

Feature selection before EEG classification supports the diagnosis of Alzheimer's disease



L.R. Trambaiolli^{a,*}, N. Spolaôr^b, A.C. Lorena^c, R. Anghinah^d, J.R. Sato^a

^a Center of Mathematics, Computing and Cognition, Universidade Federal do ABC, Santo André, Brazil

^b Laboratório de Bioinformática, Centro de Engenharia e Ciências Exatas, Universidade Estadual do Oeste do Paraná, Foz do Iguaçu, Brazil

^c Instituto de Ciência e Tecnologia, Universidade Federal de São Paulo, São José dos Campos, Brazil

^d Reference Center for Cognitive Disorders, Hospital das Clínicas, University of São Paulo, Rua Arruda Alvim 206, São Paulo, Brazil

ARTICLE INFO

Article history:

Accepted 26 June 2017

Available online 14 July 2017

Keywords:

Feature selection

Dementia

Alzheimer's disease

Electroencephalography

Pattern recognition

HIGHLIGHTS

- We achieved EEG-based diagnostic accuracy >90% with reduced data dimensionality in Alzheimer's disease.
- More than 80% of EEG features are marginal or irrelevant to support diagnosis.
- Automated selected EEG features are in agreement with previous neurobiological findings.

ABSTRACT

Objective: In many decision support systems, some input features can be marginal or irrelevant to the diagnosis, while others can be redundant among each other. Thus, feature selection (FS) algorithms are often considered to find relevant/non-redundant features.

Objective: This study aimed to evaluate the relevance of FS approaches applied to Alzheimer's Disease (AD) EEG-based diagnosis and compare the selected features with previous clinical findings.

Methods: Eight different FS algorithms were applied to EEG spectral measures from 22 AD patients and 12 healthy age-matched controls. The FS contribution was evaluated by considering the leave-one-subject-out accuracy of Support Vector Machine classifiers built in the datasets described by the selected features.

Results: The Filtered Subset Evaluator technique achieved the best performance improvement both on a per-patient basis (91.18% of accuracy) and on a per-epoch basis ($85.29 \pm 21.62\%$), after removing $88.76 \pm 1.12\%$ of the original features. All algorithms found out that alpha and beta bands are relevant features, which is in agreement with previous findings from the literature.

Conclusion: Biologically plausible EEG datasets could achieve improved accuracies with pre-processing FS steps.

Significance: The results suggest that the FS and classification techniques are an attractive complementary tool in order to reveal potential biomarkers aiding the AD clinical diagnosis.

© 2017 Published by Elsevier Ireland Ltd on behalf of International Federation of Clinical Neurophysiology.

1. Introduction

Alzheimer's disease (AD) is the main cause of dementia in many countries and with a projection to affect 682 million people until 2050 (Batsch and Mittelman, 2012). Common symptoms include memory loss and impairment of at least one cognition area (such as calculation, praxis, gnosis, executive functions or language). As

the AD definitive diagnosis can only be established with a histopathological analysis (autopsy or biopsy) (Terry, 1994), the search for biological markers to support its early diagnosis remains an open challenge.

Currently, the first aspect to be investigated in the diagnostic process is the measure of the symptoms of cognitive and functional decline. However, these neuropsychological diagnostics require long interview sessions, highly trained professionals and can be jeopardized by other factors involved in the pathological events (Hampel et al., 2010a). In recent years, many studies have explored

* Corresponding author.

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

the potential for quantification of tau protein and beta-amyloid production present in cerebrospinal fluid (CSF) (Rabinovici et al., 2016; Schierle et al., 2016; Vos et al., 2016). Both biochemical markers are historically related to AD and are obtained by extracting plasma or CSF. Using these markers, it was possible to obtain accuracy higher than 90% at the differentiation between DA patients and control subjects, as well as between DA patients and patients presenting other dementias. However, the process is invasive and may be influenced by drug treatments (Hampel et al., 2010b).

Electroencephalogram (EEG) is used as an alternative exam for the probable dementia diagnosis when it is still uncertain (Luccas et al., 1996). An EEG exam consists of recording electric potential signals at scalp, which is associated with synaptic activity (Epstein, 1995). From EEG raw data, it is possible to extract features or characteristics for supporting the automatic AD diagnosis by the data mining computational process. More specifically, classification methods can be used to build automatic diagnosis models (classifiers) (Lehmann et al., 2007).

A well-known EEG feature is the peak of the frequency spectrum (a.k.a. spectral peak), because it clearly shows consistent patterns to the medical AD literature, where the background frequency is reduced to delta and theta frequencies and the central alpha rhythm is decreased. Therefore, it is possible to observe a direct correlation between the degree of cognitive impairment and the power of low-frequency electrical activity in the EEG (Leuchter et al., 1987, 1993; Klass and Brenner, 1995; Sandmann et al., 1996).

Previous works demonstrated that the spectral peak, used as input to Machine Learning (ML) algorithms, is a simple and effective feature for differentiating healthy elderly from those with AD (Trambaiolli et al., 2011a, 2011b). However, given the large number of experimental conditions (resting, cognitive tasks, open and closed eyes, etc.), channels (up to 256 channels), montages (bipolar, referential, averaged, etc.) and frequency bands (classical delta to gamma bands, or segmented intervals such as alpha1, alpha2, beta1, beta2, beta3 and so on) in which spectral peaks may be calculated, many features are generated from the EEG exams. Some of them can have a marginal relevance for AD diagnosis and could be discarded, while others can be redundant. Feature Selection (FS) (Liu and Motoda, 2007), an active research area within the ML and DM (Data Mining) communities, aims to find a small number of features that describes the dataset as well as, or even better than, the original set of features does. There are many benefits in performing such pre-processing step, such as reducing the computational cost in classification and increasing the predictive performance. FS also allows to evidence the most prominent features related to the disease, helping to better understand its mechanisms.

The importance or quality of each feature in the data set can be measured through several criteria, which can be broadly divided into five groups (Liu and Motoda, 2007): precision, information, distance, dependency and consistency. Precision-based measures consider the performance achieved by a machine learning algorithm in data described by each subset of features under evaluation. Although these measures potentially improve the learning performance when the same algorithm is used for the FS and classification steps, they are computationally expensive. In addition, they produce biased results towards the chosen algorithm, such as not to be able to perform well with other learning algorithms.

On the other hand, information, distance, dependency and consistency criteria can be extracted from data directly, relying on statistical characteristics of the data sets which are independent of a particular ML algorithm. As a result, they are typically computationally cheaper than precision-based measures. For example, applying measures such as entropy and correlation might be faster

than training and testing a complex classifier. It should be emphasized that the four groups differ mainly in the type of statistical properties explored.

FS methods are typically organized according to the interaction between the feature evaluation procedure and machine learning algorithm. By focusing on FS methods applied during data pre-processing, two traditional approaches are identified in the literature: wrapper and filter. The former approach is related to the precision-based measures (Kohavi and John, 1997). The latter, in turn, is associated with the remaining criteria. In this work, we used only filter FS methods due to their advantages in terms of computational cost and independence of specific learners. Wrapper algorithms were not included in this study, since their results are biased towards a specific classification technique.

The literature presents few publications exploring data dimensionality reduction in the EEG-based AD diagnosis domain. In Ahmadlou et al. (2010) a graph-based visibility approach is used in the support of AD diagnosis from EEG data. To reduce data dimensionality, they employ an analysis of variance (ANOVA) statistical test followed by Principal Component Analysis (PCA). PCA is an effective technique, but it does not maintain the original features, which may be important to interpret the neurophysiological meaning of the findings. ANOVA is also used in Huang et al. (2000), selecting those features with the most distinguished values in the EEG signals, according to statistical significance. A linear discriminant analysis is then performed in such data. Despite of its wide usage for FS in biological data, parametric methods assuming a given distribution from data (such as ANOVA) are not recommended because their distributive assumptions are probably unconfirmed due to the small sample sizes available (Saeys et al., 2007).

This paper employed eight filter feature selection algorithms (using the feature importance measures previously listed) in a pre-processing step for building classifiers for AD diagnosis. Furthermore, the chosen features are analyzed regarding previous clinical findings to evaluate the ability of the filter algorithms to select biologically consistent features.

2. Materials and methods

This work was approved by the University of São Paulo Ethics Committee, and all the subjects involved agreed with the protocol and provided a written consent.

2.1. Experimental setup

2.1.1. Data acquisition

Data were collected from a total of 22 patients diagnosed with probable AD (71.5 ± 7.7 years, 17 female), and 12 age-matched healthy subjects (68.8 ± 6.7 years, 8 female). No significant difference was found between the ages of the groups ($p = 0.2323$). The AD diagnosis was performed by a physician according to NINCDS-ADRDA criteria (McKhann et al., 1984), classified as mild to moderate according to DSM-III R criteria (Spitzer et al., 1990). The subjects did not present any other diagnostic causes for the cognitive impairment, such as diabetes mellitus, kidney disease, thyroid disease, alcoholism, liver disease or vitamin B12 deficiency. For more information on this database, see Trambaiolli et al. (2011a,b).

The EEG was recorded with a *Braintech 3.0* equipment (EMSA Medical Equipments), with a 12-bit A/D converter and a sampling rate of 200 Hz. The electrodes (Fp1, Fp2, F7, F3, Fz, F4, F8, C3, Cz, C4, T5, T3, P3, Pz, P4, T4, T6, O1 and O2) were placed according to the international 10–20 system, with interconnected earlobe electrodes reference, as recommended by the Brazilian Society of Clin-

ical Neurophysiology and the American EEG Society. Records were obtained during approximately 20 min (or larger periods depending on the amount of noise identified by the technician during the collection) with subjects awake and resting with eyes closed.

Although we did not collect any channels related to ocular movements, an experienced neurologist examined all records in order to eliminate possible noise and artifacts in the EEG signals. During this examination, all intervals with eye blink or eye movements, muscle or jaw contractions and any other biological artifacts (Anghinah et al., 2006) were excluded for all channels.

After this inspection, 40 eight-second epochs were selected from the remaining parts of each subject record, making a total of 1360 epochs in our database.

2.1.2. Features

The EEG montage is the combination of electrode pairs that allow the registration of differences of potentials. For proper features interpretation derived from EEG, the montage should follow a logical and simple distribution (Rappelsberger, 1989).

Two types of montages can be used: referential or bipolar. In referential montage, a common electrode is used in all records, while, in bipolar montage, the record is made on different pairs of electrodes without a common electrode between them (Niedermeyer and da Silva, 2005). With some caveats, starting with a montage, it is possible to get virtually another type of montage by using the vector subtraction concept (Sanei and Chambers, 2008).

Therefore, based on the montages suggested by Trambaioli et al. (2011b), the following combinations of electrodes were used to obtain the spectral peaks:

- Biauricular reference (Bar): Fp1-A1, Fp2-A2, F7-A1, F8-A2, F3-A1, F4-A2, C3-A1, C4-A2, T3-A1, T4-A2, P3-A1, P4-A2, T5-A1, T6-A2, O1-A1, O2-A2, Fz-A1, Cz-A1, Pz-A1;
- Cz-reference (Czr): Fp1-Cz, Fp2-Cz, F3-Cz, F4-Cz, F7-Cz, F8-Cz, T3-Cz, T4-Cz, C3-Cz, C4-Cz, T5-Cz, T6-Cz, P3-Cz, P4-Cz, O1-Cz, O2-Cz;
- Longitudinal Bipolar (Lbp): Fp1-F3, F3-C3, C3-P3, P3-O1, O1-T5, T5-T3, T3-F7, F7-Fp1, Fp2-F4, F4-C4, C4-P4, P4-O2, O2-T6, T6-T4, T4-F8, F8-Fp2;
- Crossed Bipolar (Bcr): Fp1-Fp2, F7-F3, F3-Fz, Fz-F4, F4-F8, T3-C3, C4-T4, T5-P3, P3-Pz, Pz-P4, P4-T6, O1-O2;
- Counterpart bipolar (Bco): F7-F8, F3-F4, T3-T4, C3-C4, P3-P4, T5-T6.

We performed the virtual montage calculation on the 1360-epoch database. Afterwards, epochs with all references were filtered by an infinite impulse response elliptic filter with a zero in the frequency of 60 Hz, eliminating the interference of the power grid. A 512-point Fast Fourier Transform (FFT) was then applied with Hamming windows of 2.5 s and 90% of overlap between successive windows.

The spectral peak measure corresponds to the frequency where the EEG spectrum amplitude reaches its maximum value. For this calculation, the EEG signals were first divided into five well established bands: delta (from 0.1 to 4.0 Hz), theta (4.0–8.0 Hz), alpha (8.0–12.0 Hz), beta (12.0–30.0 Hz), and gamma (30.0–50.0 Hz).

All spectral peaks combinations for each band frequency resulted in a total of 345 available features. Therefore, our data set consists of 1360 blocks of 8-s EEG epochs (34 participants, 40 epochs each) with 345 features each.

2.1.3. Machine learning procedure

Firstly, the data set was divided according to a leave-one-subject-out (LOSO) cross-validation method. Herewith, 34 pairs of training and test sets were available for the SVM classifiers. At each training-test round, all 40 epochs from a given patient were

separated for testing, while data from all the other patients were left for training. This approach allows a better assessment of the classification performance results, by considering different data partitions. Furthermore, this evaluation procedure can be considered fairer, since all data from a given patient in the test round remains unseen by the classifier. LOSO also emulates clinical situation, where the physician consults the model, which was induced using data from previous patients, to support his/her decision for new patients (Esterman et al., 2010).

Feature selection is performed for each of the training data sets. When a feature subset is obtained, new training projections and testing data sets are created using only the selected features. SVM classifiers were trained using the original and all pre-processed training data sets. Their predictive performance is evaluated on each corresponding test set. Besides that, the percentage of data set reduction, i.e., the number of features kept after using the feature selection techniques was also recorded. Fig. 1 shows this experimental protocol.

To evaluate the feature selection and classification stability, we replied the procedure with a k-folds cross-validation with 100 permutations, in which the training and test folds have 17 subjects each (11 patients and 6 controls). This procedure is described in the Supplementary Material.

2.2. Feature selection

Feature Selection (FS) aims to find a reduced subset of important features that describes a dataset as well as, or even better than, the original set of features does. To do so, several feature importance measures have been implemented in FS algorithms. In this paper, we focused on measures that can be directly extracted from labeled data and, therefore, do not consider an external classifier to evaluate the importance of the features (supervised filter feature selection). In particular, these measures are applied in a dataset D with N instances $E_i = (x_i, y_i)$, $i = 1 \dots N$. Each instance E_i is associated with a vector of features $x_i = (x_{i1}, x_{i2}, \dots, x_{iM})$, described by M features X_j , $j = 1 \dots M$, and a label (class) y_i .

The measures used in this work can be broadly divided into four categories, inherent to the taxonomy described in Liu and Motoda (1998):

- Information measures are based on data entropy estimation;
- Distance measures, also known as separability measures, consider the dissimilarity between examples from/within one or more classes in terms of a given feature or feature subset;
- Dependency measures estimate the ability of a feature in predicting the value of another feature (e.g., their correlation);
- Consistency measures give preference to reward features that still maintain data consistent to their class values, i.e., features with similar values in data instances from the same class.

Eight feature selection algorithms, based on feature importance measures from different categories, were employed in this work:

1. **Consistency-Based Filter (CBF)**: selects small subsets of features that have high consistency, i.e., features that do not present similar or identical values for different class values (Liu and Setiono, 1996). To do so, the algorithm iterates the following steps:
 - a. The generation of a random subset of features (S) with k features, $k \leq M$;
 - b. First verification: if S is smaller than the best feature subset obtained so far, the data described by the features in S will be checked out according to an inconsistency criterion;

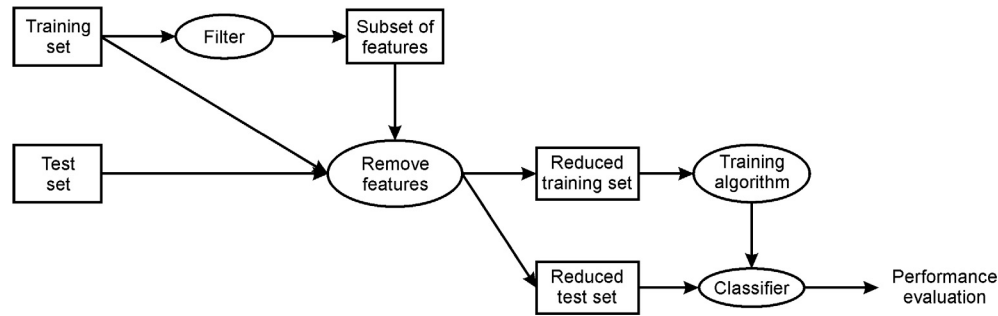


Fig. 1. Experimental protocol followed.

- c. Second verification: if the inconsistency value is also lower than the best value obtained so far, CBF will set S as the best feature subset found.
- II. **Correlation-Based Feature Selection (CFS)**: evaluates the quality of a feature subset S based on the predictive ability of each feature, i.e., the correlation between the feature and the class, as well as the degree of correlation among features within the subset. Thus, CFS rewards subsets of features highly correlated to the class and with low redundancy between each other (Hall, 1999). This importance measure herein employed fits the dependency category.
- III. **Filtered Subset Evaluator (FSE)**: runs a feature subset evaluator on the output yielded by a previously established filter that must maintain the original number and order of the features. In this work, the filter chosen was CFS, such that FSE falls within the dependency category (Hall et al., 2009).
- IV. **Chi Squared (CS)**: estimates the independence between a feature X_j and the class (Liu and Setiono, 1995). First all numeric features are discretized. Afterwards, for each class and feature value, this dependency criterion calculates the difference between the observed and the expected co-occurrences of these values. The higher the differences, the more significant the feature is.
- V. **Gain Ratio (GR)**: rewards features which minimize the data entropy, leading to a higher information gain (Hall et al., 2009). Different from the classic information gain criterion, the Gain Ratio does not favor features with larger number of values.
- VI. **Relief-F**: rewards features that have different values when considering a pair of examples from different classes, and penalizes features which have different values for examples from the same class (Kononenko, 1994). In other words, this algorithm gives higher importance value to features that lead to better class separability. It also takes into account the effect of interacting features. Relief-F fits the distance category.
- VII. **Symmetrical Uncertainty (SU)**: takes into account the correlation between each feature X_j and the class (Hall and Smith, 1998). The higher the correlation, the stronger the dependency between the feature and the class. As SU also employs a procedure based on an entropy measure, it is both a dependency and an information importance measure.
- VIII. **Ensemble Feature Selection (EFS)**: unlike the previous methods, EFS combines the results from different algorithms. In this work, EFS combines the results from the Relief-F, SU and GR algorithms (Guyon et al., 2007; Saeys et al., 2008; Abeel et al., 2010; Prati, 2012). Thus, complementarities among the techniques and different views from data can be explored. EFS follows the Borda approach (Dwork et al., 2001), in which each feature has a general rank based on the average of the ranks assigned to this feature by each FS algorithm aggregated.

The first three algorithms (CBF, CFS and FSE) deal with features according to a multivariate perspective, that is, they search for subsets of features, evaluating each possible subset at a time. In this work, they are applied according to the best first strategy (greedy hill climbing with backtracking). The last five algorithms, on the other hand, follow a univariate perspective, evaluating each feature individually and ranking all of them accordingly. Therefore, in their cases, a given threshold must be established on the number of features to be chosen or on the values achieved for the importance measures (Section 2.3). We selected the best 10–90% (with steps of 10%) features in the yielded rankings.

Table 1 summarizes important properties regarding the eight feature selection algorithms evaluated in this work.

All feature selection algorithms are implemented in the publicly available Weka software (Hall et al., 2009), with their default settings. The operating system was Windows 7 Professional on a workstation Intel DX58SO2 with Intel Core i7-990X Extreme processor, 24.0 GB RAM.

2.3. Classifier

As part of the evaluation protocol, SVM classifiers (already tested with data from AD in Lehmann et al. (2007), Trambaiolli et al. (2011a, 2011b)) were induced using the original data set (with 345 feature values) and using each of the data sets described by the subsets of features selected. The SVM is an ML technique where a training dataset (containing known labeled data examples) is used to find a hyperplane based on the features coordinates. This hyperplane should initially separate training data from two classes with a maximum margin. Afterwards, it can be used to classify new data points. When the classes are not linearly separable, feature coordinates are first mapped into a high dimension space by a kernel function. In the new space, classes become linearly separable, and the hyperplane with maximum margin can then be found (Cristianini and Shawe-Taylor, 2000).

Table 1

Important properties from the eight supervised filter feature selection algorithms evaluated in this work. The algorithms listed are: Consistency-Based Filter (CBF), Correlation-Based Feature Selection (CFS), Filtered Subset Evaluator (FSE), Chi Squared (CS), Gain Ratio (GR), ReliefF, Symmetrical Uncertainty (SU) and Ensemble.

Algorithm	Category	Perspective
CBF	Consistency	Multivariate
CFS	Dependency	Multivariate
FSE	Dependency	Multivariate
CS	Dependency	Univariate
GR	Information	Univariate
ReliefF	Distance	Univariate
SU	Dependency and Information	Univariate
Ensemble	Distance, Dependency and Information	Univariate

We used the LibSVM tool (Fan et al., 2005) in the SVMs induction. The linear kernel $K(x_i, x_j) = x_i^T x_j$ was used in the experiments, with the soft margin cost function parameter $C = 1$.

Statistical analysis was conducted to compare the accuracy achieved after each FS algorithm application with the accuracy obtained without the FS step. We evaluated the per-epochs results with a Student's t-test for repeated measures (comparison of average accuracies as continuous numbers) and the per-patient results with a Wilcoxon test for repeated measures (comparison of correctly classified subjects as binary numbers). P-values were adjusted by using the Bonferroni correction for 48 multiple comparisons (3 subsets of algorithms + 5 ranker algorithms * 9 ranker percentages). It is important to mention that the p-values of these statistical tests are only approximated, since the exact null distribution in cross-validation procedures is unknown (Noirhomme et al., 2014). However, the proper permutation test in this case of paired comparisons between two accuracies is not established.

3. Results

3.1. Classification performance

Tables 2 and 3 present the predictive results of the SVM classifiers generated in this study. Three performance metrics are presented, namely, overall accuracy (percentage of correct classifications for all subjects – AD patients and controls), sensitivity (percentage of correct classifications for AD patients) and specificity (percentage of correct classifications for healthy subjects). The best results are highlighted. While Table 2 presents the average and standard deviation of these measures for all epochs, Table 3 exposes one classification per subject based on the predictions obtained for all his/her epochs. Therefore, a subject is diagnosed according to a majority voting of the predictions for his/her 40 epochs. Thus, Table 3 does not show a standard deviation, since it represents a single proportional measure. Both tables also show the percentage of dataset reduction after each feature selection technique. For the ranking feature selection algorithms, it is necessary to define a threshold for selecting a subset of features. We varied this threshold between 10% and 90% (with steps of 10%) of the original number of features. Tables 2 and 3 present the best performance achieved among the different thresholds.

CBF selected the lowest number of features, reducing the original feature space in $97.53 \pm 0.15\%$. This technique achieved the second best performance in the per epochs classification, with $85.15 \pm 23.86\%$ of accuracy, $88.64 \pm 20.96\%$ of sensitivity and $78.75 \pm 28.29\%$ of specificity. It also showed one of the highest performances regarding per patient classification (which tied with the CFS algorithm), with 91.18% of accuracy, 95.45% of sensitivity and 83.33% of specificity.

The highest predictive performance in the per-epoch classification, namely $85.29 \pm 21.62\%$ of accuracy (p -value = 0.032), $87.27 \pm 23.12\%$ of sensitivity and $81.67 \pm 18.96\%$ of specificity, was achieved by the FSE algorithm. In addition, it also obtained the highest accuracy (91.18%) and balanced sensitivity values (90.91%) and specificity (91.67%) in the per-patient basis.

3.2. Data set reduction

Fig. 2 presents a scatter plot relating the predictive performance measures and the reduction percentage in the number of features achieved by the feature selection techniques. In these graphs, the x-axis represents a given classification performance measure (accuracy, sensitivity or specificity), while the y-axis represents the percentage of reduction in the number of features. Each feature selection technique is plotted in this graph by considering the predictive performance achieved by the SVM classifier when employing its corresponding reduced subset of features and the reduction obtained in the number of features. By examining this graph, it is possible to identify the best performing feature selection procedures by taking into account the relation between the predictive performance and the number of features used. Feature selection techniques with points closer to the up right corner of the plot represent best solutions, that is, subsets of features that lead to high performance, while using fewer features.

These plots reinforce what was presented in Table 3, in which six of the eight algorithms are in the first quadrant (with more than 50% accuracy and 50% of reduction in the number of features). Moreover, CBF, CFS and FSE algorithms are always present in the percentiles above 80% for both measures. On the other hand, the algorithms SU and GR are in the second quadrant of the graph, demonstrating good classification accuracy but a modest dataset reduction.

3.3. Selected features

We evaluated eight different FS algorithms submitted to the same 345 spectral peak input features (69 simulated channels * 5 frequency bands – delta, theta, alpha, beta, and gamma). At each LOSO iteration, each algorithm applies specific measures to estimate the importance of the features subset, leading to different optimal subsets (for more information about the experimental procedure, see Section 2.1 and Fig. 1). Fig. 3 shows a set of heat maps that are colored according to the selection frequency (the ratio between the number of LOSO iterations in which the feature is selected by the total number of LOSO iterations) of each feature and algorithm. Features are represented in the x-axis, while the feature selection techniques are in the y-axis. Features with hotter colors (closer to the red color) in the graph are those selected more times while features with cooler colors (closer to the blue color)

Table 2
Best predictive performance (mean \pm standard deviation) and percentage of features reduction on a per-epoch basis after leave-one-subject-out tests. The best results are highlighted in bold. Without FS line does not have reduction percentages or p-values since these measures are obtained comparing post-FS results versus without FS results. The algorithms listed are: Consistency-Based Filter (CBF), Correlation-Based Feature Selection (CFS), Filtered Subset Evaluator (FSE), Chi Squared (CS), Gain Ratio (GR), ReliefF, Symmetrical Uncertainty (SU) and Ensemble. *Ranker algorithms have no standard-deviation in the percentages of reduction due to experimental setup.

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	Features reduction (%)	p-value
Without FS	71.40 \pm 33.26	82.39 \pm 28.38	51.25 \pm 33.11	–	–
CBF	85.15 \pm 23.86	88.64 \pm 20.96	78.75 \pm 28.29	97.53 \pm 0.15	0.067
CFS	84.93 \pm 20.66	87.84 \pm 20.58	79.58 \pm 20.58	88.55 \pm 1.05	0.190
FSE	85.29 \pm 21.62	87.27 \pm 23.12	81.67 \pm 18.96	88.76 \pm 1.12	0.032
CS	75.22 \pm 29.38	82.84 \pm 27.16	61.25 \pm 29.18	80.00*	1.000
GR	74.19 \pm 33.18	81.70 \pm 31.73	60.42 \pm 32.56	90.00*	1.000
ReliefF	80.00 \pm 29.44	88.41 \pm 21.97	64.58 \pm 35.75	80.00*	0.626
SU	75.88 \pm 32.24	83.86 \pm 29.59	61.25 \pm 32.97	90.00*	1.000
Ensemble	73.97 \pm 32.16	83.86 \pm 28.49	55.83 \pm 31.59	60.00*	1.000

Table 3

Best predictive performance and percentage of dataset reduction on a per-subject basis after leave-one-subject-out tests. The best results are highlighted in bold. Without FS line does not have reduction percentages or p-values since these measures are obtained comparing post-FS results versus without FS results. The algorithms listed are: Consistency-Based Filter (CBF), Correlation-Based Feature Selection (CFS), Filtered Subset Evaluator (FSE), Chi Squared (CS), Gain Ratio (GR), ReliefF, Symmetrical Uncertainty (SU) and Ensemble. Ranker algorithms have no standard-deviation in the percentages of reduction due to experimental setup.

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	Features reduction (%)	p-value
Without FS	73.53	86.36	50.00	–	–
CBF	91.18	95.45	83.33	97.53 ± 0.15	0.944
CFS	91.18	95.45	83.33	88.55 ± 1.05	0.944
FSE	91.18	90.91	91.67	88.76 ± 1.12	0.944
CS	82.35	90.91	66.67	90.00*	1.000
GR	79.41	86.36	66.67	20.00*	1.000
ReliefF	85.29	90.91	75.00	90.00*	1.000
SU	82.35	90.91	66.67	40.00*	1.000
Ensemble	85.29	90.91	75.00	70.00*	1.000

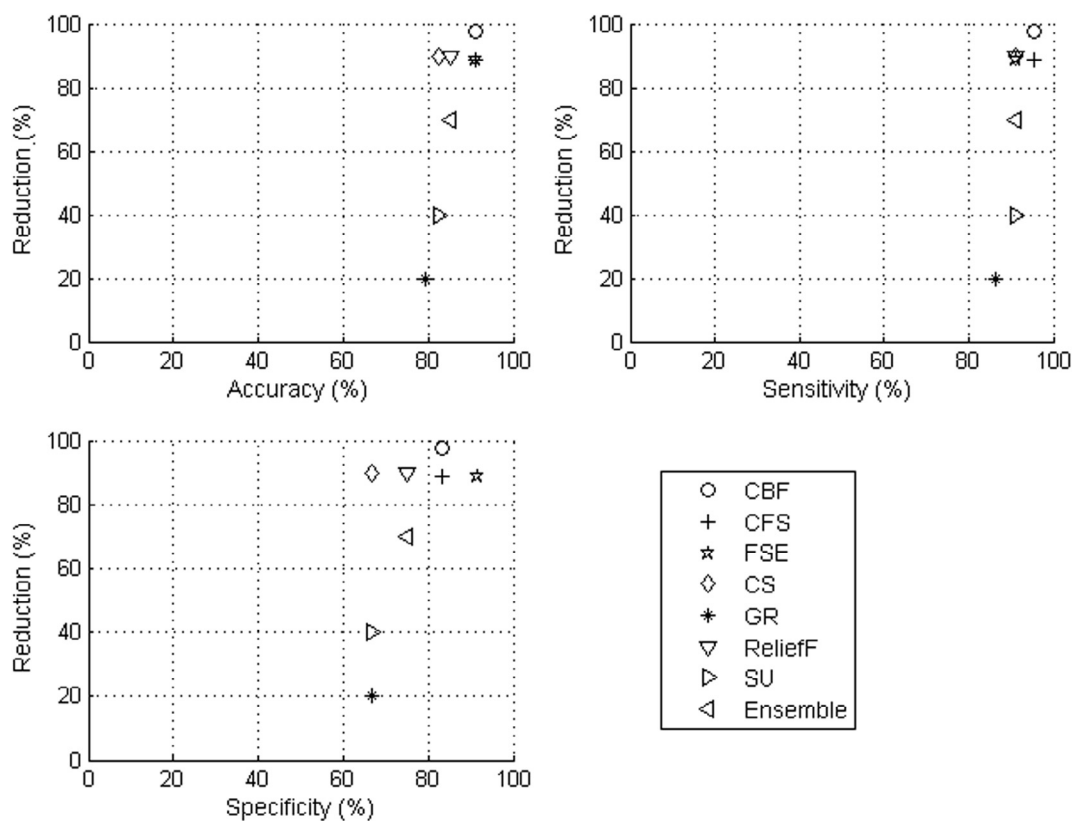


Fig. 2. Scatter plot relating the performance with the quantity of features selected by each algorithm. The y-axis presents the percentage of reduction from the original dataset, while the x-axis presents the following performance measures: (A) accuracy, (B) sensitivity and (C) specificity. Note that points closer to the up right corner of each graph achieve better performances with fewer features. The represented algorithms are: Consistency-Based Filter (CBF), Correlation-Based Feature Selection (CFS), Filtered Subset Evaluator (FSE), Chi Squared (CS), Gain Ratio (GR), ReliefF, Symmetrical Uncertainty (SU) and Ensemble.

are those selected fewer times. The heat maps were organized according to the bands where they were calculated, so that the first heat map corresponds to the delta band, followed by the theta band and so forth.

It is notable that alpha band features have the highest frequency of selection for all algorithms, followed by beta and gamma bands. Besides that, most of the selected features are arranged in the left hemisphere, especially at C3, P3, T3 and T5 positions.

Since the best predictive performance was obtained by using the features selected by the FSE algorithm, Fig. 4 presents the features selected by this technique in all LOSO rounds. For each feature, the graph presents a box-plot of the control group (blue

box) and AD group (red box). Features of the fast frequency bands (alpha, beta and gamma) have reduced spectral peak for the AD group, while the delta band has the inverse effect (reduced spectral peak in the control group).

Furthermore, by taking the two features with higher separability between groups (T5-T6 and P3-P4 – alpha band), Fig. 5 presents graphs of the points distribution in a per-epochs and per-patients (mean value of their respective epochs) basis. In both graphs, we see a grouping of control subjects with higher spectral peak values for both measures and the opposite effect in AD patients. Also, considering the patients distribution, only two AD subjects would not be correctly identified by using just these two features.

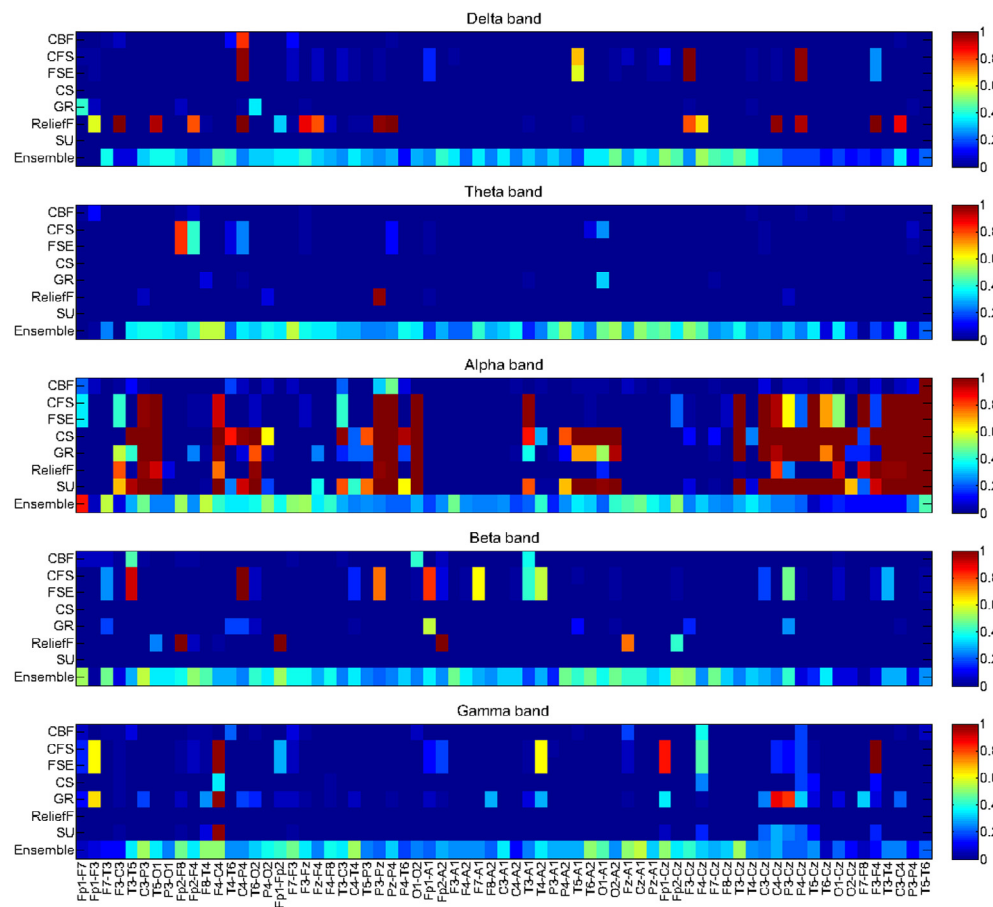


Fig. 3. Heat maps presenting the selected features for each algorithm. Points which vary their shades closer to the white color show that the feature (x-axis) was selected more times by the correspondent algorithm (y-axis) during the leave-one-subject-out iterations. The represented algorithms are: Consistency-Based Filter (CBF), Correlation-Based Feature Selection (CFS), Filtered Subset Evaluator (FSE), Chi Squared (CS), Gain Ratio (GR), ReliefF, Symmetrical Uncertainty (SU) and Ensemble.

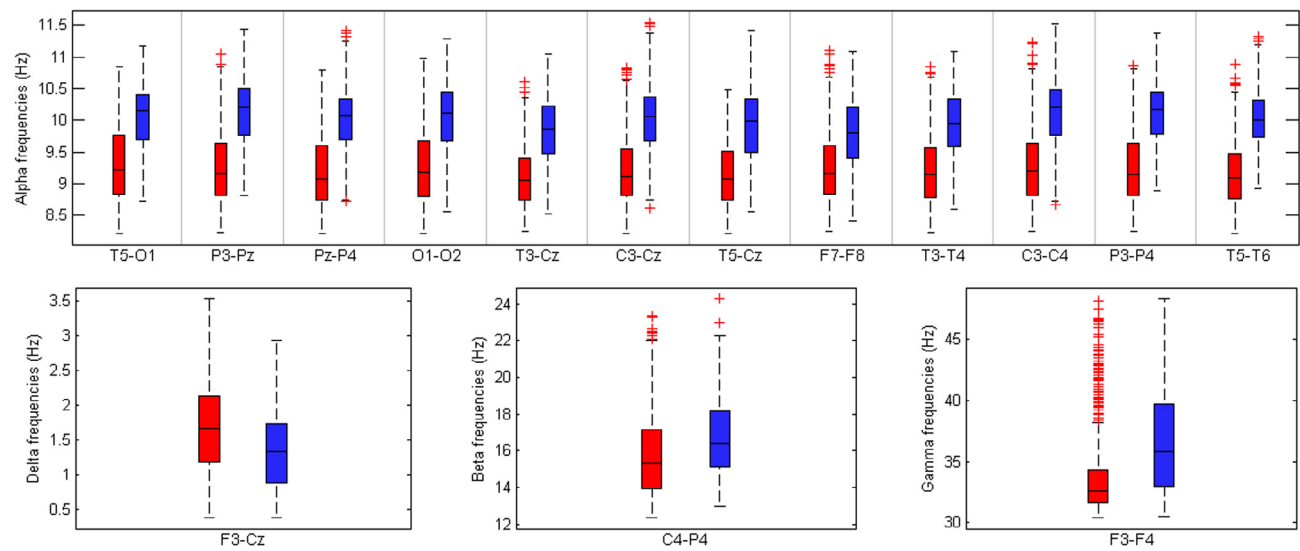


Fig. 4. Boxplots of the values of the features selected in all leave-one-subject-out interactions using Filtered Subset Evaluator algorithm. Blue boxes represent data from healthy controls and red boxes from patients. Note that high frequency features (alpha, beta and gamma features) are lower for patients, while low frequency feature (delta feature) is lower for controls. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

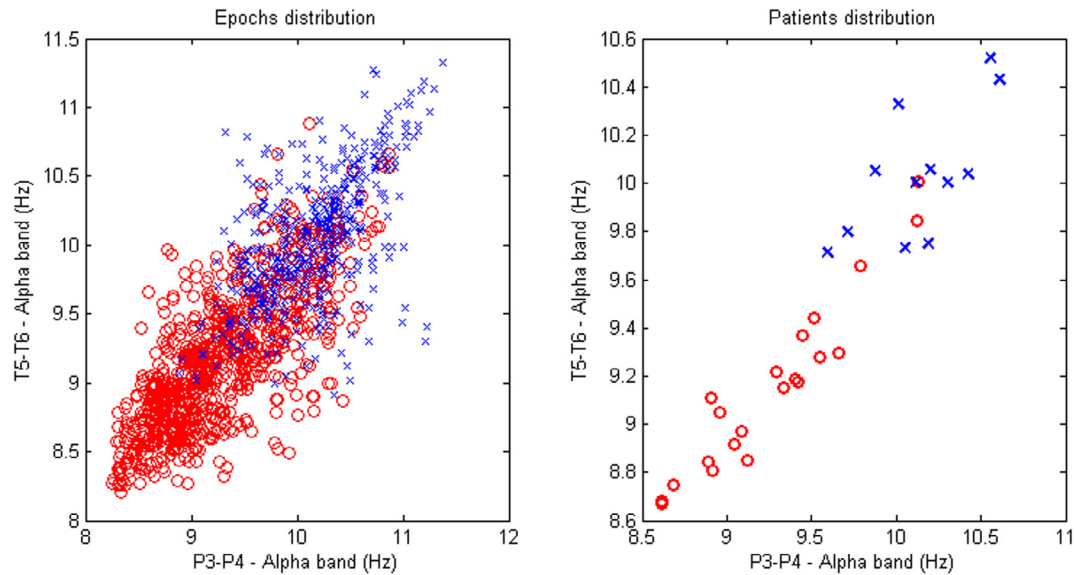


Fig. 5. Epochs and patients distribution using the two features (selected by Filtered Subset Evaluator algorithm) with the highest separability between classes. Blue crosses represent data regarding healthy controls and red circles refer to AD patients. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

4.1. Classification performance and data set reduction

Although using fewer features, all data sets after feature selection resulted in classifiers achieving similar or, in most of cases, higher performances than the ones achieved by classifiers using all features (without feature selection – WFS). These results indicate that, for computer-aided AD diagnosis, it is not always the case that a larger amount of features results in higher performance. Moreover, the lower performance achieved by WFS suggests that some features even hinder the predictor induction, which further motivates the application of feature selection.

CBF was the feature selection technique that selected the lowest number of features, achieving the second best performance in the per epochs classification and one of the highest performances regarding per patient classification (which tied with the CFS algorithm). It is noteworthy that these results are consistent with the CBF characteristics, which tends to produce subsets of features with few features and high consistency (Liu and Setiono, 1996). The fact that some of the best accuracy results came from the smallest data projections suggests that they are indeed consistent for data classification in this scenario.

On the other hand, although using a slightly larger amount of features, FSE selected subgroups of features that lead to the highest predictive performance in the per-epoch classification, as well as the highest accuracy and balanced sensitivity and specificity values in the per-patient basis. The per-patient performance is an important measure, since it simulates a real clinic situation, with some examples composing one subject diagnosis. Since FSE makes use of the CFS subset evaluator algorithm, it can be inferred that the features selected by this technique are highly correlated to the diagnosis class (controls \times AD), and show low correlation among each other (Hall, 1999; Hall et al., 2009).

Regarding the performance of ranking algorithms, it is visible that Relief-F was able to produce reduced data sets and classifiers with better predictive accuracy, using only 20% of the original features in the per-epoch classification and 10% of the original data set features in the per-patient classification. As already described in the literature, good results obtained with the algorithm Relief-F

in EEG data may be attributed to its ability to handle noisy data and correlated features, as well as by its effectiveness in dealing with data sets of high dimensionality (Aarabi et al., 2006).

The Ensemble algorithm presented the same results as Relief-F, but with lower reductions in the number of features. This is an unexpected result considering that the Ensemble combines the results from the Relief-F, SU and GR algorithms (Guyon et al., 2007; Saeys et al., 2008; Abeel et al., 2010; Prati, 2012). Thus, we are able to suppose that SU and GR methods should confuse the Ensemble approach, introducing non-important features into the optimal subset found by the Relief-F algorithm, for example.

It is also noteworthy in Fig. 2 that, with the exception of SU and GR algorithms, all other FS techniques are positioned in the right upper quadrant of the graph. Therefore, all of them were able to reduce the number of features while also achieving good classification performance. Moreover, FSE, CBF and CFS techniques were able to sharply reduce the number of features, while also achieving the highest accuracy results. CBF and CFS were the best performing techniques regarding the ratio between their sensitivity and the reduction in the number of features employed, while, for specificity, FSE is highlighted. All of them evaluate subsets of features (Liu and Setiono, 1996; Hall, 1999; Hall et al., 2009), which was found to perform better in our application than evaluating each feature individually. In fact, feature rankers tend to maintain redundant features, as they show similar ranks. Joining rankers in an ensemble did not alleviate this effect, since the ensemble result was usually worse than that of its individual counterparts.

It is important to note that the results in Table 2 present high standard deviation (SD). It is a methodological consequence of the leave-one-subject-out test, where an individual with the worst epochs accuracy reduces the group mean and increases the SD. In light of this, the k-folds cross-validation performed in the Supplementary Material helps to prove the classifier stability, since in this case all accuracy measures presented SD lower than 10%.

Finally, multivariate analysis considers subsets of features in which the sum of small individual contributions leads to the identification of different patterns from each group (Siddis et al., 2003; Haynes and Rees, 2006). In this context, FS supplements the search for an optimum subset of features able to differentiate the groups using multivariate analysis (Liu and Motoda, 2007). Thus, even if

some results after feature selection did not present an improved statistical significance, any improvement/maintenance in accuracy measures associated to the reduction of the original number of features is a contribution itself.

4.2. Selected features

Considering the heatmap (Fig. 3), it is observable that almost all algorithms select relatively similar feature subsets. Despite of the LOSO interaction, a few features were usually selected. This characteristic is more evident, for instance, for those algorithms that search for subsets of features (CBF, CFS and FSE).

On the other hand, the ranker algorithms, which select a specific number of features for each LOSO iteration, showed greater variability among the selected features. As an example, the Ensemble technique presents several features with shades of light blue in the heatmap, but few of them near to the maximum values. It seems that, contrary to other algorithms, there was not a significant consensus in the features lists, which has led to the highest variation in the feature subset chosen at each interaction.

It should be emphasized that alpha band features have the highest frequency of selection for all algorithms. Although previous work found a significant reduction in spectral power for higher frequencies of the alpha band (9–13 Hz) for patients with mild AD (Moretti et al., 2004), the slowing of this band is more commonly found in intermediate stages of AD (Jeong, 2004).

Another important finding was that the beta band features are also considered relevant, mainly by CFS and FSE algorithms, with a predominance of electrodes in the frontal, central, temporal and posterior regions. Literature shows that slowing at this band is considered the first EEG finding in AD (Coben et al., 1983; Ihl et al., 1993; Jeong, 2004), beginning with reductions in posterior regions and then in the sequence with their increase in frontal regions in more severe cases (Ihl et al., 1996).

Still concerning the electrode locations, the frequent selection of left hemisphere features is consistent with previous findings in which (mild cognitive impairment) MCI patients subsequently evolving to AD and mild AD patients have more pronounced differences settled in the left hemisphere when compared to elderly controls (Rodriguez et al., 1999; Jelic et al., 2000).

Interestingly, all algorithms also selected some features of the gamma band, whose meaning in the EEG of Alzheimer's has still no consensus. By exploring the role of gamma oscillations in memory skills, Park et al. (2012) found reduced measurements in patients with MCI compared to controls. In contrast, van Deursen et al. (2008) found increased gamma power in AD when compared to controls and to patients with MCI. In our data, the most selected gamma feature was F4-C4, which decreased in AD patients.

The features subsets that gave rise to the highest predictive performance were selected by FSE. By a specific evaluation of those features selected in all LOSO interactions (Fig. 4), there is a replication of the general findings of this paper: predominance of alpha band features, focusing mainly on electrodes in central, temporal, posterior and occipital regions from the left hemisphere. It is interesting that almost half of these top-features correspond to measures from bipolar montage (Bco). These results corroborate the findings of Trambaiolli et al. (2011b), which tested feature sets from different montages with SVM and Logistic Regression classifiers and found out that the Bco montage data set provided the best predictive performance.

We point out that these results need to be interpreted carefully. The main limitation of this study is the available sample size. Although we take care to avoid overfitting, for example, to the proportion of training set (1320 examples > 345 features) in LOSO tests with the maximum combination of attributes, small samples can bring distorted and non-generalizable results (Chu et al., 2012).

On the other hand, studies like this highlight the importance of using FS in EEG data.

Another limitation is because the original recording was performed with an interconnected earlobe electrodes reference and different montages were obtained virtually. Hence, virtual bipolar assemblies provide a standardization of records, reducing possible common physiological noise between the two electrodes. Due to this fact, information regarding regional differences, or between the hemispheres, become clearer in the case of Bco. Perhaps for this reason, there is relevance for interhemispheric features, since the dynamics of degeneration of the disease can be more accelerated in one of the hemispheres.

Finally, evaluating the features selected by the FSE algorithm, which allows a better data separability, it becomes clear how easier the identification of the two classes can be drawn. Although this is not as evident in the per epochs evaluation, it is notable that only two patients would not be correctly identified by using just this information. These two patients have extremely similar distribution to control subjects, which justifies the need for exploring other attributes. Moreover, cases like these also request a follow-up, once the diagnosis accuracy in university hospitals permeates the 85–93% interval (Parikh et al., 2005) and can only be confirmed after post-mortem examinations.

5. Conclusions

This study applies eight feature selection algorithms to evaluate variables deriving from EED data which are more important for dementia diagnosis. With the experimental results, it is possible to infer that some features together are more relevant than others and that selecting the most suitable subsets of features can reduce noise and consequently allow to induce more accurate results. Among the feature selection algorithms, it is observable a superiority for algorithms which perform a subset evaluation of the features, instead of ranking all the features individually. For the classification, the selected feature subsets gave rise to classifiers with high accuracies in AD diagnosis (more than 90%), while using less than 20% of the original feature set.

Future studies should include a greater number of individuals to validate the findings found here. Despite the fact that SVMs are mainly used in the area with many successful results, other Machine Learning classification methods could also be employed to evaluate the subsets of features.

Acknowledgements

The authors would like to thank “Universidade Federal do ABC” (UFABC), “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” (CAPES), “Fundação de Amparo à Pesquisa do Estado de São Paulo” (FAPESP) (projects 2013/10952-9, 2013/10498-6, 2013/00506-1 and 2012/22608-8) and “Conselho Nacional de Desenvolvimento Científico e Tecnológico” (CNPq) for financial support.

Conflict of interest

We have no conflict of interest to declare.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.clinph.2017.06.251>.

References

- Aarabi A, Wallois F, Grebe R. Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis. *Clin Neurophysiol* 2006;117(2):328–40.
- Abeel T, Helleputte T, Van de Peer Y, Dupont P, Saeys Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics* 2010;26(3):392–8.
- Ahmadlou M, Adeli H, Adeli A. New diagnostic EEG markers of the Alzheimer's disease using visibility graph. *J Neural Transm* 2010;117(9):1099–109.
- Anghinah R, Basile LI, Schmidt MT, Sameshima K, Gattaz WF. Biologic artifacts in quantitative EEG. *Arq Neuro-psiquiat* 2006;64:264–8.
- Batsch NL, Mittelman MS. World Alzheimer Report 2012: Overcoming the stigma of dementia. London: Alzheimer's Disease International; 2012.
- Chu C, Hsu AL, Chou KH, Bandettini P, Lin C. Alzheimer's Disease Neuroimaging Initiative. Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* 2012;60(1):59–70.
- Coben LA, Danziger WL, Berg L. Frequency analysis of the resting awake EEG in mild senile dementia of Alzheimer type. *Electroencephalogr Clin Neurophysiol* 1983;55(4):372–80.
- Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press; 2000.
- Dwork C, Kumar R, Naor M, Sivakumar D. Rank aggregation methods for the web. In: *Proceedings of the 10th international conference on World Wide Web*; 2001. p. 613–22.
- Epstein CM. Digital EEG: choosing the right equipment for your lab. In: 30th Annual Course in Clinical EEG and Electrophysiology of American EEG Society. Washington: American EEG Society; 1995. p. 1–11.
- Esterman M, Tamber-Rosenau BJ, Chiu YC, Yantis S. Avoiding non-independence in fMRI data analysis: leave one subject out. *Neuroimage* 2010;50(2):572–6.
- Fan PE, Chen PH, Lin CJ. Working set selection using second order information for training SVM. *J Mach Learn Res* 2005;6:1889–918.
- Guyon I, Li J, Mader T, Pletscher PA, Schneider G, Uhr M. Competitive baseline methods set new standards for the NIPS 2003 feature selection benchmark. *Pattern Recognit Lett* 2007;28(12):1438–44.
- Hall MA, Smith LA. Practical feature subset selection for machine learning. In: *Proc. 21st Australian computer science conf*, 1998, p. 181–91.
- Hall MA. Correlation-based feature selection for machine learning PhD Thesis. The University of Waikato.; 1999.
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. *SIGKDD Explor* 2009;11(1):10–8.
- Hampel H, Frank R, Broich K, Teipel SJ, Katz RG, Hardy J, et al. Biomarkers for Alzheimer's disease: academic, industry and regulatory perspectives. *Nat Rev Drug Discov* 2010a;9(7):560–74.
- Hampel H, Shen Y, Walsh DM, Aisen P, Shaw LM, Zetterberg H, et al. Biological markers of amyloid β -related mechanisms in Alzheimer's disease. *Exp Neurol* 2010b;223(2):334–46.
- Haynes JD, Rees G. Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 2006;7:523–34.
- Huang C, Wahlund LO, Dierks T, Julin P, Winblad B, Jelic V. Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study. *Clin Neurophysiol* 2000;111(11):1961–7.
- Ihl R, Dierks T, Frolich L, Martin E, Maurer K. Segmentation of the spontaneous EEG in dementia of the Alzheimer type. *Neuropsychobiology* 1993;27(4):231–6.
- Ihl R, Dierks T, Martin EM, Frölich L, Maurer K. Topography of the maximum of the amplitude of EEG frequency bands in dementia of the Alzheimer type. *Biol Psychiatry* 1996;39(5):319–25.
- Jelic V, Johansson SE, Almkvist O, Shigeta M, Julin P, Nordberg A, et al. Quantitative electroencephalography in mild cognitive impairment: longitudinal changes and possible prediction of Alzheimer's disease. *Neurobiol Aging* 2000;21(4):533–40.
- Jeong J. EEG dynamics in patients with Alzheimer's disease. *Clin Neurophysiol* 2004;115(7):1490–505.
- Klass DW, Brenner RP. Electroencephalography of the elderly. *J Clin Neurophysiol* 1995;12(2):116–31.
- Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell* 1997;97(1):273–324.
- Kononenko I. Estimating attributes: Analysis and extensions of RELIEF. In: Bergadano F, De Raedt L, editors. *Machine Learning: ECML-94. ECML 1994. Lecture notes in computer science (lecture notes in artificial intelligence)* 1994; vol. 784. Berlin, Heidelberg: Springer; 1994.
- Lehmann C, Koenig T, Jelic V, Prichep L, John RE, Wahlund LO, et al. Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). *J Neurosci Methods* 2007;161(2):342–50.
- Leuchter AF, Spar JE, Walter DO, Weiner H. Electroencephalographic spectra and coherence in the diagnosis of Alzheimer's-type and multi-infarct dementia: a pilot study. *Arch Gen Psychiatry* 1987;44(11):993–8.
- Leuchter AF, Cook IA, Newton TF, Dunkin J, Walter DO, Rosenberg-Thompson S, et al. Regional differences in brain electrical activity in dementia: use of spectral power and spectral ratio measures. *Electroencephalogr Clin Neurophysiol* 1993;87(6):385–93.
- Liu H, Setiono R. Chi2: Feature selection and discretization of numeric attributes. In: *Proc. 70th IEEE internat conf on tools with Artif Intell*, 1995, p. 388–8.
- Liu H, Setiono R. A probabilistic approach to feature selection—a filter solution. *ICML* 1996;96:319–27.
- Liu H, Motoda H. Feature selection for knowledge discovery and data mining. Boston: Kluwer Academic Publishers; 1998.
- Liu H, Motoda H. Computational methods of feature selection. Boca Raton: CRC Press; 2007.
- Luccas FJC, Braga NIO, Fonseca LC, Frochtengarten ML. Guidelines for recording/analyzing quantitative EEG and evoked potentials: Part I. General aspects [in portuguese]. *Braz J Epilepsy Clin Neurophysiol* 1996;2:175–82.
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984;34(7):939.
- Moretti DV, Babiloni C, Binetti G, Cassetta E, Dal Forno G, Ferrer F, et al. Individual analysis of EEG frequency and band power in mild Alzheimer's disease. *Clin Neurophysiol* 2004;115(2):299–308.
- Niedermeyer E, da Silva FL. Electroencephalography: basic principles, clinical applications, and related fields. Philadelphia: Lippincott Williams & Wilkins; 2005.
- Noirhomme Q, Lesenfants D, Gomez F, Soddu A, Schrouff J, Garraux G, et al. Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions. *NeuroImage Clin* 2014;4:687–94.
- Parikh D, Stepenosky N, Topalis A, Green D, Kounios J, Clark C, et al. Ensemble based data fusion for early diagnosis of Alzheimer's disease. In: *IEEE conf of the engineering in medicine and biology society*, 2005, p. 2479–82.
- Park JY, Lee KS, An SK, Lee J, Kim JJ, Kim KH, et al. Gamma oscillatory activity in relation to memory ability in older adults. *Int J Psychophysiol* 2012;86(1):58–65.
- Prati RC. Combining feature ranking algorithms through rank aggregation. In: 2012 Intern joint conf on neural Netw (IJCNN), 2012, p. 1–8.
- Rabinovici GD, Karlawish J, Knopman D, Snyder HM, Sperling R, Carrillo MC. Testing and disclosures related to amyloid imaging and Alzheimer's disease: Common questions and fact sheet summary. *Alzheimers Dement* 2016;12(4):510–5.
- Rappelsberger P. The reference problem and mapping of coherence: a simulation study. *Brain Topogr* 1989;2(1–2):63–72.
- Rodriguez G, Copello F, Vitali P, Perego G, Nobili F. EEG spectral profile to stage Alzheimer's disease. *Clin Neurophysiol* 1999;110(10):1831–7.
- Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics* 2007;23(19):2507–17.
- Saeys Y, Abeel T, Van de Peer Y. Robust feature selection using ensemble feature selection techniques. In: *Machine learning and knowledge discovery in databases*, 2008, p. 313–25.
- Sandmann MC, Piana ER, Sousa DS, de Bittencourt PR. Digital EEG with brain mapping in Alzheimer's dementia and Parkinson's disease: a prospective controlled study. *Arq Neuropsiquiatr* 1996;54(1):50–6.
- Sanei S, Chambers JA. EEG signal processing. West Sussex: John Wiley & Sons; 2008.
- Schierle GSK, Michel CH, Gasparini L. Advanced imaging of tau pathology in Alzheimer Disease: New perspectives from super resolution microscopy and label-free nanoscopy. *Microsc Res Techniq* 2016;79(8):677–83.
- Sidtis JJ, Strother SC, Rottenberg DA. Predicting performance from functional imaging data: methods matter. *Neuroimage* 2003;20:615–24.
- Spitzer RL, Williams JB, Gibbon M, First MB. User's guide for the structured clinical interview for DSM-III-R: SCID. US: American Psychiatric Association; 1990. p. 212.
- Terry RD. Neuropathological changes in Alzheimer disease. *Prog Brain Research* 1994;101:383.
- Trambaiolli LR, Lorena AC, Fraga FJ, Kanda PA, Anghinah R, Nittrini R. Improving Alzheimer's disease diagnosis with machine learning techniques. *Clin EEG Neurosci* 2011a;42(3):160–5.
- Trambaiolli LR, Lorena AC, Fraga FJ, Kanda PA, Nittrini R, Anghinah R. Does EEG montage influence Alzheimer's disease electroclinic diagnosis? *Int J Alzheimers Dis*. 2011b;2011:761891.
- Van Deursen JA, Vuurman EF, Verhey FR, van Kranen-Mastenbroek VH, Riedel WJ. Increased EEG gamma band activity in Alzheimer's disease and mild cognitive impairment. *J Neural Transm* 2008;115(9):1301–11.
- Vos SJ, Gordon BA, Su Y, Visser PJ, Holtzman DM, Morris JC, et al. NIA-AA staging of preclinical Alzheimer disease: discordance and concordance of CSF and imaging biomarkers. *Neurobiol Aging* 2016;44:1–8.