Multimodal-neuroimaging machine-learning analysis of motor disability in multiple sclerosis

Barbora Rehák Bučková[0000−0001−5619−3946]1,2,3, Ján
Mareš[0000−0002−4727−3685]3,4, Antonín
Škoch[0000−0002−1739−3256]3,4, Jakub
Kopal[0000−0002−1201−2872]2, Jaroslav
Tintěra[0000−0002−0685−7174]3,4, Robert
Dineen[0000−0002−9523−2546]5,6, Kamila
Řasová[0000−0002−3201−1770]7 and Jaroslav
Hlinka[0000−0003−1402−1470]2,3

[1]The Czech Technical University in Prague, Karlovo namesti 13,
121 35, Prague, The Czech Republic.
[2]Institute of Computer Science of the Czech Academy of Sciences,
Pod Vodarenskou vezi 2/271, 182 00, Prague, The Czech Republic.
[3]National Institute of Mental Health, Topolova 748, Klecany, 250
67, The Czech Republic.
[4]Institute for Clinical and Experimental Medicine, Videnska
1958, Prague, 140 21, The Czech Republic.
[5]University of Nottingham, Queen's Medical Centre, Nottingham,
NG7 2UH, United Kingdom.
[6]National Institute for Health Research, Nottingham Biomedical
Research Centre, Nottingham, NG1 5DU, United Kingdom.
[7]Charles University, Ruska 87, Prague, 100 00, The Czech
Republic.

Corresponding author: hlinka@cs.cas.cz; +420 266 05 3808;

Word count: 3342;

Abstract

Motor disability is a dominant and restricting symptom in multiple sclerosis, yet its neuroimaging correlates are not fully understood. We apply statistical and machine learning techniques on multimodal neuroimaging data to discriminate between multiple sclerosis patients and healthy controls and to predict motor disability scores in the patients. We examine the data of sixty-four multiple sclerosis patients and sixty-five controls, who underwent the MRI examination and the evaluation of motor disability scales. The modalities used comprised regional fractional anisotropy, regional grey matter volumes, and functional connectivity. For analysis, we employ two approaches: high-dimensional support vector machines run on features selected by Fisher Score (aiming for maximal classification accuracy), and low-dimensional logistic regression on the principal components of data (aiming for increased interpretability). We apply analogous regression methods to predict symptom severity. While fractional anisotropy provides the classification accuracy of 96.1% and 89.9% with both approaches respectively, including other modalities did not bring further improvement. Concerning the prediction of motor impairment, the low-dimensional approach performed more reliably. The first grey matter volume component was significantly correlated (R = 0.28-0.46, p < 0.05) with most clinical scales. In summary, we identified the relationship between both white and grey matter changes and motor impairment in multiple sclerosis. Furthermore, we were able to achieve the highest classification accuracy based on quantitative MRI measures of tissue integrity between patients and controls yet reported, while also providing a low-dimensional classification approach with comparable results, paving the way to interpretable machine learning models of brain changes in multiple sclerosis.

Keywords: Multiple Sclerosis, Machine Learning, Multimodal Analysis, Prediction, MRI

1 Introduction

Multiple sclerosis is an autoimmune disease of the central nervous system. With an increase in the age-standardized prevalence, the latest epidemiological studies report 2.2 million cases worldwide, ranking it as the most common demyelinating disease (Leray et al., 2016; Wallin et al., 2019).

Multiple sclerosis presents itself by a wide and heterogeneous spectrum of symptoms, ranging from cognitive and visual impairments to motor disabilities (Dobson and Giovannoni, 2019). Of these, motor impairment is the most dominant and restricting. Gait and postural control changes are present already in recently diagnosed minimally impaired patients and worsen over time (Burschka et al., 2012; Martin et al., 2006). Reduction in mobility together with fatigue increase risk of fall and injury, endangering the patient and reducing self-confidence (Giannì et al., 2014). More than 40% of people with multiple

sclerosis report walking difficulties, which are identified as the most challenging disease aspect. However, the majority of people with multiple sclerosis do not voluntarily seek medical advice regarding walking impairments (Asch, 2011; Heesen et al., 2008; LaRocca, 2011).

Despite motor deficit being a major concern of people with multiple sclerosis, detailed motor function evaluation is rarely performed in practice. Consequently, the literature relating motor deficits to brain imaging in multiple sclerosis is limited, often focusing on single neuroimaging modality (Jakimovski et al., 2018; Sbardella et al., 2015; Steenwijk et al., 2014; Tona et al., 2014). Moreover, the studies generally feature a small sample size, low degree of impairment, and high number of statistical tests, weakening the findings reproducibility. These limitations call for methods that not only detect group differences in neuroimaging features, but also provide continuous quantitative prediction of the level of symptoms, and ultimately may prospectively characterize and predict health development in a clinical setting. Therefore, a methodological shift from pure hypothesis testing to predictive modeling (in statistical, or even forecasting sense) is needed, along with the increase of the sample sizes; machine learning is thus a natural candidate.

Machine learning has the potential to allow integration of multiparametric datasets with identification of feature sets that promise diagnostic or prognostic relevance for individual patients. Note that apart from promising improvement of the diagnostic and prognostic capabilities by providing accurate classification and prediction models, machine learning can also help getting insight into the disease character and mechanisms in case that the estimated models are reasonably interpretable. Thus, it has become more widespread in clinical research recently (Woo et al., 2017), but its implementation is not always straightforward. Particularly for limited sample sizes, care is needed to prevent model overfitting and sustain generalization to a wider population.

Progress has been made towards the application of machine learning methods in multiple sclerosis (Stafford et al., 2020). The most common task is the classification of subjects into patients and controls, sometimes also considering the degree of impairment and the phenotype of the disease (Bendfeldt et al., 2012; Eitel et al., 2019; Kocevar et al., 2016; Marzullo et al., 2019; Tozlu et al., 2021; Zurita et al., 2018). Most of the classification approaches use the support vector machines (SVM) (Vapnik and Chapelle, 2000), a broadly used algorithm in neuroscience for its ability to deal with the common high data dimensionality (especially in cases, when the number of features exceeds the number of instances), and for its computational effectiveness (learning time is considerably shorter than in neural networks). Although theoretically, no prior dimensionality reduction is needed, it is usually desirable. Previous studies commonly employed Fisher Score (FS) (He et al., 2006), which quantifies the power of variables to distinguish between the studied classes. However, the consequences of working only with a pre-selected set of features optimized for

predictive power is often poorly discussed and, if not done carefully (i.e., blind to the test dataset), may lead to overfitting.

In this work, we build imaging-based models that provide efficient yet interpretable biomarkers for motor impairment in multiple sclerosis. We first identify imaging biomarkers across imaging modalities using machine learning for the simple task of distinguishing people with multiple sclerosis from controls. We subsequently construct predictive models for more subtle differences in motor performance among people with multiple sclerosis.

2 Materials and Methods

2.1 The Study Design

Participants with multiple sclerosis were recruited from multiple sclerosis centers across the Czech Republic. The inclusion criteria were: positive diagnosis of multiple sclerosis (Polman et al., 2011); spastic paraparesis as a prominent clinical feature; stable clinical status for at least three months preceding the study (determined by neurologist); physical ability to undergo clinical testing – consistent with the Expanded Disability Status Scale (EDSS) score $\leq$ 7.5. Participants with disturbed mobility for reasons not related to multiple sclerosis (e.g. fractures, pregnancy, stroke) were excluded. All phenotypes of multiple sclerosis were accepted: RR – Relapsing-remitting; PP – Primary Progressive; and SP – Secondary Progressive. We analyzed 64 participants with multiple sclerosis and 65 healthy controls statistically matched for age and sex (Table 1). Participants were informed about the experimental setup and provided written informed consent in accordance with the Declaration of Helsinki. The study design was approved by the Ethics Committee of the Faculty Hospital Královské Vinohrady.

2.2 Clinical Assessment

Participants with multiple sclerosis underwent a set of standardised assessments with the Berg Balance Scale (BBS) (Berg et al., 2009; Podsiadlo and Richardson, 1991); Timed Up and Go (TUG) test, The Multiple Sclerosis Impact Scale (MSIS); and Multiple Sclerosis Walking Scale (MSWS) (Fischer et al., 1999; Hobart et al., 2001,0).

2.3 MRI data Acquisition

Imaging was performed with a 3T magnetic resonance scanner (Siemens Trio Tim, Erlangen, Germany) using a 12-channel phased-array head coil. The protocol consisted of T1-weighted and T2-weighted anatomical scans, DTI and resting-state fMRI. The parameters were: DWI: TR = 9,100 ms; TE = 96 ms; FOV = 260 x 211.25 mm; 64 contiguous axial slices; 2 mm thickness; b = 0 and 1100 s/mm; 64 gradient directions; voxel size = 2.03 x 2.03 x 2mm; Resting-State fMRI: BOLD single-shot echo-planar images TR = 2500 ms, TE = 30 ms, flip angle = 70°, 64×64 matrix, FOV = 192 mm^2, 44 contiguous

axial slices, 3 mm thick, 240 volumes, acquisition time = 10 min. T1 volumetric imaging: TR = 2300 ms, TE = 4.63 ms, flip angle = 10°, matrix = 256 × 256, FOV = 256 mm^2, 156 contiguous sagittal slices, 1 mm thick.

2.4 Image processing

DTI was preprocessed using the FSL tools (FMRIB Software Library v5.0, http://www.fmrib.ox.ac.uk/fslwiki, FMRIB, Oxford, UK) and MRtrix3 v3.0_rc3 (Tournier et al., 2019). The data were denoised (`dwidenoise`) (Veraart et al., 2016a,1), Gibbs ringing artefacts were corrected (`mrdegibbs`) (Kellner et al., 2016). The volumes with low quality (visually checked) were discarded. Subsequently, eddy-current induced distortions and movement displacement was corrected by the `eddy` tool. Finally, we generated FA maps (Basser et al., 1994; Westin et al., 1997) and applied TBSS (Smith et al., 2006). We parcellated the resulting skeletonized images using the white matter ICBM-DTI-81 atlas (Mori et al., 2005) containing 48 regions and computed the mean FA for each region, which resulted in 48 FA features per subject.

The fMRI data were preprocessed using a combination of the SPM12 software package (Wellcome Department of Cognitive Neurology, London, UK); CONN toolbox (McGovern Institute for Brain Research, MIT, USA) running under MATLAB (The Mathworks) and FSL routines. Due to near-gaussianity of fMRI data (Hartman et al., 2011; Hlinka et al., 2011), we quantified FC for each pair of AAL atlas regions (Tzourio-Mazoyer et al., 2002) by the Pearson correlation coefficient (116 regions provide 6,670 FC features per subject).

Segmentation of the T1 images into white matter, gray matter, and cerebrospinal fluid was a part of the CONN pipeline employed for fMRI processing. We computed the GM region volumes using the AAL atlas (116 GMV features per subject). Additionally, the T1 scans were processed using Voxel-Based Morphometry in CAT12 (Keller and Roberts, 2008), using default settings involving bias-field and noise correction, segmentation into gray and white matter, and normalization to MNI space using DARTEL algorithm to a 1.5 mm isotropic adult template (Yassa and Stark, 2009). Total intracranial volume was estimated to be used as a covariate in statistical analyses.

2.5 Statistical Analysis

2.5.1 Classification

For the classification task, we developed two strategies:

- Fisher-score as a dimensionality reduction step, classification by support vector machines with linear kernel (FS-SVM)
- Principal component analysis as a dimensionality reduction step, classification by logistic regression (PCA-LR)

The first approach uses the SVM classifier that can, in principle, deal with high-dimensional data, although it may be prone to overfitting, especially when it is combined with informed feature selection procedure such as the commonly

used Fisher-score (He et al., 2006). In this approach, the features are selected into the classifier based on their individual classification power on the data in the training set. The box constraint (parameter C) for SVM was fixed to 1.

The latter approach uses a classical method of linear regression that can only deal with a small number of input features. To obtain these, we use PCA of the original features, a standard dimension reduction approach that avoids overfitting by being blind to the patient/control labels. Unlike in the Fisher score method, the transformation of the data is based on their explained variance in general and is not influenced by class labels.

For each strategy, we constructed four classifiers: based on FA, FC, GMV, and their combination, using leave-one-out cross-validation (see Fig. 1). All neuroimaging features were transformed to z-scores prior to the dimensionality reduction step and model fitting.

To further assess how the classification accuracy depends on the number of features, we evaluated each classifier for a range of thresholds. For the FS-SVM combination, we included 1%, 10%, 25%, 50%, 75%, and 100% of the features with the top Fisher Score. For the low-dimensional PCA-LR, we included one to ten PCA components. We consider sensitivity, specificity, and overall accuracy as measures classification quality and compare the classifiers using the McNemar test. All analyses were conducted using Matlab2018b, The MathWorks, Natick, 2018.

2.5.2 Prediction of the motor impairment

We applied an analogous approach for the prediction of motor impairment. In particular, we replaced SVM by support vector regression (SVR) and LR by linear regression (LinR). For dimensionality reduction, we sorted the features according to the Spearman correlation with the predicted scale instead of using FS. The use of PCA based on the variability explained remained unchanged. In each iteration, we performed a dimensionality reduction step on N-1 patients, constructed a model and predicted the clinical scale score for the remaining patient. Model quality was assessed by the correlation between the prediction and the original scale. Prior to the regression analysis, Box-Cox normalization was applied to the clinical scales. To avoid outlier and nongaussianity effects, Spearman correlation was employed for the final model fit assessment.

3 Results

3.1 Classification

All neuroimaging modalities allowed successful classification (albeit not necessarily at the same degree of complexity), with the best modality-specific accuracy ranging from 75 to 96 percent (Fig. 2, Table S2, Table S3).

The SVM classifier based on FA performed the best; the baseline classifier with a single FA feature reached 85% accuracy, the accuracy for the full feature set was 96%. The accuracy generally grew with feature count, although did

not show significant performance differences beyond using 25% of the features (Fig. 2A, Fig. S1).

The SVM classifiers based on FC or GMV achieved about 70 percent accuracy and were not generally improving with feature count.

We also assessed the possibility to further improve the classification by combining the features across modalities. The baseline classifier based on the combination of features from all modalities (containing 69 features) performed better than the corresponding classifier based on FA (90% vs 85% respectively), however, with a rising amount of features, the accuracy declined to 81% (full dataset).

In the low-dimensional PCA-LR approach, models using only the first PCA component of FA or GMV reached an accuracy of 78% and 69% respectively (Fig. 2B). The first component of FC provided insignificant classification; nevertheless, the third alone reached an accuracy of 70%. The lack of classification power of the first two components in fMRI suggests that, unlike the previous two modalities, the main source of variance does not relate strongly to the patients-controls differences. Notably, the three components with substantial classification power (the first FA, the first GMV, and the third FC component) were significantly correlated (Spearman correlations: $R(FA, GMV) = 0.6$, $R(FA, FC) = 0.47$, $R(GMV, FC) = 0.46$), therefore the combined model was not more effective.

Moreover, the model consisting of the first, second, and the third PCA component of FA reached an accuracy of 90%, not significantly worse than the full FA data SVM classifier ($p = 0.121$). Similarly, the classifiers based on the GMV or FC did not perform significantly worse than their counterparts in FS-SVM models (except one and two-dimensional FC models). Nevertheless, their accuracy was inferior to that of FA.

3.2 Prediction of the motor impairment

In contrast to the classification for the prediction of the motor impairment, the results were not dominated by the FA. The best results in the SVR prediction were reached when using the FC (Fig. 3, Table S4). Especially for the MSWS scale, the prediction was consistently significant, reaching up to $R = 0.79$. The FA and GMV prediction ability varied with the number of features included and did not exceed $R = 0.50$ for any combination of feature type, count and clinical scale. The results of PCA in combination with linear regression complement those reached by SVR (Fig. 4, Table S5). While FC components were only successful in predicting the MSWS scale, when using the GMV features, the low-dimensional model significantly predicted almost all scales (EDSS, BBS, TUG, and MSWS).

4 Discussion

4.1 Classification

In our study, the model using only the regional average of FA in combination with SVM reached among the highest accuracy yet reported in the literature – 96% (sensitivity = 98%; specificity = 94%). The performance improved with features added. These results justify SVM use for classification based on similar imaging datasets. However, the accuracy of low-dimensional PCA-LR is also relatively high - a simple three-dimensional model using the first three FA components reached 90%.

Interestingly, the combined models using FA, FC and GMV were not superior to individual modality models. For SVM this may be caused by the almost perfect accuracy for FA diminishing space for improvement. The lack of accuracy improvement in the low-dimensional scenario is due to high correlation between the well-performing components of each modality - the three modalities thus did not provide synergistic performance effect.

We observed similar performance of the approaches, contrasting with the commonly observed dominance of SVM in classification tasks. One might speculate concerning potential data overfitting in practice due to the class-informed choice of features in combination with a small data size - something we avoided by careful design of the pipeline. From this perspective, we would advocate the use of PCA as a class-independent data transformation method that is not prone to this type of bias. Moreover, it reduces the model dimensionality, providing more easily interpretable results.

(Zurita et al., 2018) also used the SVM approach in combination with the FS feature selection, reaching 88.9% classification accuracy between patients (EDSS higher than 1.5) and healthy controls, combining structural and functional connectivity. However, this classifier used almost 6,000 features; whereas our PCA-based classification reached 89.9% accuracy using only three. Contrary to our observation, their classification based on the FA was less accurate than using FC. Dominance of our FA-based model may stem from using robust regional averages and TBSS skeleton rather than noisy high-dimensional voxel-wise analysis. Our complementary analysis in which we applied SVM to the FC matrices generated using the Harvard-Oxford Atlas slightly outperformed the AAL atlas results, reaching 80% accuracy using 75% of features; still not matching the values reported (Zurita et al., 2018); additional investigation with liberal preprocessing only decreased the performance.

Another study (Kocevar et al., 2016) used structural connectivity matrices, to derive graph-theoretical indices. SVM classification using combined features reached 92% precision comparing 12 subjects with clinically isolated syndrome to 24 healthy controls. While a small sample was analyzed, the results line up with other literature exploring the classification potential of structural connectivity graph using other classification methods (Charalambous et al., 2019; Marzullo et al., 2019).

4.2 Neuroimaging correlates of motor disability

We did not identify any study that applied machine learning analysis to explore the relationship of specialized motor impairment scales and brain imaging, although models have been proposed combining structural, functional and cognitive impairment (Schoonheim et al., 2015; Schoonheim, M.M. et al., 2010). Tommasin et al. (Tommasin et al., 2018) reported that general disability (measured through EDSS) had a direct linear relationship with lesion load, the inverse of thalamic volume, and functional connectivity in bi-frontal region pairs. With a similar approach and atlas for FC quantification, we did not observe correlation with EDSS (or other clinical measures), after appropriate multiple comparison correction. However, the SVR consistently confirmed significant predictive power of FC, indicating a presence of a more complex, multivariate relationship. Jakimovski et al. (Jakimovski et al., 2018) studied the correlation between walking disability (represented by 25 Foot Walk and the expanded TUG) and selected MRI-derived measures and DTI-measures maps; the DTI measures were not associated with the scales. Using regression analysis, the best predictor for the expanded TUG proved to be the cortex volume, $R^2 = 0.176$, comparable with our results in TUG ($R^2 = 0.153$ with the first GMV component). Steenwijk et al. (Steenwijk et al., 2014) reported partial correlations between different measures of the corticospinal tract and disability. Although corticospinal tract FA was correlated neither with EDSS, nor with MSWS, there was a significant correlation for the cortical thickness of the cortical area connected to it.

Our results indicate that the first GMV principal component is significantly correlated with numerous motor impairment scales. The same holds for the first FA component. For FC, the first and the second components played a role in regression, with the most noticeable results in the MSWS scale. Overall, the results of PCA-LinR are more reliable than the SVR results. The prediction quality did not vary significantly with dimensionality, whereas the SVR prediction using the FA and GMV features varied with the number of features.

4.3 Classification of phenotypes

The study included numerous phenotypes of multiple sclerosis. Acknowledging that there are structural differences between the subtypes (Bozzali et al., 2002; Filippi et al., 2019; Filippi and Rocca, 2011; Rovaris et al., 2002; Schoonheim et al., 2015), we further investigated the SVM's ability to separate the relapsing-remitting and the secondary progressive group. The highest accuracy of the FS-SVM classifier was using all the GMV features – 75.9% (sensitivity 64.0%; specificity 81.6%). Considerable successes have been reported by groups which based the classifier on the graph representation of structural connectivity and combined the information with either SVM or graph convolutional neural networks (Kocevar et al., 2016; Marzullo et al., 2019), albeit the insufficient sample size calls for further research.

4.4 Limitations

The application of machine learning methods in neuroscience is generally limited by sample size. We analyzed 64 multiple sclerosis patients and 65 healthy controls, which is a considerable size in this area of research. Nevertheless, the group of multiple sclerosis patients was not homogeneous in terms of the phenotype, potentially decreasing the classification power and accuracy of the regression models. Importantly, we took transparent precautions not to overfit the models, in particular, feature selection was carried out in a manner blind to the labels in the testing set. On the other hand, the degree of patients' impairment was on average higher than in other studies, which makes this project unique, but also warrants caution when interpreting the comparison with other works in the classification task.

However, beyond the numerous advantages, there are limitations associated with machine learning application. For SVM, the biggest concern may be associated with the dimensionality of the FC. The number of support vectors rose drastically with the number of features added to the model; nevertheless, the accuracy remained unchanged and inferior to the FA. The same applies to the PCA, where there were substantial differences between the variability explained by the first component among the three modalities. While the first component of GMV explained 71.7%, for FC, it was only 8.6%. Thus, while we can generally recommend using logistic regression combined with prior dimension reduction by PCA, selecting the first component is not necessarily optimal.

Finally, in this work, we focused on the comparison of two specific analytical pipelines, for which we presented the results. However, as is the case in most data-oriented work, it is possible that better performance could be achieved using different algorithms. For example, as suggested by on of the reviewers, methods with regularization might be applied, reducing the need for dimensionality reduction, although also making it more challenging to control for the number of features included. To explore this direction, we performed logistic regression with the LASSO regularization known for the sparsity of its solutions; the observed performance is comparable to that logistic regression run on the PCA variables (Table S6).

5 Conclusions

In the current study, we confirmed the widespread decrease of white matter integrity in patients with multiple sclerosis and demonstrated that localized white matter changes are related to motor symptoms. Using state-of-the-art machine learning methods, we showed that the white matter changes are specific and sensitive enough to provide 96% accuracy in recognizing patients from healthy controls. We further proposed an approach to construct low-dimensional classifiers with competitive performance, paving the way to robust and interpretable clinical tools.

We employed multimodal machine learning models, showing that the utilization of synergy between the modalities remains a challenge in the current settings, both due to potential redundancy between the features and problems with increasing the dimension of the feature set.

Finally, we fitted predictive models for the motor disability in multiple sclerosis, suggesting that while the disease itself is most apparent in decreased white matter integrity, the functional motor changes may indeed be more reflecting the accumulated cortical atrophy and changes in functional connectivity.

Declarations

- **Competing interests:** The authors declare no competing interests.
- **Ethics approval:** The study design was approved by the Ethics Committee of the Faculty Hospital Královské Vinohrady.
- **Consent to participate:** All participants were informed about the experimental setup and provided written informed consent in accordance with the Declaration of Helsinki.
- **Consent for publication:** Not applicable.
- **Availability of data and materials:** The raw imaging data were measured in the Institute for Clinical and Experimental Medicine and are not publicly available due to restrictions imposed by the administering institution. Raw anonymized data can be shared based on a formal data sharing agreement based on a submission of a project outline, that may - depending on the nature of the data requested - require project amendment approval from the local ethics committee. Informal requests for information shall be sent to the corresponding author. Part of the preprocessed fMRI data is available for download (Bučková et al., 2021).
- **Code availability:** Not applicable.
- **Authors' contributions:** BRB: Methodology, Formal analysis, Writing - Original Draft, Visualization JM: Data Curation AS: Software, Data Curation JK: Software, JT: Investigation RAD: Conceptualization, KR: Conceptualization, Project administration, Funding acquisition, JH:

Conceptualization, Project administration, Funding acquisition, Methodology, Supervision. All authors participated on editing the manuscript and approved its final version.

Figures
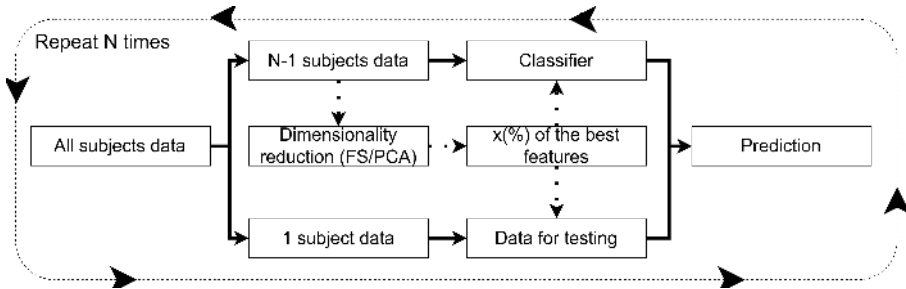
Fig. 1: A general scheme of the classification pipeline: In every iteration, one subject is left out. Either FS or PCA is computed on the training dataset, and the desired transformation (n terms of the number of components or features) is then applied. The classifier is trained using N-1 subjects and subsequently tested on the remaining subject. After N rounds, the classification accuracy is evaluated. Full lines on the image illustrate the data flow, whereas the dashed lines stand for data transformations.
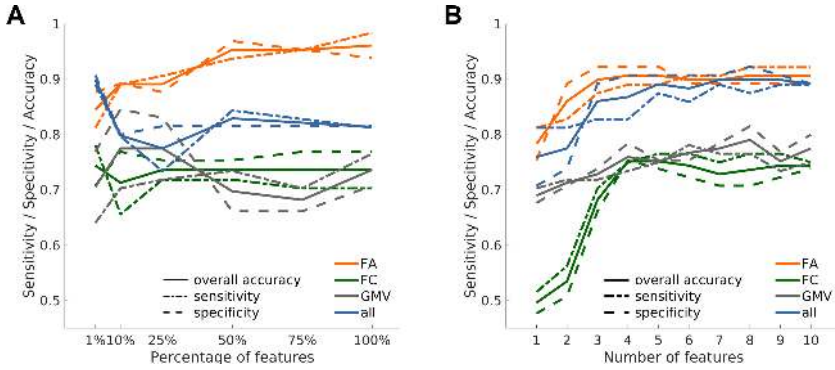
Fig. 2: Classification accuracy of Support Vector Machines classifiers (A) and Logistic Regression classifiers (B): overall accuracy (full line), sensitivity (dashed line with dots) and specificity (dashed line) for different thresholds. Classifiers based on fractional anisotropy (FA), functional connectivity (FC), and gray matter volume (GMV) correspond to orange, green and gray line respectively, the classifier based on the combination of features corresponds to the blue color.
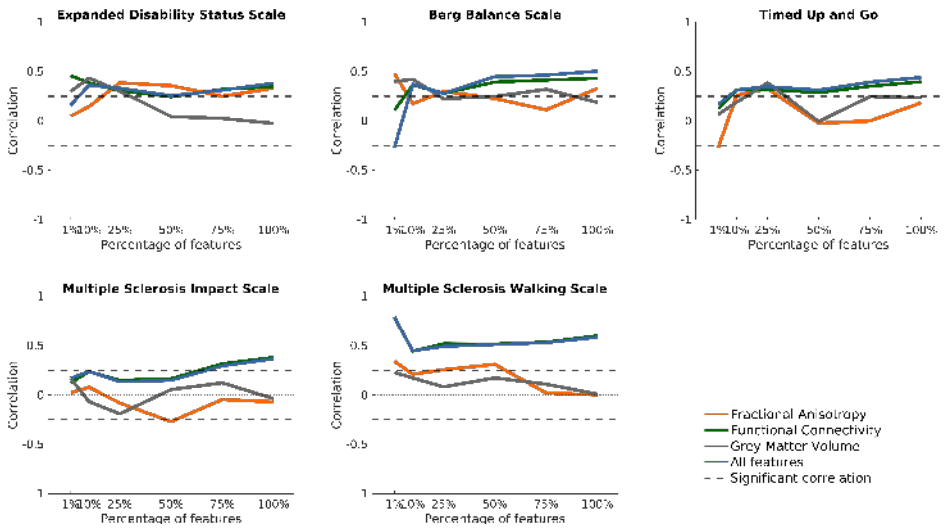
Fig. 3: The Spearman correlation of the Support Vector Regression models prediction with the clinical scales: Models based on fractional anisotropy, functional connectivity, and gray matter volume correspond to orange, green and gray line respectively, models based on the combination of all features corresponds to the blue color. EDSS – Expanded Disability Status Scale; BBS – Berg Balance Scale; TUG – Timed Up and Go Test; MSIS – Multiple Sclerosis Impact Scale; MSWS – Twelve Item Multiple Sclerosis Walking Scale.
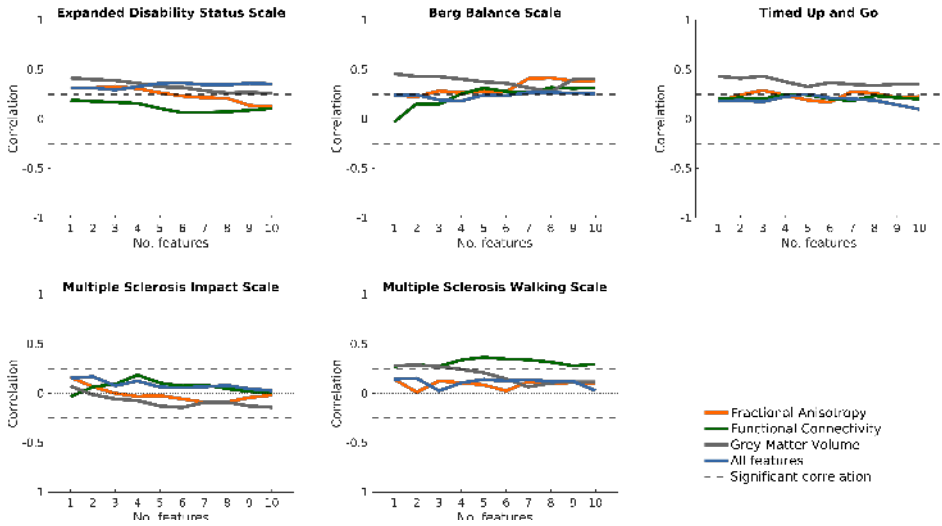
Fig. 4: The Spearman correlation of the Linear regression models prediction with the clinical scales: Models based on fractional anisotropy, functional connectivity, and gray matter volume correspond to orange, green and gray line respectively, models based on the combination of all features corresponds to the blue color. EDSS – Expanded Disability Status Scale; BBS – Berg Balance Scale; TUG – Timed Up and Go Test; MSIS – Multiple Sclerosis Impact Scale; MSWS – Twelve Item Multiple Sclerosis Walking Scale.

Tables

Healthy controls

	Women	**Men**	**All**
Number	32	33	65
Age	43 (31-68)	40 (31-68)	41 (31-68)

People with multiple sclerosis

	Women	**Men**	**All**
Number	39	25	64
RR/SP/PP	23/14/1	15/6/4	38/20/5
Age	47 (22-70)	43 (29-68)	45 (22-70)
EDSS	4.5 (1.0-6.5)	4.5 (1.5-7.5)	4.5 (1.0-7.5)
BBS	44.0 (15-56)	35.0 (5-56)	42.0 (5-56)
TUG	10.0 (5-27)	13.0 (6-109)	11.0 (5-109)
MSIS	67.0 (63-100)	80.0 (34-114)	72.5 (33-114)
MSWS	34.0 (12-59)	39.0 (25-60)	36.5 (12-60)

Table 1: Description of the dataset: All values are listed as medians and ranges; RR – Relapsing-Remitting; PP – Primary Progressive; SP – Secondary Progressive; EDSS – Expanded Disability Status Scale; BBS – Berg Balance Scale; TUG – Timed Up and Go Test; MSIS – Multiple Sclerosis Impact Scale; MSWS – Twelve Item Multiple Sclerosis Walking Scale

Appendix A

| Pair of significantly correlated regions		z-statistic	ranksum	corrected p-value	p-value
Supp Motor Area L	Cingulum Mid R	-5.73510	2,942	0.00007	0.00000
Supp Motor Area L	Cingulum Mid L	-5.38181	3,017	0.00025	0.00000
Cerebelum 9 R	Vermis 1 2	-4.90605	3,118	0.00207	0.00000
Supp Motor Area R	Cingulum Mid R	-4.82126	3,136	0.00238	0.00000
Temporal Sup R	Vermis 4 5	-4.69878	3,162	0.00349	0.00000
Amygdala R	Thalamus L	-4.54333	3,195	0.00616	0.00001
Amygdala R	Thalamus R	-4.44912	3,215	0.00820	0.00001
Rolandic Oper L	Supp Motor Area R	-4.42086	3,221	0.00820	0.00001
Cerebelum 9 L	Vermis 1 2	-4.33136	3,240	0.01068	0.00001
Frontal Sup Medial R	Cerebelum Crus2 R	-4.28425	3,250	0.01068	0.00002
Frontal Med Orb L	Cerebelum 6 R	-4.27012	3,253	0.01068	0.00002
Frontal Mid R	Cerebelum Crus2 R	-4.26070	3,255	0.01068	0.00002
Temporal Sup L	Vermis 4 5	-4.25599	3,256	0.01068	0.00002
Supp Motor Area R	Cingulum Mid L	-4.18062	3,272	0.01293	0.00003
Pallidum R	Thalamus R	-4.18062	3,272	0.01293	0.00003
Frontal Mid R	Cerebelum 7b R	-4.10525	3,288	0.01684	0.00004
Putamen R	Vermis 4 5	-4.08170	3,293	0.01754	0.00004
Rolandic Oper L	Insula R	-4.06757	3,296	0.01760	0.00005
Frontal Mid R	Cerebelum Crus1 L	-4.04872	3,300	0.01788	0.00005
Rolandic Oper L	Cingulum Mid R	-4.03930	3,302	0.01788	0.00005
Cingulum Ant L	Thalamus L	-4.02046	3,306	0.01817	0.00006
Cingulum Mid R	Heschl R	-4.01104	3,308	0.01817	0.00006
Rolandic Oper L	Supp Motor Area L	-4.00162	3,310	0.01817	0.00006
Supp Motor Area L	Cingulum Ant R	-3.98749	3,313	0.01817	0.00007
Occipital Inf R	Parietal Sup L	-3.98278	3,314	0.01817	0.00007
Frontal Sup R	Cerebelum Crus2 R	-3.95922	3,319	0.01929	0.00008
Insula R	Cingulum Ant L	-3.94038	3,323	0.02010	0.00008
Amygdala R	Cerebelum Crus2 R	-3.92625	3,326	0.02024	0.00009
Amygdala R	Putamen R	-3.92154	3,327	0.02024	0.00009
Frontal Sup Medial L	Cerebelum Crus2 R	-3.90270	3,331	0.02037	0.00010
Frontal Sup L	Precuneus L	-3.89328	3,333	0.02037	0.00010
Cingulum Post R	Temporal Sup R	-3.88385	3,335	0.02037	0.00010
Cingulum Mid L	Thalamus R	-3.87914	3,336	0.02037	0.00010
Insula L	Vermis 4 5	-3.87914	3,336	0.02037	0.00010
Frontal Mid R	Cerebelum Crus1 R	-3.87443	3,337	0.02037	0.00011
Rolandic Oper L	Cingulum Mid L	-3.85088	3,342	0.02173	0.00012
Insula L	Thalamus R	-3.84146	3,344	0.02173	0.00012
Supp Motor Area L	Cingulum Ant L	-3.83204	3,346	0.02173	0.00013
Insula R	Insula R	-3.83204	3,346	0.02173	0.00013
Frontal Sup R	Cerebelum Crus1 R	-3.80378	3,352	0.02376	0.00014
Amygdala R	Cerebelum 9 R	-3.78022	3,357	0.02514	0.00016
Cingulum Mid L	Heschl R	-3.77551	3,358	0.02514	0.00016
Frontal Inf Oper R	Frontal Sup Medial R	-3.77080	3,359	0.02514	0.00016
Temporal Mid R	Cerebelum Crus2 R	-3.76609	3,360	0.02514	0.00017
Frontal Sup Medial L	Precuneus R	-3.75667	3,362	0.02552	0.00017
Cingulum Ant L	Heschl L	-3.74254	3,365	0.02574	0.00018
Frontal Sup R	Cerebelum Crus1 L	-3.73312	3,367	0.02574	0.00019
Insula L	Cingulum Mid L	-3.73312	3,367	0.02574	0.00019
Occipital Mid L	Occipital Inf R	-3.73312	3,367	0.02574	0.00019
Heschl R	Temporal Inf R	3.71427	4,949	0.02665	0.00020
Frontal Inf Orb L	Cingulum Mid L	-3.71899	3,370	0.02665	0.00020
Cingulum Ant R	Thalamus L	-3.70014	3,374	0.02764	0.00022
Supp Motor Area L	Heschl L	-3.67659	3,379	0.02975	0.00024
Cingulum Ant L	Thalamus R	-3.66717	3,381	0.03029	0.00025
Insula R	Vermis 4 5	-3.66246	3,382	0.03029	0.00025
Frontal Sup R	Cerebelum 6 L	-3.63420	3,388	0.03157	0.00028
Frontal Inf Tri L	Vermis 4 5	-3.63420	3,388	0.03157	0.00028
Frontal Inf Orb L	Cingulum Ant L	-3.62948	3,389	0.03157	0.00028
Supp Motor Area L	Insula L	-3.62948	3,389	0.03157	0.00028
Pallidum R	Thalamus L	-3.62948	3,389	0.03157	0.00028
Frontal Mid R	Vermis 6	-3.62477	3,390	0.03162	0.00029
Putamen L	Putamen R	-3.62006	3,391	0.03169	0.00029
Cingulum Ant L	Temporal Pole Sup R	-3.60122	3,395	0.03259	0.00032

Cingulum Ant L	Cerebelum 9 L	-3.59651	3,396	0.03259	0.00032
Caudate R	Putamen R	-3.59651	3,396	0.03259	0.00032
Cingulum Mid L	Heschl L	-3.58709	3,398	0.03280	0.00033
Supp Motor Area L	Insula R	-3.58709	3,398	0.03280	0.00033
Rolandic Oper R	Temporal Inf R	3.60122	4,925	0.03259	0.00032
Cingulum Ant R	Cerebelum 9 L	-3.57767	3,400	0.03351	0.00035
Temporal Sup R	Cerebelum 4 5 L	-3.56354	3,403	0.03486	0.00037
Frontal Mid L	Cerebelum Crus1 L	-3.54941	3,406	0.03627	0.00039
Insula R	Cingulum Mid L	-3.52585	3,411	0.03910	0.00042
Precuneus R	Vermis 6	-3.52114	3,412	0.03926	0.00043
Frontal Med Orb R	Cerebelum 6 R	-3.50701	3,415	0.04030	0.00045
Frontal Sup R	Temporal Inf R	-3.50701	3,415	0.04030	0.00045
Insula L	Cingulum Ant L	-3.50230	3,416	0.04048	0.00046
Putamen L	Vermis 4 5	-3.49759	3,417	0.04067	0.00047
Occipital Inf R	Parietal Sup R	-3.48817	3,419	0.04159	0.00049
Hippocampus R	Cerebelum 9 L	-3.46933	3,423	0.04405	0.00052
Frontal Sup Medial R	Temporal Mid R	-3.45990	3,425	0.04505	0.00054
Insula R	Cingulum Mid R	-3.45519	3,426	0.04528	0.00055
Cingulum Ant L	Heschl R	-3.45048	3,427	0.04552	0.00056
Supp Motor Area L	Heschl R	-3.43164	3,431	0.04764	0.00060
Temporal Inf R	Vermis 7	-3.43164	3,431	0.04764	0.00060
Cingulum Post L	Cerebelum 6 R	-3.42693	3,432	0.04790	0.00061
Rolandic Oper L	Cerebelum Crus2 L	3.41751	4,886	0.04901	0.00063

Table S1: Pairs of regions of AAL atlas, where FC significantly differs between patients and controls

		Fisher Score and SVM percentage of features					
		1%	**10%**	**25%**	**50%**	**75%**	**100%**
FA	sensitivity	81.2%	89.1%	90.6%	93.8%	95.3%	98.4%
	specificity	87.7%	89.2%	87.7%	96.9%	95.4%	93.8%
	overall accuracy	84.5%	89.1%	89.1%	95.3%	95.3%	96.1%
	AUC	0.89	0.95	0.96	0.98	0.98	0.98
FC	sensitivity	78.1%	65.6%	71.9%	71.9%	70.3%	70.3%
	specificity	70.8%	76.9%	75.4%	75.4%	76.9%	76.9%
	overall accuracy	74.4%	71.3%	73.6%	73.6%	73.6%	73.6%
	AUC	0.83	0.79	0.79	0.82	0.83	0.82
GMV	sensitivity	64.1%	70.3%	71.9%	73.4%	70.3%	76.6%
	specificity	76.9%	84.6%	83.1%	66.2%	66.2%	70.8%
	overall accuracy	70.5%	77.5%	77.5%	69.8%	68.2%	73.6%
	AUC	0.74	0.79	0.82	0.78	0.73	0.82
all	sensitivity	89.1%	79.7%	73.4%	84.4%	82.8%	81.2%
	specificity	90.8%	80.0%	81.5%	81.5%	81.5%	81.5%
	overall accuracy	89.9%	79.8%	77.5%	82.9%	82.2%	81.4%
	AUC	0.93	0.87	0.89	0.90	0.89	0.89

Table S2: The results of the Support Vector Machine models across modalities for all thresholds: FA – Fractional Anisotropy; FC – Functional Connectivity; GMV – Grey Matter Volume; all – combination of all three modalities

Fig. S1: P-values of McNemar test between all thresholds and classifiers: Support Vector Machines (left), Logistic Regression (right). Threshold values in Support Vector Machines stand for the percentage of all features, in Logistic regression they represent the number of PCA component added to the model. FA – Fractional Anisotropy; FC – Functional Connectivity; GMV – Grey Matter Volume; all – combination of all three modalities.

Table S3: The results of the Logistic Regression models combined with **Principal Component Analysis features across modalities for all thresholds:** FA – Fractional Anisotropy; FC – Functional Connectivity; GMV – Grey Matter Volume; all – combination of all three modalities
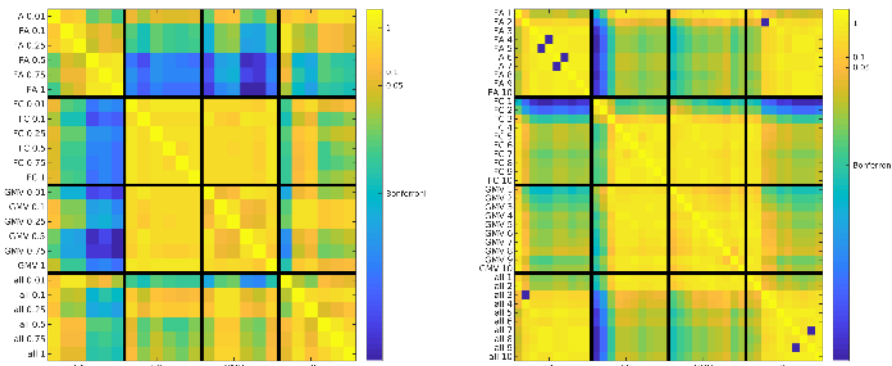
		PCA and Logistic Regression									
		number of features									
		1	2	3	4	5	6	7	8	9	10
FA	sensitivity	81.2%	82.8%	87.5%	89.1%	89.1%	90.6%	90.6%	92.2%	92.2%	92.2%
	specificity	75.4%	89.2%	92.3%	92.3%	92.3%	89.2%	89.2%	89.2%	89.2%	89.2%
	overall accuracy	78.3%	86.0%	89.9%	90.7%	90.7%	89.9%	89.9%	90.7%	90.7%	90.7%
	overall accuracy	78.3%	86.0%	89.9%	90.7%	90.7%	89.9%	89.9%	90.7%	90.7%	90.7%
	AUC	0.82	0.93	0.94	0.95	0.94	0.95	0.95	0.95	0.95	0.96
FC	sensitivity	51.6%	56.2%	70.3%	75.0%	76.6%	76.6%	75.0%	76.6%	76.6%	75.0%
	specificity	47.7%	50.8%	66.2%	75.4%	73.8%	72.3%	70.8%	70.8%	72.3%	73.8%
	overall accuracy	49.6%	53.5%	68.2%	75.2%	75.2%	74.4%	72.9%	73.6%	74.4%	74.4%
	AUC	0.48	0.64	0.63	0.79	0.79	0.78	0.78	0.76	0.77	0.79
GMV	sensitivity	70.3%	71.9%	71.9%	73.4%	75.0%	78.1%	76.6%	76.6%	73.4%	75.0%
	specificity	67.7%	70.8%	73.8%	78.5%	75.4%	75.4%	78.5%	81.5%	76.9%	80.0%
	overall accuracy	69.0%	71.3%	72.9%	76.0%	75.2%	76.7%	77.5%	79.1%	75.2%	77.5%
	AUC	0.76	0.76	0.78	0.83	0.82	0.84	0.85	0.85	0.85	0.85
all	sensitivity	81.2%	81.2%	82.8%	82.8%	87.5%	85.9%	89.1%	87.5%	89.1%	89.1%
	specificity	70.8%	73.8%	89.2%	90.8%	90.8%	90.8%	90.8%	92.3%	90.8%	89.2%
	overall accuracy	76.0%	77.5%	86.0%	86.8%	89.1%	88.4%	89.9%	89.9%	89.9%	89.1%
	AUC	0.82	0.82	0.93	0.93	0.93	0.94	0.94	0.94	0.94	0.95

		Fisher Score and SVR percentage of features					
		1%	**10%**	**25%**	**50%**	**75%**	**100%**
EDSS	**FA**	0.05	0.14	0.39	0.36	0.25	0.33
	FC	0.46	0.39	0.30	0.24	0.32	0.35
	GMV	0.30	0.43	0.30	0.04	0.02	-0.03
	all	0.16	0.36	0.33	0.25	0.31	0.38
BBS	**FA**	0.48	0.17	0.30	0.23	0.11	0.33
	FC	0.11	0.36	0.28	0.39	0.41	0.43
	GMV	0.40	0.42	0.22	0.25	0.32	0.19
	all	-0.27	0.37	0.28	0.45	0.46	0.50
TUG	**FA**	-0.27	0.26	0.33	-0.02	0.00	0.18
	FC	0.12	0.31	0.32	0.28	0.35	0.40
	GMV	0.06	0.19	0.39	-0.01	0.25	0.23
	all	0.17	0.31	0.35	0.31	0.39	0.44
MSIS	**FA**	0.02	0.08	-0.08	-0.27	-0.05	-0.07
	FC	0.12	0.24	0.15	0.17	0.32	0.38
	GMV	0.17	-0.07	-0.19	0.06	0.12	-0.04
	all	0.17	0.24	0.14	0.15	0.30	0.37
MSWS	**FA**	0.34	0.21	0.26	0.31	0.02	0.00
	FC	0.79	0.44	0.52	0.52	0.54	0.60
	GMV	0.23	0.17	0.08	0.17	0.11	0.01
	all	0.78	0.45	0.49	0.51	0.53	0.59

Table S4: The results of the Support Vector Regression models across modalities for all thresholds: FA – Fractional Anisotropy; FC – Functional Connectivity; GMV – Grey Matter Volume; all – combination of all three modalities; EDSS – Expanded Disability Status Scale; BBS – Berg Balance Scale; TUG – Timed Up and Go Test; MSIS – Multiple Sclerosis Impact Scale; MSWS – Twelve Item Multiple Sclerosis Walking Scale

Table S5: The results of the Linear Regression models combined with Principal Component Analysis components across modalities for all thresholds: FA – Fractional Anisotropy; FC – Functional Connectivity; GMV – Grey Matter Volume; all – combination of all three modalities; EDSS – Expanded Disability Status Scale; BBS – Berg Balance Scale; TUG – Timed Up and Go Test; MSIS – Multiple Sclerosis Impact Scale; MSWS – Twelve Item Multiple Sclerosis Walking Scale

| | | PCA and Linear Regression | | | | | | | | | |
| | | number of features | | | | | | | | | |
		1	2	3	4	5	6	7	8	9	10
EDSS	FA	0.31	0.31	0.33	0.31	0.27	0.23	0.21	0.21	0.14	0.13
	FC	0.19	0.18	0.17	0.16	0.11	0.06	0.07	0.07	0.09	0.11
	GMV	0.42	0.40	0.39	0.36	0.32	0.32	0.29	0.26	0.27	0.26
	all	0.31	0.32	0.29	0.33	0.36	0.37	0.35	0.34	0.36	0.35
BBS	FA	0.24	0.22	0.28	0.27	0.28	0.28	0.41	0.42	0.39	0.39
	FC	-0.03	0.15	0.15	0.26	0.31	0.28	0.27	0.32	0.31	0.32
	GMV	0.46	0.43	0.43	0.40	0.38	0.36	0.32	0.28	0.40	0.40
	all	0.24	0.25	0.19	0.18	0.24	0.24	0.27	0.27	0.26	0.25
TUG	FA	0.18	0.24	0.29	0.24	0.19	0.17	0.28	0.26	0.22	0.23
	FC	0.20	0.21	0.20	0.25	0.25	0.20	0.19	0.23	0.22	0.20
	GMV	0.43	0.41	0.43	0.38	0.33	0.37	0.35	0.34	0.35	0.35
	all	0.18	0.19	0.17	0.23	0.25	0.21	0.20	0.19	0.14	0.10
MSIS	FA	0.16	0.07	0.00	-0.03	-0.02	-0.06	-0.09	-0.09	-0.04	-0.02
	FC	-0.03	0.07	0.09	0.19	0.11	0.08	0.09	0.05	0.02	0.01
	GMV	0.07	-0.01	-0.06	-0.07	-0.13	-0.14	-0.09	-0.09	-0.13	-0.14
	all	0.15	0.17	0.08	0.13	0.06	0.06	0.07	0.08	0.05	0.03
MSWS	FA	0.14	0.01	0.13	0.11	0.09	0.03	0.12	0.10	0.11	0.10
	FC	0.28	0.29	0.28	0.34	0.37	0.35	0.34	0.32	0.28	0.30
	GMV	0.28	0.29	0.27	0.24	0.21	0.15	0.07	0.11	0.11	0.11
	all	0.14	0.15	0.03	0.11	0.14	0.13	0.14	0.12	0.12	0.03

	sensitivity	specificity	accuracy	AUC	median no. variables across folds
FA	93.8%	93.9%	93.8%	0.97	4
FC	85.9%	66.2%	76.0%	0.78	6
GMV	85.9%	73.9%	79.8%	0.85	4
all	84.4%	92.3%	88.4%	0.94	7

Table S6: The results of the LASSO Logistic Regression. LASSO had been fitted separately for every fold in leave-one-out crossvalidation, and the best model was chosen in each fold using Bayesian Information Criterion. FA – Fractional Anisotropy; FC – Functional Connectivity; GMV – Grey Matter Volume; all – combination of all three modalities;

References

Asch, P. v. (2011). Impact of mobility impairment in multiple sclerosis 2 - patients' perspectives. *European Neurological Review*, 6(2):115–120.

Basser, P. J., Mattiello, J., and LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1):259–267.

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

Berg, K., Wood-Dauphine, S., Williams, J. I., and Gayton, D. (2009). Measuring balance in the elderly: preliminary development of an instrument. *Physiotherapy Canada*, 41(6):304–311.

Bozzali, M., Cercignani, M., Sormani, M. P., Comi, G., and Filippi, M. (2002). Quantification of brain gray matter damage in different MS phenotypes by use of diffusion tensor MR imaging. *American Journal of Neuroradiology*, 23(6):985–988.

Burschka, J. M., Keune, P. M., Menge, U., Oy, U. H.-v., Oschmann, P., and Hoos, O. (2012). An exploration of impaired walking dynamics and fatigue in multiple sclerosis. *BMC Neurology*, 12(1):161.

Bučková, B., Kopal, J., Řasová, K., Tintěra, J., and Hlinka, J. (2021). Open access: The effect of neurorehabilitation on multiple sclerosis—unlocking the resting-state fMRI data. *Frontiers in Neuroscience*, 15:615.

Charalambous, T., Tur, C., Prados, F., Kanber, B., Chard, D. T., Ourselin, S., Clayden, J. D., Wheeler-Kingshott, C. A. M. G., Thompson, A. J., and Toosy, A. T. (2019). Structural network disruption markers explain disability in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 90(2):219–226.

Dobson, R. and Giovannoni, G. (2019). Multiple sclerosis–a review. *European journal of neurology*, 26(1):27–40.

Eitel, F., Soehler, E., Bellmann-Strobl, J., Brandt, A. U., Ruprecht, K., Giess, R. M., Kuchling, J., Asseyer, S., Weygandt, M., Haynes, J.-D., et al. (2019). Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional mri using layer-wise relevance propagation. *NeuroImage: Clinical*, 24:102003.

Filippi, M., Brück, W., Chard, D., Fazekas, F., Geurts, J. J. G., Enzinger, C., Hametner, S., Kuhlmann, T., Preziosa, P., Rovira, A., Schmierer, K.,

Stadelmann, C., and Rocca, M. A. (2019). Association between pathological and MRI findings in multiple sclerosis. *The Lancet Neurology*, 18(2):198–210.

Filippi, M. and Rocca, M. A. (2011). MR imaging of multiple sclerosis. *Radiology*, 259(3):659–681.

Fischer, J. S., Rudick, R. A., Cutter, G. R., and Reingold, S. C. (1999). The multiple sclerosis functional composite measure (MSFC): an integrated approach to MS clinical outcome assessment. *Multiple Sclerosis Journal*, 5(4):244–250.

Giannì, C., Prosperini, L., Jonsdottir, J., and Cattaneo, D. (2014). A systematic review of factors associated with accidental falls in people with multiple sclerosis: a meta-analytic approach. *Clinical Rehabilitation*, 28(7):704–716.

Hartman, D., Hlinka, J., Paluš, M., Mantini, D., and Corbetta, M. (2011). The role of nonlinearity in computing graph-theoretical properties of resting-state functional magnetic resonance imaging brain networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(1):013119.

He, X., Cai, D., and Niyogi, P. (2006). Laplacian score for feature selection. In Weiss, Y., Schölkopf, B., and Platt, J. C., editors, *Advances in Neural Information Processing Systems 18*, pages 507–514. MIT Press.

Heesen, C., Böhm, J., Reich, C., Kasper, J., Goebel, M., and Gold, S. (2008). Patient perception of bodily functions in multiple sclerosis: gait and visual function are the most valuable. *Multiple Sclerosis Journal*, 14(7):988–991.

Hlinka, J., Paluš, M., Vejmelka, M., Mantini, D., and Corbetta, M. (2011). Functional connectivity in resting-state fMRI: Is linear correlation sufficient? *NeuroImage*, 54(3):2218–2225.

Hobart, J., Lamping, D., Fitzpatrick, R., Riazi, A., and Thompson, A. (2001). The multiple sclerosis impact scale (MSIS-29)a new patient-based outcome measure. *Brain*, 124(5):962–973.

Hobart, J., Riazi, A., Lamping, D., Fitzpatrick, R., and Thompson, A. (2003). Measuring the impact of MS on walking ability: The 12-item MS walking scale (MSWS-12). *Neurology*, 60(1):31–36.

Jakimovski, D., Weinstock-Guttman, B., Hagemeier, J., Vaughn, C. B., Kavak, K. S., Gandhi, S., Bennett, S. E., Fuchs, T. A., Bergsland, N., Dwyer, M. G., Benedict, R. H. B., and Zivadinov, R. (2018). Walking disability measures in multiple sclerosis patients: Correlations with MRI-derived

global and microstructural damage. *Journal of the Neurological Sciences*, 393:128–134.

Keller, S. S. and Roberts, N. (2008). Voxel-based morphometry of temporal lobe epilepsy: An introduction and review of the literature. *Epilepsia*, 49(5):741–757.

Kellner, E., Dhital, B., Kiselev, V. G., and Reisert, M. (2016). Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic Resonance in Medicine*, 76(5):1574–1581.

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

LaRocca, N. G. (2011). Impact of walking impairment in multiple sclerosis. *The Patient: Patient-Centered Outcomes Research*, 4(3):189–201.

Leray, E., Moreau, T., Fromont, A., and Edan, G. (2016). Epidemiology of multiple sclerosis. *Revue Neurologique*, 172(1):3–13.

Martin, C. L., Phillips, B. A., Kilpatrick, T. J., Butzkueven, H., Tubridy, N., McDonald, E., and Galea, M. P. (2006). Gait and balance impairment in early multiple sclerosis in the absence of clinical disability. *Multiple Sclerosis Journal*, 12(5):620–628.

Marzullo, A., Kocevar, G., Stamile, C., Durand-Dubief, F., Terracina, G., Calimeri, F., and Sappey-Marinier, D. (2019). Classification of multiple sclerosis clinical profiles via graph convolutional neural networks. *Frontiers in Neuroscience*, 13.

Mori, S., Wakana, S., Zijl, P. C. M. v., and Nagae-Poetscher, L. M. (2005). *MRI Atlas of Human White Matter*. Elsevier.

Podsiadlo, D. and Richardson, S. (1991). The timed "up & go": A test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society*, 39(2):142–148.

Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F. D., Montalban, X., O'Connor, P., Sandberg-Wollheim, M., Thompson, A. J., Waubant, E., Weinshenker, B., and Wolinsky, J. S. (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology*, 69(2):292–302.

Rovaris, M., Bozzali, M., Iannucci, G., Ghezzi, A., Caputo, D., Montanari, E., Bertolotto, A., Bergamaschi, R., Capra, R., Mancardi, G. L., Martinelli, V., Comi, G., and Filippi, M. (2002). Assessment of normal-appearing white and gray matter in patients with primary progressive multiple sclerosis: A diffusion-tensor magnetic resonance imaging study. *Archives of Neurology*, 59(9):1406–1412.

Sbardella, E., Tona, F., Petsas, N., Upadhyay, N., Piattella, M., Filippini, N., Prosperini, L., Pozzilli, C., and Pantano, P. (2015). Functional connectivity changes and their relationship with clinical disability and white matter integrity in patients with relapsing–remitting multiple sclerosis. *Multiple Sclerosis Journal*, 21(13):1681–1692.

Schoonheim, M. M., Meijer, K. A., and Geurts, J. J. G. (2015). Network collapse and cognitive impairment in multiple sclerosis. *Frontiers in Neurology*, 6.

Schoonheim, M.M., Geurts, J.J.G., Barkhof, F., Radiology and nuclear medicine, Pathology, Anatomy and neurosciences, NCA - Multiple Sclerosis and Other Neuroinflammatory Diseases, and NCA - Neurodegeneration (2010). The limits of functional reorganization in multiple sclerosis. *Neurology*, 74(16):1246–1247.

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., and Behrens, T. E. J. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31(4):1487–1505.

Stafford, I. S., Kellermann, M., Mossotto, E., Beattie, R. M., MacArthur, B. D., and Ennis, S. (2020). A systematic review of the applications of artificial intelligence and machine learning in autoimmune diseases. *npj Digital Medicine*, 3(1):1–11.

Steenwijk, M. D., Daams, M., Pouwels, P. J. W., Balk, L. J., Tewarie, P. K., Killestein, J., Uitdehaag, B. M. J., Geurts, J. J. G., Barkhof, F., and Vrenken, H. (2014). What explains gray matter atrophy in long-standing multiple sclerosis? *Radiology*, 272(3):832–842.

Tommasin, S., De Giglio, L., Ruggieri, S., Petsas, N., Giannì, C., Pozzilli, C., and Pantano, P. (2018). Relation between functional connectivity and disability in multiple sclerosis: a non-linear model. *Journal of Neurology*, 265(12):2881–2892.

Tona, F., Petsas, N., Sbardella, E., Prosperini, L., Carmellini, M., Pozzilli, C., and Pantano, P. (2014). Multiple sclerosis: Altered thalamic resting-state functional connectivity and its effect on cognitive function. *Radiology*,

271(3):814–821.

Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., and Connelly, A. (2019). Mrtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage*, 202:116137.

Tozlu, C., Jamison, K., Gauthier, S. A., and Kuceyeski, A. (2021). Dynamic functional connectivity better predicts disability than structural and static functional connectivity in people with multiple sclerosis. *Frontiers in neuroscience*, 15.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289.

Vapnik, V. and Chapelle, O. (2000). Bounds on error expectation for support vector machines. *Neural Computation*, 12(9):2013–2036.

Veraart, J., Fieremans, E., and Novikov, D. S. (2016a). Diffusion MRI noise mapping using random matrix theory. *Magnetic Resonance in Medicine*, 76(5):1582–1593.

Veraart, J., Novikov, D. S., Christiaens, D., Ades-aron, B., Sijbers, J., and Fieremans, E. (2016b). Denoising of diffusion MRI using random matrix theory. *NeuroImage*, 142:394–406.

Wallin, M. T., Culpepper, W. J., Nichols, E., Bhutta, Z. A., Gebrehiwot, T. T., Hay, S. I., Khalil, I. A., Krohn, K. J., Liang, X., Naghavi, M., Mokdad, A. H., Nixon, M. R., Reiner, R. C., Sartorius, B., Smith, M., Topor-Madry, R., Werdecker, A., Vos, T., Feigin, V. L., and Murray, C. J. L. (2019). Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the global burden of disease study 2016. *The Lancet Neurology*, 18(3):269–285.

Westin, C.-F., Peled, S., Gubjartsson, H., Kikinis, R., and Jolesz, F. A. (1997). Geometrical diffusion measures for mri from tensor basis analysis.

Woo, C.-W., Chang, L. J., Lindquist, M. A., and Wager, T. D. (2017). Building better biomarkers: brain models in translational neuroimaging. *Nature Neuroscience*, 20(3):365–377.

Yassa, M. A. and Stark, C. E. L. (2009). A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. *NeuroImage*, 44(2):319–327.

Zurita, M., Montalba, C., Labbé, T., Cruz, J. P., Dalboni da Rocha, J., Tejos, C., Ciampi, E., Cárcamo, C., Sitaram, R., and 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.