent role of WM already reported with deep brain algorithms,[9] NS constitutes the best correlate of clinical disability. This is probably because it represents a better synthetic measure of disconnection for salient regions than the damage of single WM bundles.

Among the extracted features, the connection between the left middle temporal gyrus and the right superior parietal gyrus, identified in 91.4% of the runs, is anatomically implausible and represents a false positive. The identification of false positive connections represents an unsolved issue in the tractography field, and it is known to exert a nonnegligible impact on connectivity analyses.[35] The selected false positive connection was consistently present in the HC, while it was never identified in the MS patients. This can be explained by the presence of lesions that change the directions that can be followed by the tracking in particular voxels by altering the aspect of the fiber orientation distribution functions. Although the presence of false positive is a limitation of our analyses, none of the techniques currently available in the field are spared from this issue.[20] Recently, the COMMIT2 framework has been proposed with impressive results in terms of false positive removal.[36] Once validated (not only in a few healthy subjects but also in pathological brains), COMMIT2 could be used instead of COMMIT to create quantitative structural connectomes and investigate the stability of both performances and features extracted. Beyond the false positive issues, the inclusion of a small number of relapsing patients might limit the generalizability of our findings.

Despite these limitations, applying a robust FS procedure to input data derived from quantitative tractography, we were able to classify MS patients with excellent accuracy, while providing information on the WM connections and GM regions more affected by MS pathology.

ACKNOWLEDGEMENTS AND DISCLOSURE

None.

ORCID

Simona Schiavi https://orcid.org/0000-0003-1641-186X

REFERENCES

1. Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. Ann Neurol 2011;69:292-302.
2. Bester M, Petracca M, Inglese M. Neuroimaging of multiple sclerosis, acute disseminated encephalomyelitis, and other demyelinating diseases. Semin Roentgenol 2014;49:76-85.
3. Schiavi S, Petracca M, Sun P, et al. Non-invasive quantification of inflammation, axonal and myelin injury in multiple sclerosis. Brain 2021;12:213-23.
4. Inglese M, Petracca M. MRI in multiple sclerosis: clinical and research update. Curr Opin Neurol 2018;31:249-55.
5. Rocca MA, Anzalone N, Storelli L, et al. Deep learning on conventional magnetic resonance imaging improves the diagnosis of multiple sclerosis Mimics. Invest Radiol 2021;56:252-60.
6. Yoo Y, Tang LYW, Brosch T, 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 2018;17:169-78. https://www.sciencedirect.com/user/identity/landing?code=_0_SLdAGuUNvsMyuPT-0iRwZZoUj5HaCK4EAJQby&state=retryCounter%3D0%26csrfToken%3D10e9714d-68a5-4cbc-b2b7-23fc2148d434%26idpPolicy%3Durn%253Acom%253Aelsevier%253Aidp%253Apolicy%253Aproduct%253Ainst_assoc%26returnUrl%3D%252Fscience%252Farticle%252Fpii%252FS2213158217302553%253Fvia%25253Dihub%26prompt%3Dnone%26cid%3Darp-f5099617-f854-4037-b93d-815a856ba14d
7. Saccà V, Sarica A, Novellino F, et al. Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state FMRI connectivity data. Brain Imaging Behav 2019;13:1103-14.
8. Zurita M, Montalba C, Labbé T, et al. Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data. Neuroimage Clin 2018;20:724-30.
9. Eitel F, Soehler E, Bellmann-Strobl J, et al. Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. Neuroimage Clin 2019;24:102003.
10. Ruggieri S, Petracca M, Miller A, et al. Association of deep gray matter damage with cortical and spinal cord degeneration in primary progressive multiple sclerosis. JAMA Neurol 2015;72:1466-74.
11. Chard DT, Alahmadi AAS, Audoin B, et al. Mind the gap: from neurons to networks to outcomes in multiple sclerosis. Nat Rev Neurol 2021;17:173-84.
12. Petracca M, Schiavi S, Battocchio M, et al. Streamline density and lesion volume reveal a postero–anterior gradient of corpus callosum damage in multiple sclerosis. Eur J Neurol 2020;27:1076-82.

13. Ozturk A, Smith S, Gordon-Lipkin E, et al. MRI of the corpus callosum in multiple sclerosis: association with disability. Mult Scler J 2010;16:166-77.
14. Buchanan CR, Bastin ME, Ritchie SJ, et al. The effect of network thresholding and weighting on structural brain networks in the UK Biobank. Neuroimage 2020;211:116443.
15. Fornito A, Zalesky A, Bullmore ET. Fundamentals of brain network analysis. London: Academic Press Elsevier Inc; 2016. https://www.elsevier.com/books/fundamentals-of-brain-network-analysis/fornito/978-0-12-407908-3
16. Daducci A, Dal Palu A, Lemkaddem A, et al. COMMIT: convex optimization modeling for microstructure informed tractography. IEEE Trans Med Imaging 2015;34:246-57. https://ieeexplore.ieee.org/document/6884830
17. Schiavi S, Petracca M, Battocchio M, et al. Sensory-motor network topology in multiple sclerosis: structural connectivity analysis accounting for intrinsic density discrepancy. Hum Brain Mapp 2020;41:2951-63.
18. Petracca M, Zaaraoui W, Cocozza S, et al. An MRI evaluation of grey matter damage in African Americans with MS. Mult Scler Relat Disord 2018;25:29-36.
19. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 2006;31:968-80.
20. Zhang F, Daducci A, He Y, et al. Quantitative mapping of the brain's structural connectivity using diffusion MRI tractography: a review. Neuroimage 2022;249:118870.
21. Daducci A, Dal Palu A, Lemkaddem A, et al. A convex optimization framework for global tractography (abstract). IEEE Biomed Imaging 2013:524-27. https://ieeexplore.ieee.org/document/6556527
22. Pagani E, Rocca MA, De Meo E, et al. Structural connectivity in multiple sclerosis and modeling of disconnection. Mult Scler J 2019;26:220-32.
23. Steenwijk MD, Daams M, Pouwels PJW, et al. Unraveling the relationship between regional gray matter atrophy and pathology in connected white matter tracts in long-standing multiple sclerosis. Hum Brain Mapp 2015;36:1796-807.
24. Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations. Neuroimage 2010;52:1059-69.
25. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. Vol 1. New York: Springer series in statistics; 2001. https://link.springer.com/book/10.1007/978-0-387-84858-7
26. Kuncheva LI A stability index for feature selection. AIAP'07 25th IASTED International Multi-Conference: artificial intelligence and applications (abstract), February 2007, p. 390-5. https://dl.acm.org/doi/10.5555/1295303.1295370
27. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Ser B 1996;58:267-88.
28. Roffo G, Melzi S, Castellani U, et al. Infinite feature selection: a graph-based feature filtering approach. IEEE Trans Pattern Anal Mach Intell 2021;43:4396-410.
29. Roffo G, Melzi S, Cristani M. Infinite feature selection (abstract). 2015 IEEE International Conference on Computer Vision (ICCV), 2015, p. 4202-10.
30. Hao C, Qu Y & Wang X et al. Omics feature learning for cross individual ALS disease identification with EMG signal (abstract). 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021, p. 2700-5. https://ieeexplore.ieee.org/document/9669804
31. Nebli A, Rekik I. Gender differences in cortical morphological networks. Brain Imaging Behav 2020;14:1831-9.
32. Eshaghi A, Prados F, Brownlee WJ, et al. Deep gray matter volume loss drives disability worsening in multiple sclerosis. Ann Neurol 2018;83:210-22.
33. Pontillo G, Petracca M, Monti S, et al. Unraveling deep gray matter atrophy and iron and myelin changes in multiple sclerosis. AJNR Am J Neuroradiol 2021;42:1223-30.
34. Petracca M, Pontillo G, Moccia M, et al. Neuroimaging correlates of cognitive dysfunction in adults with multiple sclerosis. Brain Sci 2021;11:346.
35. Zalesky A, Fornito A, Cocchi L, et al. Connectome sensitivity or specificity: which is more important? Neuroimage 2016;142:407-20.
36. Schiavi S, Ocampo-Pineda M, Barakovic M, et al. A new method for accurate in vivo mapping of human brain connections using microstructural and anatomical information. Sci Adv 2020;6:eaba8245.

How to cite this article: Schiavi S, Azzari A, Mensi A, Graziano N, Daducci A, Bicego M, et al. Classification of multiple sclerosis patients based on structural disconnection: A robust feature selection approach. J Neuroimaging. 2022;1–9. https://doi.org/10.1111/jon.12991