Yamin, Abubakar; Valsasina, Paola; Tessadori, Jacopo; Filippi, Massimo; Murino, Vittorio; Rocca, Maria A.; Sona, Diego

Discovering functional connectivity features characterizing multiple sclerosis phenotypes using explainable artificial intelligence

Published in:
Human Brain Mapping

DOI:
10.1002/hbm.26210

Published: 15/04/2023

Document Version
Publisher's PDF, also known as Version of record

Received: 20 June 2022 | Revised: 14 December 2022 | Accepted: 2 January 2023

DOI: 10.1002/hbm.26210

RESEARCH ARTICLE

WILEY

Discovering functional connectivity features characterizing multiple sclerosis phenotypes using explainable artificial intelligence

Muhammad Abubakar Yamin[1,2,3,4] | Paola Valsasina[5] | Jacopo Tessadori[3,6] | Massimo Filippi[5,7,8,9,10] | Vittorio Murino[3,6] | Maria A. Rocca[5,7,10] | Diego Sona[3,11]

[1]Department of Electrical Engineering and Automation, Aalto University, Espoo, Finland

[2]Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland

[3]Pattern Analysis and Computer Vision, Istituto Italiano di Tecnologia, Genova, Italy

[4]Center for Autism Research, Kessler Foundation, East Hanover, New Jersey, USA

[5]Neuroimaging Research Unit, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy

[6]Dipartimento di Informatica, University of Verona, Verona, Italy

[7]Neurology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy

[8]Neurophysiology Service, IRCCS San Raffaele Scientific Institute, Milan, Italy

[9]Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy

[10]Vita Salute San Raffaele University, Milan, Italy

[11]Data Science for Health, Center for Digital Health and Wellbeing, Fondazione Bruno Kessler, Trento, Italy

Correspondence
Muhammad Abubakar Yamin, Department of Electrical Engineering and Automation, Aalto University, Espoo, Finland.
Email: abubakar.yamin@aalto.fi

Funding information
Fondazione Italiana Sclerosi Multipla, Grant/Award Number: FISM2018/R/5

Abstract

Multiple sclerosis (MS) is a neurological condition characterized by severe structural brain damage and by functional reorganization of the main brain networks that try to limit the clinical consequences of structural burden. Resting-state (RS) functional connectivity (FC) abnormalities found in this condition were shown to be variable across different MS phases, according to the severity of clinical manifestations. The article describes a system exploiting machine learning on RS FC matrices to discriminate different MS phenotypes and to identify relevant functional connections for MS stage characterization. To this end, the system exploits some mathematical properties of covariance-based RS FC representation, which can be described by a Riemannian manifold. The classification performance of the proposed framework was significantly above the chance level for all MS phenotypes. Moreover, the proposed system was successful in identifying relevant RS FC alterations contributing to an accurate phenotype classification.

KEYWORDS
Connectomics, functional connectivity, geodesic clustering, multiple sclerosis, Riemannian manifold

1 | INTRODUCTION

Multiple sclerosis (MS) is a chronic, inflammatory, demyelinating, and neurodegenerative disease of the central nervous system (CNS) which is characterized by nonuniform clinical manifestations and a variable progression. Clinical impairment is consequent to the accumulation of structural damage in the white and gray matter (GM) of the CNS (Filippi et al., 2018). Functional magnetic resonance imaging (fMRI) techniques hold great promise for the study of patients with MS, to improve our understanding of functional brain response to the progressive accumulation of disease-related injury. In the last two decades, the advent of resting state (RS) fMRI enabled the investigation of functional connectivity (FC) abnormalities also in severely disabled patients. RS fMRI studies of MS patients often showed trends towards higher RS FC at the earliest disease stages (Roosendaal et al., 2010), followed by a gradual RS FC reduction in progressive MS (PMS) phenotypes (Rocca et al., 2010; Rocca et al., 2018). However, complex patterns of regional increased/decreased RS FC in critical brain networks/regions were detected across different phenotypes (Rocca et al., 2018; Schoonheim et al., 2015).

Among the most advanced approaches introduced with the aim to analyze RS fMRI data, network-based analysis has recently received great attention. Brain network analysis is based on a mathematical framework that allows describing the brain as a graph. A graph consists of a collection of nodes (i.e., brain regions) and edges (i.e., functional connections, which are generally assessed using covariance or correlation measures) (Guye et al., 2010). Graph theoretical studies showed altered functional network properties in MS patients compared to healthy controls (HC), and reorganization of functional hubs which contributed to explaining phenotypic variability of MS patients and the presence of cognitive impairment (Guye et al., 2010; Liu et al., 2017; Rocca et al., 2016).

Within the graph theoretical framework, the RS FC of each subject is described as a covariance matrix defining the functional association between pairs of brain regions. Past works tried to directly utilize such matrices as inputs for a classifier (Crimi et al., 2019; Gao et al., 2012; Satterthwaite et al., 2013); however, these approaches failed to fully exploit the available information. Covariance matrices are, in fact, symmetric positive definite (SPD). As such, the space of these matrices forms a Riemannian manifold. In such a space, the Euclidean distance between these matrices is a suboptimal descriptor, as it fails to take into account the manifold curvature. Here, the separation between covariance matrices is expressed as the Log-Euclidean distance, which is a proper geodesic metric on a Riemannian manifold and takes into account the geometrical properties of SPD matrices (Wong et al., 2018; Yamin et al., 2019a).

Several works adopted Riemannian geometry to investigate fMRI data, generally with classification purposes: in Varoquaux et al. (2010) for instance, analysis of covariance matrices was used to discriminate between post-stroke and healthy subjects; in Zhao et al. (2018), a technique was proposed for the characterization of Riemannian trajectories in longitudinal studies. Still, in longitudinal studies, in Ng et al. (2014), transportation on the manifold of covariance matrices were employed to evaluate changes in FC following a particular task, while Qiu et al. (2015) adopted a combination of locally linear embedding and Log-Euclidean Riemannian metric for dimensionality reduction of functional brain networks. In Dodero, Minh, et al. (2015), Gaussian kernels based on Log-Euclidean and Stein divergence metrics on Riemannian manifolds have been used to classify study subjects in healthy controls and in people with pathological disorders. In Dodero, Sambataro, et al. (2015), Grassmannian geometry on graph Laplacians has been used to highlight subnetworks that, in turn, allow FC to be used for classification purposes.

Recently, machine learning approaches have been proposed to differentiate MS patients from other diseases and HC, as well as to distinguish between MS subgroups using structural connectivity (Bendfeldt et al., 2012; Eshaghi et al., 2016; Kocevar et al., 2016; Muthuraman et al., 2016; Stamile et al., 2015). In Zurita et al. (2018), a multimodal (structural and functional connectivity) approach using an SVM classifier has been proposed to classify relapsing–remitting (RR) MS patients from HC and in Saccà et al. (2019), multiple classifiers have been deployed in order to perform the classification between MS patients from control subjects. Both models used feature selection methods in order to explore the important networks involved in the considered pathology. Few approaches (Fiorini et al., 2015; Kocevar et al., 2016; Taschler et al., 2014) have been proposed to separate MS patients by phenotype using structural and clinical data. Usually, the covariance-based RS FC data is very high dimensional and indeed very complicated to be analyzed. So the minimization of this so-called curse of dimensionality is very crucial for classification tasks, but it could also lead to the risk of overfitting and misclassification (Guyon et al., 2002).

In this work, we propose a novel two-step framework exploiting a proper design of RS FC matrices to encode the high-dimensional connectivity data into a low-dimensional representation. Such encoding can be utilized to train a classification model to distinguish between different MS phenotypes and HC, which ultimately can lead to the identification of RS FC connections characterizing disease groups. Classification between subjects is not the main objective of this work; instead, the idea is to use the classification framework to identify the encoded features that better separate the groups. This information can then be used to identify the discriminative brain functional connections. As a first step towards classification, we used an unsupervised clustering approach where participating subjects were clustered into a limited number of groups with similar.

RS FC covariance matrices to reduce the impact of individual variability on MS phenotype characterization. Dominant set clustering was adopted, as it does not require the definition of the number of clusters, which is unknown a priori in this case. The only required information for clustering is the similarity matrix between subjects, easily defined once the Riemannian distances have been computed.

Once clusters have been assigned, we used the centroids of each cluster as reference points to encode RS FC and build a low-dimensional representation: RS FC of each subject, at this stage, is described by the distance from each centroid. This representation is much smaller (typically between 5 and 7 elements) than the

covariance matrix itself, which is defined by several thousand values. The resulting low-dimensional representation was then fed to a linear support vector machine (SVM) which provides the actual classification. Finally, a sensitivity analysis was carried out on the trained classifier in order to extract RS FC connections most likely to be affected and provide the definition of possible markers to monitor MS condition.

2 | MATERIAL AND METHODS

2.1 | Dataset acquisition and RS fMRI preprocessing

Subjects were enrolled in this prospective study at the Neuroimaging Research Unit (IRCCS San Raffaele Scientific Institute, Milan, Italy). Approval was received from the local ethical committee (IRCCS San Raffaele Scientific Institute, Milan, Italy; protocol ID FISM2013/S/1). All subjects gave written informed consent prior to study participation. To be included, subjects had to satisfy the following criteria: (1) right-handedness; (2) have no other major systemic, psychiatric, or neurological disorders; (3) no history of drug/alcohol abuse; (4) for patients, to be relapse- and steroid-free for at least 3 months before MRI, and have a stable disease-modifying treatment during the past 6 months. Within 48 h from MRI acquisition, MS patients underwent a complete neurological evaluation, with a rating of clinical disability using the Expanded Disability Status Scale (EDSS) score (Kurtzke, 1983).

MRI scans were collected from all study subjects using a 3.0T Philips Intera scanner (Philips Medical Systems, Eindhoven, The Netherlands) and included the following MRI sequences: (a) T2*-weighted single-shot echo-planar imaging (EPI) sequence for RS fMRI (repetition time [TR] = 3000 ms/echo time [TE] = 35 ms, field-of-view = 240 mm^2, matrix = 128 × 128, slice thickness = 4 mm, 200 sets of 40 contiguous axial slices); (b) dual echo turbo spin echo (TR/TE = 2599/16–80 ms; field-of-view = 240 × 240 mm^2; matrix = 256 × 256; echo train length [ETL] = 6; 44 contiguous, 3-mm-thick axial slices); and (c) 3D T1-weighted turbo field echo (TR/TE = 7/3.2 ms, inversion time [TI] = 900 ms, field-of-view = 256 × 240 mm^2, matrix = 256 × 240, slice thickness = 1 mm, 192 sagittal slices).

T2 lesion volumes were measured on dual-echo scans using a local thresholding segmentation technique (Jim 7.0, Xinapse Systems Ltd., Colchester, UK). RS fMRI standard preprocessing, including motion correction and registration to MNI space, was performed using fMRIPrep (Esteban et al., 2019) and selecting the nonaggressive independent component analysis (ICA)-based Automatic Removal Of Motion Artifacts (AROMA) denoising output (Pruim et al., 2015). 3D T1-weighted scans were lesion-filled and processed using FSL FAST (Zhang et al., 2001) for tissue segmentation. The gray matter (GM) tissue mask was matched with a subset of regions of interest (ROIs) derived from the AAL atlas (n = 86 regions considering cerebrum only, and excluding bilateral putamen and insula for misregistration issues) and applied to the processed RS fMRI scan of each subject to extract the mean time-series signal of all ROIs. The 86 × 86 RS FC matrices Σ_i were computed for each subject using the covariance between time series, describing in this way the brain connectivity in terms of both signal co-activation between ROIs and its amplitude.

2.2 | The Riemannian manifold of SPD matrices

Let $S = \{\Sigma_1, ..., \Sigma_K\}$ denote the set of covariance matrices representing the brain RS FC of all subjects. Covariance matrices are always guaranteed to be symmetric positive semi-definite, however, in a real scenario they are frequently also SPD, that is, for each matrix Σ_i of size $n \times n$ it holds that $x^T\Sigma_i x > 0, \forall\, x \neq 0 \in R^n$. Eventually, a small regularization to the main diagonal ($\Sigma_i = \Sigma_i + \lambda I$, with λ very small, e.g., $\lambda = 10^{-5}$) is sufficient to turn all semi-definite covariance matrices into SPDs. An important property of SPD matrices is that they always form a Riemannian manifold, which is a richer descriptor as compared to vectorial space (Pennec et al., 2006) and it allows the analysis of connectivity matrices in the manifold space (Dodero, Ha Quang, et al., 2015; Yamin et al., 2019a). Indeed, a metric based on the Euclidean distance is suboptimal when applied to SPD matrices (Tuzel et al., 2008) because it is not responsive to the geometry of manifold (Yamin et al., 2020). So, to fully exploit the advantage of the manifold geometrical structure, it is recommended to consider the notion of geodesic distance, which measures the length of the shortest curve on the manifold connecting two points (two matrices) and allows a better description of such data (Yamin et al., 2019b; Yamin et al., 2019c). Figure 1 illustrates the conceptual difference between Euclidean distance and geodesic distance on a manifold space.

Among the available geodesic metrics on the Riemannian manifold (Dodero, Ha Quang, et al., 2015; Dodero, Sambataro,

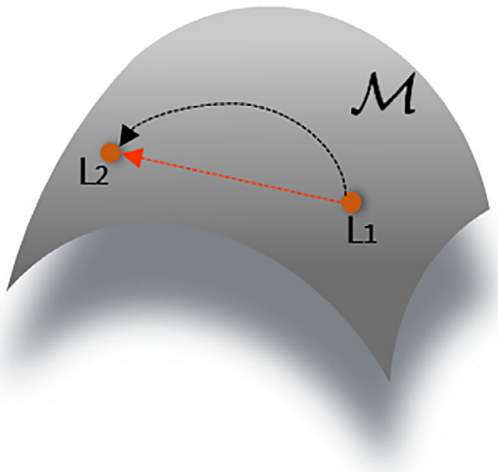

FIGURE 1 The demonstration of conceptual difference of Euclidean distance (red dashed line) and geodesic distance (black dashed line) between different point L (in orange) on a manifold M structure.

et al., 2015), we used the Log-Euclidean (Log-E) distance which is simple and fast to compute. Given two SPD matrices Σ_i and Σ_j, the Log-E distance can be computed as:

$$d_L(\Sigma_i, \Sigma_j) = \| \log(\Sigma_i) - \log(\Sigma_j) \|_F. \quad (1)$$

and the average of a set of SPD matrices is defined in a closed form (Dryden et al., 2009) as:

$$\Sigma_L = \exp\left\{ \arg \inf_{\Sigma \sum_{=1}^{K}} |\log(\Sigma_i) - \log(\Sigma)|^2 \right\} = \exp\left\{ \frac{1}{n} \sum_{i=1}^{K} \log(\Sigma_i) \right\} \quad (2)$$

2.3 | Encoding of FC matrices

In principle, RS FC alterations due to MS pathology might be helpful for classification between HC and MS patients. These alterations are concealed behind the intrinsic high variability between subjects which makes this problem hard to solve. So, the underlying hypothesis of this work is to reduce this high variability by assembling all the subjects into homogeneous groups and computing a unique representative connectome for each group. Hence, the problem can be considered as a clustering task on the RS FC matrices and the resulting centroids of each cluster become reference networks, representing groups of people with similar conditions.

The group of these reference networks represents therefore a dictionary that can be used to compress the high-dimensional connectivity patterns into a vectorial descriptor retaining the difference between groups while filtering the intrinsic variability of subjects in the same group. More specifically, the RS FC matrix of each subject can be represented by the set of distances from all cluster centroids.

2.4 | Patients classification and sensitivity analysis

With such vectorial representation of RS FC matrices, any classifier can be employed for classification. In particular, in all our experiments, we have used a linear SVM and logistic regression (LR) with LASSO regularization for comparison and analysis of features. We selected these two algorithms because they allow a simplified sensitivity analysis for further identification of relevant abnormal RS FC connections. In this way, the relative weights associated with each feature can be analyzed for their importance in characterizing and distinguishing between groups. In LR, we are using an L1 regularization, which induces a shrinking of weights associated with irrelevant features. Hence, high values are associated only with features that are important for classification. This approach is similar to Leonardi et al. (2013), where PCA was adopted to determine the representative brain networks.

Our approach is based on the speculation that being a complex data structure, RS FC matrices cannot be fully characterized using Euclidean metrics and can be better expressed by a geodesic distance on the manifold of RS FC matrices. Moreover, grouping the subjects according to their network similarity and then describing the data in terms of groups' representatives should help to minimize the noise and the great variability intrinsic to the data.

Next, we searched for alterations in brain connectivity that had the most discriminative power when comparing different groups of subjects. To this aim, we trained linear classifiers (SVM and LASSO regularized LR) on the vectorized representation of RS FC matrices of all subjects and performed a sensitivity analysis to identify the most relevant features determined by each trained classifier. The choice of the models is due to the simplicity of the sensitivity analysis, which can be carried out simply by looking at the feature weights. For each selected feature we found the threshold best separating the two analyzed classes. The difference between the geodesic means of the two groups was then used to identify the significant changes in the RS FC of the two classes.

2.5 | Geodesic dominant set clustering

We have employed Dominant Set (DS) clustering algorithm (Pavan & Pelillo, 2006), which has two main advantages: it is easily adaptable to any metric, and it automatically determines the number of clusters. Another captivating property of DS clustering is that it does not only consider the samples in isolation with respect to a reference, but it also exploits the relations with all other samples. This property generates clusters more robust-to-noise, which are completely explicit to the initialization (contrary to k-Means). Furthermore, so far DS clustering algorithm has been successfully applied in other partially related contexts (Dodero, Vascon, et al., 2015; Hou et al., 2016; Vascon et al., 2013).

The Dominant Set (DS) clustering algorithm (Pavan & Pelillo, 2006) is based on a graph theoretic concept that generalizes the maximal clique problem to weighted graphs and due to this, it iteratively computes coherent, well-separated, and compact subsets of nodes (the dominant sets) from a graph. This approach relies on an optimization algorithm to extract all clusters one by one, as needed. In DS, the dataset to be clustered is represented by a weighted undirected graph $G = (V, E, w)$, where $V = \{1, \ldots, n\}$ is the set of vertices representing the data points, the edges $E \subseteq V \times V$ represent the relationships between the data points, and w: $E \rightarrow R_+^*$ are the edge weights which reflect the similarity between the linked vertices. Indeed, DS clustering operates on a similarity matrix which can be computed as:

$$S(i,j) = 1 - \frac{d_L(i,j)}{\max(d_L)} \quad (3)$$

The similarity matrix S is a $n \times n$ non-negative, symmetric matrix, where d_L is defined in Equation (1) and $\max(d_L)$ is the maximum pairwise geodesic distance in the entire graph. In particular, an internal similarity-based nonparametric internal coherency criterion is deployed in a dominant set which analyzes the data to be included in the dominant set. The procedures of extraction of DS can be

considered as the process of maximizing the size of DS while preserving internal coherence. This is obtained by optimizing the so-called cohesiveness, a quadratic function considering the similarity matrix of the graph, and the cluster assignment of all modes. Such measures enable the data within a DS to be highly similar to each other and less similar to the data outside the DS. This allows to enable DS to formalize two important properties of all clustering methods: the intracluster homogeneity and intercluster inhomogeneity and thus a DS can be regarded as a cluster.

The cohesiveness optimization is performed through a dynamical system which is an important result from evolutionary game theory (Weibull, 1997) and is known as replicator dynamics (RD) (Bulò et al., 2011). The data to be clustered is represented in the form of a similarity matrix S. DS algorithm is nonparametric and it extracts the clusters in a sequential manner determining the number of clusters automatically. As a consequence of its ability to preserve internal coherence, the algorithm has a further advantage, it prevents the outliers to be grouped into any cluster.

3 | EXPERIMENTS

We applied our proposed algorithm to classify between HC and MS patients with different disease phenotypes according to the following scheme: HC vs. RRMS, HC vs. PMS, and RRMS vs. PMS patients. For these analyses, DS clustering resulted in different numbers of clusters ranging between 5 and 7. The working hypothesis is that RS FC matrices can be partitioned into homogeneous subgroups preserving the alterations in brain connectivity that characterize the original MS groups. So, a clustering of RS FC matrices was the first step in the proposed pipeline. In particular, we used DS clustering with geodesic distance. To achieve the double goal of both classifying HC vs. MS phenotypes and identifying discriminative abnormal connections, we designed the experiments as a composition of two processes.

3.1 | Classification

For classification purposes, we deployed a 5-fold cross-validation (CV) setup with the constraint of preserving the proportion between classes (Stone, 1974). In each iteration of CV, we used DS clustering on the training data and computed the centroid for each cluster. Then, we computed the geodesic distances of each training and testing sample from these centroids, representing all matrices as vectors of distances (features). We used these distance vectors to train and test a linear SVM. The number of features depends on the number of clusters extracted automatically by DS (usually 5 to 7 clusters). Figure 2 shows a schematic overview of the classification pipeline.

In these experiments, all distances were computed using Log-Euclidean distance (Equation 1) and the corresponding geodesic mean (Equation 2). For statistical purposes, this process was repeated 100 times and the resulting evaluation was computed as the average over these iterations. To check the significance level of classification results we also performed a permutation test on labels: a null distribution was generated by randomly permuting the labels 1000 times and for each iteration, we performed the same Linear SVM classification using the 5-folds CV approach. The resulting distribution was used to identify significance levels.

3.2 | Identification of relevant RS FC abnormalities

This part of the analysis aims at identifying brain regions having the most discriminative power when comparing different groups of

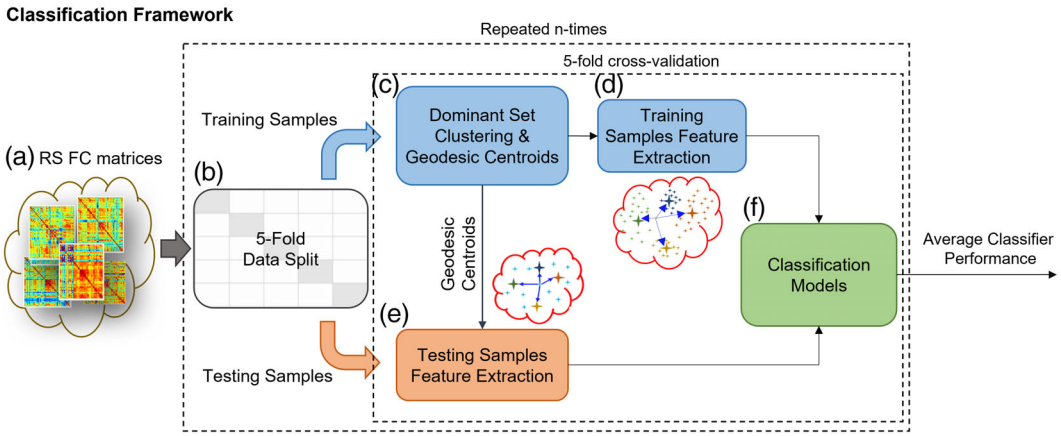

FIGURE 2 The classification framework, includes (a) input data in the form of covariance-based RS FC matrices. (b) Splitting the data into test and train samples using 5-fold cross-validation. (c) Performing dominant-set clustering on training samples and computing the centroid of each cluster using the geodesic mean formula. (d) Computing the geodesic distance of each training sample from each centroid and using this distance vector to train our classifier. (e) Computing the geodesic distance of each testing sample from each centroid obtained from training samples and using this distance vector to test the trained classifier. (f) A classification model (a linear-SVM) was used to classify MS patient phenotypes from HC and each other.

subjects. To this aim, as a first step, we used the entire dataset to compute the group representatives (cluster centroids determined by DS clustering) and then we vectorized the network representations for all subjects using the set of their distances from these representatives. Then in a 5-fold CV setup, we trained a linear classifier and performed a sensitivity analysis to identify the most relevant features determined by each trained classifier. For comparison purposes, we tested two models, the linear-SVM and the LASSO regularized logistic regression (LR). Notice that for both models the sensitivity analysis is simply carried out by looking at the feature weights.

We then made a statistical investigation repeating 100 times the CV set up to identify the relevant features as statistics of the ensemble of trained models. To select the set of significant features we used a permutation test on feature values (by randomly shuffling features value for each subject) and selected only those features with a weight value higher than the 95th percentile of permutation test values. Selected features actually represent connections having the most discriminative contribution to classification.

For each selected feature, we determined the threshold best separating the two groups of subjects. We then used the subset of subjects correctly separated by this univariate criterion to compute for each group the mean RS FC matrix using Equation (2), which actually represents the reference connectome of that group for that particular selected feature. By computing the difference of these reference connectomes for the two groups, we can actually identify the set of relevant connections which show the change of RS FC between the two groups. The flow diagram of the proposed sensitivity analysis pipeline is illustrated in Figure 3.

4 | RESULTS

Forty-three HC and 113 MS were screened for inclusion. After the exclusion of 6 HC and 13 MS patients for inadequate MR scan quality ($n = 7$ subjects), excessive head movements ($n = 9$), and bad registration output ($n = 3$), the final study cohort consisted of 37 HC and 100 MS patients. Demographic and clinical information of study subjects is reported in Table 1. There were 49 RRMS, 30 secondary progressive (SP) MS, and 21 primary progressive (PP) MS patients, defined according to Lublin et al. (2014). In order to increase the sample size and robustness of classification analysis, SPMS and PPMS patients were grouped together into a progressive MS (PMS) group. The mean T2 lesion volume in the entire patient group was 10.3 ml (interquartile range = 3.1–12.7 ml). T2 lesion volume was significantly higher ($p = .03$) in PMS (mean = 14.0 ml, interquartile range = 3.5–22.5 ml) than in RRMS patients (mean = 7.4 ml, interquartile range = 2.9–9.2 ml).

4.1 | Classification results

Figure 4 shows the boxplots for the comparison of classification accuracy achieved over 100 iterations of the 5-fold CV with full RS FC matrices. The bars represent the results for HC vs. RRMS patients, HC vs. PMS patients, and RRMS vs. PMS patients with two different classifiers (SVM in yellow and LR in green). In Figure 4, it can be seen that the achieved classification accuracy for HC vs. RRMS patients was equal to 72.51% with SVM and 70.54% with LR, for HC vs. PMS patients the average accuracy achieved was 85.19% with SVM and

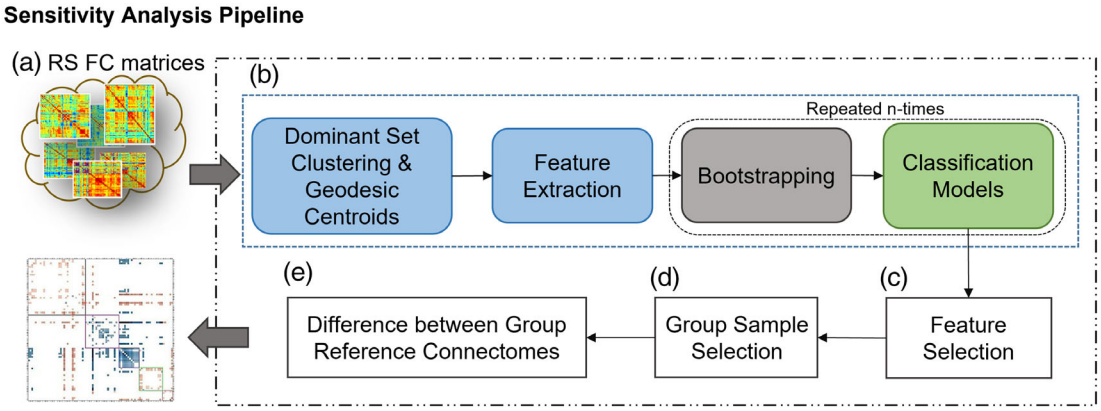

FIGURE 3 The schematic diagram of the sensitivity analysis pipeline designed to identify prominent changes in brain RS FC of MS patient phenotypes. The input data (a) in the form of covariance-based RS FC matrices is clustered with dominant-set algorithm (b) using all matrices (without any cross-validation framework). The resulting cluster representatives (geodesic centroids) were used to extract the features in the form of geodesic distances from the centroids. Then a bootstrapping mechanism was used to train multiple classifiers allowing us to analyze the weights' statistics associated with each feature. (c) Then a permutation test was used to select the set of significant features, which were then used to identify and select (d) the samples correctly separated into groups. (e) In the end, the geodesic means matrices of each group were computed and prominent connections were highlighted by simply subtracting these geodesic mean matrices (reference connectomes) from each other.

TABLE 1 Demographic and clinical features of healthy controls (HC) and patients with multiple sclerosis (MS). Patients are first considered as a whole and then divided into relapsing–remitting (RR) and progressive MS, as implemented in the classification analysis

	HC	MS patients	*p*	RRMS patients	Progressive MS patients	*p*
Subjects	37	100	-	49	51	-
Female/male [*n*]	19/18	38/62	.17[b]	31/18	31/20	.36[b]
Age, mean (IQR) [years]	36.2 (25–49)	47.4 (39–58)	<.001[a]	42.6 (35.6–49.7)	52.8 (46.5–59.7)	<.001[c]
EDSS score, median (IQR)	-	4.5 (2.0–6.0)	-	2.0 (1.5–4.0)	6.0 (5.5–7.0)	<.001[d]
Disease duration, mean (IQR) [years]	-	14.9 (3.5–20)	-	10.9 (5.8–16.3)	19.0 (12.3–25.2)	<.001[a]

Abbreviations: EDSS, expanded disability status scale; IQR, interquartile range.
[a]Two sample *t*-test.
[b]Chi-square test.
[c]ANOVA model, post hoc comparison between RRMS and PMS patients.
[d]Mann–Whitney *U* test.

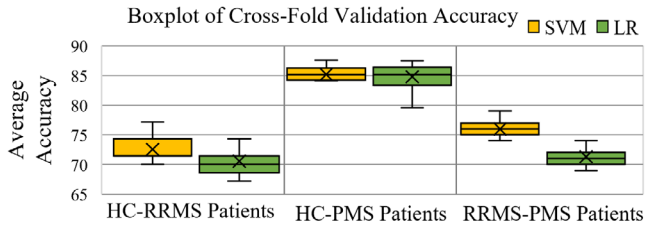

FIGURE 4 Boxplots represent the statistics over 100 repetitions of cross-fold validation accuracy with SVM (yellow) and logistic regression (green), for the classification tasks HC vs. RRMS patients, HC vs. PMS patients, and RRMS vs. PMS patients. The symbol "x" shows the mean accuracy

TABLE 2 Average classification performance obtained with SVM classifier

Group of subjects	Accuracy	Precision	Recall	F1 score
HC vs. RRMS	72.51%	72.85%	66.50%	69.40%
HC vs. PMS	85.19%	87.11%	88.38%	87.21%
RRMS vs. PMS	76.04%	77.85%	77.83%	74.65%

84.80% with LR and for RRMS vs. PMS patients, the achieved accuracy was 76.04% with SVM and 71.31% with LR. It can be perceived that accuracy with SVM was slightly better as compared to LR.

Table 2 shows the average classification performance in term of precision, recall and F1 scores for the experiments done with the SVM classifier, as the accuracy with SVM is slightly better as compared to LR. To check the statistical significance of our results achieved with the geodesic DS clustering method, we performed a permutation test on the labels. Permutation results are summarized in form of *p* values, which show the ratio of a total number of accuracy values greater than the accuracy value with the true label divided by the total number of permutations (1000 in our case). For HC vs. RRMS classification, *p* value is .014 and for HC vs. PMS classification *p* value is <.005.

4.2 | Sensitivity analysis

In order to identify RS FC connections relevant to group classification, we performed a sensitivity analysis on the trained classifiers. As explained in Section 3.2, for the given classifiers, it consists of the selection of the largest weights. As a general criterion, we performed a permutation test in order to determine a significance threshold allowing us to select the most significant weights. The boxplot in Figure 5 shows the distributions of weights of trained SVMs (yellow) and LR (green) classifiers over 100 iterations. The results are shown for all three tasks, that is, HC vs. RRMS patients, HC vs. PMS patients, and RRMS vs. PMS patients. Gray and orange horizontal lines show the maximum and minimum values of the significance threshold obtained with the permutation test for SVM and LR classifiers respectively.

It can be noticed that for all tasks the weights of the trained classifiers look quite stable, that is, there is a low variance across experiments. As described earlier in Section 2.4, due to L1 regularization, LR shrinks the weight associated with irrelevant features and assigns high values to the features which are important for classification. This thing also enables the LR classifier to have low variance across feature weights as compared to SVM and hence leads towards a more stable selection of features. So for feature selection, we considered weights obtained with LR analysis instead of SVM even if it has better performance because LR is more robust for stable feature selection.

So, for RS FC abnormalities identification, for HC vs. RRMS patients, HC vs. PMS patients, and RRMS vs. PMS patients classification, we selected clusters number 1, 2, and 4 respectively. Figure 6 illustrates the distribution of subjects around the selected feature. As described in Section 3.2, after finding the threshold over the selected feature optimally separating the two classes, we computed a reference connectome for each class as the mean of well-separated samples. We then computed the difference between reference connectomes of HC and MS phenotypes to highlight the connections mainly responsible for discrimination between the two groups.

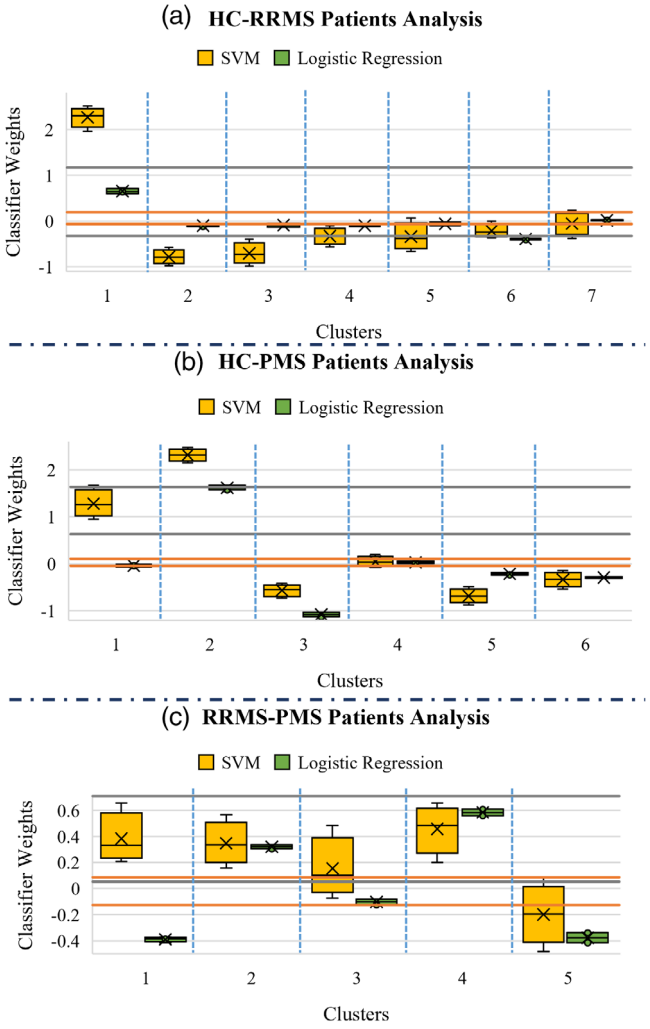

FIGURE 5 Results of sensitivity analysis for weights of SVM classifier (in yellow) and logistic regression classifier (in green) for each combination of experiments (a) HC vs. RRMS patients, (b) HC vs. PMS patients and (c) RRMS vs. PMS patients. Gray lines represent the maximum and minimum value of the threshold obtained with the permutation of features values for the SVM classifier and the orange line show the threshold values for the logistic regression classifier.

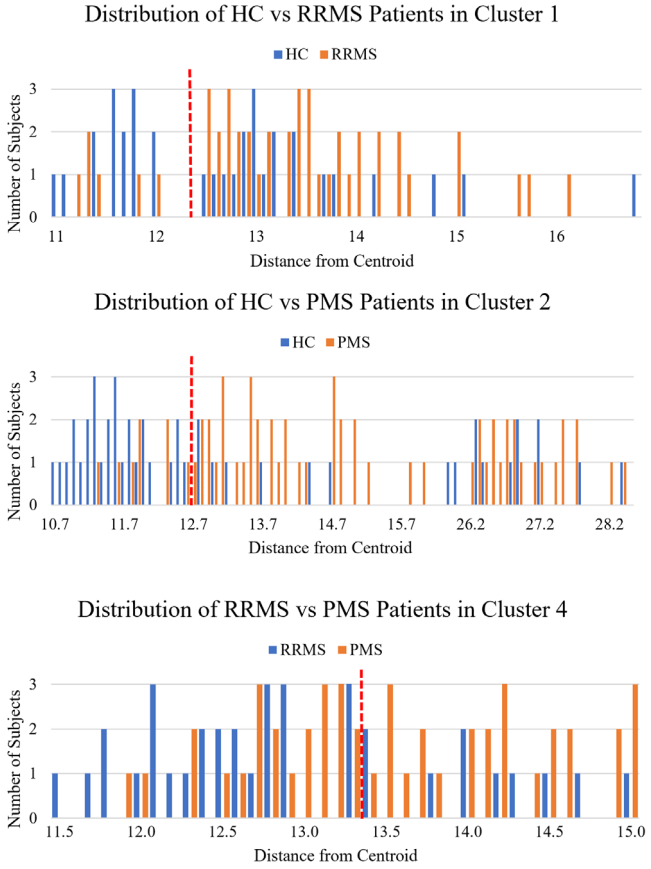

FIGURE 6 Distribution of all subjects around the selected centroid for selected cluster 1 of HC vs. RRMS (top), cluster 2 of HC vs. PMS (middle), and cluster 4 of RRMS vs. PMS (bottom). The Blue and orange color shows the subjects of two groups distributed around the centroid of the selected cluster in terms of their geodesic distance (log-E). The red dotted line shows the calculated threshold to separate the two groups

Figures 7–9 show the results obtained from the abnormalities identification analysis of HC vs. RRMS patients, HC vs. PMS patients and RRMS vs. PMS patients, respectively. The first two columns of 1st row of the figure show the reference connectome of each group along with the identification of subnetworks. The last column shows the alteration of significant connections obtained by subtracting the second reference connectome from the first one. In the bottom row, we used a connectogram to illustrate the alterations between FC differences within subnetworks and between subnetworks. Regions with blue color represent a decreased RS FC in the connectome of the second group, whereas the red color represents an increased RS FC in the second group.

As shown in Figures 7 and 8, RRMS and PMS patients were mainly characterized by increased RS FC in the basal ganglia subnetwork (especially between the bilateral thalami) vs. HC, as well as by decreased RS FC within the temporal and frontal subnetworks. In these networks, RS FC decrease mainly involved the fusiform gyrus for the temporal network and the medial frontal cortex for the frontal network, respectively. Conversely, an increased RS FC between the bilateral paracentral lobule and other regions of the frontal subnetwork was detected. Finally, both RRMS and PMS patients showed decreased RS FC within the occipital subnetwork vs. HC.

When looking at the comparison between PMS patients and HC, as highlighted in Figure 8, the increased RS FC in the basal ganglia subnetwork was more extensive than that in RRMS vs. HC and involved also the bilateral caudate nuclei. Moreover, in the parietal subnetwork, decreased RS FC among the posterior cingulate cortex, angular gyrus, and precuneus was detected. In this subnetwork, evidence of increased RS FC between the superior parietal lobule and other parietal regions was also found.

The direct comparison of PMS vs. RRMS patients showed, as highlighted in Figure 9, a stronger decrease of RS FC within occipital and temporal subnetworks in PMS vs. RRMS patients, and a higher RS FC between the posterior cingulate cortex and precuneus.

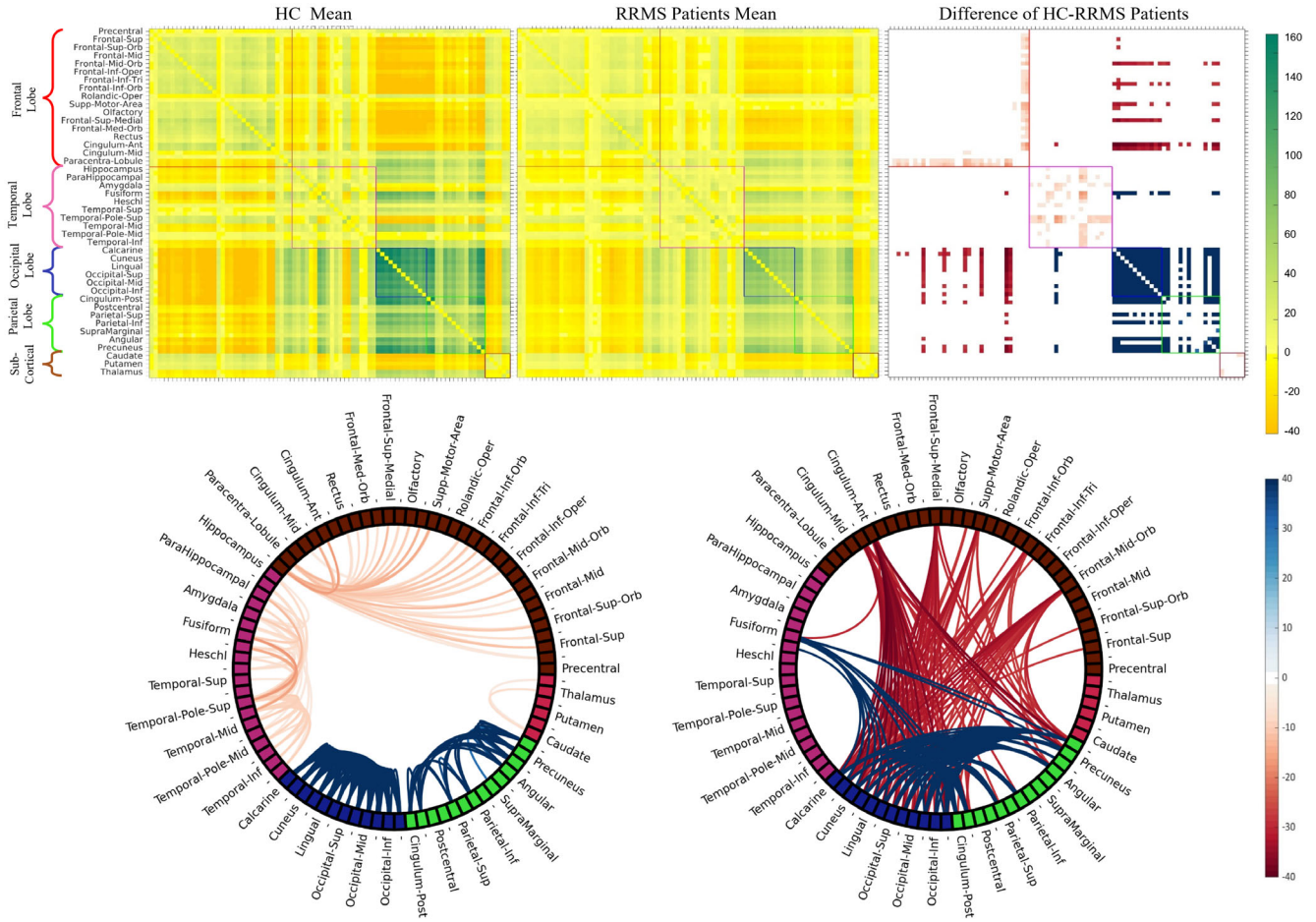

FIGURE 7 This figure illustrates the identification of RS FC abnormalities between HC and RRMS patients. Top row: The first two columns show the covariance-based reference connectomes of each group, along with the division of the full RS FC matrix into subnetworks. The last column shows the difference between reference connectomes, with significantly changing connections in terms of the difference between covariance (blue: decreased RS FC, red: increased RS FC in the second group compared to the first one). Bottom row: connectograms represent between-group RS FC differences within subnetworks (left) and between subnetworks (right)

Interestingly, as shown in Figure 7, RS FC among subnetworks was markedly increased in RRMS patients vs. HC, while between-network RS FC increase was not so evident in PMS patients vs. HC (Figure 8). This was reflected by a decreased RS FC between parietal, occipital, temporal, and frontal networks in PMS vs. RRMS (Figure 8).

5 | DISCUSSION

Analysis of RS FC reorganization in MS patients according to disease stage is a complex task, due to different plasticity mechanisms occurring in the main functional networks of the brain. According to disease phenotype, RS FC might be increased or decreased, with adaptive and maladaptive phenomena occurring with variable extents in different patients' groups (Rocca et al., 2018). This makes a classification task based on RS FC a challenging problem.

This study illustrates that a geodesic clustering-based encoding of RS FC matrices allowed a reliable differentiation of MS phenotypes from HC, and made it possible to identify relevant RS FC abnormalities most contributing to the classification. The classification results presented in Figure 4 vindicate the hypothesis that it is possible to cope with the great physiological variability across subjects grouping them by similarity, and using groups references to encode the data. These results also demonstrate that RS FC variability affecting brain networks can be presumably and effectively measured using the RS fMRI as an eminent feature of MS brain pathology and furthermore can be used to classify between healthy subjects and different phenotypes of MS.

The indexes used to evaluate the classification performance (see Table 2 and Figure 4) are significantly above the chance level for each tested task. These results strongly support the hypothesis that RS FC abnormalities can be utilized to discriminate between HC and MS phenotypes, especially when using an appropriate encoding technique which follows the manifold nature of SPD matrices. Moreover, the resulting set of clusters identifying the features was always few (between 5 and 7). This makes more robust the sensitivity analysis of feature weights after training, in order to identify the centroids responsible for the discrimination between the classes, which in turn

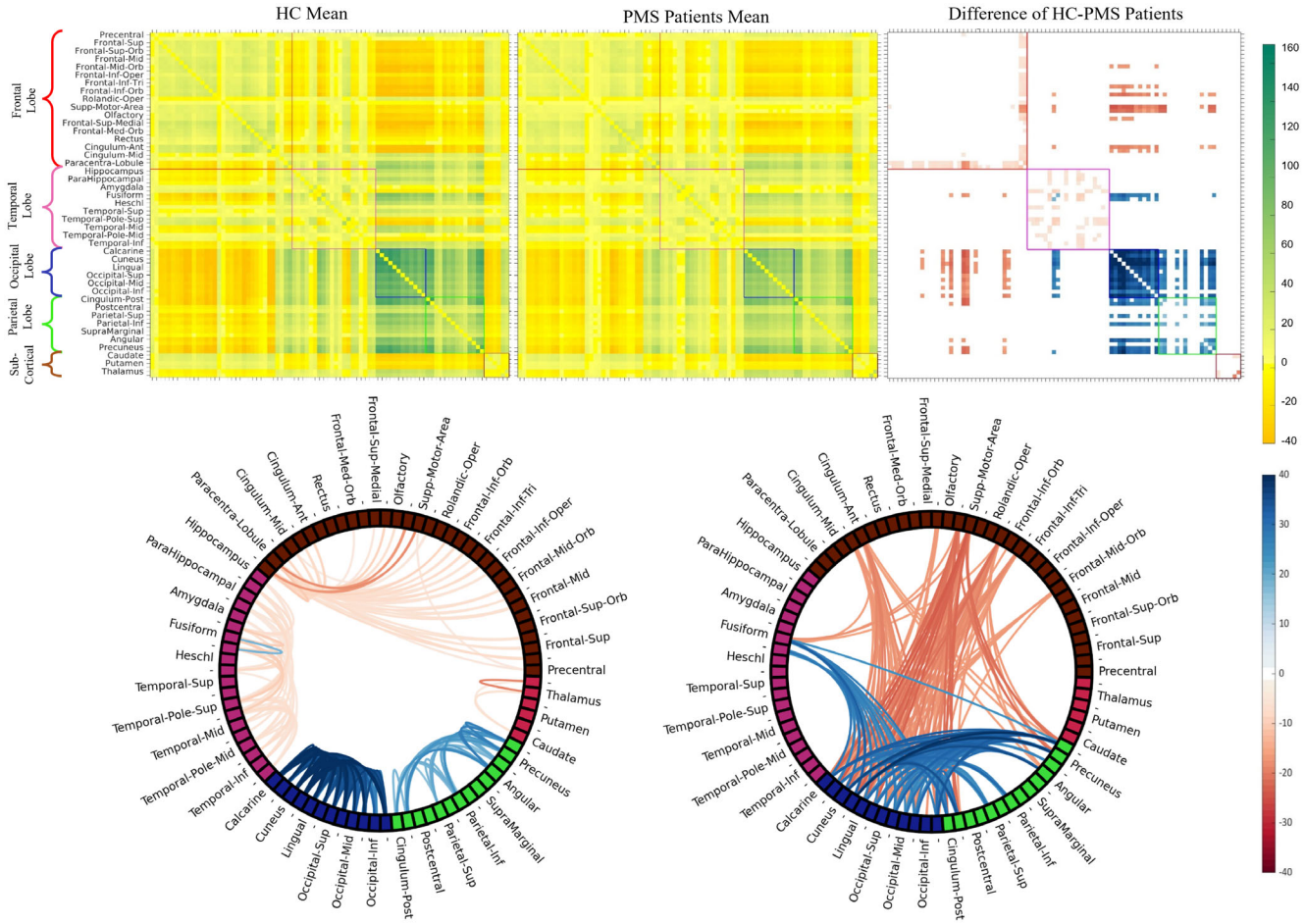

FIGURE 8 This figure illustrates the identification of RS FC abnormalities between HC and PMS patients. Top row: The first two columns show the covariance-based reference connectomes of each group, along with the division of the full RS FC matrix into subnetworks. The last column shows the difference between reference connectomes, with significantly changing connections in terms of the difference between covariance (blue: decreased RS FC, red: increased RS FC in the second group compared to the first one). Bottom row: connectograms represent between-group RS FC differences within subnetworks (left) and between subnetworks (right)

helps us to identify relevant RS FC changes (see Figures 7–9). One limitation of this study is the small sample size. To handle this issue, we have used the nested K-fold cross-validation setup which provides considerably unbiased performance estimates (Varma & Simon, 2006). Moreover, we have used linear SVMs and LASSO regularized LR, which are good with small sample set.

From a clinical point of view, we found that several functional connections in different brain lobes were clinically eloquent. Specifically, we found that increased RS FC within the subcortical network, especially involving the bilateral thalamus, and decreased FC within the temporal, parietal, and occipital subnetworks were contributing to both HC-RRMS and HC-PMS classification. The thalamus is a key deep GM structure involved in sensory, motor, and cognitive processing. These results support the notion that thalamic RS FC abnormalities occur in all disease phenotypes and might be a hallmark of MS disease (Minagar et al., 2013). This finding also confirms data from previous studies showing that abnormally high RS FC in subcortical regions, especially in the thalamus, is frequent in this condition (Hidalgo de la Cruz et al., 2021; Schoonheim et al., 2015).

The other subnetworks showing extensive RS FC abnormalities in both RRMS and PMS patients were the temporal, occipital and parietal networks.

This latter network was showing a peculiar involvement of the posterior cingulate cortex, angular gyrus, and precuneus. Such regions are well-known to be part of the default-mode network, which is one of the key networks of the brain and whose dysfunction is present in a high number of brain disorders, including MS (Rocca et al., 2010; Rocca et al., 2016; Rocca et al., 2018). Moreover, these regions, together with medial and superior temporal regions, have strong and extensive connections with several other brain regions (Rocca et al., 2016), thus facilitating functional integration. A decreased RS FC among these areas is then suggesting the presence of impaired functional integration in MS, which might be among the causes of clinical impairment (Rocca et al., 2016; Rocca et al., 2018). In addition, we found a decrease of occipital RS FC in both MS phenotypes vs. HC, which might be related to visual impairment often occurring in this disease. Finally, we would like to highlight the strong increase of RS FC among different subnetworks we found in RRMS patients

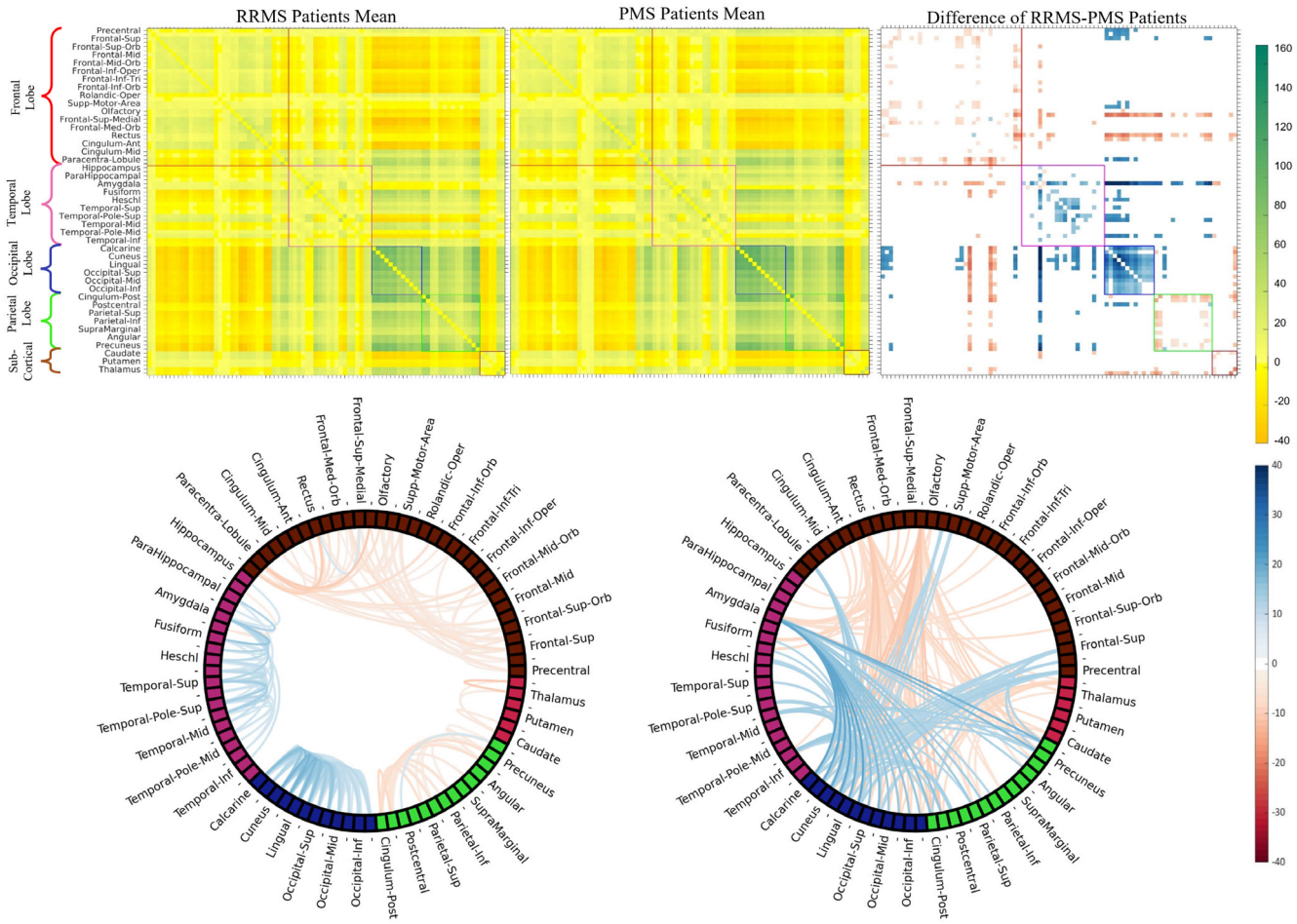

FIGURE 9 This figure illustrates the identification of RS FC abnormalities between RRMS patients and PMS patients. Top row: The first two columns show the covariance-based reference connectomes of each group, along with the division of the full RS FC matrix into subnetworks. The last column shows the difference between reference connectomes, with significantly changing connections in terms of the difference between covariance (blue: decreased RS FC, red: increased RS FC in the second group compared to the first one). Bottom row: connectograms represent between-group RS FC differences within subnetworks (left) and between subnetworks (right)

vs. HC, which tended to decline in PMS patients vs. HC. This finding suggests that, at the early disease stage, increased RS FC among regions belonging to different functional systems might be compensatory, and help to maintain moderate disability levels (Hidalgo de la Cruz et al., 2021; Liu et al., 2017; Roosendaal et al., 2010). This mechanism is likely to be lost in PMS patients, which typically show decreased long-range RS FC and severe clinical/cognitive deficits (Rocca et al., 2010; Rocca et al., 2016; Rocca et al., 2018).

6 | CONCLUSION

This work describes a novel computational framework to discriminate between HC and different phenotypes of MS patients based on FC matrices computed from RS fMRI data. Briefly, the proposed method consisted of a graph clustering framework that employs a DS clustering algorithm to cluster RS FC matrices on the Riemannian manifold and considering the properties of SPD matrices we used geodesic metric (more specifically, Log-Euclidean distances), which, in this context, proved to be superior to the Euclidean metric. Besides this clustering approach, a novel idea of data encoding was also suggested. The hypothesis we intended to test with this approach is that grouping subjects with similar FCs help compensating for intrinsic intersubject variability, thus facilitating the extraction of features significant for disease phenotype characterization. The framework has been conceived to study how the overall connectivity changes according to the different disease phenotypes. Classification is only used to detect the most discriminant features. Nevertheless, it can be used in clinical applications where the disease phenotype for new subjects can be determined using the already trained framework, processing the subjects as test samples.

More in detail, classification features have been obtained as the geodesic distances between each subject's RS FC matrix and the centroids of the clusters defined through DS clustering. Both clustering and classifier training occurred on training folds, in order to prevent double-dipping. In testing, we computed the distances of each test sample from the centroids defined on the training set and then used these distance values to test the performance of the trained classifier.

In order to minimize the impact of fold selection, the results presented in the article are relative to 100 repetitions of the training algorithm. This study also proved the viability of the encoding scheme of RS FC matrices by using DS clustering and then defining a vector space representation according to their distance from cluster centroids.

The results of this study proved that MS patients at different stages of the disease, observed as a whole, were well discriminated from HC in terms of RS FC. This method might be beneficial in monitoring disease development and improving patient management. Ideally, the results presented here can be extended in the future to longitudinal studies, so that, instead of a classification task, this approach can be used to constantly monitor the evolution of single subjects and provide a pharmacological therapy tailored to the predicted changes.

ACKNOWLEDGMENTS

This study was partially supported by FISM with a research grant (FISM2018/R/5), and financed or co-financed with the "5 per mille" public funding.

DATA AVAILABILITY STATEMENT

The anonymized dataset used and analyzed during the current study are available from the corresponding author upon reasonable request.

ORCID

Muhammad Abubakar Yamin https://orcid.org/0000-0003-3188-709X
Massimo Filippi https://orcid.org/0000-0002-5485-0479
Maria A. Rocca https://orcid.org/0000-0003-2358-4320
Diego Sona https://orcid.org/0000-0003-2883-5336

REFERENCES

Bendfeldt, K., Klöppel, S., Nichols, T. E., Smieskova, R., Kuster, P., Traud, S., Mueller-Lenke, N., Naegelin, Y., Kappos, L., Radue, E.-W., & Borgwardt, S. J. (2012). Multivariate pattern classification of gray matter pathology in multiple sclerosis. *NeuroImage*, *60*(1), 400–408.

Bulò, S. R., Pelillo, M., & Bomze, I. M. (2011). Graph-based quadratic optimization: A fast evolutionary approach. *Computer Vision and Image Understanding*, *115*(7), 984–995.

Crimi, A., Giancardo, L., Sambataro, F., Gozzi, A., Murino, V., & Sona, D. (2019). Multilink analysis: Brain network comparison via sparse connectivity analysis. *Scientific Reports*, *9*(1), 1–13.

Dodero, L., Ha Quang, M., San Biagio, M., Murino, V., & Sona, D. (2015). Kernel-based classification for brain connectivity graphs on the Riemannian manifold of positive definite matrices. In *Proceedings – International symposium on biomedical imaging (04 2015)*. IEEE. https://doi.org/10.1109/ISBI.2015.7163812

Dodero, L., Minh, H. Q., San Biagio, M., Murino, V., & Sona, D. (2015). Kernel-based classification for brain connectivity graphs on the Riemannian manifold of positive definite matrices. In *2015 IEEE 12th international symposium on biomedical imaging (ISBI)* (pp. 42–45). IEEE.

Dodero, L., Sambataro, F., Murino, V., & Sona, D. (2015). Kernel-based analysis of functional brain connectivity on Grassmann manifold. In *International conference on medical image computing and computer-assisted intervention* (pp. 604–611). Springer.

Dodero, L., Vascon, S., Murino, V., Bifone, A., Gozzi, A., & Sona, D. (2015). Automated multi-subject fiber clustering of mouse brain using dominant sets. *Frontiers in Neuroinformatics*, *8*, 87.

Dryden, I. L., Koloydenko, A., & Zhou, D. (2009). Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, *3*(3), 1102–1123.

Eshaghi, A., Wottschel, V., Cortese, R., Calabrese, M., Sahraian, M. A., Thompson, A. J., Alexander, D. C., & Ciccarelli, O. (2016). Gray matter MRI differentiates neuromyelitis optica from multiple sclerosis using random forest. *Neurology*, *87*(23), 2463–2470.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116.

Filippi, M., Bar-Or, A., Piehl, F., Preziosa, P., Solari, A., Vukusic, S., & Rocca, M. A. (2018). Multiple sclerosis. *Nature Reviews Disease Primers*, *4*, 43.

Fiorini, S., Verri, A., Tacchino, A., Ponzio, M., Brichetto, G., & Barla, A. (2015). A machine learning pipeline for multiple sclerosis course detection from clinical scales and patient reported outcomes. In *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 4443–4446). IEEE.

Gao, S., Tsang, I. W.-H., & Chia, L.-T. (2012). Sparse representation with kernels. *IEEE Transactions on Image Processing*, *22*(2), 423–434.

Guye, M., Bettus, G., Bartolomei, F., & Cozzone, P. J. (2010). Graph theoretical analysis of structural and functional connectivity MRI in normal and pathological brain networks. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *23*(5–6), 409–421.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1), 389–422.

Hidalgo de la Cruz, M., Valsasina, P., Mesaros, S., Meani, A., Ivanovic, J., Martinovic, V., Drulovic, J., Filippi, M., & Rocca, M. A. (2021). Clinical predictivity of thalamic sub-regional connectivity in clinically isolated syndrome: A 7-year study. *Molecular Psychiatry*, *26*(6), 2163–2174.

Hou, J., Gao, H., & Li, X. (2016). DSets-DBSCAN: A parameter-free clustering algorithm. *IEEE Transactions on Image Processing*, *25*(7), 3182–3193.

Kocevar, G., Stamile, C., Hannoun, S., Cotton, F., Vukusic, S., Durand-Dubief, F., & Sappey-Marinier, D. (2016). Graph theory-based brain connectivity for automatic classification of multiple sclerosis clinical courses. *Frontiers in Neuroscience*, *10*, 478.

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, *33*(11), 1444.

Leonardi, N., Richiardi, J., Gschwind, M., Simioni, S., Annoni, J.-M., Schluep, M., Vuilleumier, P., & Van De Ville, D. (2013). Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest. *NeuroImage*, *83*, 937–950.

Liu, Y., Wang, H., Duan, Y., Huang, J., Ren, Z., Ye, J., Dong, H., Shi, F., Li, K., & Wang, J. (2017). Functional brain network alterations in clinically isolated syndrome and multiple sclerosis: A graph-based connectome study. *Radiology*, *282*(2), 534–541.

Lublin, F. D., Reingold, S. C., Cohen, J. A., Cutter, G. R., Sørensen, P. S., Thompson, A. J., Wolinsky, J. S., Balcer, L. J., Banwell, B., Barkhof, F., Bebo, B., Calabresi, P. A., Clanet, M., Comi, G., Fox, R. J., Freedman, M. S., Goodman, A. D., Inglese, M., Kappos, L., … Polman, C. H. (2014). Defining the clinical course of multiple sclerosis: The 2013 revisions. *Neurology*, *83*(3), 278–286.

Minagar, A., Barnett, M. H., Benedict, R. H., Pelletier, D., Pirko, I., Sahraian, M. A., Frohman, E., & Zivadinov, R. (2013). The thalamus and multiple sclerosis: Modern views on pathologic, imaging, and clinical aspects. *Neurology*, *80*(2), 210–219.

Muthuraman, M., Fleischer, V., Kolber, P., Luessi, F., Zipp, F., & Groppa, S. (2016). Structural brain network characteristics can differentiate CIS from early RRMS. *Frontiers in Neuroscience*, *10*, 14.

Ng, B., Dressler, M., Varoquaux, G., Poline, J. B., Greicius, M., & Thirion, B. (2014). Transport on Riemannian manifold for functional connectivity-

based classification. In *International conference on medical image computing and computer-assisted intervention* (pp. 405–412). Springer.

Pavan, M., & Pelillo, M. (2006). Dominant sets and pairwise clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(1), 167–172.

Pennec, X., Fillard, P., & Ayache, N.(2006). A Riemannian framework for tensor computing. *International Journal of Computer Vision*, *66* (1), 41–66.

Pruim, R. H., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J. K., & Beckmann, C. F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage*, *112*, 267–277.

Qiu, A., Lee, A., Tan, M., & Chung, M. K. (2015). Manifold learning on brain functional networks in aging. *Medical Image Analysis*, *20*(1), 52–60.

Rocca, M. A., Valsasina, P., Absinta, M., Riccitelli, G., Rodegher, M. E., Misci, P., Rossi, P., Falini, A., Comi, G., & Filippi, M. (2010). Default-mode network dysfunction and cognitive impairment in progressive MS. *Neurology*, *74*(16), 1252–1259.

Rocca, M. A., Valsasina, P., Leavitt, V. M., Rodegher, M., Radaelli, M., Riccitelli, G. C., Martinelli, V., Martinelli-Boneschi, F., Falini, A., Comi, G., & Filippi, M. (2018). Functional network connectivity abnormalities in multiple sclerosis: Correlations with disability and cognitive impairment. *Multiple Sclerosis Journal*, *24*(4), 459–471.

Rocca, M. A., Valsasina, P., Meani, A., Falini, A., Comi, G., & Filippi, M. (2016). Impaired functional integration in multiple sclerosis: A graph theory study. *Brain Structure and Function*, *221*(1), 115–131.

Roosendaal, S. D., Schoonheim, M. M., Hulst, H. E., Sanz-Arigita, E. J., Smith, S. M., Geurts, J. J., & Barkhof, F. (2010). Resting state networks change in clinically isolated syndrome. *Brain*, *133*(6), 1612–1621.

Saccà, V., Sarica, A., Novellino, F., Barone, S., Tallarico, T., Filippelli, E., Granata, A., Chiriaco, C., Bruno Bossio, R., Valentino, P., & Quattrone, A. (2019). Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state FMRI connectivity data. *Brain Imaging and Behavior*, *13*(4), 1103–1114.

Satterthwaite, T. D., Wolf, D. H., Ruparel, K., Erus, G., Elliott, M. A., Eickhoff, S. B., Gennatas, E. D., Jackson, C., Prabhakaran, K., Smith, A., Hakonarson, H., Verma, R., Davatzikos, C., Gur, R. E., & Gur, R. C. (2013). Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. *NeuroImage*, *83*, 45–57.

Schoonheim, M. M., Hulst, H. E., Brandt, R. B., Strik, M., Wink, A. M., Uitdehaag, B. M., Barkhof, F., & Geurts, J. J. (2015). Thalamus structure and function determine severity of cognitive impairment in multiple sclerosis. *Neurology*, *84*(8), 776–783.

Stamile, C., Kocevar, G., Hannoun, S., Durand-Dubief, F., & Sappey-Marinier, D. (2015). A graph based classification method for multiple sclerosis clinical forms using support vector machine. In *Medical learning meets medical imaging* (pp. 57–64). Springer.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *36*(2), 111–133.

Taschler, B., Ge, T., Bendfeldt, K., Müller-Lenke, N., Johnson, T. D., & Nichols, T. E. (2014). Spatial modeling of multiple sclerosis for disease subtype prediction. In *International conference on medical image computing and computer-assisted intervention* (pp. 797–804). Springer.

Tuzel, O., Porikli, F., & Meer, P. (2008). Pedestrian detection via classification on Riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(10), 1713–1727.

Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, *7*(1), 1–8.

Varoquaux, G., Baronnet, F., Kleinschmidt, A., Fillard, P., & Thirion, B. (2010). Detection of brain functional-connectivity difference in post-stroke patients using group-level covariance modeling. In *International conference on medical image computing and computer-assisted intervention* (pp. 200–208). Springer.

Vascon, S., Cristani, M., Pelillo, M., & Murino, V. (2013). *Using dominant sets for k-NN prototype selection* (pp. 131–140). Springer.

Weibull, J. W. (1997). *Evolutionary game theory*. MIT Press.

Wong, E., Anderson, J. S., Zielinski, B. A., & Fletcher, P. T. (2018). Riemannian regression and classification models of brain networks applied to autism. In *International workshop on connectomics in neuroimaging* (pp. 78–87). Springer.

Yamin, A., Dayan, M., Squarcina, L., Brambilla, P., Murino, V., Diwadkar, V., & Sona, D. (2019a). *Comparison of brain connectomes using geodesic distance on manifold: A twins study. In 2019 IEEE International Symposium on Biomedical Imaging (ISBI2019)* (pp. 1797–1800). IEEE.

Yamin, A., Dayan, M., Squarcina, L., Brambilla, P., Murino, V., Diwadkar, V., & Sona, D. (2019c). Analysis of dynamic brain connectivity through geodesic clustering. In *International conference on image analysis and processing* (pp. 640–648). Springer.

Yamin, M. A., Dayan, M., Squarcina, L., Brambilla, P., Murino, V., Diwadkar, V., & Sona, D. (2019b). Investigating the impact of genetic background on brain dynamic functional connectivity through machine learning: A twins study. In *2019 IEEE EMBS international conference on biomedical & health informatics (BHI)* (pp. 1–4). IEEE.

Yamin, M. A., Tessadori, J., Akbar, M. U., Dayan, M., Murino, V., & Sona, D. (2020). Geodesic clustering of positive definite matrices for classification of mental disorder using brain functional connectivity. In *2020 international joint conference on neural networks (IJCNN)* (pp. 1–5). IEEE.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57.

Zhao, Q., Kwon, D., & Pohl, K. M. (2018). A Riemannian framework for longitudinal analysis of resting-state functional connectivity. In *International conference on medical image computing and computer-assisted intervention* (pp. 145–153). Springer.

Zurita, M., Montalba, C., Labbé, T., Cruz, J. P., da Rocha, J. D., Tejos, C., Ciampi, E., Cárcamo, C., Sitaram, R., & Uribe, S. (2018). Characterization of relapsing-remitting multiple sclerosis patients using support vector machine classifications of functional and diffusion MRI data. *NeuroImage: Clinical*, *20*, 724–730.

How to cite this article: Yamin, M. A., Valsasina, P., Tessadori, J., Filippi, M., Murino, V., Rocca, M. A., & Sona, D. (2023). Discovering functional connectivity features characterizing multiple sclerosis phenotypes using explainable artificial intelligence. *Human Brain Mapping*, 1–13. https://doi.org/10.1002/hbm.26210