Magnetoencephalography-based interpretable automated differential diagnosis in neurodegenerative diseases

Klepachevskyi D[1,2], Romano A[3], Aristimunha B[4,5], Angiolelli M[6], Trojsi F[7], Bonavita S[7], Sorrentino G[8,9,10], Andreone V[11], Minino R[3], Troisi Lopez E[9], Polverino A[10], Jirsa V[2], Saudargienė A[1], Corsi M.-C.[12]*, Sorrentino P[2,9,13]*

1. Neuroscience Institute, Lithuanian University of Health Sciences, Kaunas, Lithuania
2. Institut de Neurosciences des Systèmes, Aix-Marseille Université, 13005 Marseille, France
3. Department of Medical Motor and Wellness Sciences, University of Naples "Parthenope", Naples, Italy
4. Université Paris-Saclay, Inria TAU, LISN-CNRS, France
5. Universidade Federal do ABC, Santo André, Brazil
6. Department of Engineering, Università Campus Bio-Medico di Roma, Rome, Italy
7. Department of Advanced Medical and Surgical Sciences, University of Campania "Luigi Vanvitelli", Naples, Italy
8. Department of Economic, Legal, Informatics and Motor Sciences, University of Naples Parthenope, Nola, Italy
9. Institute of Applied Sciences and Intelligent Systems, National Research Council, Pozzuoli, Italy
10. ICS Maugeri Hermitage Napoli, Via Miano 69 - 80145
11. Department of Neurology, Cardarelli Hospital, Naples, Italy
12. Sorbonne Université, Institut du Cerveau – Paris Brain Institute -ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France
13. University of Sassari, Department of Biomedical Sciences, Viale San Pietro, 07100, Sassari, Italy

*These authors supervised equally

† Correspondance to: pierpaolo.SORRENTINO@univ-amu.fr

Abstract

Automating the diagnostic process steps has been of interest for research grounds and to help manage the healthcare systems. Improved classification accuracies, provided by ever more sophisticated algorithms, were mirrored by the loss of interpretability on the criteria for achieving accuracy. In other words, the mechanisms responsible for generating the distinguishing features are typically not investigated. Furthermore, the vast majority of the classification studies focus on the classification of one disease as opposed to matched controls. While this scenario has internal validity, concerning the appropriateness toward answering scientific questions, it does not have external validity. In other words, differentiating multiple diseases at once is a classification problem closer to many real-world scenarios. In

this work, we test the hypothesis that specific data features hold most of the discriminative power across multiple neurodegenerative diseases. Furthermore, we perform an explorative analysis to compare metrics based on different assumptions (concerning the underlying mechanisms). To test this hypothesis, we leverage a large Magnetoencephalography dataset (N=109) merging four cohorts, recorded in the same clinical setting, of patients affected by multiple sclerosis, amyotrophic lateral sclerosis, Parkinson's disease, and mild cognitive impairment. Our results show that it is possible to reach a balanced accuracy of 67,1% (chance level = 35%), based on a small set of (non-disease specific) features. We show that edge metrics (defined as statistical dependencies between pairs of brain signals) perform better than nodal metrics (considering region while disregarding the interactions. Moreover, phase-based metrics slightly outperform amplitude-based metrics. In conclusion, our work shows that a small set of phase-based connectivity metrics applied to MEG data successfully distinguishes across multiple neurological diseases.

Introduction

In the last twenty-five years, the widespread availability of large-scale brain functional data in health and disease has brought great hope toward discovering the mechanisms underpinning brain diseases and the appearance of neurological symptoms. Cognitive functions emerge from the coordinated interactions among brain regions, manifesting as statistical dependencies among the corresponding brain signals. The overall statistical dependencies between all pairs of signals are often referred to as “functional connectivity” (Friston, 1994). Functional connectivity (FC) is subject-specific and allows subject identification (Finn & Rosenberg, 2021), is altered during the execution of tasks (Corsi et al., 2020), in different environmental conditions (Shine et al., 2016), as well as in neurological diseases (Sorrentino et al., n.d., 2018, 2019). The commonest and most straightforward approach to assessing statistical dependencies has been using descriptive metrics (e.g., Pearson’s correlation). This approach has no underlying assumptions concerning the mechanism underlying the observed statistical dependencies. Other techniques take a more mechanisms-driven approach. As an example, the hypothesis of communication through coherence posits that the occurrence of communication between regions might occur via more or less synchronization (Fries, 2015). Then, metrics such as the Phase Locking value (PLV) were developed to quantify communication via the synchronization between brain signals (such as electroencephalography-EEG and magnetoencephalography-MEG) (Bastos & Schoffelen, 2016). These metrics have been classically used to characterize multiple neurodegenerative diseases (Stam, 2010). More recently, it was shown that large-scale brain activity is far from stationary, and instead, it is characterized by aperiodic, scale-free bursts of activity (Haldeman & Beggs, 2005; Shriki et al., 2013; Tagliazucchi et al., 2012). Then, borrowing from statistical mechanics, the dependencies among brain regions were understood as the presence of scale-free bursts of activities, named *“neuronal avalanches*”, that describe the presence of aperiodic, non-linear bursts of activities spreading brain regions. Intriguingly, in several neurological diseases (such as Parkinson’s disease, Amyotrophic lateral sclerosis, and Mild Cognitive Impairment), brain dynamics spread differently with respect to healthy controls (Polverino et al., 2024; Romano et al., 2023; Sorrentino et al., 2019), and, more importantly, changes in the way aperiodic waves spread proved to be strongly predictive of individual clinical disability (Polverino et al., 2024; Romano et al., 2023).

Despite extensive efforts, there has been a lack of replicability of the studies, regardless of the particular technique adopted to estimate functional connectivity (Kelly & Hoptman, 2022). In other words, the measurements and metrics devised to this day might fail to optimally capture disease-relevant mechanisms comprehensively. As a consequence, automatic

classification among multiple neurological diseases cannot be achieved with high accuracy based on functional data alone.

In this paper, we take a different approach and start from the assumption that the way pathophysiological processes spread across the brain has some aspects to it that are specific to a given disease and can be best measured in a set of features that are (spatially) shared among multiple diseases. As a direct consequence, functional connectivity should show specific elements that distinguish various diseases. Hence, the first hypothesis of our study is that it is possible to identify a (small) set of features that can classify multiple neurological diseases.

To test our hypothesis, we leveraged a vast cohort of source-reconstructed MEG data from patients affected by mild cognitive impairment (MCI), amyotrophic lateral sclerosis (ALS), Parkinson's disease (PD), and Multiple Sclerosis (MS).

First, we compared the classification performance of four FC metrics that capture different properties of the signals (AEC, PLV, Pearson's correlation coefficient, and ATM) associated with a four-class problem (i.e. MCI, PD, MS, and ALS). We considered the features that can differentiate the considered neurological diseases for each FC metric taken separately.

Furthermore, we compared nodal and edge metrics, under the hypothesis that edges, which more directly represented the interactions among brain regions, would outperform local (i.e.) metrics. We compared the classification performance using three different machine learning algorithms (i.e., XGBoost, Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA), to demonstrate that the performance of a given feature-set is algorithm-independent.

Finally, for each FC metric, we identified the most informative features used by the classifier, under the hypothesis that such relevant features were linked to the neurophysiology of the considered neurological diseases. Such a study would make the classification results more interpretable and would enable us to identify clusters of brain interactions sensitive to the neurophysiological mechanisms associated with the considered diseases.

The purpose of this work is to explore a diverse set of connectivity metrics to propose an interpretable automated pipeline for differentiated diagnosis of neurodegenerative diseases.

1. Materials and Methods

2.1 Participants

One hundred nine patients with different neurological diseases (ALS, MCI, PD, MS) were recruited from Hermitage Capodimonte Clinic in Naples (Polverino et al., 2022; Romano et al., 2023; Sorrentino et al., 2019, 2022). Specifically, Thirty-two MCI patients (18 males and 14

females; mean age 71.31; SD ± 6.83; mean education 10.54; SD ± 4.33) were recruited from the Center of Cognitive and Memory Disorders of the Hermitage Capodimonte Clinic in Naples, Italy. The MCI diagnosis was done according to the National Institute on Ageing-Alzheimer's Association criteria (Albert et al., 2011). Thirty-nine ALS patients (29 males and 10 females; mean age 59.63; SD ± 12.87; mean education 10.38 years SD ± 4.3) were selected in collaboration with the ALS Center of the First Division of Neurology of the University of Campania "Luigi Vanvitelli" (Naples, Italy). The ALS diagnosis was performed according to the El-Escorial criteria (Brooks, 1994). Twenty patients (14 males and 6 females; mean age 64.5; SD ± 12.18; mean education 11 years SD ± 3.9) with a confirmed diagnosis of Parkinson's disease according to the United Kingdom Parkinson's Disease Brain Bank criteria (Gibb & Lees, 1988) were recruited in collaboration with the Movement Disorder Unit of Cardarelli hospital in Naples. Finally, eighteen patients (6 males and 12 females; mean age 45.05; SD ± 9.92; mean education 14-11 years SD ± 4.89) with Multiple Sclerosis were recruited in collaboration with University of Campania Luigi Vanvitelli. The diagnosis was performed following the 2017 revision of the McDonald criteria (Thompson et al., 2018). Each participant underwent a specific motor and/or neuropsychological evaluation according to the clinical characteristics of each disease. A complete summary of the cohort description is available in Table 1. The study protocol was approved by the ''Comitato Etico Campania Centro'' (Prot.n.93C.E./Reg. n.14-17OSS) and all participants provided written informed consent in accordance with the Declaration of Helsinki.

Type of disease	***Number of participants (109)***	***Age (mean ± SD)***	***Years of education (mean ± SD)***	***Gender (ratio)***
Mild Cognitive Impairment (MCI)	32	71.31 (SD ± 6.83)	10.54 (SD ± 4.33)	18 m / 14 f
Multiple Sclerosis (MS)	18	45.05 (SD ± 9.92)	14.11 (SD ± 4.89)	6m /12 f
Parkinson's Disease (PD)	20	64.5 (SD ± 12.18)	11 (SD ± 3.9)	14 m / 6 f
Amyotrophic Lateral Sclerosis (ALS)	39	59.63 (SD ± 12.87)	10.38 (SD ± 4.3)	29 m/10 f

Table 1: Demographic features of the cohort: m: males; f: females; SD: Standard Deviation

2.2 MEG and MRI acquisition, pre-processing, and source reconstruction

MEG and MRI acquisition, preprocessing, and source reconstruction were performed similarly to previous studies (Cipriano et al., 2024, p. 20; Romano et al., 2022). Briefly, all patients underwent an MRI scan using a 3T Biograph mMR tomograph (Siemens HealthcareErlangen, Germany) equipped with a 12 channels head coil. Specifically, 3 dimensional T1-weighted images (gradient-echo sequence inversion recovery prepared fast spoiled gradient recalled-echo, time repetition = 6,988 ms, inversion time = 1,100 ms, echo time = 3.9 ms, flip angle = 10, voxel size = 1 × 1 × 1.2 mm3) were acquired. The MEG acquisition was performed using a 163-magnetometer system placed in a magnetically shielded room (AtB Biomag UG, Ulm, Germany). Fastrack (Polhemus®) was used to define the position of the head under the helmet and to digitalize the position of four anatomical landmarks (nasion, right, and left preauricular and apex) and four reference coils. Each patient performed two recordings of 3.5 minutes each, with a one-minute break, during a resting state, with eyes closed. Electrocardiographic and electrooculographic signals were recorded to remove physiological artifacts. Data were acquired with a sampling frequency of 1024 Hz. A Principal component analysis (PCA) was used to reduce the environmental noise, and an independent component analysis (ICA) was used to remove physiological artifacts (namely ocular and cardiac artifacts). Finally, to obtain the source-reconstructed time series of the patients, according to the Automated Anatomical Labeling (AAL) atlas, we used a beamformer algorithm and the volume conduction model proposed by Nolte (Nolte, 2003). The time series were filtered between 0.5 and 48 Hz.

2.3 Connectivity Metrics

Phase Locking Value (PLV)

The PLV measures the phase synchronization between two narrowband signals, and it is computed as : (Lachaux et al., 1999).

$$PLV = \left| E\left[e^{j\Delta\Phi xy(t)} \right] \right|,$$

where $\Delta\Phi_{xy}(t)$ represents the difference between $\Phi_x(t)$ - $\Delta\Phi_y(t)$, $[E\ ]$ is the statistical expectation, and $\Delta\Phi_{x,y}(t)$ are the instantaneous phases of the analytical signals.

Correlation Coefficient (CC)

We computed Pearson's correlation coefficient to estimate the pairwise synchronization between signals of different brain regions.

Avalanche Transition Matrix (ATM)

The ATM describes the probability that after the activation of region *i* at the time *t* , the region *j* will be active at the time t +δ (Sorrentino et al., 2021). The ATMs are computed starting from neuronal avalanches, which are defined as events that start when at least one region is above the threshold and end when all the regions return to their baseline activity. Hence, there is one ATM for each avalanche. More specifically, the ATM contains, in the ijth position, the probability that region j is active at time t+1 given that region t is active at time t.ATMs were then averaged element-wise over all the avalanches for a subject, and finally symmetrized.

Amplitude Envelope Correlation (AEC)

The amplitude envelope is used to estimate the statistical interdependencies between brain regions. It is computed as the correlation coefficient between the analytical amplitude of two signals. High values of amplitude correlation between the envelopes indicate that two brain regions display a coordinated behavior (Brookes et al., 2011, 2012).

Nodal analysis

Each of the connectivity metrics yields an adjacency matrix. We have compared directly a subset of the entries of the matrices (see section 3.2), that is “edge-metrics” or nodal metrics. Three different edge-specific metrics were used: betweenness centrality, eigenvector centrality, and the degree. Betweenness centrality is a centrality measure that is equal to the number of the shortest paths passing through a given node. Another centrality measure is eigenvector centrality, which determines a node's relative importance within a network. Lastly, the degree of a node is the sum of the weights of the edges incident upon the node.

2.4 Classification Algorithms

To evaluate the discriminative ability of different feature sets (PLV, CC, ATM, AEC) and compare them with each other, we applied three different Machine Learning (ML) algorithms. Balanced accuracy was used as an evaluation metric, since we have imbalanced classes. ML algorithms include Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost). The general modelling workflow is summarized in Fig. 1

Linear Discriminant Analysis

LDA is a widely used approach for solving multi-class classification problems. The algorithm separates multiple classes (in our study - 4 classes) with multiple features through a data dimensionality reduction approach. LDA aims to find a hyperplane that best separates the classes while minimizing the overlap within each class. Related work has revealed that LDA performs well with multiclass diagnosis problems (Lin et al., 2021).

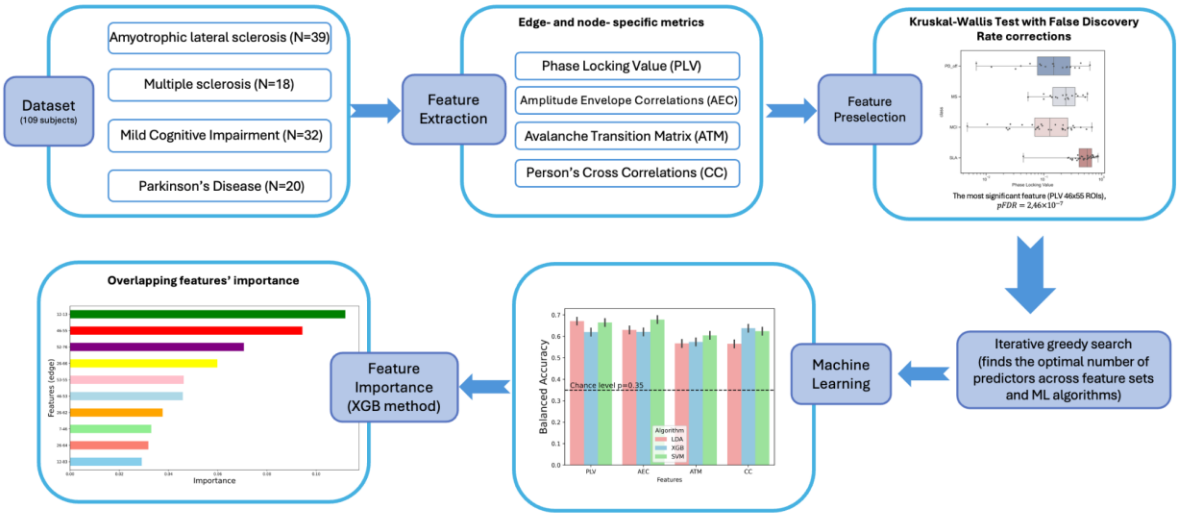


Figure 1 The general workflow of modeling

Support Vector Machines

SVM is another widely used technique for solving supervised tasks with multiple classes. Several studies identified SVM as an outstanding algorithm for solving tasks with multiple classes (Maqsood et al., 2022). SVM performs complex data transformations (according to the selected kernel function) and maximizes the separation boundaries between the data points depending on the classes.

Extreme Gradient Boosting

Recent studies showed that XGBoost is a state-of-the-art tree-based machine learning model that outperforms many other algorithms, including deep learning models (Grinsztajn et al., 2022). Moreover, the XGBoost algorithm provides an assessment of the relative importance of individual predictors, which allows us to interpret our findings (Manju et al., n.d.)
XGBoost is an ensemble method that builds a predictive model by combining predictions of multiple individual decision trees. It uses weak learner trees, these are decision trees with a single split, called decision stumps. The algorithm works by sequentially adding weak learners to the ensemble, with each new learner focusing on correcting the errors made by the previous one.
XGBoost is known for its high accuracy and has been shown to outperform other machine learning algorithms in many predictive modeling tasks. In addition, it is highly scalable and can handle large datasets.

2.5 Statistical Analysis

Kruskal-Wallis test

For each connectivity metric taken separately, to identify the most statistically significant different features among the four groups (PD, MCI, SLA, MS) to be considered for the

classification, we used the Kruskal-Wallis test. Since the brain is a non-linear dynamic system, we relied on a non-parametric statistical test, checking the null hypothesis that two or more independent groups were drawn from the same underlying distribution. The same approach was used for both edge and nodal metrics.

Multiple comparison correction

Since we have numerous features to be considered for a given FC metric, we used the false discovery rate to correct for inflated significance. The False Discovery Rate (FDR) is used to control the expected proportion of false positives. The FDR is the expected ratio of the number of false positive classifications, or false “discoveries”, to the total number of positive classifications (rejections of the null hypothesis). The p-values of the Kruskal-Wallis test were corrected accordingly. Finally, we sorted the features according to the corrected p-values in ascending order.

Spearman Correlation

We used Spearman correlation to evaluate the correlation between the features’ ranks. Spearman’s rank correlation coefficient is a non-parametric measure of statistical dependence between two variables. This way, we evaluated the relation between the ranks of the nodal or edge features across different FC metrics (PLV, AEC, ATM).

Repeated Stratified K-fold splits

To get valid results and avoid overfitting, we applied Repeated Stratified K-fold cross-validation, which repeats k-folds n times with different randomization for each repetition (J.-H. Kim, 2009). Then, for each fold, we have pooled our results across multiple randomization.
First, our whole dataset was split into two parts. For the first part, we use Stratified K-folds cross-validation to tune hyperparameters and find an optimal set that gives the best result. After tuning the hyperparameters on the first part of the dataset, then we used Stratified K-fold cross-validation 10 times for the second part. Then, accuracies obtained by each set are averaged. This way, we prevent data leakage, and it helps to get a more robust estimation of the accuracy by averaging over all repetitions and all folds. We used 10 repetitions of 10-fold cross-validations, and therefore we ensure that our evaluation is not affected by the specific choice of the validation set.

2.7 Evaluation metrics

Balanced accuracy

We used balanced accuracy as an evaluation metric for the classification algorithms since our dataset is imbalanced (NALS=39, NMCI=32, NPD=20, NMS=18) . The balanced accuracy is calculated by taking the average of the recalls obtained in each class (Thölke et al., 2023).

Recall

Recall is an evaluation metric that measures how often a classification algorithm correctly identifies positive instances among all the actual positive samples in the dataset.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Receiver Operating Characteristic (ROC) curve

ROC curve is a graph that displays the performance of a binary classification algorithm of predicting a positive class at all possible thresholds. The lower the classification threshold, the more observations are successfully classified. ROC curve uses False Positive Rate on the x-axis and True Positive Rate on its y-axis.
The area under the ROC curve (AUC) is an evaluation measure that measures the area underneath the ROC curve, and its maximum possible value equals one. In this manuscript, we compute the ROC curve for each class separately.

Confusion Matrix

A confusion matrix is an N x N matrix, where N is the number of classes. It has the true labels on the rows and the predicted labels on the second axis. This way, a confusion matrix shows how many times each class was classified correctly and also how often it was misclassified (and how).
We used a confusion matrix for 4 classes, therefore we have a 4 x 4 matrix, where TP_i represents the observations that were correctly classified for class i , and E_{ij} represents where true class j was misclassified with predicted class i. After that, we took the relative percentages across columns to see the whole picture in percentages, therefore each column's values will sum up to 100%. This is done by dividing each element of each column by the sum of all elements of that column and multiplying by 100. For example, for column 4 and its third element, it is done as follows:

$$\frac{E_{43}}{(E_{41} + E_{42} + E_{43} + TP_4)} \times 100$$

Code availability

The code used to perform the analysis of this study is publicly available at https://github.com/dklpp/multiclass_meg_features_analysis

2. Results

3.1 Kruskal-Wallis Test

Each adjacency matrix obtained from a given FC metric (namely PLV, AEC, ATM, or CC) is a square matrix with the dimension of $n_{regions} \times n_{regions}$, where $n_{regions}$is equal to 116 regions of interest. All matrices are symmetric and contain ones on the main diagonal., Hence, we take the triangular matrix, excluding the main diagonal elements, leading to 6670 edge-wise features. Given the high dimensionality of the feature space, we identified the most statistically significant different features among the four groups to be considered for the classification.

A non-parametric statistical Kruskal-Wallis test was performed for each feature to compare the four independent groups (PD, SLA, MS, MCI). After applying Kruskal-Wallis Test and False Discovery Rate correction, we found that there were more than 120 statistically significant edge features (pFDR < 0.002) for each of the 4 edge-specific FC metrics. The lowest corrected with FDR p-value p<0.0001 (pFDR = $2{,}46 \times 10^{-7}$) was obtained with the edge-wise PLV metric between the right frontal superior gyrus and the right postcentral gyrus (see Fig 2)

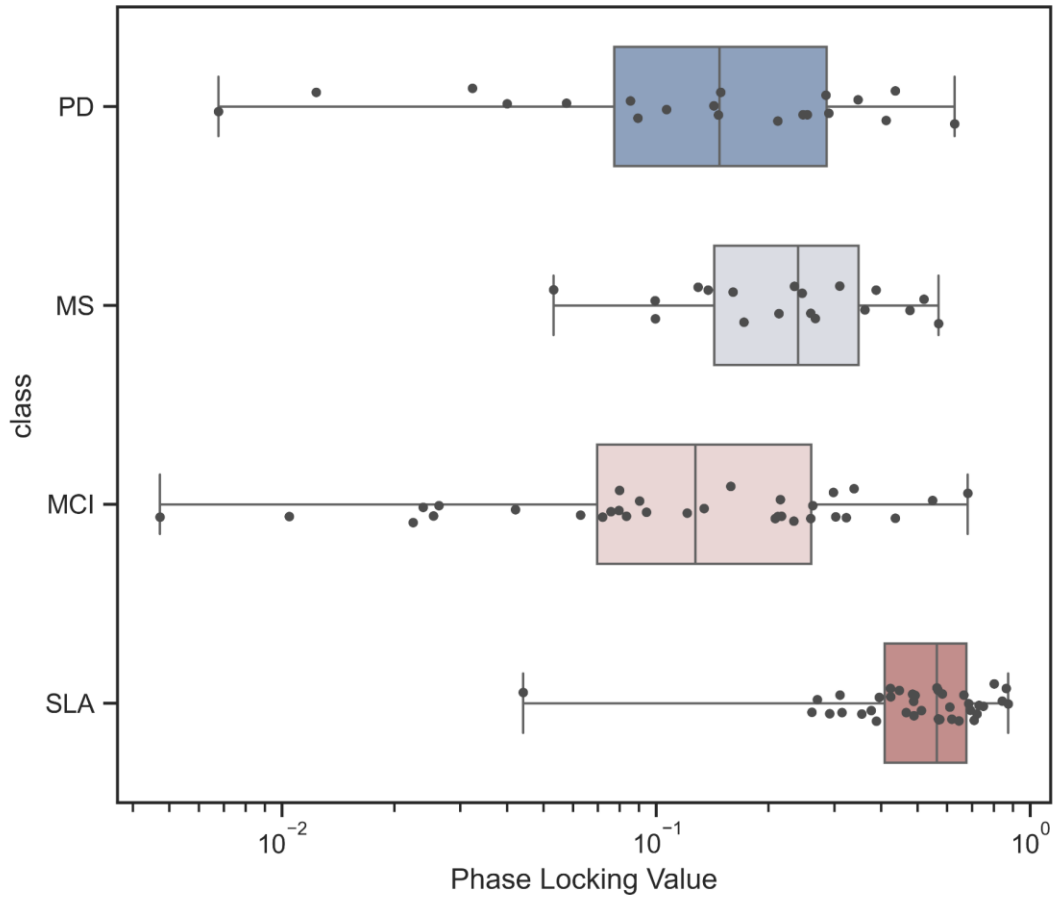


Figure 2 The most significant features' (PLV values between the right frontal superior gyrus and the right postcentral gyrus) boxplots with the observations for 4 classes with FDR p-value (pFDR = 2,46 x 10^{-7})

3.2 Classification algorithms

Based on the significant edges, we have then classified the participants. We used a consecutive iterative search technique, starting with the 15 best features (according to their corrected p-values), and sequentially added features and compared the accuracies, and this procedure was repeated until 39 features were fed to the classifier (Table 2). Stability had been reached at this point, and further increasing the number of features led to a slight worsening of the performance (not shown). Furthermore, given the relatively small size of our sample, we kept a lower number of features to reduce overfitting.

Num pred / Metrics	**n=15**	**n=16**	**n=18**	**n=22**	**n=25**	**n=28**	**n=30**	**n=31**	**n=34**	**n=36**	**n=38**	**n=39**
Edge metrics												
PLV	0.643	0.629	0.639	0.611	0.647	0.653	0.639	0.631	**0.671**	0.651	0.627	0.635
AEC	0.568	0.588	0.584	0.605	0.611	0.614	0.613	0.610	0.581	0.604	**0.630**	0.608
ATM	**0.567**	0.554	0.520	0.514	0.470	0.515	0.529	0.500	0.504	0.512	0.514	0.517
CC	0.466	0.477	0.521	0.511	0.495	0.539	**0.565**	0.560	0.556	0.545	0.529	0.525
Nodal metrics												
AEC (eign. centr.)	0.481	0.478	0.491	0.468	0.468	0.459	0.434	**0.492**	0.454	0.446	0.418	0.399
AEC (betw. centr.)	0.358	0.345	0.350	**0.381**	0.362	0.312	0.329	0.334	0.314	0.333	0.341	0.348
AEC (degree)	0.293	0.286	0.296	0.311	**0.366**	0.313	0.321	0.315	0.337	0.330	0.356	0.348
PLV (degree)	**0.407**	0.389	0.376	0.345	0.377	0.347	0.321	0.312	0.311	0.302	0.323	0.317
PLV (betw. centr.)	0.436	**0.454**	0.443	0.392	0.401	0.384	0.371	0.384	0.373	0.372	0.345	0.363
PLV (eign. centr.)	0.350	0.354	0.346	0.384	0.414	0.408	0.387	0.378	0.436	**0.443**	0.435	0.432
ATM (eign. centr.)	0.479	0.475	**0.489**	0.475	0.460	0.454	0.438	0.480	0.431	0.427	0.409	0.403
ATM (betw. centr.)	0.358	0.350	0.350	**0.381**	0.362	0.312	0.329	0.334	0.314	0.333	0.341	0.349
ATM (degree)	0.251	0.251	0.263	0.289	0.302	0.276	0.279	0.269	0.289	0.274	0.308	**0.317**
CC (betw. centr.)	0.506	**0.532**	0.512	0.530	0.523	0.486	0.468	0.447	0.452	0.468	0.452	0.453
CC (degree)	0.396	0.389	0.396	0.447	0.446	**0.478**	0.455	0.443	0.421	0.399	0.421	0.417

Table 2:*Balanced accuracy for the number of features which contain the best accuracy across different metrics (PLV, AEC, ATM, CC). For visual purposes, it is demonstrated only with the LDA algorithm. The accuracies obtained with the XGBoost and the SVM algorithms are in the supplementary material (see S4 and S5).*

We decided to display the balanced accuracies for each FC metric taken separately (both edge-based and node-based metrics) to see a clearer picture of different sets' performances (Fig 3).

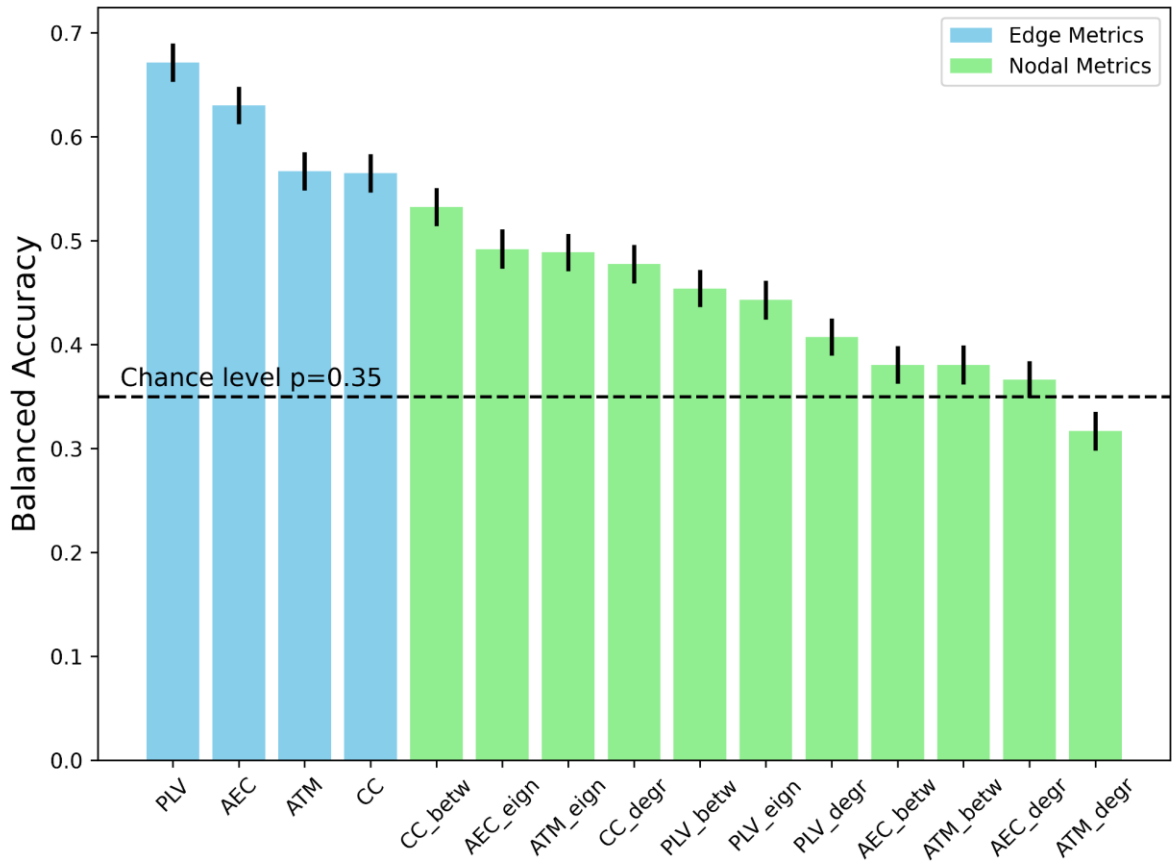


Figure 3 The balanced accuracies for all feature sets with LDA classifier. Each bar plot displays the averaged accuracy with its standard errors. The boxplots for other algorithms are available in the supplementary material (Fig S1-S2).

As shown in Fig3, we observed that all edge metrics consistently outperformed nodal metrics. Consistently, we observed that the standard deviations (over different repetitions of the K-folds) were higher for nodal metrics. Note that the results refer to the best-performing feature selection (i.e. the number of features is not fixed across different metrics).

Finally, we identified the optimal number of features (i.e. the features that showed the lowest corrected p-values and which led to the highest balanced accuracy) for each of the 3 different Machine Learning classification algorithms considered (namely XGBoost, SVM, LDA)(Fig. 3). The exhaustive search algorithm yielded the feature sets (which nodes/ edges) with the best-balanced accuracies for each algorithm across different FC metrics.

For the sake of simplicity, we discuss here the two best-performing FC metrics per classification algorithm (Fig. 4). In the case of the SVM algorithm, the AEC showed a balanced accuracy of 67.8% with a total of 31 top features, the PLV presented a balanced accuracy of 66.5 % with 36 top features. With the XBoost classifier, the CC showed a balanced accuracy of 63.8% with 35 top features), and the PLV presented a balanced accuracy of 62.8%, with 15 top features. Finally, in the case of the LDA classifier, the PLV showed a balanced accuracy

of 67.1%, with 34 top features and the AEC presented a performance of 63.0% with 38 top features.

Given the lower performance obtained with the nodal metrics, we shall proceed with the analyses exclusively on the edges.

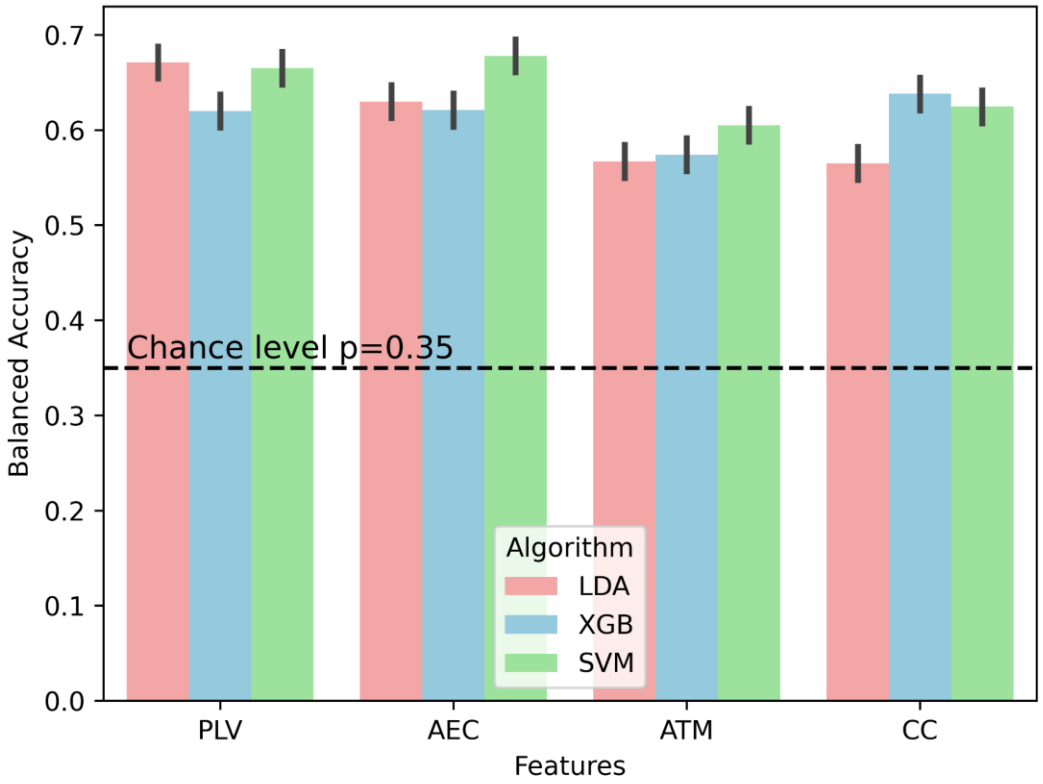


Figure 4 Balanced accuracies for different metrics across 3 Machine Learning algorithms with its standard errors. We observe consistent higher performance of PLV and AEC, in comparison to ATM and CC. The chance level for our dataset equals 35%.

The chance level for the balanced problem with 4 classes is equal to 25%. However, since we have a dataset with unbalanced classes, in such tasks the chance level is usually assumed to be the probability of predicting the most frequent class label in the target.In our case, SLA, which contains 38 patients out of 108, is the most numerous class. Therefore, the chance level is calculated as follows:

$$p = 38 / 108 \simeq 0{,}35 \; or \; 35\% \qquad (1)$$

In this task, a more objective evaluation metric is the balanced accuracy. Nevertheless, it is also useful to compare and evaluate overall accuracies. The trend remains the same – i.e. the same edge-specific metrics stay as the top features sets. Still, the accuracies are slightly higher: AEC (73.3%, 28 features), PLV (72.7%, 36 features), ATM (67.5%, 32 features), CC (69.1%, 36 features) for SVM classifier; AEC (68.2%, 38 features), PLV (69.5%, 34 features), ATM (63.2%, 15 features), CC (60.4%, 30 features) for LDA classifier; AEC (68.5%, 26 features), PLV (68.7%, 15 features), ATM (64.2%, 39 features), CC (70.5%, 35 features) for XGBoost classifier.

Since the PLV is the most performant metric, we now focus on the PLV for sensitivity analyses.

3.3 ROC curves

After repetitive stratified K-folds, we can estimate probabilities for each class to be correctly predicted (i.e. the probabilities sum up to 1) with different classification algorithms (LDA, SVM, XGB). We applied the One-vs-All technique, where we fix one desired class and all other classes are treated as one class. This way, we can replace our multi classification task to a binary class, and it enables us to build the ROC curves.
We calculated the average ROC curve for each repetition of the 10 folds, and the red curve displays the overall average ROC curve across 10 repetitions. The ROC curves were built for each class separately (Fig 5) which display the trade-off between False Positive Rate on x axis, and True Positive Rate on y axis.
ALS patients display the best results in terms of classification accuracy, and PD patients the worst results. Accordingly, it is worth mentioning that deviations of the ROC curves for PD patients are also much higher in comparison to other classes.

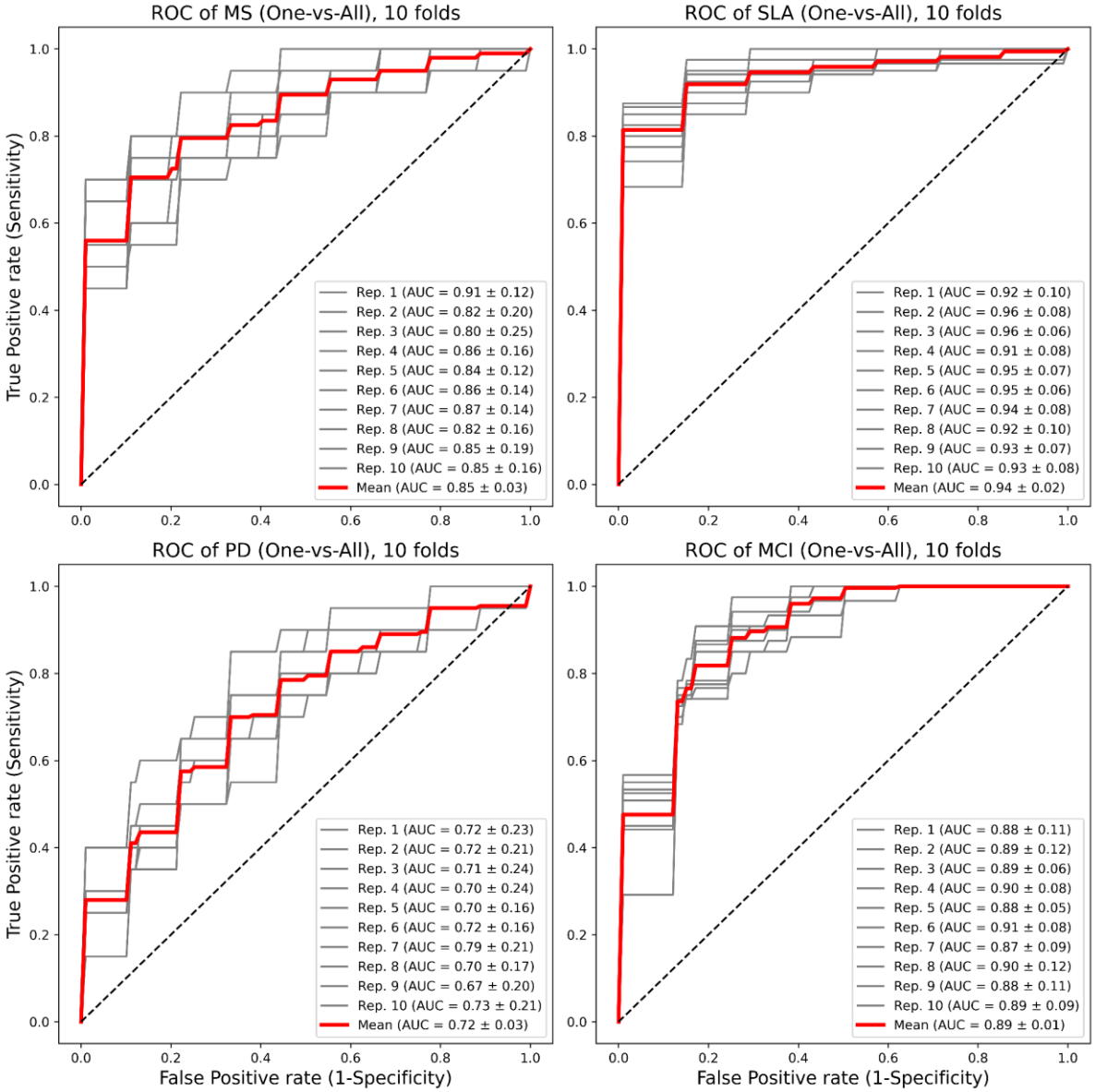


Figure 5 ROC curves for 4 classes (MS, SLA, PD, MCI) for LDA machine learning classifier with overall mean curve and mean curves for each repetition of Stratified K-folds**.** ROC curves are built with PLV edge-based features with 10 repetitions over 10 k-folds.

3.4 Confusion Matrix

The confusion matrix was built to depict the whole picture of the classifier's performance: it allows seeing what percentage of each class was classified correctly, and where mistakes were made (Fig 6).

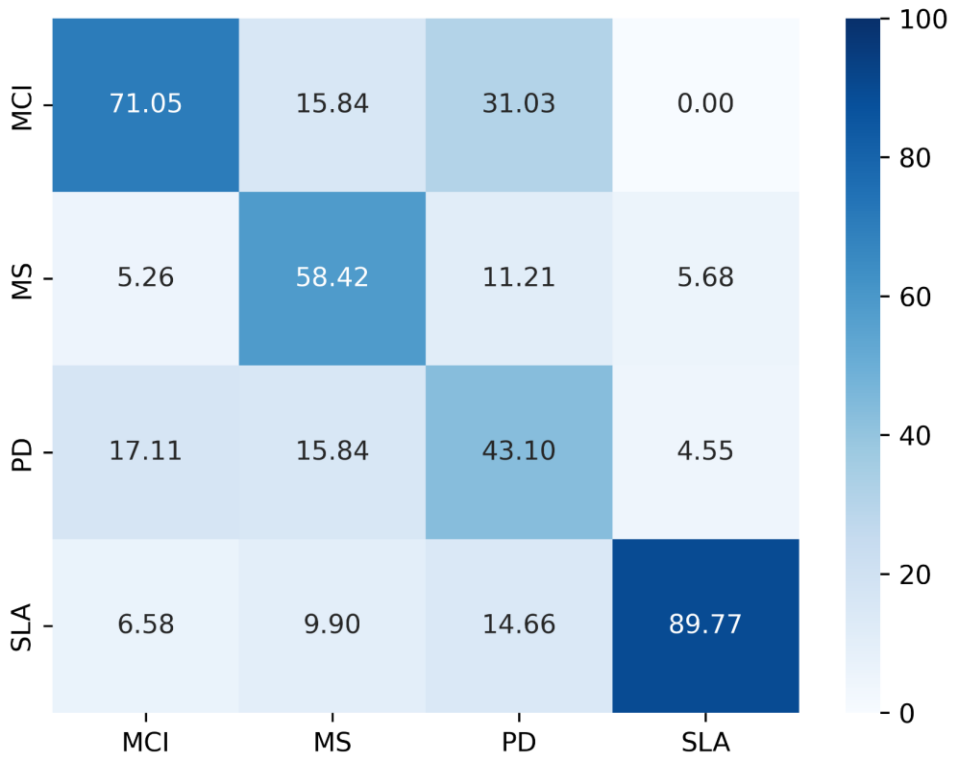


Figure 6 Confusion Matrix with relative percentage representation for 4 classes (SLA, PD, MS, MCI) with true values on y axis and predicted values on x axis for PLV features with LDA machine learning classifier.

Again, one can observe that the accuracy for PD patients is the worst, while the results for ALS subjects are the best. While these results might be affected by imbalanced classes, the results for all classes are well above chance level (which is equal to 35%).

3.5 Feature importance

The XGBoost classification algorithm allows us to quantify and compare the relative importance of the features during the classification process. To get valid results, we have defined 10 cross-validations with stratified KFolds. That is, the balance of the classes in the train and test splits are preserved. We run 40 times these stratified cross-validation iterations to reach convergence and then validate our results. After these steps, we obtained the feature importance for each of the features obtained from each FC metric taken separately (we focus here on the PLV, AEC, and ATM). Then, we evaluated and compared the features that showed the highest importance values for each of the sets considered (Fig. 7A)

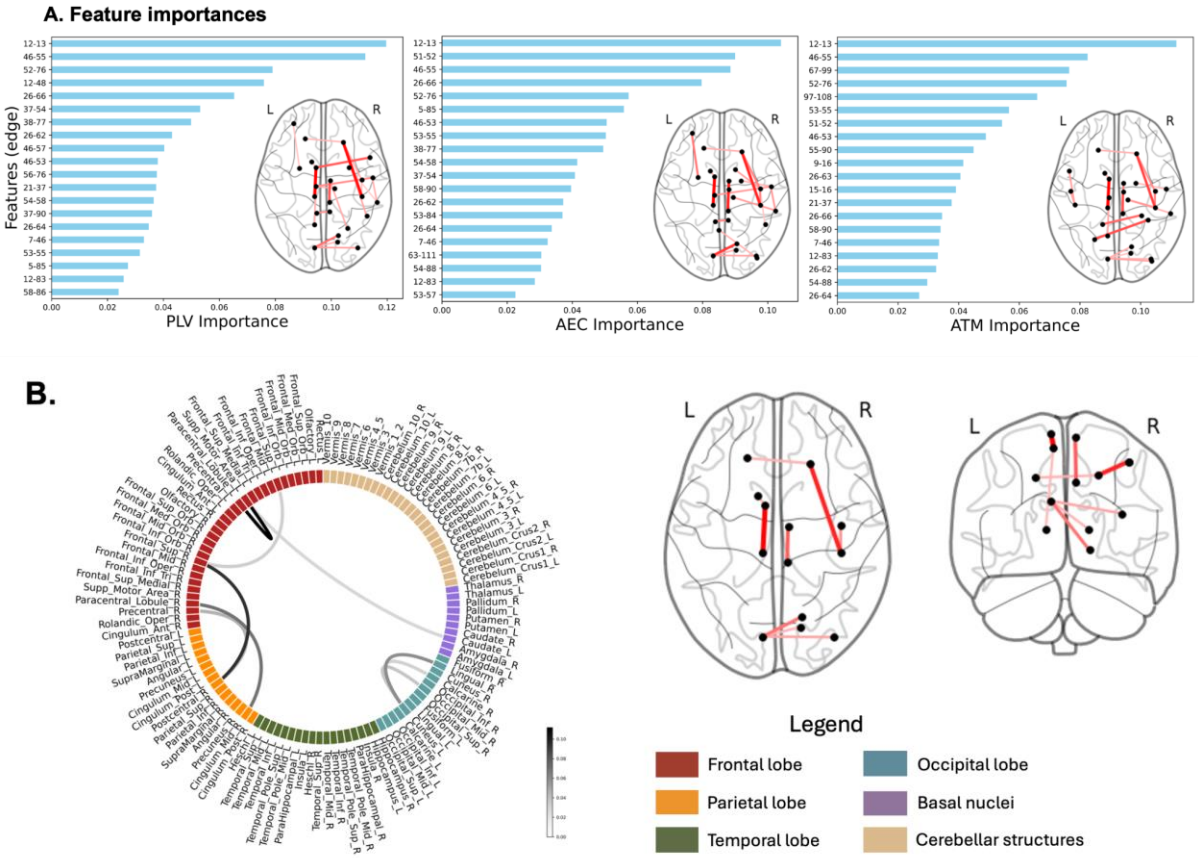


Figure 7. A. Feature importance for XGBoost across FC metrics along with these features on the brain plots. B. Connectome of overlapping features and its brain plot. The list of the edges is reported in the supplementary materials (see S6-S7-S8).

There are 9 overlapping features (i.e. edges) across 3 FC metrics (PLV, ATM, AEC), where we focused on the first 20 features with the largest feature importance according to XGBoost evaluation.

As an example of the consistency of our results across different metrics, the edge between the left supplementary motor area and the left paracentral lobule is the most important in all 3 feature sets. At the same time, the edge between the right frontal superior gyrus and right post central gyrus is also among the top 3 features across three feature sets.

Another interesting observation is that the left cuneus appears three times among the top pairs of edges (pairs of edges left cuneus and right lingual, left cuneus and Occipital middle gyrus, the left cuneus and the right calcarine cortex), meanwhile the left supplementary motor area, the right frontal superior gyrus and the right postcentral gyrus appear twice (for the exhaustive list of the region of interest see the supplementary material S3) .

We selected 20 of the most significant features, according to their corrected pFDR values, and ranked them across feature sets (where rank 1 means the most significant, and rank 20 indicates the least significant feature). This way, we identified that edge features for the next pairs of ROIs: left supplementary motor area and left paracentral lobule, right frontal superior gyrus and right postcentral gyrus, right paracentral lobule and right middle cingulum are ranked as the top features (ranks 1 and 2) for three edge-specific FC metrics (AEC, PLV, ATM), therefore we see a strong overlap of features across metrics.

We noticed that 9 edges out of 20 selected were the same for 3 different metrics – AEC, PLV and ATM (Fig 7B)

To systematically test these findings, we applied pairwise Spearman correlation for the ranks of three metrics (Fig 8). This revealed that, indeed, PLV and AEC features' ranks are highly positively correlated with a correlation coefficient equal to 0.89.

We observed that AEC, PLV and ATM have 9 overlapping features out of the 20 best features according to the p-values, including the Cross Correlations leaves only one overlapping feature.

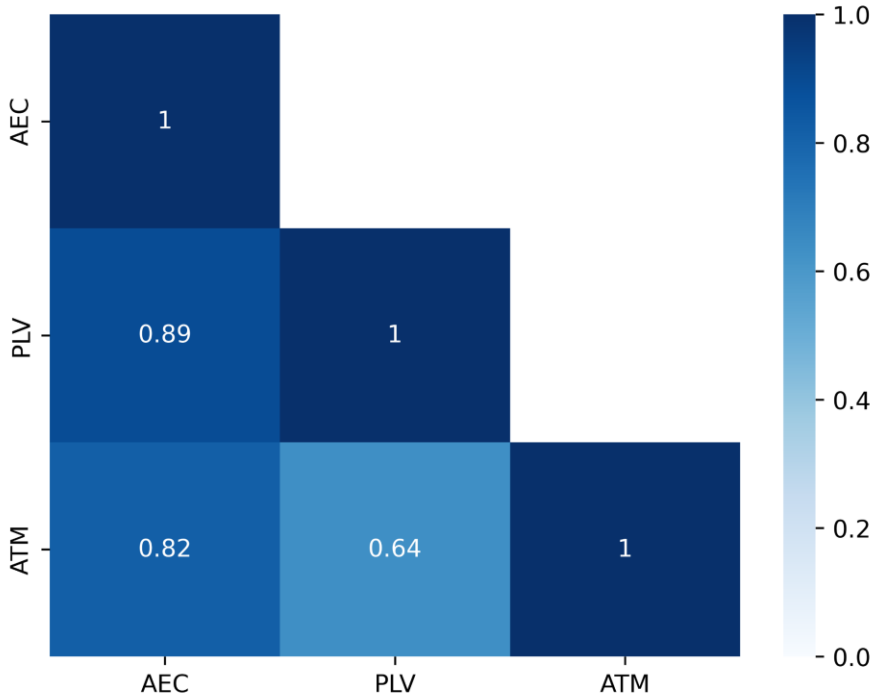


Figure 8. Feature importance for XGBoost across metrics

3. Discussion

In this study, we set out to identify a set of functional biomarkers to perform automated differential diagnosis (among MS, MCI, PD, and ALS) from MEG data. Our work focuses on the interpretability of the biomarkers, which is why we compared multiple connectivity metrics (AEC, PLV, Pearson's correlation coefficient, and ATM) that are different in terms of interpretation. We tested the robustness of our analyses by feeding the data features to multiple classification algorithms (i.e., XGBoost, SVM, LDA). In particular, we used a vast amount of MEG data from a total number of 109 subjects affected by four different neurological diseases: ALS, PD, MCI, and MS. Firstly, from each cohort, we extracted different feature sets from four different FC metrics (PLV, ATM, AEC, and CC), each of which was obtained starting from a symmetric matrix, leading to a total number of 6670 edge-wise features. Due to the high dimensionality of our sample, we performed a Kruskal-Wallis test to reduce the dimensionality and to consider only the most discriminative features.

Firstly, our results showed more than 120 significant edge features among the four different feature sets (PLV, AEC, ATM, CC) for all the diseases. In particular, the Kruskal-Wallis test showed statistical significance ($pFDR<0.0001$) of edge-wise PLV values between the right frontal superior gyrus and the right postcentral gyrus. This finding might be related to the fact that the frontal lobe is involved in physiological processes related to motor function (which are notably impaired in SLA, PD, and MS) as well as to cognitive function, such as long-term memory (which is often impaired in MCI).

We then move on to estimate the accuracy of the selected features by adding the features in an iterative manner according to their p values. Such an approach enabled us to determine the optimal number of top features for each FC metric taken separately. The AEC reached a balanced accuracy of 63.07% with 38 features added, while the edge-wise PLV displayed the best-balanced accuracy (67,14%) with a total number of 34 added features. To our knowledge, no previous research has combined MEG data from patients with MCI, MS, PD, and ALS. However, some research has focused on these conditions individually. As an example, López ME et al. (López et al., 2014a) examined 105 subjects (36 controls and 69 MCI cases). They identified spectral bio-marked changes in the theta, alpha, and beta frequency bands in MEG data in MCI. Kim MJ et al. (M.-J. Kim et al., 2023) utilized a large EEG dataset including 417 MCI cases and applied a neural network that detected dementia with 81.1% accuracy. In the closely related work, Giovannetti A. et al. (Giovannetti et al., 2021) presented the Deep-MEG neural network, which was tested on 54 Alzheimer's patients, each undergoing a five-minute resting state task. Similarly to our study, they used functional connectivity indices, phase locking values for classification. They reported an 87.4% AUC-ROC in identifying early MCI symptoms. For multiple sclerosis disease, using EEG, Kiiski H. et al. (Kiiski et al., 2018) assessed the responses of 35 subjects with multiple sclerosis during event-related potential cognitive tasks over three years. They found significant correlations between ERP visual components and cognitive function, identified using machine learning techniques. In related research, Karaca et al. employed a continuous wavelet transform to differentiate nine multiple sclerosis patients from 11 controls, achieving accuracy rates between 80%-88% in their best-performing models (Karaca et al., 2021). Furthermore, Ahmadi A. et al. analyzed five MS patients, developing a detection model using phase locking values and an online sequential

595 extreme learning classifier, with performance scores ranging from 82% to 96% across different
596 tasks (Ahmadi et al., 2019).
597
598 Our results are consistent across the three different ML algorithms (see Fig 3) with PLV and
599 AEC showing the best accuracy with respect to ATM and CC. These results are in agreement
600 with Chaturvedi et al., who showed that features extracted from the Phase lagIndex (PLI) were
601 able to better discriminate between PD patients with and without MCI as compared to spectral
602 features (Chaturvedi et al., 2019). These results might suggest that phase-based metrics
603 might be more suitable in classification performance analysis as compared to amplitude-based
604 metrics (i.e., power spectra). Phase-based metrics specifically capture synchronization among
605 brain signals (defined as a bounded average phase difference). Synchronization is typically
606 measured in the framework of the communication-through-coherence hypothesis (Fries,
607 2015), whereby communication among brain regions might be captured by the coherent
608 activities of the corresponding brain signals. On the other hand, non-periodic activities, as well
609 as simple correlation coefficients, seem to perform less well in this context. Regardless of the
610 chosen FC metric, we see a consistent trend where the estimates at the edge level outperform
611 those at the nodal level in terms of disease classification. On the one hand, nodal metrics
612 capture the local activities and are predominantly sensitive to the dynamics of the local
613 activations. On the other hand, edge metrics focus primarily on how brain regions interact
614 among themselves. Therefore, our results might be interpreted as evidence that
615 neurodegenerative diseases primarily alter how regions interact with each other at the large-
616 scale level. In particular, since the PLV is the best-performing metric, this might be interpreted
617 as the neurodegenerative diseases altering the ability of brain regions that are far apart to
618 synchronize their activities.
619 While the PLV is sensitive to volume conduction, volume conduction does not offer a
620 reasonable explanation for the ability to classify different subjects according to diagnosis.
621 Furthermore, the edges of the ATM (that are more robust to volume conduction artifacts) also
622 confirm the ability to correctly diagnose patients well above chance level.
623 Furthermore, the set of edges that contributed more to the classification were FC metric-
624 independent. In general, it is interesting to note that the edges that are relevant to classification
625 irrespective of the metrics are typically longer range connections, either in the antero-posterior
626 direction of cross-hemispheric. Again, these results are representative of significant
627 involvement and, as a consequence, impairment, of the frontal lobe in PD, ALS, MS and MCI
628 respectively (Foong et al., 1997; Kendi et al., 2008; Trojsi et al., 2012; Wang et al., 2012).
629
630 Not many studies explain the rationale behind the feature extraction and selection method
631 choice. It usually consists in a trade-off between enriching the information of interest and the
632 risk of adding irrelevant inputs that could reduce the classification performance. Two types of
633 approaches have been proposed. The first one consists in considering that fusing features will
634 result in an improvement of the classification performance. For instance, Geraedts et al fused
635 features obtained from the estimation of the power spectra in seven frequency bands (resulting
636 in 16 674 features per EEG) before selecting them to discriminate cognitive functions in
637 patients with Parkinson's Disease during Deep Brain Stimulation (Geraedts et al., 2021).
638 Similarly, López et al extracted spectral and non-linear metrics before fusing them and
639 selecting them via their fast correlation-based filter to discriminate early Alzheimer's disease
640 and its prodromal form from healthy subjects (López et al., 2014b).

Another approach consists in fusing the classifiers' output rather than the different types of features. Fusing the classifiers' outputs confers a higher reliability and robustness through redundancy and facilitates the integration of heterogeneous data without normalizing them (Roli, 2009; Roli & Fumera, 2002; Ruta & Gabrys, 2000). In a recent work, we proposed a framework that was based on Riemannian geometry extended to functional connectivity measures through an ensemble learning method. We validated it on numerous publicly available datasets (Corsi et al., 2022) . Such an approach notably ranked 1st in a clinical challenge that consisted in discriminating mental states from data obtained from stroke patients (Corsi et al., 2021). Future work will consist in considering this type of approach to enrich the information of interest used to discriminate diseases.

In conclusion, our is the first study investigating automated differential diagnosis in several neurological diseases, based on different connectivity metrics as well as on different classification algorithms. Our results demonstrate the existence of a common set of edges that drive the classification performance, irrespective of the particular metric chosen or the algorithm. These results demonstrate the existence of a robust set of long-range connections that are altered in neurodegeneration, across multiple diseases, and valid in terms of distinguishing specific diseases. Future studies will have to confirm the external validity of our results to different datasets and extend our analyses to more neurodegenerative diseases.

Data availability statement

The magnetoencephalography data and the reconstructed avalanches are available upon request to the corresponding author, conditional on appropriate ethics approval at the local site. The availability of the data was not previously included in the ethical approval, and therefore data cannot be shared directly. In case data are requested, the corresponding author will request an amendment to the local ethical committee.

Competing interests

The authors report no competing interests

Funding

This work was financially supported by Ministero Sviluppo Economico (Contratto di sviluppo industriale "Farmaceutica e Diagnostica" [CDS 000606]); European Union "NextGenerationEU," (Investimento 3.1. M4. C2), project IR0000011, EBRAINS-Italy of PNRR and Contratto di sviluppo industriale- "Progetto CDS000904 - agevolazioni ex DM del 09/12/2014"

References

Ahmadi, A., Davoudi, S., & Daliri, M. R. (2019). Computer Aided Diagnosis System for multiple sclerosis disease based on phase to amplitude coupling in covert visual attention. *Computer Methods and Programs in Biomedicine*, *169*, 9–18. https://doi.org/10.1016/j.cmpb.2018.11.006

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

Bastos, A. M., & Schoffelen, J.-M. (2016). A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00175

Brookes, M. J., Woolrich, M., Luckhoo, H., Price, D., Hale, J. R., Stephenson, M. C., Barnes, G. R., Smith, S. M., & Morris, P. G. (2011). Investigating the electrophysiological basis of resting state networks using magnetoencephalography. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(40), 16783–16788. https://doi.org/10.1073/pnas.1112685108

Brookes, M. J., Woolrich, M. W., & Barnes, G. R. (2012). *Measuring functional connectivity in MEG: A multivariate approach insensitive to linear source leakage*. https://doi.org/10.1016/j.neuroimage.2012.03.048

Brooks, B. R. (1994). El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and th. *Journal of the Neurological Sciences*, *124*, 96–107.

704 Chaturvedi, M., Bogaarts, J. G., Kozak (Cozac), V. V., Hatz, F., Gschwandtner, U., Meyer,
705 A., Fuhr, P., & Roth, V. (2019). Phase lag index and spectral power as QEEG
706 features for identification of patients with mild cognitive impairment in Parkinson's
707 disease. *Clinical Neurophysiology*, *130*(10), 1937–1944.
708 https://doi.org/10.1016/j.clinph.2019.07.017
709 Cipriano, L., Minino, R., Liparoti, M., Polverino, A., Romano, A., Bonavita, S., Pirozzi, M. A.,
710 Quarantelli, M., Jirsa, V., Sorrentino, G., Sorrentino, P., & Troisi Lopez, E. (2024).
711 Flexibility of brain dynamics is increased and predicts clinical impairment in
712 relapsing–remitting but not in secondary progressive multiple sclerosis. *Brain*
713 *Communications*, *6*(2), fcae112. https://doi.org/10.1093/braincomms/fcae112
714 Corsi, M.-C., Chavez, M., Schwartz, D., George, N., Hugueville, L., Kahn, A. E., Dupont, S.,
715 Bassett, D. S., & De Vico Fallani, F. (2020). Functional disconnection of associative
716 cortical areas predicts performance during BCI training. *NeuroImage*, *209*, 116500.
717 https://doi.org/10.1016/j.neuroimage.2019.116500
718 Corsi, M.-C., Chevallier, S., de Vico Fallani, F., & Yger, F. (2022). Functional connectivity
719 ensemble method to enhance BCI performance (FUCONE). *IEEE Transactions on*
720 *Biomedical Engineering*, 1–1. https://doi.org/10.1109/TBME.2022.3154885
721 Corsi, M.-C., Yger, F., Chevallier, S., & Noûs, C. (2021). Riemannian Geometry on
722 Connectivity for Clinical BCI. *ICASSP 2021 - 2021 IEEE International Conference on*
723 *Acoustics, Speech and Signal Processing (ICASSP)*, 980–984.
724 https://doi.org/10.1109/ICASSP39728.2021.9414790
725 Finn, E. S., & Rosenberg, M. D. (2021). Beyond fingerprinting: Choosing predictive
726 connectomes over reliable connectomes. *NeuroImage*, *239*, 118254.
727 https://doi.org/10.1016/j.neuroimage.2021.118254
728 Foong, J., Rozewicz, L., Quaghebeur, G., Davie, C. A., Kartsounis, L. D., Thompson, A. J.,
729 Miller, D. H., & Ron, M. A. (1997). Executive function in multiple sclerosis. The role of
730 frontal lobe pathology. *Brain*, *120*(1), 15–26. https://doi.org/10.1093/brain/120.1.15
731 Fries, P. (2015). Rhythms For Cognition: Communication Through Coherence HHS Public

Access. *Neuron*, *88*(1), 220–235. https://doi.org/10.1016/j.neuron.2015.09.034

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping*, *2*(1–2), 56–78. https://doi.org/10.1002/hbm.460020107

Geraedts, V. J., Koch, M., Contarino, M. F., Middelkoop, H. A. M., Wang, H., van Hilten, J. J., Bäck, T. H. W., & Tannemaat, M. R. (2021). Machine learning for automated EEG-based biomarkers of cognitive impairment during Deep Brain Stimulation screening in patients with Parkinson's Disease. *Clinical Neurophysiology*, *132*(5), 1041–1048. https://doi.org/10.1016/j.clinph.2021.01.021

Gibb, W. R. G., & Lees, A. J. (1988). A comparison of clinical and pathological features of young- and old-onset Parkinson's disease. *Neurology*, *38*(9), 1402–1402. https://doi.org/10.1212/WNL.38.9.1402

Giovannetti, A., Susi, G., Casti, P., Mencattini, A., Pusil, S., López, M. E., Di Natale, C., & Martinelli, E. (2021). Deep-MEG: Spatiotemporal CNN features and multiband ensemble classification for predicting the early signs of Alzheimer's disease with magnetoencephalography. *Neural Computing and Applications*, *33*(21), 14651–14667. https://doi.org/10.1007/s00521-021-06105-4

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). *Why do tree-based models still outperform deep learning on tabular data?* (arXiv:2207.08815). arXiv. https://doi.org/10.48550/arXiv.2207.08815

Haldeman, C., & Beggs, J. M. (2005). Critical Branching Captures Activity in Living Neural Networks and Maximizes the Number of Metastable States. *Physical Review Letters*, *94*(5), 058101. https://doi.org/10.1103/PhysRevLett.94.058101

Karaca, B. K., Akşahin, M. F., & Öcal, R. (2021). Detection of multiple sclerosis from photic stimulation EEG signals. *Biomedical Signal Processing and Control*, *67*, 102571. https://doi.org/10.1016/j.bspc.2021.102571

Kelly, R. E., & Hoptman, M. J. (2022). Replicability in Brain Imaging. *Brain Sciences*, *12*(3), 397. https://doi.org/10.3390/brainsci12030397

Kendi, A. T. K., Lehericy, S., Luciana, M., Ugurbil, K., & Tuite, P. (2008). Altered Diffusion in

760 the Frontal Lobe in Parkinson Disease. *American Journal of Neuroradiology*, *29*(3),
761 501–505. https://doi.org/10.3174/ajnr.A0850
762 Kiiski, H., Jollans, L., Donnchadha, S. Ó., Nolan, H., Lonergan, R., Kelly, S., O'Brien, M. C.,
763 Kinsella, K., Bramham, J., Burke, T., Hutchinson, M., Tubridy, N., Reilly, R. B., &
764 Whelan, R. (2018). Machine Learning EEG to Predict Cognitive Functioning and
765 Processing Speed Over a 2-Year Period in Multiple Sclerosis Patients and Controls.
766 *Brain Topography*, *31*(3), 346–363. https://doi.org/10.1007/s10548-018-0620-4
767 Kim, J.-H. (2009). Estimating classification error rate: Repeated cross-validation, repeated
768 hold-out and bootstrap. *Computational Statistics & Data Analysis*, *53*(11), 3735–
769 3745. https://doi.org/10.1016/j.csda.2009.04.009
770 Kim, M.-J., Youn, Y. C., & Paik, J. (2023). Deep learning-based EEG analysis to classify
771 normal, mild cognitive impairment, and dementia: Algorithms and dataset.
772 *NeuroImage*, *272*, 120054. https://doi.org/10.1016/j.neuroimage.2023.120054
773 Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase
774 synchrony in brain signals. *Human Brain Mapping*, *8*(4), 194–208.
775 https://doi.org/10.1002/(sici)1097-0193(1999)8:4<194::aid-hbm4>3.0.co;2-c
776 Lin, W., Gao, Q., Du, M., Chen, W., & Tong, T. (2021). Multiclass diagnosis of stages of
777 Alzheimer's disease using linear discriminant analysis scoring for multimodal data.
778 *Computers in Biology and Medicine*, *134*, 104478.
779 https://doi.org/10.1016/j.compbiomed.2021.104478
780 López, M. E., Cuesta, P., Garcés, P., Castellanos, P. N., Aurtenetxe, S., Bajo, R., Marcos,
781 A., Delgado, M. L., Montejo, P., López-Pantoja, J. L., Maestú, F., & Fernandez, A.
782 (2014a). MEG spectral analysis in subtypes of mild cognitive impairment. *Age*
783 *(Dordrecht, Netherlands)*, *36*(3), 9624. https://doi.org/10.1007/s11357-014-9624-5
784 López, M. E., Cuesta, P., Garcés, P., Castellanos, P. N., Aurtenetxe, S., Bajo, R., Marcos,
785 A., Delgado, M. L., Montejo, P., López-Pantoja, J. L., Maestú, F., & Fernandez, A.
786 (2014b). MEG spectral analysis in subtypes of mild cognitive impairment. *Age*
787 *(Dordrecht, Netherlands)*, *36*(3), 9624. https://doi.org/10.1007/s11357-014-9624-5

788 Manju, N., Harish, B. S., & Prajwal, V. (n.d.). Ensemble Feature Selection and Classification
789 of Internet Traffic using XGBoost Classifier. *International Journal of Computer*
790 *Network and Information Security*, *11*(7), 37.
791 Maqsood, S., Damaševičius, R., & Maskeliūnas, R. (2022). Multi-Modal Brain Tumor
792 Detection Using Deep Neural Network and Multiclass SVM. *Medicina (Kaunas,*
793 *Lithuania)*, *58*(8), 1090. https://doi.org/10.3390/medicina58081090
794 Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its
795 use for magnetoencephalography forward calculation in realistic volume conductors.
796 *Physics in Medicine and Biology*, *48*(22), 3637–3652. https://doi.org/10.1088/0031-
797 9155/48/22/002
798 Polverino, A., Lopez, E. T., Minino, R., Liparoti, M., Romano, A., Trojsi, F., Lucidi, F., Gollo,
799 L., Jirsa, V., Sorrentino, G., & Sorrentino, P. (2022). Flexibility of Fast Brain
800 Dynamics and Disease Severity in Amyotrophic Lateral Sclerosis. *Neurology*.
801 https://doi.org/10.1212/WNL.0000000000201200
802 Polverino, A., Troisi Lopez, E., Liparoti, M., Minino, R., Romano, A., Cipriano, L., Trojsi, F.,
803 Jirsa, V., Sorrentino, G., & Sorrentino, P. (2024). Altered spreading of fast aperiodic
804 brain waves relates to disease duration in Amyotrophic Lateral Sclerosis. *Clinical*
805 *Neurophysiology*, *163*, 14–21. https://doi.org/10.1016/j.clinph.2024.04.003
806 Roli, F. (2009). Multiple Classifier Systems. In S. Z. Li & A. Jain (Eds.), *Encyclopedia of*
807 *Biometrics* (pp. 981–986). Springer US.
808 http://link.springer.com/referenceworkentry/10.1007/978-0-387-73003-5_148
809 Roli, F., & Fumera, G. (2002). Analysis of Linear and Order Statistics Combiners for Fusion
810 of Imbalanced Classifiers. *Multiple Classifier Systems*, 252–261.
811 https://doi.org/10.1007/3-540-45428-4_25
812 Romano, A., Troisi Lopez, E., Cipriano, L., Liparoti, M., Minino, R., Polverino, A., Cavaliere,
813 C., Aiello, M., Granata, C., Sorrentino, G., & Sorrentino, P. (2023). Topological
814 changes of fast large-scale brain dynamics in mild cognitive impairment predict early
815 memory impairment: A resting-state, source reconstructed,

magnetoencephalography study. *Neurobiology of Aging*, *132*, 36–46. https://doi.org/10.1016/j.neurobiolaging.2023.08.003

Romano, A., Trosi Lopez, E., Liparoti, M., Polverino, A., Minino, R., Trojsi, F., Bonavita, S., Mandolesi, L., Granata, C., Amico, E., Sorrentino, G., & Sorrentino, P. (2022). The progressive loss of brain network fingerprints in Amyotrophic Lateral Sclerosis predicts clinical impairment. *NeuroImage: Clinical*, *35*, 103095. https://doi.org/10.1016/j.nicl.2022.103095

Ruta, D., & Gabrys, B. (2000). An Overview of Classifier Fusion Methods. *Computing and Information Systems*, *7*(1), 1–10.

Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., Moodie, C. A., & Poldrack, R. A. (2016). The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron*, *92*(2), 544–554. https://doi.org/10.1016/j.neuron.2016.09.018

Shriki, O., Alstott, J., Carver, F., Holroyd, T., Henson, R. N. A., Smith, M. L., Coppola, R., Bullmore, E., & Plenz, D. (2013). Neuronal avalanches in the resting MEG of the human brain. *Journal of Neuroscience*, *33*(16), 7079–7090. https://doi.org/10.1523/JNEUROSCI.4286-12.2013

Sorrentino, P., Petkoski, S., Sparaco, M., Troisi Lopez, E., Signoriello, E., Baselice, F., Bonavita, S., Pirozzi, M. A., Quarantelli, M., Sorrentino, G., & Jirsa, V. (2022). Whole-brain propagation delays in multiple sclerosis, a combined tractography—Magnetoencephalography study. *The Journal of Neuroscience*, JN-RM-0938-22. https://doi.org/10.1523/JNEUROSCI.0938-22.2022

Sorrentino, P., Rucco, R., Baselice, F., Micco, R. D., Tessitore, A., Hillebrand, A., Mandolesi, L., Breakspear, M., Gollo, L. L., & Sorrentino, G. (2019). Extensive functional repertoire underpins complex behaviours: Insights from Parkinson's disease. In *bioRxiv* (p. 823849). Cold Spring Harbor Laboratory. https://doi.org/10.1101/823849

Sorrentino, P., Rucco, R., Jacini, F., Trojsi, F., Lardone, A., Baselice, F., Femiano, C., Santangelo, G., Granata, C., Vettoliere, A., Monsurrò, M. R., Tedeschi, G., &

Sorrentino, G. (2018). Brain functional networks become more connected as amyotrophic lateral sclerosis progresses: A source level magnetoencephalographic study. *NeuroImage: Clinical*, *20*, 564–571. https://doi.org/10.1016/j.nicl.2018.08.001

Sorrentino, P., Rucco, R., Lardone, A., Liparoti, M., Lopez, E. T., Cavaliere, C., Soricelli, A., Jirsa, V., Sorrentino, G., & Amico, E. (n.d.). *Clinical connectome fingerprints of cognitive decline*. https://doi.org/10.1101/2020.10.09.332635

Sorrentino, P., Seguin, C., Rucco, R., Liparoti, M., Troisi Lopez, E., Bonavita, S., Quarantelli, M., Sorrentino, G., Jirsa, V., & Zalesky, A. (2021). The structural connectome constrains fast brain dynamics. *eLife*, *10*, e67400. https://doi.org/10.7554/eLife.67400

Stam, C. J. (2010). Use of magnetoencephalography (MEG) to study functional brain networks in neurodegenerative disorders. *Journal of the Neurological Sciences*, *289*(1–2), 128–134. https://doi.org/10.1016/j.jns.2009.08.028

Tagliazucchi, E., von Wegner, F., Morzelewski, A., Brodbeck, V., & Laufs, H. (2012). Dynamic BOLD functional connectivity in humans and its electrophysiological correlates. *Frontiers in Human Neuroscience*, *6*(DEC). https://doi.org/10.3389/fnhum.2012.00339

Thölke, P., Mantilla-Ramos, Y.-J., Abdelhedi, H., Maschke, C., Dehgan, A., Harel, Y., Kemtur, A., Mekki Berrada, L., Sahraoui, M., Young, T., Bellemare Pépin, A., El Khantour, C., Landry, M., Pascarella, A., Hadid, V., Combrisson, E., O'Byrne, J., & Jerbi, K. (2023). Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage*, *277*, 120253. https://doi.org/10.1016/j.neuroimage.2023.120253

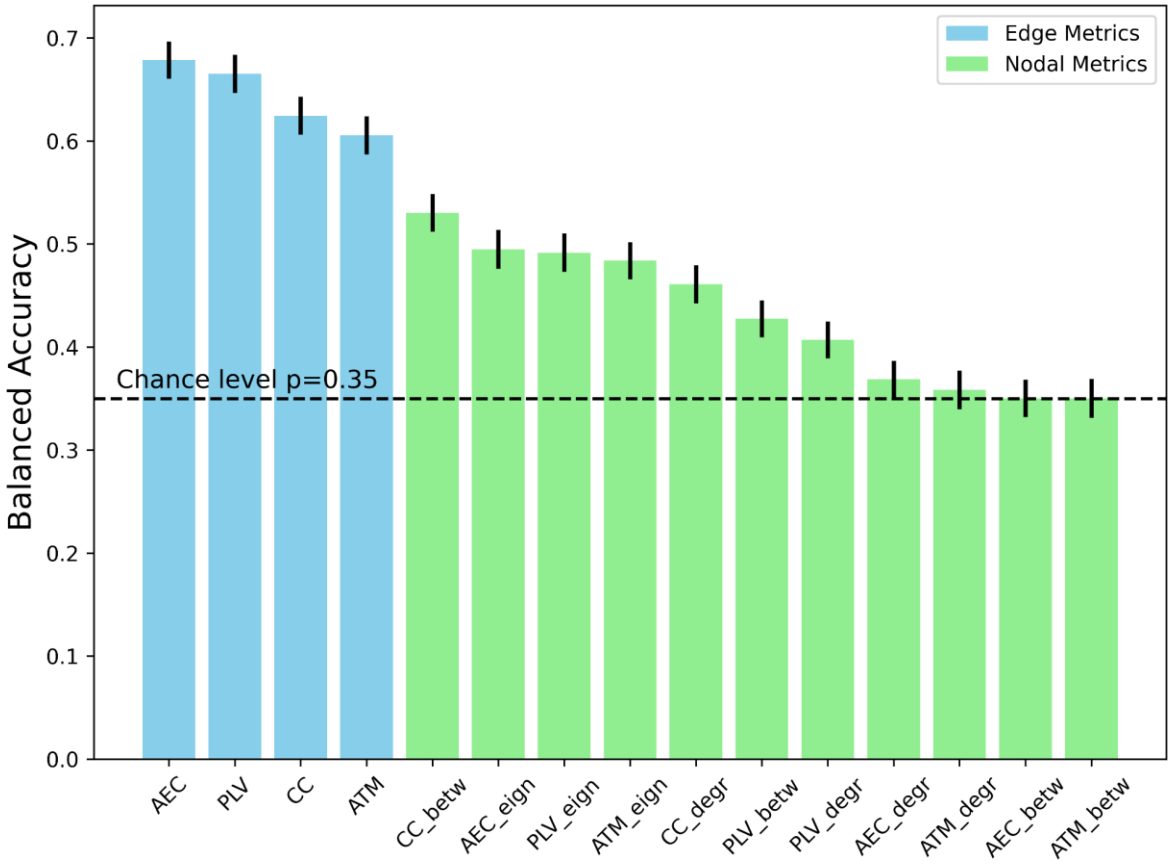
Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., … Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, *17*(2), 162–173. https://doi.org/10.1016/S1474-4422(17)30470-2

Trojsi, F., Monsurro, M. R., Esposito, F., & Tedeschi, G. (2012). Widespread structural and functional connectivity changes in amyotrophic lateral sclerosis: Insights from advanced neuroimaging research. *Neural Plasticity*, *2012*.

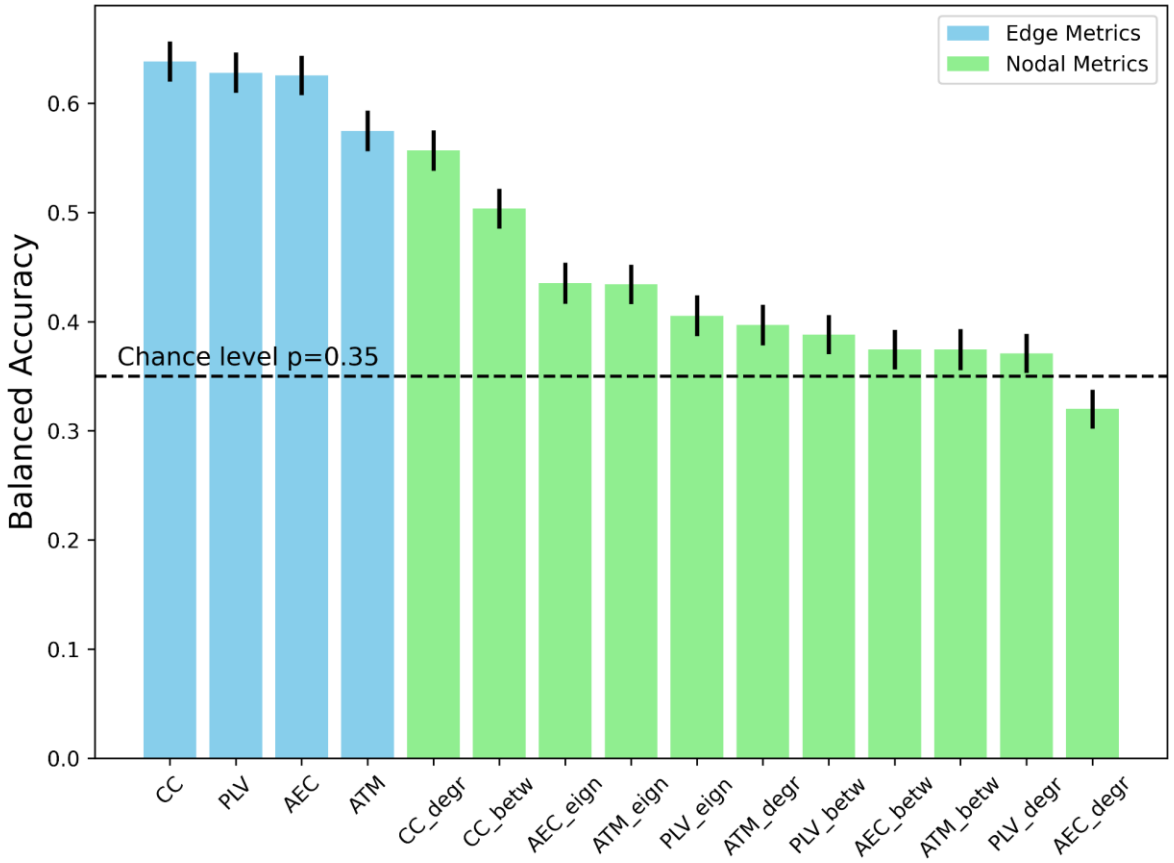
Wang, Z., Jia, X., Liang, P., Qi, Z., Yang, Y., Zhou, W., & Li, K. (2012). Changes in thalamus connectivity in mild cognitive impairment: Evidence from resting state fMRI. *European Journal of Radiology*, *81*(2), 277–285. https://doi.org/10.1016/j.ejrad.2010.12.044

Supplementary materials

Supplementary materials 1: The balanced accuracies for all feature sets with SVM classifier. Each bar plot displays the averaged accuracy with its standard errors.



Supplementary materials 2: The balanced accuracies for all feature sets with XGBoost classifier. Each bar plot displays the averaged accuracy with its standard errors.



Supplementary materials 3: Exhaustive list of regions of interest

N° ROIS	Anatomical correspondance	N° ROIS	Anatomical correspondance	N° ROIS	Anatomical correspondance
1	Rectus L	40	Rectus R	79	Hippocampus L
2	Olfactory L	41	Olfactory R	80	Hippocampus R
3	Frontal Superior Orbital L	42	Frontal Superior Orbital R	81	Amygdala L
4	Frontal Medial Orbital L	43	Frontal Medial Orbital R	82	Amygdala R
5	Frontal Medial Orbital L	44	Frontal Medial Orbital R	83	Caudate L
6	Frontal Inferior Orbital L	45	Frontal Inferior Orbital R	84	Caudate R
7	Frontal Superior L	46	Frontal Superior R	85	Putamen L
8	Frontal Medial L	47	Frontal Medial R	86	Putamen R
9	Frontal Inferior Opercolum L	48	Frontal Inferior Opercolum R	87	Pallidum L
10	Frontal Inferior Triangular L	49	Frontal Inferior Triangular R	88	Pallidum R
11	Frontal Superior Medial L	50	Frontal Superior Medial R	89	Thalamus L
12	Supplementary Motor area L	51	Supplementary Motor area R	90	Thalamus R
13	Paracentral Lobule L	52	Paracentral Lobule R	91	Cerebellum Crus1 L
14	Precentral L	53	Precentral R	92	Cerebelum Crus1 R
15	Rolandic Opercolum L	54	Rolandic Opercolum R	93	Cerebelum Crus2 L
16	Postcentral L	55	Postcentral R	94	Cerebelum Crus2 R

17	Parietal Superior L	56	Parietal Superior R	95	Cerebellum 3 L
18	Parietal Inferior L	57	Parietal Inferior R	96	Cerebellum 3 R
19	Supra Marginal L	58	Supra Marginal R	97	Cerebellum 4 5 L
20	Angular L	59	Angular R	98	Cerebellum 4 5 R
21	Precuneus L	60	Precuneus R	99	Cerebellum 6 L
22	Occipital Superior L	61	Occipital Superior R	100	Cerebellum 6 R
23	Occipital Medial L	62	Occipital Medial R	101	Cerebellum 7b L
24	Occipital Inferior L	63	Occipital Inferior R	102	Cerebellum 7b R
25	Calcarine L	64	Calcarine R	103	Cerebellum 8 L
26	Cuneus L	65	Cuneus R	104	Cerebellum 8 R
27	Lingual gyrus L	66	Lingual gyrus R	105	Cerebellum 9 L
28	Fusiform gyrus L	67	Fusiform gyrus R	106	Cerebellum 9 R
29	Heschl L	68	Heschl R	107	Cerebellum 10 L
30	Temporal Superior L	69	Temporal Superior R	108	Cerebellum 10 R
31	Temporal Medial L	70	Temporal Medial R	109	Vermis 1 2
32	Temporal Inferior L	71	Temporal Inferior R	110	Vermis 3
33	Temporal Pole Superior L	72	Temporal Pole Superior R	111	Vermis 4 5
34	Temporal Pole Medial L	73	Temporal Pole Medial R	112	Vermis 6
35	ParaHippocampal L	74	ParaHippocampal R	113	Vermis 7
36	Cingulum Anterior L	75	Cingulum Anterior R	114	Vermis 8
37	Cingulum Medial L	76	Cingulum Medial R	115	Vermis 9
38	Cingulum Posterior L	77	Cingulum Posterior R	116	Vermis 10
39	Insula L	78	Insula R		

Supplementary materials 4: Balanced accuracy for the number of features which contain the best accuracy across different metrics (PLV, AEC, ATM, CC) with XGBoost.

	n=15	n=17	n=18	n=19	n=24	n=25	n=26	n=27	n=28	n=35	n=38	n=39
Edge metrics												
PLV	0.628	0.606	0.597	0.618	0.575	0.577	0.573	0.563	0.572	0.544	0.559	0.557
AEC	0.554	0.606	0.597	0.601	0.594	0.611	0.626	0.622	0.617	0.602	0.602	0.622
ATM	0.513	0.533	0.533	0.551	0.55	0.535	0.524	0.519	0.531	0.561	0.554	0.575
CC	0.529	0.565	0.559	0.574	0.591	0.586	0.595	0.594	0.591	0.638	0.612	0.608
Nodal metrics												
CC (betw.)	0.496	0.489	0.504	0.483	0.468	0.482	0.48	0.467	0.476	0.47	0.459	0.448
PLV (betw.)	0.379	0.385	0.388	0.384	0.376	0.364	0.357	0.356	0.374	0.354	0.341	0.341
PLV (eign.)	0.378	0.372	0.372	0.373	0.378	0.376	0.401	0.405	0.402	0.383	0.399	0.4
AEC (eign.)	0.414	0.421	0.435	0.404	0.401	0.407	0.401	0.403	0.413	0.405	0.399	0.403
AEC (betw.)	0.374	0.359	0.358	0.356	0.334	0.323	0.327	0.321	0.332	0.342	0.35	0.352

ATM (betw.)	0.374	0.359	0.358	0.356	0.334	0.323	0.327	0.321	0.332	0.342	0.35	0.352
ATM (eign.)	0.422	0.427	0.423	0.415	0.424	0.414	0.418	0.419	0.434	0.41	0.408	0.412
PLV (degree)	0.333	0.343	0.328	0.327	0.341	0.371	0.363	0.364	0.348	0.361	0.335	0.339
AEC (degree)	0.298	0.308	0.289	0.289	0.297	0.299	0.31	0.298	0.32	0.284	0.27	0.262
ATM (degree)	0.373	0.377	0.372	0.378	0.397	0.376	0.368	0.375	0.376	0.346	0.373	0.368
CC (degree)	0.537	0.557	0.548	0.539	0.522	0.529	0.535	0.533	0.534	0.531	0.528	0.544

Supplementary materials 5: Balanced accuracy for the number of features which contain the best accuracy across different metrics (PLV, AEC, ATM, CC) with SVM.

	n=16	**n=19**	**n=20**	**n=21**	**n=22**	**n=25**	**n=28**	**n=30**	**n=31**	**n=32**	**n=34**	**n=35**	**n=36**	**n=37**	**n=39**
Edge metrics															
PLV	0.614	0.64	0.635	0.636	0.628	0.634	0.633	0.601	0.631	0.64	0.651	0.607	0.665	0.648	0.61
AEC	0.541	0.574	0.611	0.601	0.613	0.648	0.652	0.661	0.679	0.665	0.665	0.644	0.632	0.654	0.647
ATM	0.536	0.544	0.541	0.535	0.532	0.454	0.544	0.586	0.597	0.605	0.602	0.596	0.597	0.599	0.58
CC	0.505	0.483	0.543	0.537	0.501	0.558	0.577	0.575	0.568	0.584	0.6	0.624	0.621	0.625	0.559
Nodal metrics															
AEC (eign. centr.)	0.486	0.493	0.495	0.489	0.49	0.472	0.459	0.459	0.459	0.461	0.461	0.455	0.454	0.454	0.439
AEC (betw. centr.)	0.293	0.288	0.293	0.299	0.304	0.317	0.25	0.346	0.342	0.342	0.35	0.341	0.341	0.334	0.34
CC (betw. centr.)	0.487	0.48	0.497	0.49	0.512	0.495	0.53	0.453	0.492	0.496	0.486	0.479	0.495	0.474	0.457
PLV (betw. centr.)	0.412	0.427	0.401	0.394	0.378	0.361	0.33	0.358	0.344	0.344	0.363	0.354	0.349	0.328	0.369
PLV (eign. centr.)	0.397	0.426	0.432	0.422	0.458	0.467	0.458	0.462	0.487	0.488	0.454	0.436	0.402	0.423	0.492
ATM (betw. centr.)	0.293	0.288	0.293	0.299	0.304	0.317	0.339	0.346	0.342	0.342	0.35	0.341	0.341	0.334	0.34
ATM (Eign. centr.)	0.475	0.475	0.483	0.484	0.483	0.454	0.463	0.454	0.456	0.459	0.448	0.449	0.442	0.435	0.423
PLV (degree)	0.391	0.362	0.38	0.4	0.407	0.389	0.387	0.379	0.377	0.376	0.406	0.398	0.343	0.394	0.353
AEC (degree)	0.365	0.343	0.304	0.35	0.337	0.369	0.327	0.326	0.324	0.327	0.323	0.325	0.322	0.334	0.322
ATM (degree)	0.31	0.315	0.358	0.341	0.327	0.31	0.292	0.293	0.315	0.292	0.331	0.321	0.302	0.327	0.314
CC (degree)	0.461	0.414	0.359	0.404	0.391	0.379	0.419	0.388	0.4	0.394	0.408	0.382	0.389	0.379	0.37

Supplementary materials 6: Features importance. List of associated edges for PLV.

12-13	Supplementary Motor area L and Paracentral Lobule L
46-55	Frontal Superior R and Postcentral R
52-76	Paracentral Lobule R and Cingulum Medial R
12-48	Supplementary Motor area L and Frontal Inferior Opercolum R
26-66	Cuneus L and Lingual gyrus R
37-54	Cingulum Medial L and Rolandic Opercolum R
38-77	Cingulum Posterior L and Cingulum Posterior R
26-62	Cuneus L and Occipital Medial R
46-57	Frontal Superior R and Parietal Inferior R
46-53	Frontal Superior R and Precentral R
56-76	Parietal Superior R and Cingulum Medial R
21-37	Precuneus L and Cingulum Medial L
54-58	Rolandic Opercolum R and Supra Marginal R
37-90	Cingulum Medial and L Thalamus R
26-64	Cuneus L and Calcarine R
7-46	Frontal Superior L and Frontal Superior R
53-55	Precentral R and Postcentral R
5-85	Frontal Medial Orbital L and Putamen L
12-83	Supplementary Motor area L and Caudate L
58-86	Supra Marginal R and Putamen R

Supplementary materials 7: Features importance. List of associated edges for AEC.

12-13	Supplementary Motor area L and Paracentral Lobule L
51-52	Paracentral Lobule R and Paracentral Lobule R
46-55	Frontal Superior R and Postcentral R
26-66	Cuneus L and Lingual gyrus R
52-76	Paracentral Lobule R and Cingulum Medial R
5-85	Frontal Medial Orbital L and Putamen L
46-53	Frontal Superior R and Precentral R
53-55	Precentral R and Postcentral R
38-77	Cingulum Posterior L and Cingulum Posterior R
54-58	Rolandic Opercolum R and Supra Marginal R
37-54	Cingulum Medial L and Rolandic Opercolum R
58-90	Supra Marginal R and Thalamus R
26-62	Cuneus L and Occipital Medial R
53-84	Precentral R and Caudate R
26-64	Cuneus L and Calcarine R
7-46	Frontal Superior L and Frontal Superior R
63-111	Occipital Inferior R and Vermis 4 5
54-88	Rolandic Opercolum R Pallidum R
13-83	Paracentral Lobule L and Caudate L
53-57	Precentral R and Parietal Inferior R

Supplementary materials 8: Features importance. List of associated edges for ATM.

12-13	Supplementary Motor area L and Paracentral Lobule L
46-55	Frontal Superior R and Postcentral R
67-99	Fusiform gyrus R and Cerebelum 6 L
52-76	Paracentral Lobule R and Cingulum Medial R
97-108	Cerebelum 4 5 L and Cerebelum 10 R
53-55	Precentral R and Postcentral R
51-52	Paracentral Lobule R and Paracentral Lobule R
46-53	Frontal Superior R and Precentral R
55-90	Postcentral R and Thalamus R
9-16	Frontal Inferior Opercolum L and Postcentral L
26-63	Cuneus L and Occipital Inferior R
15-16	Rolandic Opercolum L and Postcentral L
21-37	Precuneus L and Cingulum Medial L
26-66	Cuneus L and Lingual gyrus R
58-90	Supra Marginal R and Thalamus R
7-46	Frontal Superior L and Frontal Superior R

12-83	Supplementary Motor area L and Caudate L
26-62	Cuneus L and Occipital Medial R
54-88	Rolandic Opercolum R Pallidum R
26-64	Cuneus L and Calcarine R