Brain Imaging and Behavior (2019) 13:1103–1114
https://doi.org/10.1007/s11682-018-9926-9

ORIGINAL RESEARCH

Evaluation of machine learning algorithms performance for the prediction of early multiple sclerosis from resting-state FMRI connectivity data

Valeria Saccà[1] · Alessia Sarica[2] · Fabiana Novellino[2] · Stefania Barone[3] · Tiziana Tallarico[3] · Enrica Filippelli[3] · Alfredo Granata[3] · Carmelina Chiriaco[2] · Roberto Bruno Bossio[4] · Paola Valentino[3] · Aldo Quattrone[2,3]

Published online: 11 July 2018

Abstract

Machine Learning application on clinical data in order to support diagnosis and prognostic evaluation arouses growing interest in scientific community. However, choice of right algorithm to use was fundamental to perform reliable and robust classification. Our study aimed to explore if different kinds of Machine Learning technique could be effective to support early diagnosis of Multiple Sclerosis and which of them presented best performance in distinguishing Multiple Sclerosis patients from control subjects. We selected following algorithms: Random Forest, Support Vector Machine, Naïve-Bayes, K-nearest-neighbor and Artificial Neural Network. We applied the Independent Component Analysis to resting-state functional-MRI sequence to identify brain networks. We found 15 networks, from which we extracted the mean signals used into classification. We performed feature selection tasks in all algorithms to obtain the most important variables. We showed that best discriminant network between controls and early Multiple Sclerosis, was the sensori-motor I, according to early manifestation of motor/sensorial deficits in Multiple Sclerosis. Moreover, in classification performance, Random Forest and Support Vector Machine showed same 5-fold cross-validation accuracies (85.7%) using only this network, resulting to be best approaches. We believe that these findings could represent encouraging step toward the translation to clinical diagnosis and prognosis.

Keywords Resting state fMRI · Support vector machine · Random Forest · Naïve Bayes · K-nearest-neighbor · Artificial neural network

Introduction

In the last few years, several studies applied Machine Learning techniques to analyze Magnetic Resonance Imaging (MRI) (Douglas et al. 2011; Formisano et al. 2008; LaConte et al. 2005; Lemm et al. 2011; Misaki et al. 2010; Mourão-Miranda et al. 2005; Pereira et al. 2009; Wang et al. 2007). This increased interest of the scientific community derived from the possibility to extract important, clinically useful and novel knowledge from neuroimaging data, especially thanks to the high performing hardware and the innovative algorithms implemented (Chan et al. 2001; Pereira et al. 2009; Sivapriya et al. 2015). Among the different kinds of MRI acquisition, the functional MRI (fMRI) can show a measure of brain activity, by detecting changes associated with blood flow. Cerebral blood flow and neuronal activation were coupled and when an area of the brain was in use, blood flow to that region increases. The detection of the activated regions allowed to understand how the healthy brain works, and in a growing number of studies it was applied for understanding which brain function was disrupted by a particular disease. There were two kinds of investigations that could be done by using the fMRI: the resting state fMRI (rs-fMRI) and the

V. Saccà and A. Sarica contributed equally to this work

✉ Fabiana Novellino
fabiana.novellino@cnr.it

1 Department of Medical and Surgical Sciences, University "Magna Graecia", Catanzaro, Italy

2 National Research Council, Institute of Bioimaging and Molecular Physiology (IBFM), Catanzaro, Italy

3 Institute of Neurology, University Magna Graecia, Catanzaro, Italy

4 Neurology Operating Unit Serraspiga, Provincial Health Authority, Cosenza, Italy

task-based fMRI. In rs-fMRI, the study was focused on the activation of specific brain networks in rest condition and on how they interact with each other. The task-based fMRI enables to monitor how the blood flow varies in the different networks during a specific movement or through a specific exercise. These techniques were used to understand the evolution of neurodegenerative diseases in order to support their early diagnosis by discovering new biomarkers. In particular, rs-fMRI has catalyzed the attention of scientific community with the aim to explore spontaneous neuronal activity in the human brain during rest. A number of rs-fMRI studies have consistently reported the formation of functionally linked resting-state networks during rest. Although using different groups of subjects, different methods and different types of MRI acquisition protocols, these studies have consistently identified the same resting-state networks (i.e. default mode network, visual network), thus demonstrating the robustness of these functionally linked networks within the brain.

Multiple Sclerosis (MS) was a demyelinating disease that affects the central nervous system and disrupts the flow of information within the brain, resulting in a range of signs and symptoms, including motor, cognitive, and sometimes psychiatric problems. Several MS forms exist, with new symptoms either occurring in isolated attacks, with complete symptoms disappearance between attacks (relapsing-remitting forms, R-R) or building up over time (progressive forms). The cause of MS was still unknown, although it was hypothesized the disease was triggered by as-yet- unidentified environmental factor in persons who were genetically predisposed to respond. The course of MS was highly varied and unpredictable, therefore the progress over time and the severity of MS in any one person cannot yet be predicted. A timely and fast diagnosis was crucial in this pathology. In the last years, several classification approaches for distinguishing MS from healthy controls have been applied on the histological and RNA data (Briggs et al. 2010; Goldstein et al. 2010; Keller et al. 2009; Ulrich et al. 2010), on gray matter (GM) alterations (Bendfeldt et al. 2012), on MR parameters (Yamamoto et al. 2010) and diffusion tensor (DT) MRI tractography (Mesaros et al. 2012) or on Magnetic Resonance Spectroscopic Imaging (Ion-Mărgineanu et al. 2017). However, investigations about MS were limited to relatively few publications with fMRI data, which analysis was fundamental to describe alterations in brain functional architecture and their role in disease progression and clinical impairment. In particular, these alterations could discriminate between patients and healthy controls in a predictive setting, helping clinicians in the early diagnosis. Based on these concepts, rs-fMRI offers a promising method to further investigate the functional impact of early stages of MS, in which long-range connectivity can be altered by both inflammatory processes and damages (Richiardi et al. 2012).

The application of the classification task on fMRI data has continually evolved and modified with the interest in this field. The knowledge of these classification methods was fundamental to implement the right and suitable approach. In fact, according to the features number or to the data distribution, the selection of the methodology was different and the application of the wrong algorithm could bring to incorrect results and bad interpretations. Moreover, the choice of the right algorithm could be crucial for the right diagnosis, allowing to distinguish among different neurodegenerative diseases and to avoid important and progressive sides effects, which could be fatal for the patients. The feature selection, performed prior to the classification, allows to explore possible new biomarkers involved in the considered pathology. For example, the weights of the features could be used to understand mechanisms behind the pathology and its clinical course.

This work aimed at evaluating whether several kinds Machine Learning techniques could be effective to support early diagnosis of MS from resting-state functional (rs-fMRI) connectivity data and which of them presented the best performance on these data. In particular, we explored the ability in distinguishing between healthy controls and patients with MS of mean signals extracted from Independent Component Analysis (ICA) corresponding to 15 well-known networks, using five different methods: Random Forest (RF) (Breiman 2001), Support Vector Machine (SVM) (Cortes and Vapnik 1995), Naïve Bayes (NB) (Murphy 2006), K-nearest-neighbor (kNN) (Peterson 2009) and an Artificial Neural Network (ANN) (Dayhoff and DeLeo 2001). Features selection has been performed to each method in order to obtain the ranking of features importance and features weights were then calculated. We trained the different classifiers on the most important features and we evaluated the 5-fold cross-validation accuracies (with and without feature selection), in order to verify the differences between the various techniques.

The present paper was organized as follows: a presentation of materials, methods and of concepts of machine learning classification in which we described the different methodologies applied and their parameters. Finally, we reported the results and we discussed about the comparison among the algorithms performance.

Materials & methods

Patients

Eighteen consecutive MS patients (18 early-RR) were recruited from the Neurological Unit of the University 'Magna Graecia' of Catanzaro and enrolled in this study, matched for demographic variables with 19 healthy controls.

The difference in sex distribution between groups was evaluated with the chi-square test, while the differences in the mean of age between MS and control groups was evaluated using unpaired two sample t tests. Statistical analysis was performed with R language vs3.3.2 (Team 2000).

The RR-MS patients fulfilled the revised McDonald diagnostic criteria (Polman et al. 2011). Each patient was assessed clinically by an experienced neurologist to determine the Expanded Disability Status Scale (EDSS) score.

Eligibility criteria were a diagnosis of Relapsing–Remitting Multiple Sclerosis with an onset of symptoms no more than 36 months before the time of screening (The CAMMS223 Trial Investigators 2008); no clinical relapses for at least three months prior to study entry; no assumption of steroids, or disease-modifying therapies in the three months before recruitment; no history of traumatic brain injury, past or current history of substance abuse, or other coexisting medical conditions. Inclusion criteria for healthy subjects were: no previous history of neurological or psychiatric diseases; normal MRI of the brain (as assessed by structural MRI scanning) and no assumption of drugs acting on the central nervous system. All participants provided written informed consent and the study was approved by the local institutional review board.

The demographic and clinical characteristics of MS group were summarised in Table 1.

MRI acquisition

MRI were acquired by a 3 T scanner with 8 channel head coils (Discovery MR-750, GE, Milwaukee, WI, USA), including: (a) whole-brain T1-weighted (SPGR; TE/TR = 3.7/9.2 ms, flip angle 12°, voxel-size $1 \times 1 \times 1$ mm^3); (b) conventional T2-weighted; (c) resting-state functional MRI (rs-FMRI), 200 volumes of a repeated gradient-echo echo planar imaging sequence (TR/TE: 2000/25 msec; thickness/gap = 3/0.8 mm).

MRI pre-processing

LST-toolbox v.2.0.15 was used for the automated white matter (WM) lesion segmentation and filling on the corrected T1 and FLAIR, non-linearly co-registered, with an initial threshold kappa = 0.15 and the belief map set to GM. Estimated total lesion load (*ml*) and number of lesions were calculated.

All rs-FMRI were pre-processed using tools from the FMRIB's Software Library v.5.0 as follows: (1) discarding of the first 5 volumes to remove T1 equilibrium effects, (2) skull-stripping of images, (3) motion and slice-time correction, (4) denoising with high-pass temporal filter (128 s), (5) spatial smoothing with a Gaussian kernel of FWHM 8 mm. No subject had more than 1 mm maximum translation or 1 mm of maximum rotation in each axis. No statistical significant differences (t-test, $p < 0.05$) were found between the two groups in the mean Euclidian distance and the mean Euler angle.

WM and CSF masks were created by T1 segmentation and then applied to each rs-FMRI to extract the signals. WM and CSF signal together with 6 motion parameters were regressed out from the time series. Following the removal of these nuisance variables, the residual time series for each scan were demeaned and co-registered to the T1 normalized on the MNI 152 standard space.

We applied an independent component analysis (ICA) on this pre-processed rs-FMRI with the GIFT toolbox v.3.0b using the Infomax approach (Rachakonda et al. 2007). Previous studies demonstrated the reliability of the ICA analysis for exploring subtle alterations in MS (Sbardella et al. 2015). Indeed, it has been widely demonstrated that ICA networks changes correlated with clinical MS parameters (Sbardella et al. 2015). The strength of ICA, which is a data-driven and whole-brain approach, lies in its intrinsic characteristics to reveal dynamics for which temporal model is not available (Calhoun et al. 2009). Moreover, ICA was designed to separate a multivariant signal in its sub-components, thus allowing to simplify the analysis of a complex of signals (Sbardella et al. 2015). Compared to the region-of-interest (ROI) approach, ICA overcame the relative arbitrariness of the ROI selection, since it is used without any a prior hypothesis, (Sbardella et al. 2015).

An intermediate model order (number of components = 21) was chosen to achieve a balance between robustness of component spatial maps and the number of components extracted. All these components were stable across multiple runs of

Table 1 Demographic and clinical characteristics of MS and CTRL groups

	MS ($N = 18$)	CTRL ($N = 19$)	*P-Value*
Gender (M/F)	9/9	9/10	1[a]
Mean Age (Years)	34.17 (22–49)	37.42 (21–64)	0.43[b]
Mean Age At Onset (Years)	32.5 (20–45)	–	
Disease Duration (Months)	17.83 (3–36)	–	
Mean EDSS	2.13 (1–4.5)	–	
Mean Lesion Load MRI (Ml)	6.37(28.35–0.141)	–	

MS Multiple Sclerosis *CTRL* Controls *EDSS* Expanded Disability Status Scale *MRI* magnetic resonance imaging

[a] χ^2 -test;

[b] Unpaired two sample t-tests

independent component decomposition. In particular, the plugin ICASSO,[1] provided in the GIFT toolbox, was used for identifying the most reliable and stable components across 20 iterations with different initial conditions and bootstrapped data sets. The ICASSO quality index obtained was of 0.97, indicating a reliable solution (Himberg et al. 2004).

Fifteen mean components were visually identified by two experts. We first applied voxelwise one-sample t-test on all the 15 mean components including all the subjects, both RR-MS and healthy controls. The t-maps obtained from on-sample t-test were automatically thresholded by GIFT with a default value of 1.5, and then they were binarized, thus resulting in masks for extracting the mean signal of the 15 networks for each subject. For automatically extracting these data, we used the toolbox MarsBaR for Matlab 2013a (Brett et al. 2002).

Machine learning analysis

In a machine learning classification task, the data were used to predict a category or a label. This process is also called supervised learning, in which an algorithm uses information received in input to create a model that is able to assign the right label to each observation. An algorithm that implements classification is known as a classifier. A classifier allows to group observations in classes and to identify differences between them. By defying classes labels, a model can be designed to assign data samples into classes. An example of how a Machine Learning classifier works is reported in Fig. 1.

The choice of the right Machine Learning algorithm to use depends by several factors, i.e. size, quality and nature of features. Each methodology creates a different classification method in according to its nature. In the present study, machine learning analysis were conducted on a dataset composed of 37 rows (subjects) and 15 features (mean signal in the network) with Matlab 2013a and R language (3.3.2). Five different classifiers were build: Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes (NB), k-Nearest Neighbor (kNN) and Artificial Neural Network (ANN).

We implemented first a RF algorithm to consider the tree decision category and a SVM algorithm with a Radial Kernel to consider a non-linear category. Subsequently, we performed a linear classifier, NB, since it is one of the more popular statistical technique and a kNN classifier, which is based on the Euclidean distance between points and it is a non-linear approach. Finally, we built an Artificial Neural Network, which represents a non-linear structure of statistical data organized as modeling tools able to simulate complex relationship between inputs and outputs that other analytic functions could not represent.

[1] ICASSO Toolbox, http://research.ics.aalto.fi/ica/icasso/

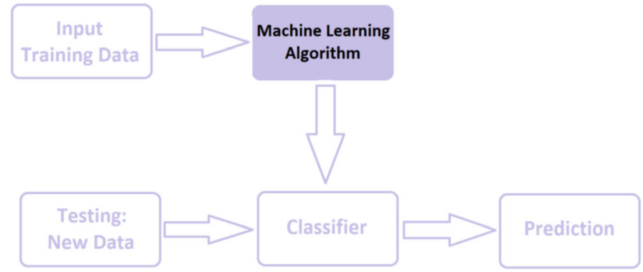

Fig. 1 Example of machine learning classifier. Starting by the input dataset with class labels, an algorithm of Machine Learning is applied to create the right model to classify the different labels (Classifier block). To evaluate the accuracy of the methodology, external testing data is provided to classifier in order to evaluate its performance

Random Forest

Random Forest (Breiman 2001) is a Machine Learning technique is composed by a collection of Classification and Regression Trees (CART) and returns as outcome the class corresponding to the one with the majority of votes (see Fig.2 (a)).

In this work, RF was built with 10.000 trees in the forest (ntree) and a recommended value for the number of variables considered at each split of mtry = √ *features number*. Number of features was 15 and thus mtry was 4. Out-of-bag (OOB) error was also calculated.

Support vector machine

Support Vector Machine (Cortes and Vapnik 1995) is machine learning tool applied in many fields like EEG signal classification, cancer identification, bioinformatics, seizure prediction, face recognition and speech disorder (Bhuvaneswari and Kumar 2013). SVM allows to construct the optimal hyperplane with largest margin for separating data between two groups (see Fig. 2b). For two-dimensional data, a single hyperplane is enough to separate the data into two groups such as +1 or − 1. When data were nonlinear and they show more complexity, such as biomedical data, this approach could be not adequate, since the data distribution were casual and a line fails to separate classes. Moreover, using a linear approach on complex data leads to several errors in the classification step.

In case of data not linearly separable, these were converted into higher dimensional mapping for classification, called Feature Mapping and its mapping has a particular function Φ(xi), as in (1). To find the value of Φ, a Kernel functions can be used.

$$\mathrm{xiT\ xj} = \mathrm{K(xi, xj)} = \Phi(\mathrm{xi})\mathrm{T}\ \Phi(\mathrm{xj}) \quad (1)$$

$K(x_i,x_j)$ is the Kernel functions which is based on the inner product of two variants x_i and x_j. In original space dot product

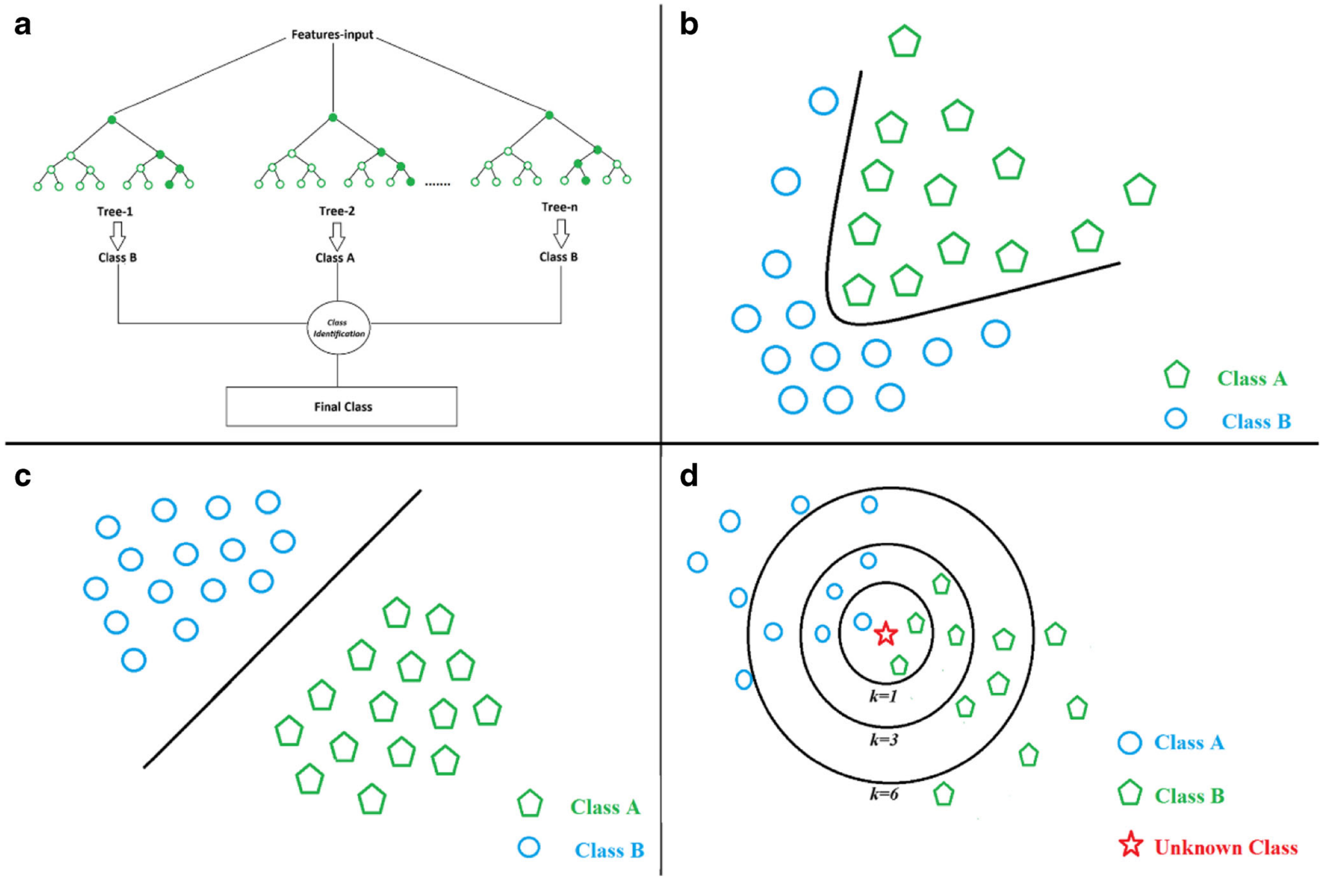

Fig. 2 **(a)** example of decision trees composed a RF algorithm; **(b)** example of a kernel radial in SVM classification; **(c)** example of NB in which the data are linearly separable; **(d)** example of kNN, in which are showed the different classification based on the k value

of $x_{i,}.x_j$ is used for calculation and it is converted into higher space can be replaced dot products as kernel function.

One of most popular function kernel is the radial kernel function (RBF). Considering two data samples defined with x and μ, which represented as feature vectors in some input space, a RBF kernel is defined as:

$$\mathrm{K}(\mathrm{xi},\mathrm{xj}) = e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{2}$$

Where $(x\text{-}\mu)^2$ is the squared Euclidean distance between the two features vectors.

To implement a RBF-SVM algorithm, two parameters have been established: boxcostraint C and gamma γ. The parameter γ defines how far the influence of a single training example reaches, while the boxconstraint C controls the classification and the misclassification, due to data overfitting.

For SVM algorithm performed for our study, a radial kernel was used and an iterative algorithm have set the better values for boxcostraint C and gamma γ as equal to 7.

Naïve Bayes

Naïve Bayes (Murphy 2006) classifier is the one of easier, since it consists into draw a simple line to discriminate the considered classes (see Fig. 2c). Moreover, it is a more popular statistical technique applied for the classification in data mining issues. It is based on Bayes theorem with independence assumption between the features:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3}$$

Bayes theorem allows to calculate the probability of a random event A attends knowing that an event B occurred. The probability a priori of A, B and B influenced by A must be known. NB classifiers were highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. The Naive Bayes Classifier technique is based on the so-called Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Despite its simplicity, NB can often outperform more sophisticated classification methods. NB has proven its effective application, often reported as "surprisingly" accurate, in text classification, medical diagnosis and systems performance management (Pang et al. 2002). Generally, a NB classifier is not suitable for non-linear problem since the classes were not linearly separable.

K-nearest neighbors

K-nearest-neighbor classification is one of the simplest classification methods. It should be one of the first choices for a classification study when there is little or no prior knowledge about the distribution of the data. K-nearest-neighbor classification was developed from the need to perform discriminant analysis when reliable parametric estimates of probability densities were difficult to determine (Peterson 2009).

KNN methods consist to compare every unseen point x through a distance function dist (x, x_i) to all points x_i of the training set. The k minimal distances were performed and the majority over the corresponding labels y_i is taken as the resulting label for x. An example of kNN algorithm is reported in Fig. 2d. The principal issue of the kNN method is the number of data points and the distance. In fact, a kNN method work well, if a reasonable distance (typically the Euclidean one) is available and if the number of data points in the training set is not huge.

Efficiency of the kNN algorithm largely depends on the value of k (i.e., choosing the number of nearest neighbors). The optimal value for k can be selected by cross-validation, but it depends also by first inspecting the data. In general, a large k value is more precise as it reduces the overall noise. In fact, a low signal-to-noise ratio requires larger values of k (Lemm et al. 2011). The k values usually have been between 3 and 10 for most datasets.

In this case, kNN algorithm was developed with a k value of 7, which represented the value with the better classification accuracy.

Artificial neural network

An artificial neural network is a learning algorithm, inspired by the structure and functional aspects of biological neural networks (Dayhoff and DeLeo 2001). In a biological network, neurons receive inputs from other neurons (or sensing cells) and then they send an output to the other cells that were connected to it. In a computational ANN, the process is the same and the elaborations were structured as an interconnected group of artificial neurons, which process information. Elaborated information is transmitted to linked neurons.

An artificial neuron usually has many inputs and one output. The neuron has two modes of operation: the training mode and the using mode. In the training mode, the neuron could be trained to fire, i.e. it is ready to transmit commands and information, while in the using mode, in presence of input pattern, its associated output becomes the current output. If the input pattern does not belong in the taught list of input patterns, the firing rule is used to determine whether to fire or not.

Neurons were connected in layers, and signals travel from the first (input) to the last (output) layer. Modern neural network projects typically have a few thousand to a few million neural units and millions of connections.

ANN could be used in several fields and one of its most important application is the pattern recognition, implemented by using a feed-forward neural network. In this configuration, in the training phase, the network is trained to associate outputs with input patterns (see Fig. 3). When the network is used, it identifies the input pattern and tries to output the associated output pattern. The implemented ANN in this paper was built with 3 hidden layers and a weights decay of 0.1.

Feature selection

Variable and feature selection represent a powerful tool for improving the prediction performance and for better discovering the knowledge that hides in high- dimensional datasets (Guyon and Elisseeff 2003). Indeed, a known issue in classification task is to minimize the so-called curse of dimensionality, which could lead to the risk of overfitting and misclassification of records (Guyon et al. 2002). This problem gets worse when the feature space is larger than the number of records. Given a particular classification model, it is very useful to select the best subset of features, based on variables ranking, which improves the overall performance of the prediction.

For all these reasons, we applied different feature selection approaches, depending on the classifier itself, to obtain a ranking of the most predictive networks. We used the intrinsic feature selection of RF, based on the *Gini index*, the recursive feature elimination (*rfe*) for the SVM and kNN and the features weights calculation with the function "*varImp*" present in *caret* package for the other algorithm.

Recursive feature elimination (Guyon et al. 2002; Guyon and Elisseeff 2003) consists to assign weights to features (e.g. the coefficients of a linear model), given an external estimator. It is an iterative method to select features by repetitively considering smaller and smaller sets of features. Initially, the estimator is trained on the all set of features, calculating weights for each one of them. The predictors were ranked and the less

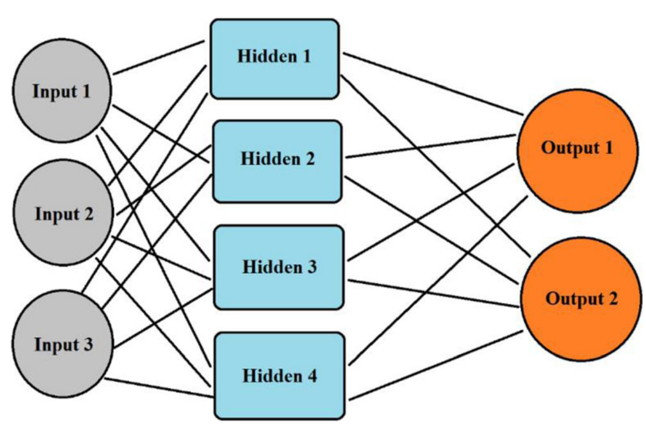

Fig. 3 internal architecture of an ANN with the hidden layers

important ones were sequentially eliminated prior to modeling. The goal is to find a subset of predictors that can be used to produce an accurate model.

For the linear classifier (NB), method returns the absolute value of the t–statistic for each model while for ANN, method is based on Gevrey et al. (2006), which uses combinations of the absolute values of the weights. Finally, we re-built each of the five classifiers only on the most important features.

Cross validated accuracy

For evaluating both the feature selection task and the classification performance, we used a 5-fold cross-validation, splitting the dataset into 80% training set (30 subjects) and 20% testing set (7 subjects). A k-fold cross-validation is a method of evaluating and comparing learning algorithms by dividing data into two parts: one used to learn or train a model and the other used to validate the model. Dataset is first partitioned into k equally (in our case 5) folds and k iterations of training were performed. Within each of these iterations, a different fold of the data is held-out for the testing operation while the remaining $k-1$ folds were used for training the classification model (Refaeilzadeh et al. 2009). Thus, variance of the data has reduced, avoiding possible overfitting or bias errors. Moreover, methods of k-fold-cross-validation could be used to compare the performance of different machine learning models on the same data set, allowing to choose the better algorithm for the considered data.

In this paper, we used the 5-fold cross-validation to calculate and to compare the different classification performance both using all features and using only the most important features obtained through the feature selection.

Results

Clinical findings

Demographic and clinical characteristics of patients have been elaborated to investigate the possible dissimilarity between MS patients and healthy subjects. Significant differences have not been detected, in particular, age and gender distributions were not statistically different between the two groups.

MRI findings

The mean lesion load for the MS was 5.25 ml, with a mean number of lesions of 15. By visually inspecting ICA-derived components of rs-fMRI, we identified 15 networks, reported in Fig. 4.

In particular, we found the following networks: (i) sensori motor I and II (see Fig. 4A and B), including somatosensory (postcentral gyrus) and motor (precentral gyrus) regions and extends to the supplementary motor areas; (ii) primary and secondary visual (see Fig. 4c and d) composed by occipital and lingual gyri, the parietal lobe, the fusiform and inferior temporal gyri; (iii) auditory (see Fig. 4e), including mostly primary auditory cortex; (iv) working memory right and left that were lateralized networks (see fig. 4f and g) consisting of superior parietal and superior frontal regions; (v) executive (see fig. 4h) including the midline frontal areas, with the anterior cingulate gyrus, SMA, and portions of the basal ganglia; (vi) attention (see Fig. 4j) consisting of temporal and parietal cortices; salience (see Fig. 4k) composed by the anterior cingulate cortex, presupplementary motor area, and anterior insulae; (vii) anterior and posterior default mode (see fig. 4l and m) composed by precuneus, medial frontal and posterior cingulate cortex; (viii) ventral stream (see Fig. 4n) including mostly medial temporal lobes; (ix) cerebellum (see Fig. 4o) composed by cerebellar lobes; (x) basal ganglia (see Fig. 4p) that contains the subcortical nuclei, situated at the base of the forebrain.

Machine learning algorithms on all features

The five classifiers were then trained on all 15 ICA networks mean signal as previous described. The 5-fold cross-validation accuracies were reported in Table 2. The classification parameters (accuracy, sensibility, specificity, positive predictive and negative predictive value) were calculated as the mean obtained in the different five folds. SVM, RF and kNN showed a comparable accuracy (SVM: 63.3%, RF: 56.5% and kNN 63.2%). ANN represented the algorithm with the major accuracy (69.9%), while NB presented the worse performance (46.6%).

Feature selection

To increase these results, we applied the feature selection operation to calculate the features weights. In all five cases in our study, we obtained the same feature, which represent the *sensori-motor I network*. In SVM, NB and kNN the weight of this network was 0.8 (while the second feature was cerebellum with 0.68), in ANN it was 11.178 (the second feature was the salience network with 10.287) and finally in RF it was 2.8 (the second feature was the sensori-motor II with 1.3). These results were reported in Fig. 5. It could be also noticed that similar importance values were obtained for the sensori-motor II, cerebellum and working memory networks.

Machine learning algorithms on the sensori-motor I

In presence of these results, we have performed new classifiers using only the sensori-motor I to evaluate its possible predominance and effect in early MS classification using fMRI data. The classification parameters were reported in

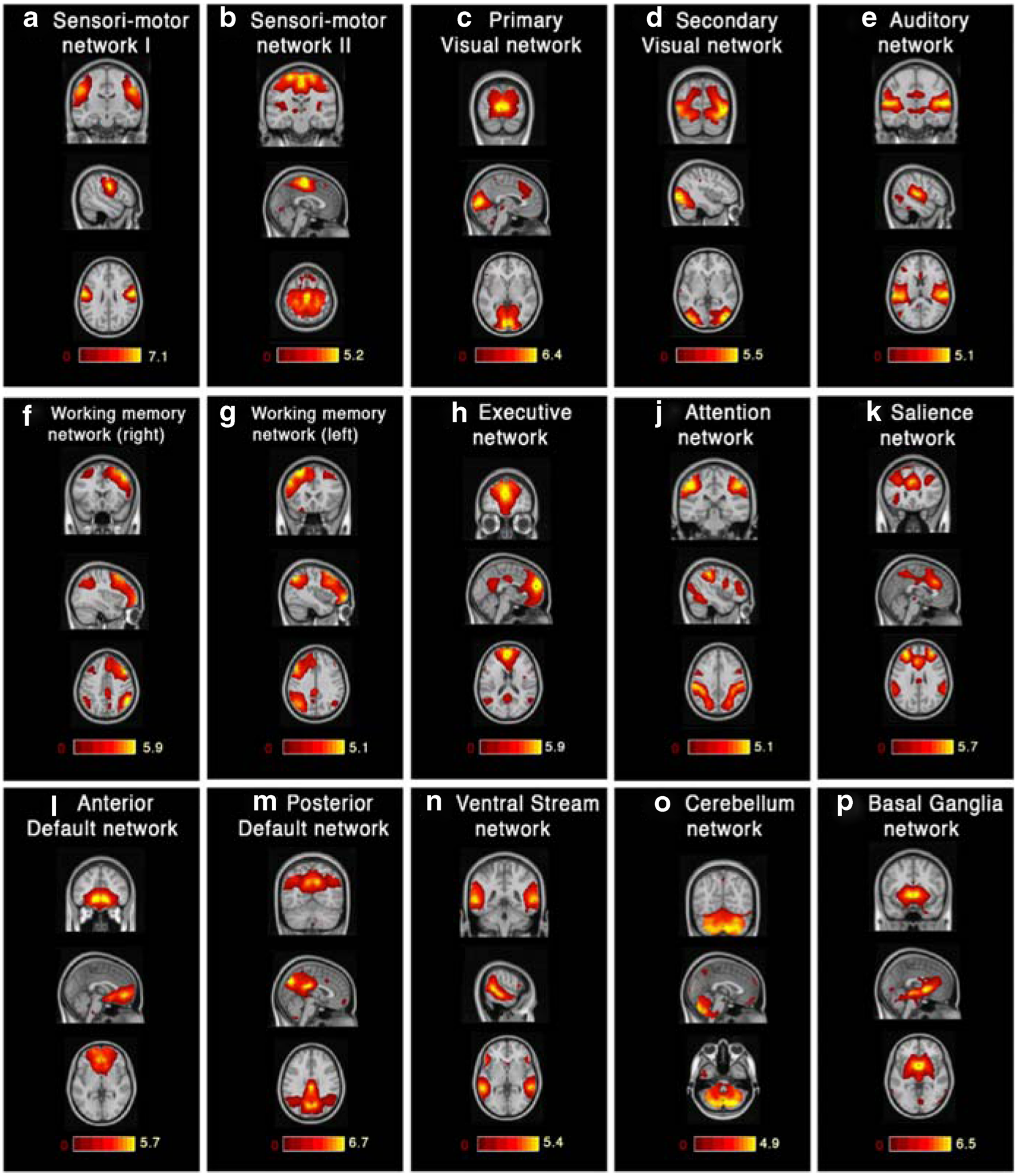

Fig. 4 The 15 networks generated by GIFT: **(a)** sensori-motor network I; **(b)** sensori-motor network II; **(c)** Primary Visual network; **(d)** Secondary Visual network; **(e)** Auditory network; **(f)** working memory network (right); **(g)** working memory network (left); **(h)** executive network; **(j)** attention network; **(k)** salience network; **(l)** anterior default network; **(m)** posterior default network; **(n)** ventral stream network; **(o)** cerebellum network; **(p)** basal ganglia network. The color bars show z-score

Table 3. SVM and RF presented about the same 5-fold cross-validation accuracies of the classifiers with all features. On the other hand, NB, kNN and ANN presented a slightly increase of the accuracies using only a feature. All accuracies were increased compared to the classifiers with all features. NB, kNN and ANN showed the same accuracy of 71.42%. However, in particular, with only this network, RF and SVM classifiers reached an accuracy of 85.7% to demonstrate the strength of the sensori-motor I in discriminating between the two groups. More interestingly, the only misclassified patient, with these two classifiers, resulted to have the lowest value of lesion volume (0.141 *ml* and 2 lesions).

Discussion

In this paper, we trained five different Machine Learning algorithms on functional connectivity data extracted from ICA

Table 2 Five-Fold Cross Validation Accuracies, Sensibilities, Specificities, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) calculated for each classifier, using all networks

Classification performance using all networks				
Algorithm	Accuracies	Sensibility	Specificity	PPV-NPV
RF	56.5%	53.3%	60%	3–3
SVM	63.3%	60%	80%	4–2
NB	46.6%	46.6%	46.6%	3–3
kNN	63.2%	66.6%	60%	4–2
ANN	69.9%	53.3%	53.3%	3–3

Table 3 Accuracies, sensibilities, specificities, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) calculated for each classifier, using only Sensori Motor I network

Classification performance using only sensori motor I network				
Algorithm	Accuracies	Sensibility	Specificity	PPV-NPV
RF	85.7%	100%	66.7%	6–1
SVM	85.7%	100%	66.7%	6–1
NB	71.42%	50%	100%	5–2
kNN	71.42%	50%	100%	5–2
ANN	71.42%	50%	100%	2–3

networks, with the aim of evaluating how well they performed in discriminating between healthy controls and MS patients for the support of the early diagnosis of MS from rs-fMRI data. In particular, we applied RF, SVM, NB, kNN and ANN approaches on the same dataset consisting of ICA networks values. Features selection was then performed in each classifier and results were compared.

All five classifiers, trained on all the features, showed very poor 5-fold cross-validation accuracies (see Table 2). The performance values were very different among classifiers since they derived from different algorithm approaches. RF and

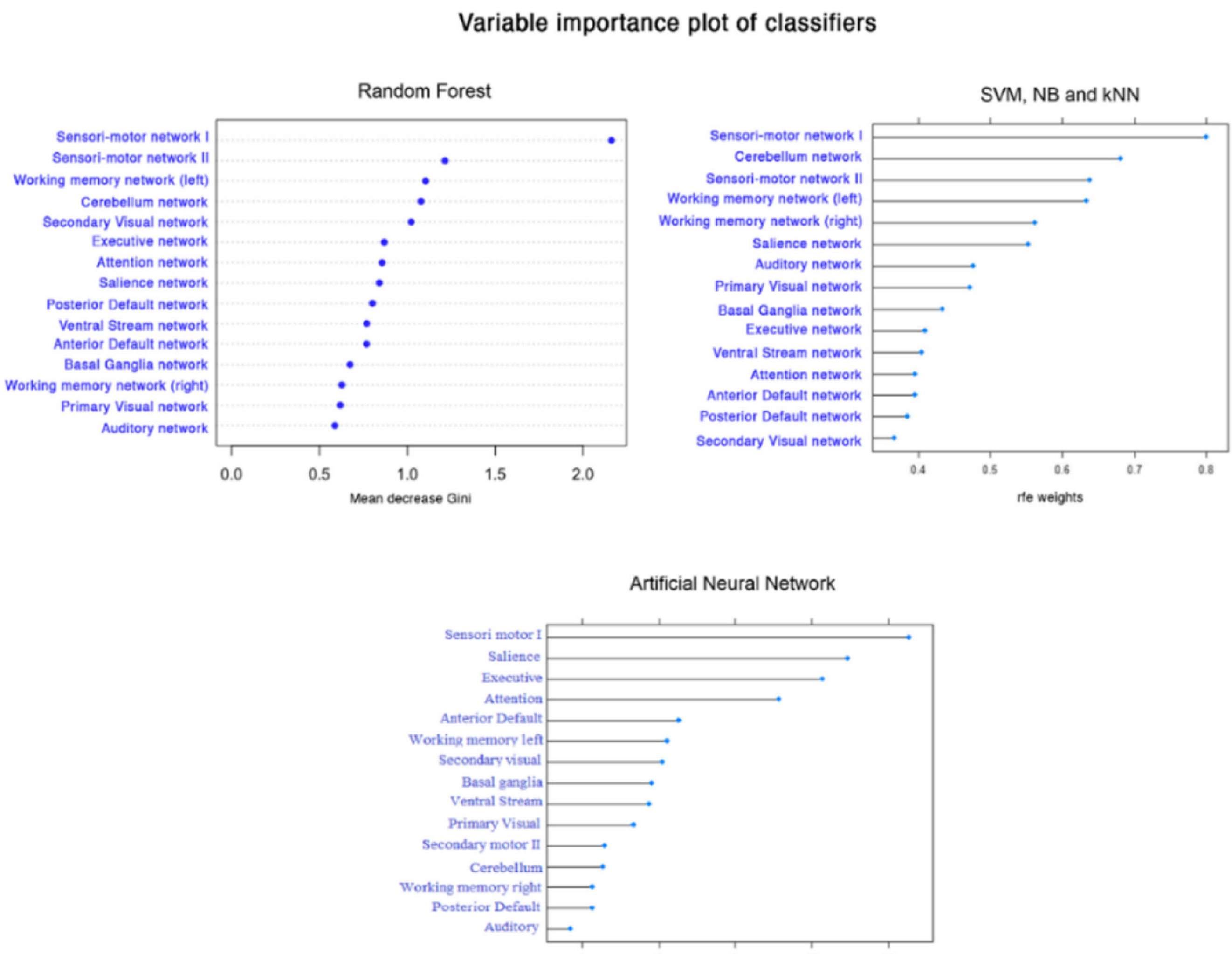

Fig. 5 features selection applied on the five algorithms. In each case, the most important network is the sensori-motor I

SVM presented similar accuracies. It was very interesting since they were the most popular Machine Learning algorithms used in practice (Kumar and Thenmozhi 2006; Liu et al. 2013; Statnikov et al. 2008). Usually, SVM was used in presence of a small amount of data, which could be reasonably clean, outlier free and when the features were correlated, while RF works better in presence of a larger number of instances to have a well randomization to reduce the presence of any outliers. In our case, on the 5-fold cross-validation accuracies of SVM and RF were comparable thanks to the pre-processing operations in which the signals were submitted to several denoising steps in order to eliminate various artefacts, i.e. WM and CSF masks and head motion. Moreover, the features amount was small and probably RF was slightly disadvantaged. On the contrary, NB presented a poor result (46.6%), probably linked to the linear nature of NB, while kNN (63.2%) and ANN (69.9%) were comparable to SVM and RF performance.

Moreover, we applied the feature selection in order to obtain the weights of features and to increase these accuracies. The most important features for each classifier was *sensori motor I* network. Thus, we build new classifiers using only this network for the classification. All accuracies were increased respect to the previous analyses demonstrating the strength of the sensori-motor I in discriminating between the two groups. NB, kNN and ANN showed the same accuracy (71.4%). In particular, with only this network, RF and SVM classifiers reached an accuracy of 85.7%. It was worth of noting that the only misclassified subject was the patient with the lowest value in the lesion load (0.141 ml). Interestingly, this misclassification could derive from the early stage of the disease in this patient, since a so *low lesion load* might cause imperceptible changes in functional connectivity, which results undetectable for the algorithms.

Considering these results on the rs-fMRI connectivity data, we found that algorithms as NB, kNN and ANN did not represent the most accurate approach could be used with this type of data, for different reasons derived by the nature of them. Linear classification (NB) was not suitable to elaborate complex data as biomedical ones. kNN suffered the change subject number to classify. ANN instead worked better in presence of very big data and with only 37 subjects it was not able to find the right classification strategy. Moreover, these three algorithms showed a good specificity in recognizing all patients, but their application was not efficient to identify healthy controls. Instead, sensibility of classifying controls values were poor. On the other hand, RF and SVM algorithms showed comparable and good results in each classification parameters, which could be useful in clinical practice. Usually, these approaches were not in tune and in particular several studies showed that SVM worked better with unimodal data and on the contrary RF gave a better accuracy with multimodal data. Moreover, RF usually improved its performance with the features selection phase, building the classifier on the most important discriminant characteristics. Thus initially, in this study we thought that SVM could give a better performance, using only the rs-fMRI data.

Interestingly, all five considered algorithms showed the sensori-motor I as most important feature. This finding was in according to the early manifestation of motor/sensorial deficits in MS and to recent studies found in literature. In fact, several studies showed an abnormal connectivity in this network in presence of MS disease (Lowe et al. 2008; Rocca et al. 2009). It depended probably by the robustness of pre-processing methodology, consisting in the removal of WM and CSF signal together with 6 motion parameters from the time series. The signal cleaning allowed to represent the significant network for the pathology, in according to the characteristic early symptoms of the disease, which were motor/sensorial deficits. In particular, by using only the sensori-motor I as feature to create the classification, RF and SVM showed the same accuracy. These values demonstrate both the robustness of pre-processing methodology and the possibility to facilitate the diagnosis operation in the case of early MS. Moreover, the ranking of variables importance, calculated with this approach, has an anatomo-clinical correspondence with a specific biological meaning in the MS evolution.

We believe that our robust approach could reveal subtle changes in functional connectivity and that can be in the future translated into the clinical realm for diagnosis and prognosis.

The results of this study should be interpreted in the context of some limitations, in particular the relatively small sample size. We recognized that this aspect could be a restriction for the study, in fact this work could be considered as a preliminary investigation in this field. Despite the low dimension of our cohort, we applied important steps for the signals cleaning. In fact, we used WM and CSF masks, calculated for each subject, in order to remove their negative effect on the analysis. Moreover, we performed a fundamental evaluation on the subject involuntary head motion parameters, calculating for each subject several characteristics, as root mean squared (rms) and Eular Angle, and performing an outlier analysis in order to investigate any high motion value that could distort the ICA components. After these operations, we extracted temporal signal noise ratio (tSNR) for each subject and its mean value was 408.76, which showed a good signal cleaning. Moreover, our results were in according to the literature, since the importance of sensori motor I network in MS patients was previously demonstrated (Mezzapesa et al. 2008, et Rocca et al. 2009). Moreover, we should address another possible weakness of our study, related to the extraction of the mean signal of the networks. It was well known that simply averaging across a region, with a large number of voxels as in our masks, could include noise derived by the small number of activated/deactivated voxels (Poldrack 2007). This was mostly true when classical statistical methods were used for comparing groups. On the other hand, our machine learning approaches have shown their capability of handling noisy data from functionally coherent regions and this was demonstrated by the good accuracies we reached also in a small sample.

Conclusion

In the last few years, the application of Machine Learning in Resting State data became very popular and interesting. However, a better knowledge of these algorithms was necessary to implement the correct approach for the considered problem in according to the features number and to the data distribution. Thus, based on the considered disease, the right algorithm should be different. The present study was designed for evaluating whether several kinds of Machine Learning techniques could support early diagnosis of MS from rs-fMRI connectivity data. For this goal, we implemented five different classifiers with the following algorithms: SVM, RF, NB, kNN and ANN. The features considered derived from the analysis of 15 well-known networks from the resting-state fMRI using the tool MELODIC together with GIFT. We showed that, with all different classification algorithms in association with the feature selection, the best discriminant network between controls and early MS, was the sensori-motor I. In particular, RF and SVM seemed to represent the better approach in this case, since they presented the same accuracies using only this variable, highlighting the robustness of preprocessing methodology. It could be also noticed that similar importance values were obtained for the sensori-motor II, cerebellum and working memory networks. These findings were in according to the early manifestation of motor/sensorial deficits in MS.

In conclusion, we believe that our robust approach could reveal subtle changes in functional connectivity and that can be in the future translated into the clinical realm for diagnosis and prognosis.

Compliance with ethical standards

Disclosure This study was not financially supported.

Conflict of interest All of the authors reported no biomedical financial interests or potential conflicts of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Bendfeldt, K., Klöppel, S., Nichols, T. E., Smieskova, R., Kuster, P., Traud, S., et al. (2012). Multivariate pattern classification of gray matter pathology in multiple sclerosis. *Neuroimage, 60*(1), 400–408.

Bhuvaneswari, P., & Kumar, J. S. (2013). Support vector machine technique for EEG signals. *International Journal of Computer Applications, 63*(13).

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). Region of interest analysis using an SPM toolbox. In *8th international conference on functional mapping of the human brain*. Japan: Sendai.

Briggs, F. B., Bartlett, S. E., Goldstein, B. A., Wang, J., McCauley, J. L., Zuvich, R. L., et al. (2010). Evidence for CRHR1 in multiple sclerosis using supervised machine learning and meta-analysis in 12 566 individuals. *Human Molecular Genetics, 19*(21), 4286–4295.

Calhoun, V. D., Liu, J., & Adali, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage, 45*, S163–S172.

Chan, J. C. W., Chan, K. P., & Yeh, A. G. O. (2001). Detecting the nature of change in an urban environment: A comparison of machine learning algorithms. *Photogrammetric Engineering and Remote Sensing, 67*(2), 213–226.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Dayhoff, J. E., & DeLeo, J. M. (2001). Artificial neural networks. *Cancer, 91*(S8), 1615–1635.

Douglas, P. K., Harris, S., Yuille, A., & Cohen, M. S. (2011). Performance comparison of machine learning algorithms and number of independent components used in fMRI decoding of belief vs. disbelief. *Neuroimage, 56*(2), 544–553.

Formisano, E., De Martino, F., Valente, G. (2008). Multivariate analysis of fMRI time series: Classification and regression of brain responses using machine learning. Magnetic Resonance Imaging, 26(7), 921–934.

Gevrey, M., Dimopoulos, I., & Lek, S. (2006). Two-way interaction of input variables in the sensitivity analysis of neural network models. *Ecological Modelling, 195*(1), 43–50.

Goldstein, B. A., Hubbard, A. E., Cutler, A., & Barcellos, L. F. (2010). An application of random forests to a genome-wide association dataset: Methodological considerations & new findings. *BMC Genetics, 11*(1), 49.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*(Mar), 1157–1182.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning, 46*(1–3), 389–422.

Himberg, J., Hyvärinen, A., & Esposito, F. (2004). Validating the independent components of neuroimaging time-series via clustering and visualization. *NeuroImage, 22*(3), 1214–1222.

Ion-Mărgineanu, A., Kocevar, G., Stamile, C., Sima, D. M., Durand-Dubief, F., Van Huffel, S., Sappey-Marinier, D. (2017). Machine learning approach for classifying multiple sclerosis courses by combining clinical data with lesion loads and magnetic resonance metabolic features. Frontiers in Neuroscience, 11, 398.

Keller, A., Leidinger, P., Lange, J., Borries, A., Schroers, H., Scheffler, et al. (2009). Multiple sclerosis: microRNA expression profiles accurately differentiate patients with relapsing-remitting disease from healthy controls. PLoS One, 4(10), e7440.

Kumar, M., & Thenmozhi, M. (2006). Forecasting stock index movement: A comparison of support vector machines and random forest. Indian Institute of Capital Markets 9th Capital Markets Conference. http://ssrn.com/abstract, =876544.

LaConte, S., Strother, S., Cherkassky, V., Anderson, J., & Hu, X. (2005). Support vector machines for temporal classification of block design fMRI data. *NeuroImage, 26*(2), 317–329.

Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K. R. (2011). Introduction to machine learning for brain imaging. *Neuroimage, 56*(2), 387–399.

Liu, M., Wang, M., Wang, J., & Li, D. (2013). Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical, 177*, 970–980.

Lowe, M. J., Beall, E. B., Sakaie, K. E., Koenig, K. A., Stone, L., Marrie, R. A., & Phillips, M. D. (2008). Resting state sensori-motor functional connectivity in multiple sclerosis inversely correlates with transcallosal motor pathway transverse diffusivity. *Human Brain Mapping, 29*(7), 818–827.

Mesaros, S., Rocca, M. A., Kacar, K., Kostic, J., Copetti, M., Stosic-Opincal, T., et al. (2012). Diffusion tensor MRI tractography and cognitive impairment in multiple sclerosis. *Neurology, 78*(13), 969–975.

Mezzapesa, D. M., Rocca, M. A., Rodegher, M., Comi, G., et al. (2008). Functional cortical changes of the sensorimotor network are associated with clinical recovery in multiple sclerosis. *Human Brain Mapping, 29*(5), 562–573.

Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage, 53*(1), 103–118.

Mourão-Miranda, J., Bokde, A. L., Born, C., Hampel, H., & Stetter, M. (2005). Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. *NeuroImage, 28*(4), 980–995.

Murphy, K. P. (2006). *Naive bayes classifiers*. University of British Columbia.

Pang, B., Lee, L., Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10 (pp. 79–86). Association for Computational Linguistics, https://doi.org/10.3115/1118693.1118704

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage, 45*(1), S199–S209.

Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia, 4*(2), 1883.

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience, 2*(1), 67–70.

Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., et al. (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology, 69*(2), 292–302.

Rachakonda, S., Egolf, E., Correa, N., Calhoun, V. (2007). Group ICA of fMRI toolbox (GIFT) manual. Dostupné z http://www.nitrc.org/docman/view. Php/55/295/v1. 3d_ GIFTManual pdf [cit. 2011-11-5].

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). *Cross-validation* (pp. 532–538). Springer US: In Encyclopedia of database systems.

Richiardi, J., Gschwind, M., Simioni, S., Annoni, J. M., Greco, B., Hagmann, P., et al. (2012). Classifying minimally disabled multiple sclerosis patients from resting state functional connectivity. *Neuroimage, 62*(3), 2021–2033.

Rocca, M. A., Absinta, M., Valsasina, P., Ciccarelli, O., Marino, S., Rovira, A., et al. (2009). Abnormal connectivity of the sensorimotor network in patients with MS: A multicenter fMRI study. *Human Brain Mapping, 30*(8), 2412–2425.

Sbardella, E., Petsas, N., Tona, F., and Pantano, P. (2015). Resting-State fMRI in MS: General Concepts and Brief Overview of Its Application. *Biomed Res Int*, 212693.

Sivapriya, T. R., Kamal, A. R., Thangaiah, P. (2015). Ensemble Merit Merge Feature Selection for Enhanced Multinomial Classification in Alzheimer's Dementia. *Computational and mathematical methods in medicine*, 2015.

Statnikov, A., Wang, L., & Aliferis, C. F. (2008). A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics, 9*(1), 319.

Team, R. C. (2000). R language definition. Vienna, Austria: R foundation for statistical computing. http://www.R-project.org/

The CAMMS223 Trial Investigators. (2008). Alemtuzumab vs. interferon Beta-1a in early multiple sclerosis. *The New England Journal of Medicine, 359*, 1786–1801.

Ulrich, R., Kalkuhl, A., Deschl, U., & Baumgärtner, W. (2010). Machine learning approach identifies new pathways associated with demyelination in a viral model of multiple sclerosis. *Journal of Cellular and Molecular Medicine, 14*(1–2), 434–448.

Wang, Z., Childress, A. R., Wang, J., & Detre, J. A. (2007). Support vector machine learning-based fMRI data group analysis. *NeuroImage, 36*(4), 1139–1151.

Yamamoto, D., Arimura, H., Kakeda, S., Magome, T., Yamashita, Y., Toyofuku, F., et al. (2010). Computer-aided detection of multiple sclerosis lesions in brain magnetic resonance images: False positive reduction scheme consisted of rule-based, level set method, and support vector machine. *Computerized Medical Imaging and Graphics, 34*(5), 404–413.