Journal of Affective Disorders 222 (2017) 49–56

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

The relevance of feature selection methods to the classification of obsessive-compulsive disorder based on volumetric measures

Lucas R. Trambaiolli[a,*], Claudinei E. Biazoli Jr.[a], Joana B. Balardin[a], Marcelo Q. Hoexter[b], João R. Sato[a]

[a] Center of Mathematics, Computing and Cognition, Universidade Federal do ABC, Rua Santa Adélia, 166, Santo André, SP 09210-170, Brazil
[b] Department and Institute of Psychiatry, University of São Paulo Medical School, Rua Dr. Ovídio Pires de Campos, 785, São Paulo 01060-970, SP, Brazil

ARTICLE INFO

Keywords:
Magnetic resonance imaging
Machine learning
Obsessive-compulsive disorder
Feature selection

ABSTRACT

Background: Magnetic resonance images (MRI) show detectable anatomical and functional differences between individuals with obsessive-compulsive disorder (OCD) and healthy subjects. Moreover, machine learning techniques have been proposed as tools to identify potential biomarkers and, ultimately, to support clinical diagnosis. However, few studies to date have investigated feature selection (FS) influences in OCD MRI-based classification.

Methods: Volumes of cortical and subcortical structures, from MRI data of 38 OCD patients (split into two groups according symptoms severity) and 36 controls, were submitted to seven feature selection algorithms. FS aims to select the most relevant and less redundant features which discriminate between two classes. Then, a classification step was applied, from which the classification performances before and after different FS were compared. For the performance evaluation, leave-one-subject-out accuracies of Support Vector Machine classifiers were considered.

Results: Using different FS algorithms, performance improvement was achieved for Controls vs. All OCD discrimination (19.08% of improvement reducing by 80% the amount of features), Controls vs. Low OCD (20.10%, 75%), Controls vs. High OCD (17.32%, 85%) and Low OCD vs. High OCD (10.53%, 75%). Furthermore, all algorithms pointed out classical cortico-striato-thalamo-cortical circuitry structures as relevant features for OCD classification.

Limitations: Limitations include the sample size and using only filter approaches for FS.

Conclusions: Our results suggest that FS positively impacts OCD classification using machine-learning techniques. Complementarily, FS algorithms were able to select biologically plausible features automatically.

1. Introduction

Obsessions are defined as a set of intrusive and unpleasant thoughts or mental images that cause significant anxiety. Compulsions, on the other hand, are characterized as a series of repetitive behaviors or mental acts performed to suppress anxiety, distress and/or discomfort. In general, a clinical diagnosis of obsessive-compulsive disorder (OCD) is performed when obsessions and/or compulsions start to interfere with an individual social, occupational or academic functioning and occupy more than 1 h per day (Abramovitch et al., 2013). The heterogeneous clinical presentation of OCD is a challenge to elucidate the neurobiology underlying this complex condition, as well as for biomarkers investigation and development new treatments. The identification vulnerability markers of OCD would be of great value for the understanding of its neurobiology and for early clinical detection. It has been suggested that early detection of OCD might improve effectiveness of treatment (Cherian et al., 2014; Mancebo et al., 2014) and the use of multivariate techniques applied to neuroimaging data have been proved to be an important field of investigation in this context (Mourão-Miranda et al., 2011; Klöppel et al., 2012).

The potential role of structural brain features to the diagnosis and classification of OCD has been evaluated by recent studies using machine learning (ML) techniques. These techniques can actually emulate future clinical situations with leave-one-subject-out (LOSO) approaches, where a new subject is compared to a database to support diagnosis. Moreover, although it has greater variability than methods such as k-folds cross-validation, LOSO is more recommended in experiments with few subjects to avoid effect size inflation (Esterman et al., 2010). ML algorithms extract highly relevant features from neuroimaging data to classify individual diagnostic status (Bishop and

* Corresponding author.
E-mail address: lucasrtb@gmail.com (L.R. Trambaiolli).

http://dx.doi.org/10.1016/j.jad.2017.06.061
Received 15 April 2017; Received in revised form 1 June 2017; Accepted 26 June 2017
Available online 27 June 2017

Nasrabadi, 2006). For example, structural magnetic resonance imaging (MRI) measures of gray and white-matter volumes were used to assess the feasibility of discriminating individual subjects with OCD from healthy controls with classification accuracy (CA) ranging from 75% to 93% across studies (Soriano-Mas et al., 2007; Hu et al., 2016). Hoexter et al. (2013) applied Support Vector Regression to cortical volumes of OCD patients in order to predict symptom's severity and found nodes within the cortico–striato–thalamo–cortical circuit, mainly the left medial orbitofrontal cortex and the left putamen as the most discriminative regions. These preliminary studies were important steps suggesting that ML methods may identify neurobiological markers of OCD based only on structural MRI datasets.

However, given the large number of structures measured by MRI, usually many features are automatically extracted from the imaging data. Some of them can have a marginal relevance or generate redundant information for OCD prediction that could reduce the accuracy of many multivariate approaches. Feature Selection (FS) (Liu and Motoda, 2007), an active research area within the ML communities, aims to find a small number of features that describes the dataset at least as well as the original set of features (Chu et al., 2012). In other words, FS selects the most relevant features according to given criterion, maintaining their original values. Benefits in performing such a pre-processing step in neuroimaging data include a lower computational cost, better predictive performance and better interpretability of ML, as the most prominent features should be related to the disorder biological basis.

The importance of each feature can be measured by several approaches, which are classified according to its interaction with the ML algorithm. One method commonly used in FS algorithms is called filter, usually relying on statistical characteristics of the relations between one (univariate) or more (multivariate) features and classes. This approach is also defined as algorithm-independent (Habeck and Stern, 2010). Alternatively, wrapper algorithms are classifier-dependent approaches, using precision measures achieved by a particular ML algorithm for a given subset of features (Kohavi and John, 1997). Considering this range of possibilities, Tohka et al. (2016) compared different methods for FS of MRI data, and reported that the result of each algorithm depends on the application and sample in question.

In this context, we focused the current study on the employment of seven feature selection algorithms (two multivariate and five univariate methods) for building classifiers for OCD (high severity OCD, low severity OCD and controls) based on structural MRI data. Further we explored how the feature selection outputs were consistent with anatomical sites thought to be implicated in the neurobiology of OCD.

2. Methods

2.1. Subjects

Patients with OCD, diagnosed according to DSM-IV-TR criteria and treatment-naïve, were referred from primary psychiatric services, from the Brazilian Association of Tourette Syndrome and OCD - ASTOC, or were recruited through announcements in the local media. Healthy subjects were recruited among college students, hospital and university staff. The Ethics Committee of the University of São Paulo Medical School, Brazil approved this study and informed consent was obtained from all participants.

The original dataset was divided into two groups with similar age and gender distributions: the first one consisted of 38 OCD patients and the second group was composed of 36 controls (demographic information is presented in Table 1).

Recent studies have found correlations between structural abnormalities and the severity of obsession and compulsion symptoms (for example, using the Yale-Brown Obsessive-Compulsive Scale - Y-BOCS) (Gilbert et al., 2008a; Venkatasubramanian et al., 2012; Hoexter et al., 2013; Tang et al., 2015). In this context, the sample of patients was divided into two sub-groups of 19 individuals each according to the median (25.5) of the symptoms scores (Table 1), allowing to compare between symptoms burden. The first group comprised individuals with a Y-BOCS score > 25.5 (High OCD group) while the second group had a Y-BOCS score < 25.5 (Low OCD group).

Table 1
Demographic data, including ages (mean ± standard-deviation), number of males and females and Y-BOCS scores (mean ± standard-deviation) of each group and subgroup.

Group	*n*	*Ages*	*Gender*	*y-BOCS*
OCD	38	31.53 ± 10.23	23F/15M	25.13 ± 5.17
Low OCD	19	29.63 ± 8.33	9F/10M	21.00 ± 3.20
High OCD	19	33.42 ± 11.76	14F/5M	29.26 ± 2.96
Controls	36	27.78 ± 7.80	23F/13M	–

We understand the restrictions about the median-split method. However, the choice for this approach considered the histogram presented in Fig. 1A, where despite five subjects with similar scores around the median value, it represent two sufficiently clinically distinct groups ($p < 0.001$). Moreover, this method also reduces the loss of sample examples, which could occur with other approaches (e.g. percentile-split) and cause a negative impact during the machine learning procedures.

All groups were tested for statistical differences between each pair of groups (Control Vs OCD; control vs high OCD; control vs low OCD; high vs low OCD). The Mann-Whitney-Wilcoxon test was applied to evaluate differences of age, and the Chi-squared test was used to compare gender. The results were: Controls versus OCD: ($w = 545.5$, p-value = 0.135 for ages and x-squared = 0.003, p-value = 0.954 for genders); Controls versus Low OCD: ($w = 144.5$, $p = 0.299$ and $x = 1.762$, $p = 0.184$); Controls versus High OCD: ($w = 298.5$, $p = 0.446$ and $x = 0.799$, $p = 0.372$); Low OCD versus High OCD: ($x = 247$, $p = 0.094$ and $x = 0.188$, $p = 0.664$).

Exclusion criteria were: any medical disorder possibly affecting the central nervous system, MRI contraindications (including claustrophobia, pacemakers, aneurysm clips, artificial heart valve, cochlear implants, metal fragments or foreign objects in the eye, skin or body), substance abuse/dependence, or previous or current use of psychotropic medication (benzodiazepines, antipsychotics, antidepressants, stimulants, mood stabilizers).

2.2. Image acquisition

Images were acquired by using a 1.5 T GE Sigma MRI scanner (General Electric, Milwaukee WI, USA). Throughout the brain, 248 images were obtained using a T1-3D-SPGR sequence (axial acquisition, TE = 4.20 ms, TR = 10.5 ms, flip angle = 15°, 256 × 192 matrix, 248 slices with thickness = 1.6 mm).

2.3. Image processing

Several previous studies have used regional volume measures to evaluate structural differences between OCD patients and controls (Piras et al., 2015). Similarly, it was also employed by Soriano-Mas et al. (2007) and Hu et al. (2016) when applying ML techniques to identify OCD patients. Therefore, this study used the same measures as inputs to each ML algorithm.

The image processing was performed using the FreeSurfer package (version 4.3.1) available in http://surfer.nmr.mgh.harvard.edu/ (Dale et al., 1999). FreeSurfer pipeline is based on the following steps: cortical surface modeling, spherical coordinate transformation, curvature (sulci and gyri) registration to standard space, and automated parcellation and labeling of cortical and subcortical structures (Fischl et al., 2002, 2004). Further details about recon-all pipeline can be found at Fischl et al. (1999). The volume of 117 labeled regions (Supplementary

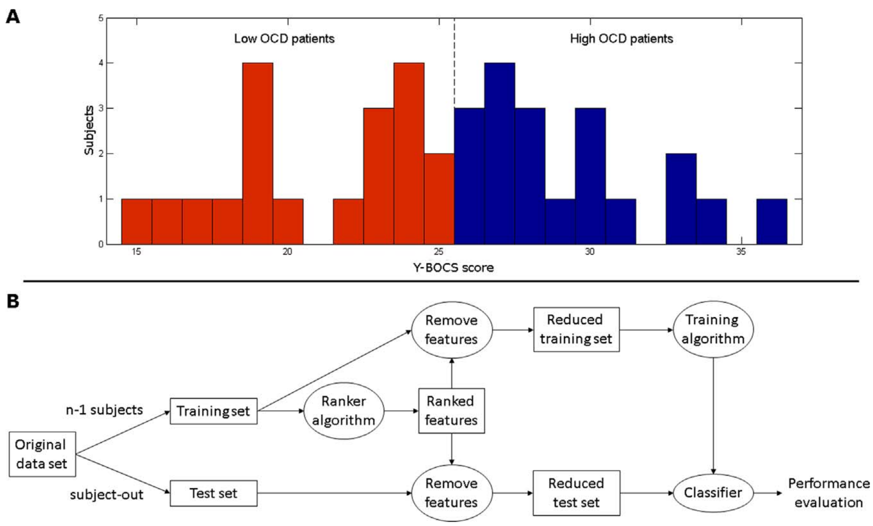

Fig. 1. (A) Histogram of OCD patients distribution according to Y-BOCS scores; (B) The experimental protocol followed.

material) was used as the features used to predict the group labels.

2.4. Machine learning

The data set was divided according to a leave-one-subject-out (LOSO) cross-validation method (Mitchell et al., 2004; Pereira et al., 2009; Esterman et al., 2010). Due to the limited sample size, a leave-one-out procedure was used instead of K-fold cross-validation. Then, the protocol illustrated in Fig. 1B was repeated for all pairs of training and test sets available (due to the LOSO approach, the number of pairs is equal to the sample size).

Firstly, at each training-test round, volumes from a given subject were separated for testing, while data from the other subjects were left for training. This approach allows a better assessment of the classification performance results, by considering different partitions of data. LOSO also emulates clinical situation where a model, induced using data from previous patients, is utilized by the physician to support his decision for new patients (Esterman et al., 2010).

Subsequently, ranker-based feature selection was performed for each one of the training data sets. When a feature subset was obtained, projections of the training and testing datasets described by the selected features only were built. Thus the SVM classifiers were performed using these projections. Their predictive performance was evaluated on each corresponding test set.

2.4.1. Feature selection

In order to preselect brain regions with the most predictive power, feature selection was performed before the classification process using the Weka toolbox (Hall et al., 2009; Witten et al., 2011) and the R software (R Core Team, 2011). Feature sets were evaluated using leave-one-subject-out (LOSO) tests; in this procedure, each algorithm is successively learned on *n-1* samples of images and tested on the remaining one. This is repeated *n* times so that every subject was left out once. For each algorithm, the features are ranked accordingly to the corresponding importance measure and we selected the best 5–95% (with steps of 5%) features in the yielded rankings.

Although there are several feature selection methods in the literature, we considered two multivariate and five univariate FS algorithms in this paper, as follows:

I. *Recursive Support Vector Machine* (rSVM): trains a linear SVM classifier (Cristianini and Shawe-Taylor, 2000) for two classes and orders the weights assigned to each feature (Chu et al., 2012). The higher the weight, the more important the feature is.
II. *Summed Correlations Feature Selection* (SCFS): evaluates the redundancy between features by calculating the Pearson's correlation of each variable with everyone else. Features with the lowest sum of correlations are considered less redundant than the others.
III. *T-test based* (TT): performs a two-sample *t*-test for each feature comparing two classes and sort the p-values in ascending order (Chu et al., 2012). Thus, the smaller the p-value, the most significant feature is.
IV. *Chi Squared (CS):* estimates the independence between each feature and the class (Liu and Setiono, 1995). The first step is a discretization of all numeric features followed by a calculation of the difference between the observed and the expected co-occurrences of these values, for each class and feature value. The higher the differences, the more significant the feature is.
V. *Gain Ratio* (GR): rewards features which minimize the entropy of data, leading to higher information gain (Hall et al., 2009). However, different from the classic information gain criterion, the GR does not favor features with larger number of values.
VI. *Relief-F*: features with different values for a pair of examples from different classes are gratified, while features with different values for examples from the same class are penalized (Hall et al., 2009). In other words, this algorithm gives higher importance to features that lead to better class separability and takes into account the effect of interacting features.
VII. *Symmetrical Uncertainty* (SU): measures the correlation between each feature and the class (Hall and Smith, 1998), where the higher the correlation, the stronger the dependence between the feature and the class. As SU also employs a procedure based on an entropy measure, it takes into account a dependency and an information measure.

2.4.2. Classifier

As an evaluation protocol, we adopted the Support Vector Machines (SVMs) classifier, such as other computational studies previously did (Chen and Lin, 2006). Moreover, SVM is commonly applied to the analysis of neuroimaging data (Orrù et al., 2012) and already tested with neuroimaging data from OCD (Li et al., 2014; Parrado-Hernández et al., 2014; Hu et al., 2016).

SVM were induced using the original data set (with 117 vol features) and using each of the data sets described by the subsets of features selected. In SVMs, a set of training data was used to draw a maximum margin hyperplane separating on the basis of the coordinates of the features that separate the two classes. Subsequently, the coordinates of this hyperplane were used to test a new set of data and assess the accuracy of this model (Cristianini and Shawe-Taylor, 2000). A linear kernel was used in the experiments, with the trade-off parameter C assuming the values from 10^{-5} to 10^{5}. Classifications were performed comparing the four available population groups, which were

Table 2
Higher performance for balanced accuracy (sum of the percentage of subjects correctly classified in each class divided by two) achieved by each FS algorithm and the respective differences (Δ accuracy) against classification without feature selection. The respective percentage of features and *C*-values of SVM are also presented. The algorithms listed are: recursive Support Vector Machines (rSVM), Sum of Correlations Feature Selection (SCFS), T-test (TT), Chi Squared (CS), Gain Ratio (GR), ReliefF, Symmetrical Uncertainty (SU). Significant results are presented in bold.

Algorithm	Balanced accuracy (%)	Δ Accuracy (%)	Reduction of features (%)	C-value
Controls vs. All OCD				
Without FS	52.56	–	–	1.00E − 04
rSVM	63.67	11.11	90	1.00E − 01
SCFS	66.45	13.89	85	1.00E + 04
TT	71.64	19.08	80	1.00E − 04
CS	70.10	17.54	65	1.00E − 04
GR	70.10	17.54	65	1.00E − 04
Relief-F	68.86	16.30	95	1.00E − 02
SU	70.10	17.54	65	1.00E − 04
Controls vs. Low OCD				
Without FS	57.02	–	–	1.00E − 04
rSVM	62.72	5.70	95	1.00E − 03
SCFS	67.98	10.96	70	1.00E − 04
TT	63.82	6.80	90	1.00E − 02
CS	77.12	20.10	75	1.00E − 04
GR	77.12	20.10	75	1.00E − 04
Relief-F	66.30	9.28	85	1.00E − 03
SU	77.12	20.10	75	1.00E − 04
Controls vs. High OCD				
Without FS	54.53	–	–	1.00E − 04
rSVM	66.59	12.06	75	1.00E − 04
SCFS	71.86	17.32	85	1.00E − 03
TT	65.06	10.53	95	1.00E + 03
CS	62.43	7.89	60	1.00E − 04
GR	62.43	7.89	60	1.00E − 04
Relief-F	63.82	9.28	95	1.00E − 02
SU	62.43	7.89	60	1.00E − 04
Low OCD vs. High OCD				
Without FS	63.16	–	–	1.00E − 04
rSVM	65.79	2.63	95	1.00E − 04
SCFS	73.68	10.53	75	1.00E − 04
TT	*55.26*	*−7.89*	*40*	*1.00E − 04*
CS	71.05	7.89	90	1.00E + 06
GR	68.42	5.26	85	1.00E + 03
Relief-F	65.79	2.63	90	1.00E − 04
SU	68.42	5.26	85	1.00E + 03

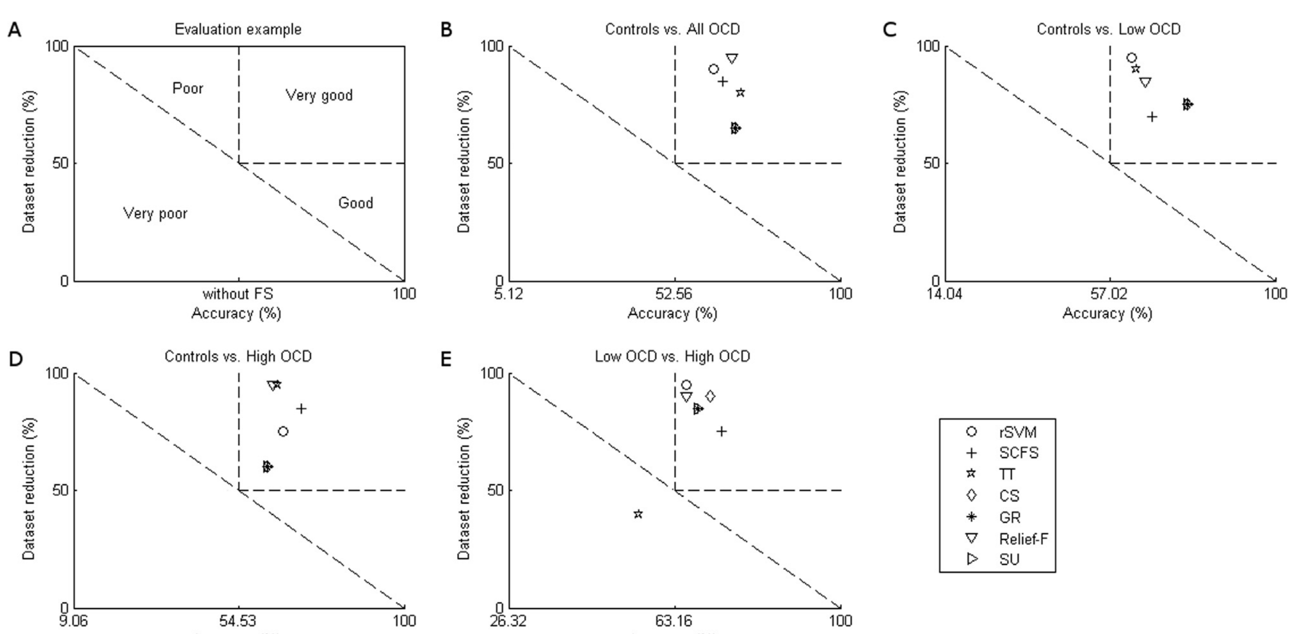

Fig. 2. Scatter plot relating the performance with the reduction of data set reached by each algorithm. The y-axis presents the percentage of reduction from the original dataset, while the x-axis presents the following performance measures: (A) Evaluation example, (B) Controls vs. All OCD, (C) Controls vs. Low OCD, (D) Controls vs. High OCD, (E) Low OCD vs. High OCD. Note that points closer to the up right corner of each graph achieve better performances with fewer features. The represented algorithms are: recursive Support Vector Machines (rSVM), Sum of Correlations Feature Selection (SCFS), T-test (TT), Chi Squared (CS), Gain Ratio (GR), Relief-F, Symmetrical Uncertainty (SU).

“All OCD patients versus Controls”, "Low OCD vs. Controls", "High OCD vs. Controls" and "Low OCD vs. High OCD." SVM was performed using the "e1071" package of the R software (R Core Team, 2014).

To evaluate the performance of the classifier in each test, its accuracy was defined as the sum of the percentage of subjects correctly classified in each class divided by two. This measure, called here as “balanced accuracy”, avoids any influence of unbalanced groups in the final result (Sokolova and Lapalme, 2009). To evaluate post-FS results, all balanced accuracies were submitted to and adaptation of the graphical-based evaluation proposed by Lee et al. (2006), which focuses on the ratio of accuracy to dataset reduction. With this, an FS result can be considered as “Very good”, “Good”, “Poor” and “Very poor”, according to its Cartesian coordinates.

3. Results

3.1. Classification

Classification accuracies before FS were 52.56% for “All OCD vs. Controls”, 57.02% for “Controls vs. Low OCD” tests, 54.53% for “Controls vs. High OCD” and 63.16% for “Low OCD vs. High OCD”. Table 2 presents the higher classification accuracies after each FS algorithms, and the respective differences from the “without FS” test in each comparison. This table also lists the percentage of features and the trade-off parameter *C* values where best accuracy results were achieved.

Fig. 2 presents the graphical-based evaluation of the FS results. As exemplified in Fig. 2A, a “Very good” result is in the first quadrant of the graph, with accuracy higher than the accuracy without FS and more than 50% of reduction from the original dataset. Otherwise, a “Very poor” result is under the diagonal formed by the points (100% of accuracy minus two times the classification error without FS; 100% of data set reduction) and (100% of accuracy; 0% of data set reduction).

3.2. Feature selection

For better observing the features that are selected at least in 50% of iterations by the different techniques, Fig. 3 shows four heat maps (for each classification comparison) that are colored according to the number of times a given feature is selected. Features are presented in the y-axis, while the FS algorithms are in the x-axis. Features with their shades closest to the white color in the graph are those selected most times. Heat maps with all features are presented in Supplementary material.

Considering the features listed in Fig. 3, Fig. 4 shows boxplots for features selected in at least four algorithms during one or more comparisons.

4. Discussion

4.1. Classification performance and dataset reduction

In this study, we applied different feature selection methods to predict clinical diagnosis based on individual structural MRI scans obtained from OCD patients as a whole, and split up into different levels of severity (high and low OCD) versus healthy controls. Several studies have reported differences in brain structures at a group level between OCD and controls, especially in regions of the frontal–subcortical circuits (for a review see Menzies et al., 2008). However, these group level abnormalities have been difficult to replicate at individual level (Orrù et al., 2012; Pauls et al., 2014). Here, we report that classification performance of OCD based on volumetric measures varies with specific FS approaches.

First, it is important to clarify that the univariate analyzes, which is commonly used in studies that search for differences between two data groups, consider only one attribute independently of the others. However, in many situations, a single feature is not enough to provide information which differentiates both groups. On the other hand, multivariate analyzes consider sets of features that can, from small individual contributions, be summed up and lead to the identification of different patterns from each group (Sidtis et al., 2003; Haynes and Rees, 2006). Hence, the feature selection might be considered as a complementary tool in the search for optimum subset of features in order to discriminate groups using multivariate analysis (Liu and Motoda, 2007).

Considering the results of the graphical evaluation procedure, all methods presented a “very good” balance between accuracy after feature selection and the size of the final set of features, except the TT algorithm in the Low OCD vs. High OCD comparison. Besides, although up to 95% dataset reduction, classifiers also achieved similar or, in most of the cases, higher performances after FS for the other three comparisons. These results suggest that, for potential computer-aided biomarkers of OCD, it is not always the case that more information would lead to higher accuracy. Moreover, the lower performance achieved without FS suggests that some features even hinder the predictor induction, further motivating the application of FS.

The same improvements in accuracy were obtained by the CS, GR and SU algorithms for three of the four comparisons, using the same percentages of features and C-values. CS and SU methods measure the relationships of correlation and independence between the features and their classes (Liu and Setiono, 1995; Hall and Smith, 1998), while the GR searches for features with low entropies (Hall et al., 2009). Thus, considering the similar performances after FS using these techniques, as well as the quantity and similarity between subsets generated in all tests (including Controls vs. Low OCD and Controls vs. High OCD tests), we

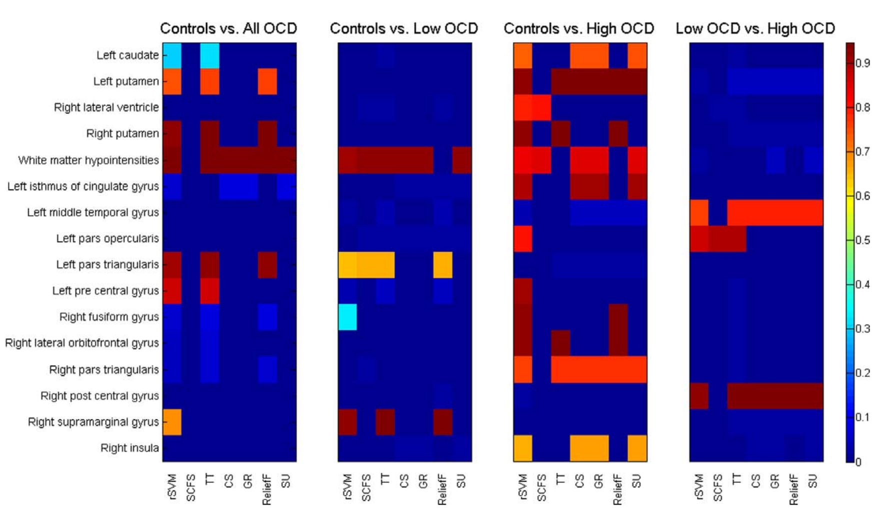

Fig. 3. Heat maps presenting the features selected at least in 50% of iterations for each algorithm. Points which vary their colors closest to the hotest color show that the feature (y-axis) was selected most times by the correspondent algorithm (x-axis) during the leave-one-subject-out iterations. The represented algorithms are: recursive Support Vector Machines (rSVM), Sum of Correlations Feature Selection (SCFS), T-test (TT), Chi Squared (CS), Gain Ratio (GR), ReliefF, Symmetrical Uncertainty (SU). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article)

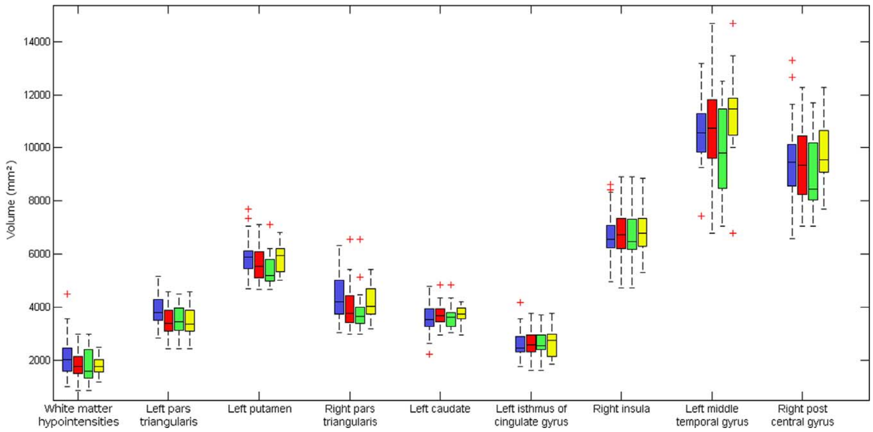

Fig. 4. Boxplots presenting volumes (mm^3) of the features selected in Fig. 3 in at least four feature selection algorithms. Blue (first) boxes correspond to Control group, red (second) boxes to Low OCD group, green (third) boxes to High OCD group and yellow (fourth) boxes to the All OCD group. Red crosses present the outliers. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article)

have features which are both low entropic and dependent of the respective class in the resulting subsets.

SCFS was the FS algorithm that reached 10% of accuracy improvement for all comparisons leastwise. This method selected subgroups of features using up to 30% of the original dataset that leads to up to 17.32% higher performances. Although not providing the highest level of dataset reduction, SCFS generates subsets with little correlation between features, reducing the redundancy and providing subsets with objective information about the classes (Yu and Liu, 2003).

ReliefF and TT, otherwise, were the worst algorithms, especially in the comparison between OCD groups (TT presented a "Very poor" performance in the graphical evaluation). Although the Relieff algorithm can better deal with noisy data (Robnik-Šikonja and Kononenko, 2003), it was considered inferior to other techniques in previous studies using MRI (Luts et al., 2007). Similarly, despite the fact that the TT algorithm is a commonly used FS technique in neuroimaging studies, its performance varies according to the sample size since it relies on statistical comparisons (*t*-tests) between the classes/groups (Chu et al., 2012). Here, the performances of both algorithms were no significant in comparisons involving OCD sub-groups, probably due to the relatively smaller sample size in those tests.

Considering the values of the C parameter presented in Table 1 and the logical configuration of SVMs, it is notable that significant results were predominantly obtained for smaller values of C, showing that the data do not require very strict decision boundaries (Cristianini and Shawe-Taylor, 2000). Additionally, all relevant results in Table 1 were using less than 50% of the original dataset. These aspects contribute to reducing the complexity of the system and, consequently, the computational cost and the processing time.

A higher accuracy in Low vs. High OCD classification than those achieved in other combinations was unexpected. We speculate that, in this case, the most relevant variables for group differentiation are probably those structural changes associated with disease progression. Thus, among all available features, a particular subset contributes more than other to the final classification, even without the selection of attributes. It is indeed supported by the little performance differences after feature selection, even with the elimination of noisy data. On the other hand, tests comparing controls with any OCD configuration probably received initial weights dissipated among many features. Therefore, the feature selection filters unrelated data allowing the classifier to adjust its weights and, consequently, increase classification performance. It becomes evident at the three OCD vs. controls setups with great improvements in accuracy after feature selection (with the Δ accuracy measure approaching to 20% in all three tests).

4.2. Relevant features selected

Among the selected features shown in Fig. 3, there is a little predominance of findings in the right hemisphere (8) compared to the left hemisphere (7) and only one feature of the whole brain. In fact, volumetric differences between OCD and controls are frequently reported to be lateralized (Friedlander and Desrocher, 2006; Piras et al., 2015). Moreover, only the putamen and the pars triangularis were taken bilaterally.

Previous neuroimaging studies comparing OCD and controls found bilateral abnormalities in the putamen, showing increased volumes either bilaterally (Szeszko et al., 2008), left (Yoo et al., 2008) or right lateralized (Gilbert et al., 2008b) in OCD patients when compared to control groups. We have instead observed a volume reduction in left putamen in Low OCD group, but a slight increase in the High OCD group when compared with control group. This region is related to various motor activities, motor control learning, preparation and motor performance tasks (Jenkins et al., 1994; Marchand et al., 2008) and have already been linked to symptom's severity in OCD (Hoexter et al., 2013), which would explain the variation found between the Low and High OCD groups. Putamen is part of the Cortico-Striato-Thalamo-Cortical (CSTC) circuitry, classically associated with the psychopathology of OCD (Li and Mody, 2016).

Three other regions from the CSTC were considered relevant by our FS algorithms: left caudate, right orbitofrontal cortex and the pars triangularis of the inferior frontal gyrus (ptIFG). An increased volume in OCD compared to controls was the previously described in the left caudate (Zarei et al., 2011). In the right lateral orbitofrontal cortex (rOFC), however, contradictory results have been reported. Szeszko et al. (2008) found an increased volume of in OCD, while Yoo et al. (2008) and Lázaro et al. (2009) described decreased rOFC volumes. Volume increases in these two regions were proposed to be related to the imbalance of the direct and indirect pathways in CSTC circuitry (Pauls et al., 2014).

The ptIFG is related with inhibitory control processes (Bari and Robbins, 2013). In our data, both hemispheres show ptIFC with smaller volumes in OCD patients than in controls (Fig. 4), and the left pars triangularis decreases according to the group level (controls > OCD low > high OCD). A neighboring region and also component of IFG, which also participates in inhibitory control process, is the pars opercularis, considered relevant in left hemisphere by our algorithms. Previous studies in OCD portray reduced WM volumes in this region (van den Heuvel et al., 2009; Lázaro et al., 2009; Gonçalves et al., 2015). These regional volume changes could be related to the failure in the inhibitory control of patients with OCD (Penades et al., 2007).

Another area commonly related to OCD is the postcentral gyrus with reduced white matter (WM) volume bellow the right postcentral gyrus (Lázaro et al., 2009, 2011) in patients. Close to it, the left precentral gyrus also presented reduced WM volume in female populations (Yoo et al., 2008). In turn, Yoo et al. (2008) found a bilateral reduction in insula volume, contradictory to Kim et al. (2001) which showed increased volume in the right insula (Kim et al., 2001).

Unilaterally selected as relevant, the left isthmus cingulate is part of the cingulate cortex. The cingulate as a whole presented reduced volume billateraly (Carmona et al., 2007) and WM reduction in the left (Togao et al., 2010) in OCD patients when compared to healthy subjects. Other regions with volume reduction associated with OCD are also present in Fig. 3, such as the left middle temporal gyrus (Kopřivová et al., 2009), the right fusiform gyrus (Kopřivová et al., 2009) and right supramarginal gyrus (Yoo et al., 2008).

In general, two conclusions can be drawn from our findings. First, the accuracy improvement obtained using only relevant features in a leave-one-subject-out approach suggests that feature selection is a relevant step to individual level analysis. Second, there was an overall congruence between the features selected by the different FS methods with the previous findings on OCD using statistical group-level analysis. The complementary nature of these two approaches further validates the previous evidence on the relation between structural variance of these specific brain regions and OCD manifestation.

4.3. Limitations and perspectives

Although we take attention to emulating a clinical situation with LOSO approach (Esterman et al., 2010), the main limitation of this study is the available sample size. Thus, future studies should include a greater number of individuals to validate the findings found here. On the other hand, despite the fact that SVMs are mainly used in the area with many successful results, other multivariate pattern recognition methods could also be employed to evaluate the subsets of features and improve the classification accuracy. For instance, Soriano-Mas et al. (2007) applied a Euclidian-distance-based method to identify OCD patients, achieving an excellent accuracy of 93.1% in a first database and 76.7% of accuracy in a validation database. Hu et al. (2016), on the other hand, using a Gaussian process classifier achieved 77.27% and 80.30% of accuracies using gray and white matter features, respectively. Thus, considering the great results from both studies, the combination of FS algorithms discussed here with Euclidian-distance-based and GPC methods should be considered in future work.

More FS techniques should also be explored in OCD context. Wrapper and embedded methods, for example, may promote valuable contributions about the combination of classifiers and relevance measures such that explored here (Kohavi and John, 1997; Habeck and Stern, 2010; Tohka et al., 2016). These techniques were previously explored in fMRI data proving to be able to detect and accurately classify spatial patterns (De Martino et al., 2008). Finally, other features are also relevant to be explored as input to ML algorithms, such as the cortical thickness (Fischl and Dale, 2000), which already presented relevant results to OCD patients during group analysis (Nakamae et al., 2012).

5. Conclusion

This study applies seven feature selection algorithms to structural MRI data in order to evaluate which features are more relevant for discriminating between high, low OCD and healthy controls. In summary, our results suggest that not all the volumetric information available in MRI recordings is useful for OCD classification. Among the feature selection algorithms, we found that the optimal method would depend on the specific comparison to be performed. Nevertheless, feature selection positively impacts classification procedures, as using selected feature subsets resulted in higher than 15% improvement on classifiers accuracies, even using less than 30% of the original feature set.

Role of the funding source

This study received financial support in the form of grants provided by the “Fundação de Amparo à Pesquisa do Estado de São Paulo” (FAPESP, Grant 2011/21357-9). L.R.T. received a FAPESP scholarship (2013/10498-6 and 2015/17406-5) and a "Universidade Federal do ABC" (UFABC) scholarship. J.B.B. received a "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES) scholarship. M.Q.H. received a FAPESP scholarship (2013/16864-4). The funding sources had no involvement in research procedures or during the preparation of the article.

Acknowledgements

The authors thank all patients and volunteers who participated in this investigation.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jad.2017.06.061.

References

Abramovitch, A., Abramowitz, J.S., Mittelman, A., 2013. The neuropsychology of adult obsessive–compulsive disorder: a meta-analysis. Clin. Psychol. Rev. 33, 1163–1171.
Bari, A., Robbins, T.W., 2013. Inhibition and impulsivity: behavioral and neural basis of response control. Prog. Neurobiol. 108, 44–79.
Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning. Springer, New York.
Carmona, S., Bassas, N., Rovira, M., Gispert, J.D., Soliva, J.C., Prado, M., Tomas, J., Bulbena, A., Vilarroya, O., 2007. Pediatric OCD structural brain deficits in conflict monitoring circuits: a voxel-based morphometry study. Neurosci. Lett. 421 (218e223).
Chen, Y.W., Lin, C.J., 2006. Combining SVMs with various feature selection strategies. In: Guyon, I., Nikravesh, M., Gunn, S., Zadeh, L.A. (Eds.), Feature Extraction: Foundations and Applications. Springer, Berlin, pp. 315–324.
Cherian, A.V., Math, S.B., Kandavel, T., Reddy, Y.C., 2014. A 5-year prospective follow-up study of patients with obsessive–compulsive disorder treated with serotonin reuptake inhibitors. J. Affect. Disord. 152, 387–394.
Chu, C., Hsu, A.L., Chou, K.H., Bandettini, P., Lin, C., Alzheimer's Disease Neuroimaging Initiative, 2012. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. Neuroimage 60, 59–70.
Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines: and Other Kernel-based Learning Methods. University Press, Cambridge.
Dale, A., Fischl, B., Sereno, M., 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. Neuroimage 9, 179–194.
De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., Formisano, E., 2008. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. Neuroimage 43, 44–58.
Esterman, M., Tamber-Rosenau, B.J., Chiu, Y.C., Yantis, S., 2010. Avoiding non-independence in fMRI data analysis: leave one subject out. Neuroimage 50, 572–576.
Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207.
Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. 97, 11050–11055.
Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.
Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22.
Friedlander, L., Desrocher, M., 2006. Neuroimaging studies of obsessive–compulsive disorder in adults and children. Clin. Psychol. Rev. 26, 32–49.
Gilbert, A.R., Mataix-Cols, D., Almeida, J.R.C., Lawrence, N., Nutche, J., Diwadkar, V., Keshavan, M.S., Phillips, M.L., 2008a. Brain structure and symptom dimension relationships in obsessive–compulsive disorder: a voxel-based morphometry study. J. Affect. Disord. 109, 117–126.
Gilbert, A.R., Keshavan, M.S., Diwadkar, V., Nutche, J., Macmaster, F., Easter, P.C., Buhagiar, C.J., Rosenberg, D.R., 2008b. Gray matter differences between pediatric obsessive-compulsive disorder patients and high-risk siblings: a preliminary voxel-based morphometry study. Neurosci. Lett. 435 (45e50).
Gonçalves, Ó.F., Sousa, S., Maia, L., Carvalho, S., Leite, J., Ganho, A., Fernandes-Gonçalves, A., Frank, B., Pocinho, F., Carracedo, A., Sampaio, A., 2015. Inferior frontal gyrus white matter abnormalities in obsessive–compulsive disorder. Neuroreport 26, 495–500.
Haynes, J.D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. 7, 523–534.
Habeck, C., Stern, Y., Alzheimer’s Disease Neuroimaging Initiative, 2010. Multivariate data analysis for neuroimaging data: overview and application to Alzheimer's disease. Cell Biochem. Biophys. 58, 53–67.
Hall, M.A., Smith, L.A., 1998. Practical feature subset selection for machine learning. P.

Aust. Comput. Sci. Conf. 181–191.
Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I., 2009. The weka data mining software: an update. SIGKDD Explor. 11, 10–18.
Hoexter, M.Q., Miguel, E.C., Diniz, J.B., Shavitt, R.G., Busatto, G.F., Sato, J.R., 2013. Predicting obsessive-compulsive disorder severity combining neuroimaging and machine learning methods. J. Affect. Disord. 150, 1213–1216.
Hu, X., Liu, Q., Li, B., Tang, W., Sun, H., Li, F., Yang, Y., Gonga, Q., Huang, X., 2016. Multivariate pattern analysis of obsessive–compulsive disorder using structural neuroanatomy. Eur. Neuropsychopharmacol. 26, 246–254.
Jenkins, I.H., Brooks, D.J., Nixon, P.D., Frackowiak, R.S., Passingham, R.E., 1994. Motor sequence learning: a study with positron emission tomography. J. Neurosci. 14, 3775–3790.
Kim, J.J., Lee, M.C., Kim, J., Kim, I.Y., Kim, S.I., Han, M.H., Chang, K.H., Kwon, J.S., 2001. Grey matter abnormalities in obsessive-compulsive disorder: statistical parametric mapping of segmented magnetic resonance images. Br. J. Psychiatry 179 (330e334).
Klöppel, S., Abdulkadir, A., Jack, C.R., Koutsouleris, N., Mourão-Miranda, J., Vemuri, P., 2012. Diagnostic neuroimaging across diseases. Neuroimage 61, 457–463.
Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. Artif. Intel. 97, 273–324.
Kopřivová, J., Horáček, J., Tintěra, J., Praško, J., Raszka, M., Ibrahim, I., Höschl, C., 2009. Medial frontal and dorsal cortical morphometric abnormalities are related to obsessive-compulsive disorder. Neurosci. Lett. 464, 62–66.
Lázaro, L., Bargalló, N., Castro-Fornieles, J., Falcón, C., Andrés, S., Calvo, R., Junqué, C., 2009. Brain changes in children and adolescents with obsessive–compulsive disorder before and after treatment: a voxel-based morphometric MRI study. Psychiatry Res. 172, 140–146.
Lázaro, L., Castro-Fornieles, J., Cullell, C., Andrés, S., Falcón, C., Calvo, R., Bargalló, N., 2011. A voxel-based morphometric MRI study of stabilized obsessive–compulsive adolescent patients. Prog. Neuro-Psychopharmacol. 35, 1863–1869.
Lee, H.D., Monard, M.C., Voltolini, R.F., Prati, R.C., Chung, W.F., 2006. A simple evaluation model for feature subset selection algorithms. Iberoam. J. Artif. Intell. 32, 09–17.
Li, F., Huang, X., Tang, W., Yang, Y., Li, B., Kemp, G.J., Mechelli, A., Gong, Q., 2014. Multivariate pattern analysis of DTI reveals differential white matter in individuals with obsessive-compulsive disorder. Hum. Brain Mapp. 35, 2643–2651.
Li, B., Mody, M., 2016. Cortico-striato-thalamo-cortical circuitry, working memory, and obsessive–compulsive disorder. Front. Psychiatry 7, 78.
Liu H, Setiono R., 1995. Chi2: Feature selection and discretization of numeric attributes. In: Proc. IEEE Int. Conf. Artif. Intell, pp. 388–388.
Liu, H., Motoda, H., 2007. Computational Methods of Feature Selection. CRC Press, Boca Raton.
Luts, J., Heerschap, A., Suykens, J.A., Van Huffel, S., 2007. A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection. Artif. Intell. Med. 40, 87–102.
Mancebo, M.C., Boisseau, C.L., Garnaat, S., Eisen, J.L., Greenberg, B., Sibrava, N.J., Stout, R.L., Rasmussen, S.A., 2014. Long-term course of pediatric obsessive-compulsive disorder: three years of prospective follow-up. Compr. Psychiatry 55, 1498–1504.
Marchand, W.R., Lee, J.N., Thatcher, J.W., Hsu, E.W., Rashkin, E., Suchy, Y., Gordon, C., Jennifer, S., Sharon, B.S., 2008. Putamen coactivation during motor task execution. Neuroreport 19, 957–960.
Menzies, L., Chamberlain, S., Laird, A., Thelen, S., Sahakian, B., Bullmore, E., 2008. Integrating evidence from neuroimaging and neuropsychological studies of obsessive-compulsive disorder: the orbitofronto-striatal model revisited. Neurosci. Biobehav. Rev. 32, 525–549.
Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M., Newman, S., 2004. Learning to decode cognitive states from brain images. Mach. Learn. 57, 145–175.
Mourão-Miranda, J., Hardoon, D.R., Hahn, T., Marquand, A.F., Williams, S.C., Shawe-Taylor, J., Brammer, M., 2011. Patient classification as an outlier detection problem: an application of the one-class support vector machine. NeuroImage 58, 793–804.
Nakamae, T., Narumoto, J., Sakai, Y., Nishida, S., Yamada, K., Kubota, M., Miyata, J., Fukui, K., 2012. Reduced cortical thickness in non-medicated patients with obsessive-compulsive disorder. Prog. Neuropsychopharmacol. Biol. Psychiatry 37 (1), 90–95.
Orrù, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G., Mechelli, A., 2012. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci. Biobehav. Rev. 36, 1140–1152.
Parrado-Hernández, E., Gómez-Verdejo, V., Martínez-Ramón, M., Shawe-Taylor, J., Alonso, P., Pujol, J., Menchón, J.M., Cardoner, N., Soriano-Mas, C., 2014. Discovering brain regions relevant to obsessive–compulsive disorder identification through bagging and transduction. Med. Image Anal. 18, 435–448.
Pauls, D.S.L., Abramovitch, A., Rauch, S.L., Geller, D.A., 2014. Obsessive-compulsive disorder: an integrative genetic and neurobiological perspective. Nat. Rev. Neurosci. 15, 410–424.
Penades, R., Catalan, R., Rubia, K., Andres, S., Salamero, M., Gasto, C., 2007. Impaired response inhibition in obsessive compulsive disorder. Eur. Psychiatry 22, 404–410.
Pereira, F., Mitchell, T., Botvinick, M., 2009. Machine learning classifiers and fMRI: a tutorial overview. NeuroImage 45, S199–S209.
Piras, F., Piras, F., Chiapponi, C., Girardi, P., Caltagirone, C., Spalletta, G., 2015. Widespread structural brain changes in OCD: a systematic review of voxel-based morphometry studies. Cortex 62, 89–108.
R Core Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL 〈http://www.R-project.org/〉. (Accessed 4 November 2016).
Robnik-Šikonja, M., Kononenko, I., 2003. Theoretical and empirical analysis of ReliefF and RReliefF. Mach. Learn. 53, 23–69.
Sidtis, J.J., Strother, S.C., Rottenberg, D.A., 2003. Predicting performance from functional imaging data: methods matter. Neuroimage 20, 615–624.
Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inform. Process. Manag. 45, 427–437.
Soriano-Mas, C., Pujol, J., Alonso, P., Cardoner, N., Menchón, J.M., Harrison, B.J., Deus, J., Vallejo, J., Gaser, C., 2007. Identifying patients with obsessive–compulsive disorder using whole-brain anatomy. NeuroImage 35, 1028–1037.
Szeszko, P.R., Christian, C., MacMaster, F., Lencz, T., Mirza, Y., Taormina, S.P., Easter, P., Rose, M., Michalopoulou, G.A., Rosenberg, D.R., 2008. Gray matter structural alterations in psychotropic drug-naive pediatric obsessive-compulsive disorder: an optimized voxel-based morphometry study. Am. J. Psychiatry 165, 1299–1307.
Tang, W., Huang, X., Li, B., Jiang, X., Li, F., Xu, J., Yang, Y., Gong, Q., 2015. Structural brain abnormalities correlate with clinical features in patients with drug-naïve OCD: a DARTEL-enhanced voxel-based morphometry study. Behav. Brain Res. 294, 72–80.
Togao, O., Yoshiura, T., Nakao, T., Nabeyama, M., Sanematsu, H., Nakagawa, A., Noguchi, T., Hiwatashi, A., Yamashita, K., Nagao, E., Kanba, S., Honda, H., 2010. Regional gray and white matter volume abnormalities in obsessive-compulsive disorder: a voxel-based morphometry study. Psychiatry Res. 184 (29e37).
Tohka, J., Moradi, E., Huttunen, H., Alzheimer's Disease Neuroimaging Initiative, 2016. Comparison of feature selection techniques in machine learning for anatomical brain MRI in dementia. Neuroinformatics 14, 279–296.
van den Heuvel, O.A., Remijnse, P.L., Mataix-Cols, D., Vrenken, H., Groenewegen, H.J., Uylings, H.B., van Balkom, A.J.L.M., Veltman, D.J., 2009. The major symptom dimensions of obsessive-compulsive disorder are mediated by partially distinct neural systems. Brain 132 (853e868).
Venkatasubramanian, G., Zutshi, A., Jindal, S., Srikanth, S.G., Kovoor, J.M., Kumar, J.K., Reddy, Y.J., 2012. Comprehensive evaluation of cortical structure abnormalities in drug-naive, adult patients with obsessive-compulsive disorder: a surface-based morphometry study. J. Psychiatry Res. 46, 1161–1168.
Witten, I., Frank, E., Hall, M., 2011. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco.
Yoo, S.Y., Roh, M.S., Choi, J.S., Kang, D.H., Ha, T.H., Lee, J.M., Kim, I.Y., Kim, S.I., Kwon, J.S., 2008. Voxel-based morphometry study of gray matter abnormalities in obsessive-compulsive disorder. J. Korean Med. Sci. 23 (24e30).
Yu, L., Liu, H., 2003. Feature selection for high-dimensional data: a fast correlation-based filter solution. ICML 3, 856–863.
Zarei, M., Mataix-Cols, D., Heyman, I., Hough, M., Doherty, J., Burge, L., Winmillf, L., Nijhawanf, S., Matthews, P.M., James, A., 2011. Changes in gray matter volume and white matter microstructure in adolescents with obsessive-compulsive disorder. Biol. Psychiatry 70 (1083e1090).