



ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
“Antonio Ruberti”
CONSIGLIO NAZIONALE DELLE RICERCHE

G. Fiscon, E. Weitschek, P. Bertolazzi, M.C. De Cola, S. De Salvo, P. Bramanti, G. Felici

**EEG SIGNALS ANALYSIS TO DETECT
ALZHEIMERS DISEASE PATIENTS**

R. 14-10 2014

Giulia Fiscon – Istituto di Analisi dei Sistemi ed Informatica del CNR, via dei Taurini 19 - 00185 Roma, Italy. Email: giulia.fiscon@iasi.cnr.it.

Emanuel Weitschek – Istituto di Analisi dei Sistemi ed Informatica del CNR, via dei Taurini 19 - 00185 Roma, Italy. Email: emanuel.weitschek@iasi.cnr.it.

Paola Bertolazzi – Istituto di Analisi dei Sistemi ed Informatica del CNR, via dei Taurini 19 - 00185 Roma, Italy. Email: bertola@iasi.cnr.it.

Maria Cristina De Cola – IRCCS Centro Neurolesi “Bonino-Pulejo”, Messina, Italy.
Email: crisrina.decola@gmail.com.

Simona De Salvo – IRCCS Centro Neurolesi “Bonino-Pulejo”, Messina, Italy.
Email: simo.desalvo@hotmail.it.

Placido Bramanti – IRCCS Centro Neurolesi “Bonino-Pulejo”, Messina, Italy.
Email: bramanti@irccsneurolesiboninopulejo.it.

Giovanni Felici – Istituto di Analisi dei Sistemi ed Informatica del CNR, via dei Taurini 19 - 00185 Roma, Italy. Email: felici@iasi.cnr.it.

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", CNR
viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161

fax ++39-06-7716461

email: iasi@iasi.cnr.it

URL: <http://www.iasi.cnr.it>

Abstract

Alzheimer's disease (AD) is the most widespread form of dementia. It is a neurodegenerative disorder, for which actually no cure is known. Furthermore, a particular disease called Mild Cognitive Impairment (MCI) affects patients that suffer of some isolated cognitive deficit due to which they could develop AD. Advances in bioinformatics and clinical informatics assist the medical doctors to manage the exponential growth of demented patient data. Several clinical data sets are available from electronic health records in medical environments. In particular, Electroencephalography (EEG) appears as non-invasive and repeatable technique to diagnose brain abnormalities. The open challenge is to perform clinical studies in order to shed light on biological and medical questions related to AD and MCI. Despite of technological advances, the analysis of EEG continues to be carried out by experts, who are subject to laborious interpretation of the spectrum. Computational methods may lead to a quantitative analysis of these signals and hence to characterize EEG time series. The aim of this work is to achieve an automatic patients classification from the EEG biomedical signals involved in AD and MCI in order to support medical doctors in the right diagnosis formulation. The analysis of the biological EEG signals requires effective and efficient computer science methods to extract relevant information. Data mining, which guides the automated knowledge discovery process, is a natural way to approach EEG data analysis. Specifically, in our work we apply the following analysis steps: (i) pre-processing of EEG data; (ii) processing of the EEG signals by the application of time-frequency transforms; and (iii) classification by means of new and well-known machine learning methods. We obtain promising results from the classification of AD, MCI and control patient samples and we plan to extend the analysis and the pre-processing step by using different Time-frequency Transform and dedicated tools.

1. Introduction

Alzheimer's disease (AD) is regarded to be the most widespread form of dementia [5]. Alzheimer is a neurodegenerative disease whose symptoms are the loss of memory and cognitive impairments able to affect social and occupational activities. Currently, no drugs are known to cure AD and the average survival times from onset of dementia is about 4.5 years, with a peak of 11 years for younger patients (65-70 years at diagnosis) [38]. Three different phases characterize the AD progression: mild, moderate, and severe. Moreover, a preclinical stage, called Mild Cognitive Impairment (MCI), affects patients that suffer from some isolated cognitive deficit due to which they could develop AD [30], [31]. Diagnosing MCI and mild AD is hard, because most symptoms are often ascribed to normal consequences of ageing. Nowadays, the diagnosis requires a combination of physical, neurological, and neuropsychological evaluations, and a variety of other diagnostic tests including imaging techniques. Nevertheless, the final diagnosis can only be defined with a histopathological analysis of the brain [7]. However, early diagnosis of potential AD is crucial for the adoption of therapeutic strategies able to slow the progression of the disease.

In recent years, several studies investigated Electroencephalography (EEG) as prominent candidate for diagnosing AD [17], [13], [26]. The EEG is a non-invasive recording of the electrical spontaneous activity of the brain measured at different sites of the scalp [24] and hence the EEG signals can indirectly contain information about physiological conditions related to the brain. The brain activity is characterized by several rhythms that are distinguished by their different frequency ranges. The main five brain rhythms are the following: *delta* rhythm (0.5-4 Hz), *theta* rhythm (4-7 Hz, with an amplitude greater than 20 μV), *alpha* rhythm (8-13 Hz, with an amplitude of 30-50 μV), *beta* rhythm (13-30 Hz, with a low amplitude ranging from 5 to 30 μV) that is usually related to active thinking and attention, outside world, and problems solving, and finally the *gamma* rhythm (≥ 30 Hz), generally related to the mechanism of consciousness. Then, assuming that an EEG recorded under resting conditions can be representative of the brain global state, particular frequency patterns are supposed to be meaningful for recognizing abnormalities [16]. Therefore, we can formulate as working hypothesis that EEG time series related to healthy controls are different from those corresponding to patients affected by neurodegenerative diseases (e.g., Alzheimer) or other pathologies (e.g., epilepsy).

Different studies have shown that AD has (at least) three major effects on EEG: reduce complexity, slow-down of signals, and perturbations in EEG synchrony [13]. On one hand, studies on spectral analysis have shown that AD patients increase activity in the delta and theta frequency bands, while decrease activity in the alpha and beta bands [9], [2], [8], suggesting a slow-down of the EEG signal [25] (i.e., the signal frequency is lower than would be expected for the patient's age and level of alertness). On the other hand, studies on nonlinear dynamics analysis have agreed that AD leads to changes in the architecture of the neural network with a decrease of complexity in EEG temporal patterns [4], [28]. In fact, EEG of AD and MCI patients seems to be more regular than in age-matched control subjects, suggesting that MCI and AD can induce loss of neurons, alter their functional interactions, and hence make neural activity dynamics simpler and more predictable [12]. Furthermore, it has been recently suggested [12] that the slow-down and the loss of complexity in the EEG signals are strongly correlated. However, since there tends to be large variability among AD and MCI patients, the main issue in the analysis of EEG signals relies on the artifacts and patterns similarities to those corresponding to normal activities. Therefore, there has been growing interest of the scientific community in developing automated EEG analysis systems in order to improve the sensitivity of EEG for detecting AD and reducing the analysis times, especially when long recording are required [35].

In this work, we present how a computer aided technique based on EEG-signal preprocessing, spectrum analysis, and automatic classification with supervised machine learning methods may support the diagnosis of different dementia related pathologies (e.g., AD and MCI) and help to discriminate the healthy control subjects (later referred to as CT).

2. Materials and Methods

In dealing with EEG processing, an important issue resides on the huge number of dimensions used to describe each sample (also referred to as *features*). It is due firstly to the non-stationarity of the

Table 1: Data set Description

| Clinical Condition | N. of subjects [%] | | | Age (mean \pm std.dev. [years]) | | |
|--------------------|--------------------|----------|-----|-----------------------------------|----------------|----------------|
| | Female | Male | Tot | Female | Male | Tot |
| CT | 5 (36%) | 9 (64%) | 14 | 57.4 \pm 8.4 | 67.9 \pm 7.9 | 64.1 \pm 9.4 |
| AD | 29 (59%) | 20 (41%) | 49 | 78.2 \pm 7.6 | 78.6 \pm 4.1 | 78.4 \pm 6.4 |
| MCI | 20 (54%) | 17 (46%) | 37 | 72.7 \pm 9.1 | 75.7 \pm 9.7 | 74.1 \pm 9.4 |
| Tot | 54 (54%) | 46 (46%) | 100 | 74.2 \pm 10.1 | 75.4 \pm 8.2 | 74.8 \pm 9.2 |

EEG signals that leads to the features computation in a time-dependent way, and secondly to the large number of EEG channels. Therefore, EEG recognition procedure mainly includes three main steps: data collection and pre-processing, feature extraction, and classification.

2.1. Participants

We enrolled 86 patients referred for dementia to IRCCS Centro Neurolesi Bonino-Pulejo from April 2012 to July 2013 (49 women and 37 men), and 14 healthy controls (CT) (5 women and 9 men). According to the World Health Organization definition, an expert neurologist classified the patients in affected by MCI or affected by AD. The mean age of the enrolled AD, MCI, and CT patients (average \pm standard deviation) was 78.4 \pm 6.4, 74.1 \pm 9.4, 64.1 \pm 9.4 years, respectively. All three subject categories (MCI, AD, CT) were homogeneous for age and gender, indeed, the chi-square test did not show statistically significant association between gender and clinical condition ($p > 0.05$), as well as the two-tailed t-test did not find statistically significant difference in age between women and men ($p > 0.05$ for each subject's category). A more detailed description of the subject characteristics is reported in Table 1. All recruited subjects signed an informed consent to enter the study, which was approved by the Ethical Committee of the IRCCS Centro Neurolesi "Bonino-Pulejo". Inclusion (enrolment) criteria were a negative anamnesis for neurological comorbid disease, and the ability to undergo to the EEG recording; whereas, the only exclusion criterion was the assumption of a pharmacological treatment that should modify the electrical brain activity.

2.2. Data Collection and Pre-processing

Multi-channel EEG signals have been recorded using monopolar connections with earlobe electrode landmark [24]. The electrodes position over the scalp has been obtained according to the 10-20 International System for the electrode placement (i.e., Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2). An example of the EEG electrode placement is given in Figure 1. The electrodes measure the weak electrical potentials produced by the brain activities in the range of μ Volt. Recordings have been performed in the resting state with closed eyes. In this way, it may be assumed that different brain regions are governed by the same dynamic process.

Data have been collected with a signal duration of 300 seconds and at a sampling frequency of 1024 and 256 samples per second. In order to reduce the EEG background artifacts, we extract only 180 seconds for each signal (i.e., from 60 to 240 seconds) and we transform each one at 256 samples per second. Summarizing, the EEG potentials have been recorded from human samples belonging to three categories:

1. patients affected by AD;
2. patients affected by MCI;
3. healthy control samples (CT).

An example of the extracted EEG recordings of 180 seconds for each category are drawn in Figure 2, Figure 3, and Figure 4, respectively.

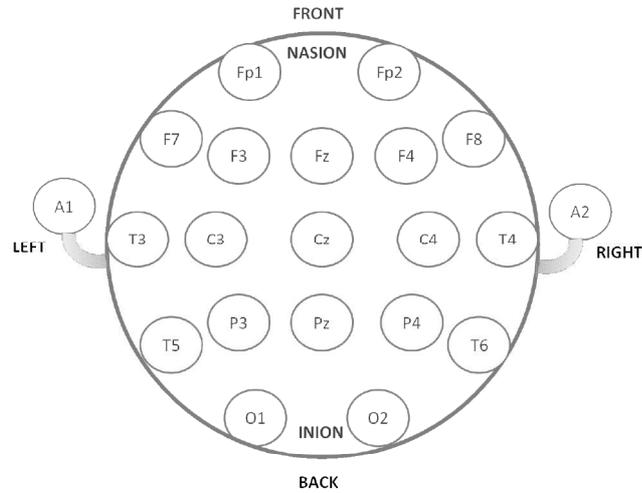


Figure 1: Placement of the EEG electrodes according to the 10-20 International System.

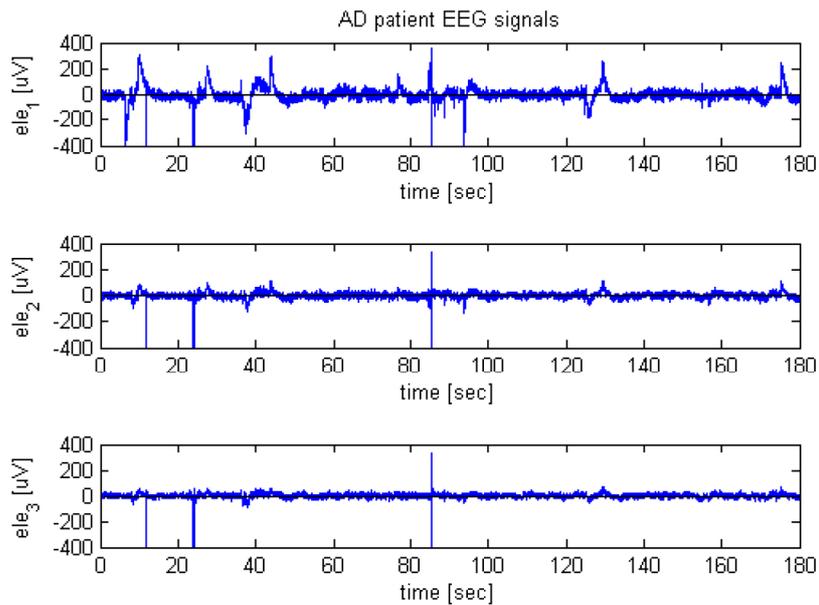


Figure 2: First three EEG electrode recordings of 180 seconds for a patient affected by AD.

2.3. Feature Extraction

The Feature Extraction step aims to extract desirable features from the rough time domain EEG signal. It has been shown that features extracted in frequency domain are one of the best solution to recognize the mental tasks starting from EEG signals [1]. We analyze the biomedical signal by means of the Fast Fourier Transform (FFT) which is based on the application of the Discrete Fourier Transform (DFT) to the signal in order to estimate its spectrum [33]. The DFT formula is the following:

$$X[k] = \sum_{s=0}^{S-1} x[s]e_k[s] \quad (1)$$

6.

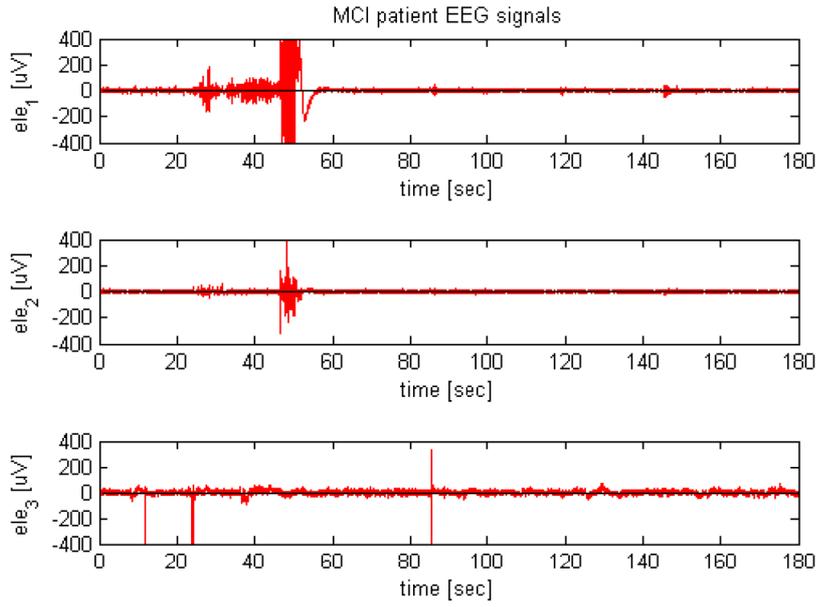


Figure 3: First three EEG electrode recordings of 180 seconds for a MCI diseased patient.

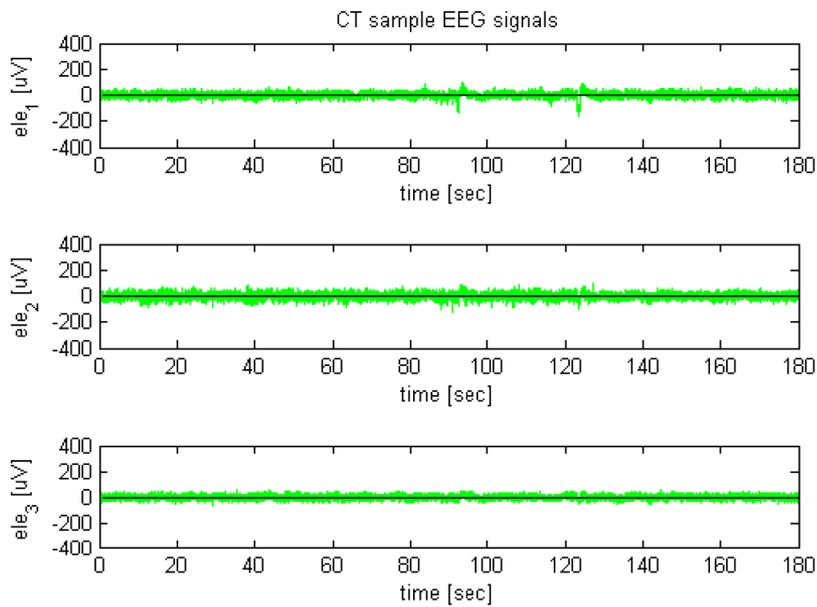


Figure 4: First three EEG electrode recordings of 180 seconds for a CT sample.

with:

- x : the time series signal (the data), $s = 0, 1, \dots, S - 1$;
- S : the total number of samples in signal x ;

- X : the frequency domain representation of the time-series signal x ;
- k : the k -th frequency component, $k = 0, 1, \dots, S - 1$;
- s : the s -th sample in the time domain;
- $e_k[s] = e^{-\frac{jk_s 2\pi}{S}}$: the k -th basis function.

$e_k[s]$ is computed at the same times s when the recorded signal $x[s]$ is sampled. Such a formula yields as output one complex number $X[k]$ for each k component.

Since EEG signal is non stationary, its spectrum changes with time and thus this signal can be approximated as an array of independent stationary signal components. In order to perform the next step, data collected and processed by extracting the Fourier Coefficients (N) require to be converted in a comma-separated matrix file. A schema of the matrix data representation is given in Table 2.

We have implemented the spectral analysis in MATLAB[®] [29]. MATLAB[®] is an high-level technical computing language whose key feature relies on its interactive environment for algorithm development, data visualization and analysis, signal processing, numerical integration, and several additional application fields. Specifically, we use the Fast Fourier Transform functions (FFT) to perform the EEG signals processing.

2.4. Classification

The classification step aims to assign an unknown instance into a given class based on its features [15] (e.g., AD-MCI). An additional goal is also to compute a clear and compact classification model that fits the data: for example the “if-then” rules (e.g., if $featureX > 6$ and $EEG - signal01 < 0.4$ then the patient sample is AD). The classification model can be a valid aid for the medical doctors and investigators to extract the clinical variables associated to neuropathologic diseases and to formulate a diagnosis for new patients.

We adopt an automatic classification approach, known as supervised learning: unknown objects are automatically assigned to a class by analyzing their attributes (features) by using a classification model computed from objects with a known class (training set). Each object is composed of a tuple (x_i, c_i) where x_i is the n -dimensional attribute set and c is the class of the i -th object (assume $c_i \in C = c_1, \dots, c_m$, with m number of classes). A formulation of the classification problem is the following: given t objects $\{x_1 \dots x_t\}$ whose class is known, build a function and use it to classify new objects, whose class is unknown. This function is called also *classifier* and is defined as $\bar{c} : R^t \rightarrow C$. The goal is to have $\bar{c}(x_i) = c_i$, for each x_i .

For evaluating the classification performances we adopt:

- *Accuracy* (A), or correct rate

$$A = \frac{TP + TN}{TP + TN + FP + FN}; \quad (2)$$

- *Precision* (P), or Positive Predictive Value (PPV)

$$P = \frac{TP}{TP + FP}; \quad (3)$$

Table 2: Example of the experiment data set with $P=n^o$ of patients, $M=n^o$ of electrodes, $N=n^o$ of Fourier Coefficients

| Patient | $Fourier_{(1,1)}$ | \dots | $Fourier_{(M,MN)}$ | Class |
|------------|-------------------|---------|--------------------|---------|
| $sample_1$ | $value_{(1,1)}$ | \dots | $value_{(1,MN)}$ | AD |
| $sample_2$ | $value_{(2,1)}$ | \dots | $value_{(2,MN)}$ | MCI |
| \dots | \dots | \dots | \dots | \dots |
| $sample_P$ | $value_{(P,1)}$ | \dots | $value_{(P,MN)}$ | Control |

8.

- *Recall (R)*, or True Positive Rate (TPR) or sensitivity

$$R = \frac{TP}{TP + FN}; \quad (4)$$

- *Specificity (S)*, or True Negative Rate (TNR)

$$S = \frac{TN}{TN + FP}; \quad (5)$$

- *F-measure (F)*

$$F = \frac{2P \cdot R}{P + R}; \quad (6)$$

where:

- *True Positives (TP)*: objects of that class recognized in the same class;
- *False Positives (FP)*: objects not belonging to that class recognized in that class;
- *True Negatives (TN)*: objects not belonging to that class and not recognized in that class;
- *False Negatives (FN)*: object belonging to that class and not recognized in that class.

In order to increase the robustness of our classification analysis, we use the cross validation approach. Cross validation is a standard sampling technique that splits the data set in a random way in k disjoint sets; then, the classification procedure is run k times with different sets. At a generic run k the $k - th$ subset is used as test set and the remaining $k - 1$ sets are merged and used as training set for building the model. Each of the k sets contains a random distribution of the data. The cross validation sampling procedure builds k models and each of this model is validated with a different set of data. Classification statistics are computed for each model and the average of these represents an accurate estimation of the classifier's performance. Specifically, we use the *leave-one-out* approach (L-1-out), that represents the degenerate case of a k -fold cross validation, where k is equal to the whole number of the given samples. For a data set with t samples, it performs t experiments by using $t - 1$ samples for the training set, while the remaining one is left for testing.

Currently many classification methods are present in the literature and for our analysis we choose Support Vector Machines, Decision Trees, and Rule-based classifiers. This choice stems from the fact that we aim to test and compare the performances of several supervised approaches, which rely on different mathematical and computational techniques, i.e., functions, trees, and logic formulas (rules).

2.4.1. Support Vector Machines

Support Vector Machines (SVM) [11] build a separating hyperplane, which maximizes the minimum distance between the data of different classes in a new space that has been obtained by applying a kernel function to the original data. SVM are particularly suited for binary classification tasks; in this case, the input data are two sets of n dimensional vectors. More formally, given a training data T consisting of t points in n input vectors where $T = \{(x_i, c(x_i)) | x_i \in R^n, c(x_i) \in \{-1, 1\}\}_{i=1}^t$ (in this case, the class for each point is labeled -1 or 1) we want to find a hyperplane $ax = b$ which separates the two classes. Ideally, we want to impose the constraints $ax_i b \geq 1$ for x_i in the first class and $ax_i b \leq -1$ for x_i in the second class. These constraints can be relaxed by the introduction of non-negative slack variables s_i and then rewritten as $c(x_i)(ax_i - b) \geq 1 - s_i$ for all x_i .

A training data with $s_i = 0$ is closest to the separating hyperplane $\{x : ax = b\}$ and its distance from the hyperplane is called margin. The hyperplane with the maximum margin is the optimal separating hyperplane. It turns out that the optimal separating hyperplane can be found through a quadratic optimization problem, i.e., minimize $\frac{1}{2}||a||^2 + C \sum_{i=1}^t s_i$ over all (a, b, s) which satisfy the constraint

$c(x_i)(ax_i - b) \geq 1 - s_i$. Here, C is a margin parameter, used to set a trade off between the maximization of the margin and minimization of classification error. The points that satisfy the previous constraint with the equality are called support vectors. The same reasoning can be applied when the data is transformed with a kernel function, and some theoretical results state that the computational complexity of the method is not affected by the complexity of the kernel function adopted.

Support Vector Machines can also be used less effectively for multi-class classification. Basically, in the multi-class case, the standard approach is to reduce the classification to a series of binary decisions, decided by standard Support Vector Machines.

Potential limitations of the method are:

- the high computational requirements for multi-class problems;
- the classification model cannot be interpreted easily in terms of the original variables by domain experts;
- the choice of the right type of kernel function.

In this study, we address these limitations by running the experiments in parallel, by not considering the classification model of SVMs as additional knowledge for AD / MCI, and by testing different kernel functions (i.e., linear and polynomial with coefficients 2 and 3). Indeed, the best results are obtained by setting the coefficient of the polynomial kernel function to 2.

2.4.2. Decision Trees

Decision Trees (e.g., C4.5 [34]) are used to model sequential decision problems. They are composed of nodes and edges: internal nodes represent the predicate of the objects in the data set, whereas each edge represents a splitting rules over one attributes (typically, binary splitting rules). Indeed, every node has two (or more) outgoing branches: one is associated with objects whose attributes satisfy the predicate, whereas the other to the ones which do not. The attribute classes are represented in the tree by leaf nodes. The classification is given by a model that predicts the class of the object by learning simple decision rules inferred from the data features. The class attribute is then assigned to the object by means of a path from the root to the output leaf node, where the predicates are applied to the object attributes and each node defines the path split. The widespread tree decision classifiers, such as C4.5 [34], rely on entropy rule or information gain rule which finds at each node a predicate that optimizes an entropy function of the defined partition.

Potential limitations of the method are:

- the computation of an optimal decision tree is an NP-complete problem;
- decision trees can be very sensitive to changes in the training data and outliers;
- complexity (too many splits and large trees).

In this work, the limitations of decision trees are handled mainly by proper parameters selection. In particular, we set the minimum instances per leaf to 10, which limits the growth of the tree and the too many splits that may lead to over-fitting, and we use a leave-one-out cross validation sampling scheme for testing all the different training data sets.

2.4.3. Rule-based classifiers

Rule-based classifiers [27] assign a given class to each object according to a specific function $r : condition \rightarrow c$ (called *classification rule*), such that the rule r covers an object x if the attributes of x satisfy the *condition* of r . Therefore, in this type of classification the classifier uses logic propositional formulas in disjunctive or conjunctive normal form (“if then rules”) for classifying the given samples.

A logic formula-based classifier classifies on the basis of the formula triggered by the sample. For extracting a set of classification formulas there are two main classes of methods: direct extraction from data and indirect extraction, which extract the formulas from other classification models, like Decision Trees. As example for indirect method we can derive from a decision tree the logic formulas whose clauses are represented by the paths from the root to the leaves.

Direct methods partition the attribute space into smaller subspace so that all the sample that belong to a subspace can be classified using a single classification formula. Examples of direct methods for computing separating formulas are RIPPER [10], LSQUARE [18], RIDOR [21] and PART [20]. These approaches normally produce sub optimal formulas because the formulas are generated in a greedy way. The major strength of classification formulas is the expressiveness of the models, that are very easy to interpret.

Potential limitations of the method are:

- it is susceptible to noise;
- it is high computationally demanding;
- the length of the logic formulas may be not interpretable by humans.

To deal with these limitations, we set a maximum number of 10 attributes (Fourier coefficients) chosen by a feature selection algorithm to perform rules extraction and classification.

2.5. Classification software

Two supervised machine learning software programs have been used to classify patient from the EEG processing signals:

2.5.1. DMB

DMB (Data Mining Big) [37], [3], [19] is a collection of data mining tools engineered for the classification of biological data, characterized by an underlying logic formalization and several optimization problems that are used to formulate different steps of data mining process. The main characteristic of the DMB system is the production of logic formulas as a model to characterize the data. DMB takes as input a matrix containing the elements and their attributes, plus a class label for each element. Regardless of the form of the input data it returns as output an explanation in terms of logic formulas of the type if [(X is in Ax) and (Y is in Ay) or (Z is in Az)] then CLASS = C, where X, Y, Z are attributes of the elements, Ax, Ay, Az are set of possible values that the attributes can take, and C is one of the classes in which the elements are partitioned.

DMB is composed of five main steps:

1. discretization: transformation of numerical features into discrete ones;
2. discrete cluster analysis: clustering of the features with the same behaviours;
3. feature selection: selection of the features that are considered to be more relevant;
4. logic formulas extraction;
5. classification.

DMB is composed of different multi-purpose tools such as BLOG, MALA, and DMIB designed to address the analysis of several biological data (e.g., DNA barcode and gene expression profiles). DMB is available at <http://dmb.iasi.cnr.it/>.

2.5.2. Weka

Weka (Waikato Environment for Knowledge Analysis) [22] is a Java open source package that collects the most widespread algorithms to handle mostly classification, numeric prediction, or clustering problems (available at cs.waikato.ac.nz/ml/weka). Among the several packages collected in Weka, the “Weka.classifier” package includes the implementation of classification and prediction algorithms, including the most important “Classifier” class. The latter defines the structure of any schema of classification or prediction assessment and it is made up by two methods, *buildClassifier()* and *classifyInstances()*, whose implementation is necessary for all supervised machine learning algorithms. Weka presents an integrated user friendly interface and its methods can be easily applied to the EEG signals after the effective preprocessing and input conversion into the comma separated matrix file format.

2.6. Design of experiments

In the following, we describe the experimental settings of the EEG signals processing.

2.6.1. Data Collection and Pre-preprocessing

Multichannel EEG signal are collected and recorded by

- using 19 electrodes;
- examining 100 subjects that belong to AD, MCI, and control classes;
- applying a sampling frequency of 256 or 1024 samples per second;
- taking into account 300 seconds of the signal duration.

Then, the recorded EEG signals are pre-processed by

- converting each one at a frequency sampling of 256 samples per second;
- extracting 180 seconds from the total recorded original signal (from 60 seconds to 240 seconds).

2.6.2. Feature Extraction

We apply the FFT functions to the collected data in the following way:

- taking into account the 180-second signal and extracting N Fourier Coefficients (N equal to 16 and 32);
- dividing the signal into 6 epochs of 30 seconds and extracting for each one 16(32) Fourier Coefficients.

Afterwards, the 16(32) Fourier Coefficients are arranged as follows:

- building up a matrix composed of 304(608) columns (i.e., columns refer to the 19 electrodes multiplied by the 16(32) Fourier Coefficients) and 100 rows (i.e., the number of patients);
- building up 19 matrices composed of 16(32) columns (i.e., number of Fourier Coefficient) and 100 rows (i.e., number of patients);
- building up 16(32) matrices composed of 19 columns (i.e., number of electrodes) and of 100 rows (i.e., number of patients).

2.6.3. Classification

Finally, we apply the following classifiers of Weka 3.6.9 environment:

- Support Vector Machines (SMO that implements the SVM);
- Decision Tree (J48 that implements the C4.5 algorithm);

and the DMB rule-based machine learning algorithm, in order to handle the following three classification problems:

- AD *vs* CT;
- MCI *vs* CT;
- AD *vs* MCI.

The choice of using three 2-class classification models instead of a single 3-class one is motivated by two main considerations: first, we want to identify if the 3 sets have specific characteristics that single them out with respect to the rest of the data; second, the nature of some adopted classifiers is intrinsically binary (i.e., DMB and SVM) and therefore they are expected to perform better. This is indeed the case of the experiments presented in the following (results with poorer performances obtained when 3-classes are used are not shown in the tables).

3. Results and Discussion

All the steps of the EEG processing analysis have been applied in order to classify the experimental samples either in AD or MCI patients or controls. In this section, we report and explain the results of the classification analysis by taking into account the FFT with $N = 16$ components and a unique file for all the electrodes (see subsection 2.6.2).

The classification algorithms are tested with a leave-one-out cross validation (i.e., $k = 63, 51,$ and 86 folds for AD *vs* CT, MCI *vs* CT, and AD *vs* MCI, respectively).

Table 3 presents the results of SVM (with polynomial kernel coefficient set to 2), Decision Tree, and DMB classifiers that concern the EEG signals of 180 seconds and $N = 16$ extracted Fourier Coefficients. It is worth noting that the adopted classification metrics have been shown as weighted averages because they are balanced between the analyzed classes. Additionally, we also test the MultiLayer Perceptron classification method [6] whose performances are not satisfying and hence are not reported.

When dealing with the MCI and CT identification, Decision Tree J48 outperforms SVM and DMB in the percentage of correct classification achieving a 90% of accuracy and a 87% of specificity, as well as high performances in all the other metrics, using a leave-one-out cross validation (51-folds). Even when dealing with the identification of AD and CT, J48 achieves better classification results (73% of accuracy) than SVM (68% of accuracy) and DMB (71% of accuracy) in the leave-one-out cross validation (63-folds). Finally, when distinguishing AD from MCI patients, J48 achieves the highest performances (80% of accuracy, 79% of specificity) with respect to SVM and DMB classifiers.

Table 3: Classification performances [%] for EEG signals of 180-seconds with $N = 16$ in a L-1-out scheme ($-63,51,86$ folds for AD *vs* CT, MCI *vs* CT, AD *vs* MCI, respectively)

| Sampling | AD <i>vs</i> CT | | | MCI <i>vs</i> CT | | | AD <i>vs</i> MCI | | |
|-------------|-----------------|-----------|-----|------------------|-----------|-----|------------------|-----------|-----|
| | SVM | J48 | DMB | SVM | J48 | DMB | SVM | J48 | DMB |
| Accuracy | 68 | 73 | 71 | 61 | 90 | 65 | 58 | 80 | 52 |
| Precision | 59 | 72 | 64 | 57 | 90 | 65 | 57 | 80 | 54 |
| Recall | 68 | 73 | 71 | 61 | 90 | 65 | 58 | 80 | 52 |
| Specificity | 19 | 46 | 26 | 32 | 87 | 47 | 54 | 79 | 54 |
| F-measure | 63 | 72 | 67 | 59 | 90 | 65 | 57 | 80 | 52 |

Table 4 presents the results of SVM (with polynomial kernel coefficient set to 2), Decision Tree and DMB classifiers concerning the EEG signals divided in 6 epochs of 30 seconds and $N = 16$ extracted Fourier Coefficients. Also in this case, the MultiLayer Perceptron classification method have very low performances and hence are not reported.

For what concerns the leave-one-out sampling, Decision Tree classifier (J48) outperforms both SVM and DMB in the percentage of correct classification, achieving a 86% of accuracy for the AD identification from CT, a 88% of accuracy for the MCI identification from CT, and a 83% of accuracy for the AD identification with respect to the MCI patients, as well as high performances in the corresponding other metrics. Also here the adopted classification metrics have been shown as weighted averages because they are balanced between the analyzed classes.

Table 4: Classification performances [%] for the 30-seconds EEG epochs with $N = 16$ in a L-1-out scheme (=63,51,86 folds for AD *vs* CT, MCI *vs* CT, AD *vs* MCI, respectively)

| Sampling | AD <i>vs</i> CT | | | MCI <i>vs</i> CT | | | AD <i>vs</i> MCI | | |
|-------------|-----------------|-----------|-----|------------------|-----------|-----|------------------|-----------|-----|
| | SVM | J48 | DMB | SVM | J48 | DMB | SVM | J48 | DMB |
| Accuracy | 54 | 86 | 62 | 51 | 88 | 61 | 49 | 83 | 51 |
| Precision | 62 | 85 | 57 | 59 | 88 | 57 | 49 | 83 | 50 |
| Recall | 54 | 86 | 62 | 51 | 88 | 61 | 49 | 83 | 51 |
| Specificity | 36 | 60 | 18 | 46 | 82 | 32 | 47 | 81 | 48 |
| F-measure | 57 | 85 | 59 | 54 | 88 | 59 | 49 | 82 | 50 |

Moreover, we tested other rule-based classification methods [27], whose performances are comparable to the ones of DMB, and lower than J48. Indeed, rule-based classification methods and SVM, appear to be more prone to over-fitting and noise, while J48 is able to handle these issues thanks to the control parameter, which sets the minimum instances per leaf. In this way, the size of the tree is bound, and over-fitting is avoided. Additionally, we remark that DMB and J48 provide the investigator with a classification model containing the involved EEG electrodes, which are taken into account for performing the patient identification.

An example of the classification tree model is drawn in Figure 5. To conclude, the partition of EEG



Figure 5: J48 tree for AD *vs* CT

signals into 6 epochs of 30 seconds effectively identifies Alzheimer’s disease experimental samples with better accuracy than the consideration of the whole 180-second signal. This fact could be explained by the higher number of features that are available to the classification algorithm.

Furthermore, since the EEG signals may hold back some hidden artifacts or background noise, we expect to obtain higher classification performances by investigating the epoch divisions with a more effective EEG signal pre-processing and by applying other consolidated time-frequency analysis such as the Wavelet Transform [36]. Moreover, our results confirm and reinforce those presented in related works that concern the frequency analysis and classification of EEG signals of AD and MCI patients (e.g., [26], [32], [23]).

4. Conclusions

In this work, patients with Alzheimer’s disease and Mild Cognitive Impairment, as well as healthy control samples, were investigated. Their EEG signals were analyzed and processed by applying the time-frequency analysis through the consolidated Fourier Transform. Well-known supervised learning methods (i.e., SVM, Decision Trees, and Rule-based classifiers) allowed an accurate classification of the human samples, obtaining promising results. In particular, Decision Tree methods stood out among the other ones in terms of classification performances.

We plan to extend the analysis of the EEG signals by applying the Wavelet Transform in order to obtain a more efficient time-frequency decomposition of EEG and an optimal timefrequency resolution of the time-evolution of frequency patterns. Furthermore, we plan to select off-line artifact-free epochs by means of some suitable and *ad-hoc* tools (e.g., EEGLab [14] for MATLAB[®]) in order to perform a more effective EEG pre-processing.

Acknowledgement

The authors were partially supported by the FLAGSHIP “InterOmics” project (PB.P05) and “EPIGEN” project funded by the Italian MIUR and CNR institutions and by the GenData 2020 project.

References

- [1] A. Akrami, S. Solhjoo, A. Motie-Nasrabadi, and M.-R. Hashemi-Golpayegani, "Eeg-based mental task classification: linear and nonlinear classification of movement imagery," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pp. 4626–4629, IEEE, 2006.
- [2] A. Arenas, R. Brenner, and C. F. Reynolds, "Temporal slowing in the elderly revisited," *Am J EEG Technol*, vol. 26, pp. 105–114, 1986.
- [3] P. Bertolazzi, G. Felici, and E. Weitschek, "Learning to classify species with barcodes," *BMC bioinformatics*, vol. 10, no. Suppl 14, p. S7, 2009.
- [4] C. Besthorn, H. Förstl, C. Geiger-Kabisch, H. Sattel, T. Gasser, and U. Schreiter-Gasser, "Eeg coherence in alzheimer disease," *Electroencephalography and clinical neurophysiology*, vol. 90, no. 3, pp. 242–245, 1994.
- [5] T. Bird, "Alzheimer's disease and other primary dementias," *Harrisons principles of internal medicine*, vol. 2, pp. 2391–2398, 2001.
- [6] C. M. Bishop, *Neural networks for pattern recognition*. Oxford university press, 1995.
- [7] H. Braak and E. Braak, "Neuropathological staging of alzheimer-related changes," *Acta neuropathologica*, vol. 82, no. 4, pp. 239–259, 1991.
- [8] D. Cibils, "Dementia and qeeg (alzheimer's disease)," *Supplements to Clinical Neurophysiology*, vol. 54, pp. 289–294, 2002.
- [9] L. A. Coben, W. L. Danziger, and L. Berg, "Frequency analysis of the resting awake eeg in mild senile dementia of alzheimer type," *Electroencephalography and clinical neurophysiology*, vol. 55, no. 4, pp. 372–380, 1983.
- [10] W. W. Cohen, "Fast effective rule induction," in *In Proceedings of the Twelfth International Conference on Machine Learning*, pp. 115–123, Morgan Kaufmann, 1995.
- [11] N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- [12] J. Dauwels, K. Srinivasan, M. Ramasubba Reddy, T. Musha, F.-B. Vialatte, C. Latchoumane, J. Jeong, and A. Cichocki, "Slowing and loss of complexity in alzheimer's eeg: two sides of the same coin?," *International journal of Alzheimer's disease*, vol. 2011, 2011.
- [13] J. Dauwels, F. Vialatte, and A. Cichocki, "Diagnosis of alzheimers disease from eeg signals: Where are we standing?," *Current Alzheimer Research*, vol. 7, no. 6, pp. 487–505, 2010.
- [14] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.
- [15] S. Dulli, S. Furini, and P. E., *Data Mining*. Springer, 2009.
- [16] T. Elbert, W. Lutzenberger, B. Rockstroh, P. Berg, and R. Cohen, "Physical aspects of the eeg in schizophrenics," *Biological psychiatry*, vol. 32, no. 7, pp. 595–606, 1992.
- [17] T. H. Falk, F. J. Fraga, L. Trambaiolli, and R. Anghinah, "Eeg amplitude modulation analysis for semi-automated diagnosis of alzheimers disease," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–9, 2012.
- [18] G. Felici and K. Truemper, "A minsat approach for learning in logic domains," *INFORMS Journal on Computing*, vol. 13, no. 3, pp. 1–17, 2002.

- [19] G. Felici and K. Truemper, “A minsat approach for learning in logic domains,” *INFORMS Journal on computing*, vol. 14, no. 1, pp. 20–36, 2002.
- [20] E. Frank and I. H. Witten, “Generating accurate rule sets without global optimization,” in *In proc. of the 15th int. conference on machine learning*, Morgan Kaufmann, 1998.
- [21] B. R. Gaines and P. Compton, “Induction of ripple-down rules applied to modeling large databases,” *Journal of Intelligent Information Systems*, vol. 5, no. 3, pp. 211–228, 1995.
- [22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The weka data mining software: an update,” *SIGKDD Explor. Newsl.*, vol. 11, pp. 10–18, Nov. 2009.
- [23] C. Huang, L.-O. Wahlund, T. Dierks, P. Julin, B. Winblad, and V. Jelic, “Discrimination of alzheimer’s disease and mild cognitive impairment by equivalent eeg sources: a cross-sectional and longitudinal study,” *Clinical Neurophysiology*, vol. 111, no. 11, pp. 1961–1967, 2000.
- [24] H. H. Jasper, “The ten twenty electrode system of the international federation,” *Electroencephalography and clinical neurophysiology*, vol. 10, pp. 371–375, 1958.
- [25] J. W. Kowalski, M. Gawel, A. Pfeffer, and M. Barcikowska, “The diagnostic value of eeg in alzheimer disease: correlation with the severity of mental impairment,” *Journal of clinical neurophysiology*, vol. 18, no. 6, pp. 570–575, 2001.
- [26] C. Lehmann, T. Koenig, V. Jelic, L. Prichep, R. E. John, L.-O. Wahlund, Y. Dodge, and T. Dierks, “Application and comparison of classification algorithms for recognition of alzheimer’s disease in electrical brain activity (eeg),” *Journal of neuroscience methods*, vol. 161, no. 2, pp. 342–350, 2007.
- [27] T. Lehr, J. Yuan, D. Zeumer, S. Jayadev, and M. Ritchie, “Rule based classifier for the analysis of gene-gene and gene-environment interactions in genetic association studies,” *BioData Mining*, vol. 4, no. 1, p. 4, 2011.
- [28] T. Locatelli, M. Corsi, D. Liberati, M. Franceschi, and G. Comi, “Eeg coherence in alzheimer’s disease,” *Electroencephalography and clinical neurophysiology*, vol. 106, no. 3, pp. 229–237, 1998.
- [29] MATLAB, *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc., 2010.
- [30] R. C. Petersen, “Early diagnosis of alzheimers disease: is mci too late?,” *Current Alzheimer Research*, vol. 6, no. 4, p. 324, 2009.
- [31] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, “Mild cognitive impairment: clinical characterization and outcome,” *Archives of neurology*, vol. 56, no. 3, pp. 303–308, 1999.
- [32] K. Polat and S. Güneş, “Classification of epileptiform eeg using a hybrid system based on decision tree classifier and fast fourier transform,” *Applied Mathematics and Computation*, vol. 187, no. 2, pp. 1017–1026, 2007.
- [33] G. Powell and I. Percival, “A spectral entropy method for distinguishing regular and irregular motion of hamiltonian systems,” *Journal of Physics A: Mathematical and General*, vol. 12, no. 11, p. 2053, 1979.
- [34] J. R. Quinlan, “Improved use of continuous attributes in c4. 5,” *arXiv preprint cs/9603103*, 1996.
- [35] O. A. Rosso, A. Mendes, R. Berretta, J. A. Rostas, M. Hunter, and P. Moscato, “Distinguishing childhood absence epilepsy patients from controls by the analysis of their background brain electrical activity (ii): a combinatorial optimization approach for electrode selection,” *Journal of neuroscience methods*, vol. 181, no. 2, pp. 257–267, 2009.
- [36] A. Subasi, “Eeg signal classification using wavelet feature extraction and a mixture of expert model,” *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.

- [37] E. Weitschek, A. L. Presti, G. Drovandi, G. Felici, M. Ciccozzi, M. Ciotti, and P. Bertolazzi, “Human polyomaviruses identification by logic mining techniques,” *Virology journal*, vol. 9, no. 1, pp. 1–6, 2012.
- [38] J. Xie, C. Brayne, and F. E. Matthews, “Survival times in people with dementia: analysis from population based cohort study with 14 year follow-up,” *bmj*, vol. 336, no. 7638, pp. 258–262, 2008.