A hybrid feature selection approach for the early diagnosis of Alzheimer's disease

Esteve Gallego-Jutglà[1], Jordi Solé-Casals[1]*, François-Benoît Vialatte[2], Mohamed Elgendi[3,4], Andrzej Cichocki[5], Justin Dauwels[4]

[1] Data and Signal Processing Group, University of Vic – Central University of Catalonia, 08500 Vic, Catalonia, Spain

Telf: +34 938815519 Fax: +34 938814307

esteve.gallego@uvic.cat jordi.sole@uvic.cat

[2] ESPCI ParisTech, Laboratoire SIGMA, 75231 Paris Cedex 05, France

Telf: +33 (0)140794541 Fax: +33 (0)140794553

francois.vialatte@espci.fr

[3] University of Alberta, Edmonton, Alberta, Canada T6G 2E8

Telf: +1 (780)8074141, Fax: +1 (780)4926393

moe.elgendi@gmail.com

[4] Nanyang Technological University, Singapore 639798

Telf: +65 67905410

jdauwels@ntu.edu.sg

[5] BSI RIKEN ABSP Lab, Wako Saitama, 351-0198 Japan

Telf: +81 484679668 Fax: +81 484679686

cia@brain.riken.jp

* Corresponding author

Abstract

Objective. Recently, significant advances have been made in the early diagnosis of Alzheimer's disease from EEG. However, choosing suitable measures is a challenging task. Among other measures, frequency Relative Power and loss of complexity have been used with promising results. In the present study we investigate the early diagnosis of AD using synchrony measures and frequency Relative Power on EEG signals, examining the changes found in different frequency ranges. *Approach*. We first explore the use of a single feature for computing the classification rate, looking for the best frequency range. Then, we present a multiple feature classification system that outperforms all previous results using a feature selection strategy. These two approaches are tested in two different databases, one containing MCI and healthy subjects (patients age: 71.9 ± 10.2, healthy subjects age: 71.7 ± 8.3), and the other containing Mild AD and healthy subjects (patients age: 77.6 ± 10.0; healthy subjects age: 69.4± 11.5). *Main Results*. Using a single feature to compute classification rates we achieve a performance of 78.33% for the MCI data set and of 97.56 % for Mild AD. Results are clearly improved using the multiple feature classification, where a classification rate of 95% is found for the MCI data set using 11 features, and 100% for the Mild AD data set using 4 features. *Significance*. The new features selection method described in this work may be a reliable tool that could help to design a realistic system that does not require prior knowledge of a patient's status. With that aim, we explore the standardization of features for MCI and Mild AD data sets with promising results.

Keywords: Alzheimer's disease; Mild cognitive impairment; Electroencephalography; Synchrony; Relative power; Granger causality; Gram-Schmidt orthogonal forward regression

1. Introduction

Alzheimer's disease (AD) is a neurodegenerative disease and the most common form of dementia. AD is a progressive and irreversible deterioration of brain functions that starts with loss of memory and leads to other cognitive impairments, such as language and judgment deficits. Currently no cure exists for Alzheimer's, but administering certain medications in the early stages may delay the onset of symptoms [1,2]. Therefore, developing methods for detecting the pathology in its earliest stages is a critical task.

The progression of AD is classified into four stages. The first, or "preclinical," stage is Mild Cognitive Impairment (MCI). MCI patients usually present some memory impairment, but retain their abilities in other cognitive domains and functional activities [3,4]. Some MCI patients (between 6% and 25%) later develop AD. The next steps are characterized by growing cognitive deficits, which cause a reduction of independence. The second and third stages are known as Mild AD and Moderate AD, while the last stage is known as Severe AD, entailing complete dependence on caregivers [1]. MCI and Mild AD are key stages: An early diagnosis of AD in either stage may confer several benefits [2].

Electroencephalography (EEG) has been suggested as a potential diagnostic tool for AD. Compared to other systems like functional Magnetic Resonance Imaging (fMRI) or Positron Emission Tomography (PET), EEG systems are inexpensive and easy to transport. Studies have repeatedly found AD cause three major perturbations in EEG data: slowing of EEG, reduction in the complexity of EEG signals, and changes in EEG synchrony (see [2] and [5] for an extended review). These changes in the EEG data have been used as a discriminative feature to diagnose AD. Early diagnosis is, however, by no means a simple task, as these perturbations in the EEG data tend to vary across subjects, and therefore have insufficient specificity [6].

Recently, a strong relationship between the slowing of EEG and a reduction in the complexity of the signals has been reported. The results presented in [7] demonstrate that frequency Relative Power (RP), a measure used to parameterize the slowing of EEG, and loss of complexity are strongly anti-correlated at low frequencies. As two of the main perturbations in EEG data are closely related, the present study investigates the early diagnosis of AD using the

two other changes in EEG: slowing of EEG and changes in EEG synchrony. Earlier research [8] has shown that a small group of synchrony measures may suffice to quantify EEG synchrony in AD patients, due to the high correlation observed between some synchrony measures. Consequently, in this paper, we select a reduced group of synchrony measures and a power measure to study the difference between healthy subjects and AD patients.

An EEG recording is usually characterized by the presence of activity on specific frequency bands: 0.1-4 Hz (δ), 4-8 Hz (θ), 8-13 (α), 13-30 (β) and 30-100 (γ) [9,10]. To distinguish between AD patients and healthy patients, studies traditionally analyse the standard frequency bands [11,12], or extend the analysis to the entire frequency range between 4 and 30 Hz [8]. Some studies have analysed all the frequency bands between 1 and 30 Hz, for instance, using a power measure [13] or a set of synchrony measures [14]. The present study investigates whether the diagnosis of AD can be improved by analysing all possible frequency ranges in the 1-30Hz frequency range (e.g., 1-2 Hz, 1-3Hz, 1-4Hz…. 29-30Hz) using power and synchrony measures. To the best of our knowledge, no study so far has conducted out such an analysis.

The present study analyses two different data sets, one consisting of MCI patients and another of Mild AD patients. Classification is evaluated in the entire set of frequency ranges. First, each measure is used independently as an input feature, and individual classification results are presented. Then, a multiple feature classification is performed using the measures that best characterize the data set and the optimal frequency range. The optimal measures and the optimal frequency range are selected with the Orthogonal Forward Regression (OFR) algorithm.

The paper is organized as follows: Section 2 introduces the two data sets (MCI and Mild AD data sets), explains the measures used to characterize AD patients, and details the methods used to apply those measures. Section 3 presents the results, which are further discussed in Section 4. Finally, Section 5 concludes the paper.

2. Material and methods

This Section presents the methods applied to two different data sets. Data sets are presented in Section 2.1. Synchrony and power measures are presented in Section 2.2. Selected measures were studied in all possible sets of frequency ranges between 1 and 30 Hz. The computation of

those measures is detailed in Section 2.3. Differences between measures were statistically analysed (defined in Section 2.4) and classified, first independently and then concurrently using the OFR algorithm (defined in Section 2.6). Some of the methods used in this work are detailed in [8]. However, we present a number of novel analytical methods, based on a new frequency approach and the optimal selection of measures, which aim to improve the rate at which MCI/Mild AD patients are distinguished from healthy subjects.

2.1 Data sets

In this study, we consider two data sets. One data set contains EEG recordings of MCI patients and healthy subjects, and the other contains EEG recordings of Mild AD patients and healthy subjects.

2.1.1 The MCI data set: MCI patients and control subjects

The EEG data contained in this follow-up data set have been previously analysed in a number of studies evaluating the early diagnosis of AD [7,8,15–18].

The MCI data set originally consisted of fifty-three patients. Initially patients who only complained of memory impairment were recruited. They underwent thorough neuropsychological testing that revealed a quantified and objective evidence of memory impairment with no apparent loss in either general cognitive, behavioural, or functional status. The patients did not suffer from other neurological diseases. The classification of very mild dementia impairment required a Mini-Mental Status Exam (MMSE) $\geq$24 and a Clinical Dementia Rating (CDR) scale score of 0.5 with memory performance less than one standard deviation below the normal reference (Wechsler Logical Memory Scale and Paired Associates Learning subtests, IV and VII, $\leq$9, and/or $\leq$5 on the 30 min delayed recall of the Rey-Osterreith figure test). Each patient underwent Single-Photon Emission Computed Tomography (SPECT) at initial evaluation and was followed clinically for 12–18 months. Twenty-five of these fifty-three very mild AD patients developed probable or possible AD according to the criteria defined by the National Institute of Neurological and Communicative Disorders and Stroke, and the Alzheimer's disease and Related Disorders Association (NINDS-ADRDA). These subjects formed our group of patients of the MCI data set (age: 71.9 ± 10.2 years old), which MMSE

scores were 28.5 ± 1.6). EEG recordings were conducted at the MCI stage. The control group consisted of healthy subjects who had no memory or other cognitive impairments. This control group was formed for 56 age-matched healthy subjects (age: 71.7 ± 8.3 years old), which MMSE scores were 26 ± 1.8.

For the recording, 21 Ag/AgCl electrodes (discs with a diameter of 8 mm) were placed on the scalp according to the 10-20 international system, with the reference electrode on the right earlobe. The EEG was recorded with Biotop 6R12 (NEC San-ei, Tokyo, Japan) at a sampling rate of 200 Hz, with analog bandpass filtering in the frequency range 0.5-250 Hz and online digital bandpass filtering between 0.5 and 30 Hz, using a third-order Butterworth filter (forward and reverse filtering).

2.1.2 The Mild AD data set: Mild AD patients and healthy subjects

The second EEG data set has also been analysed previously [13,14,16,19,20].

The EEGs were recorded during a resting period containing various states: awake, drowsy, alert and resting with eyes closed and open. All recording sessions and experiments proceeded after informed consent was obtained of the subjects or the caregivers and were approved by local institutional ethics committees. The EEG data is composed of 24 healthy control subjects (age: 69.4± 11.5; 10 males) and 17 patients with mild AD (age: 77.6 ± 10.0; 9 males). The patient group underwent a full battery of cognitive tests (MMSE, Rey Auditory Verbal Learning Test, Benton Visual Retention Test, and memory recall tests). The results from the psychometric tests were scored and interpreted by a psychologist, and all clinical and psychometric findings were discussed at a multidisciplinary team meeting. All controls were healthy volunteers and had normal EEGs (confirmed by a Consultant Clinical Neurophysiologist). The patients did not suffer from other neurological diseases. The EEG time series were recorded using 19 electrodes positioned according to the Maudsley system, similar to the 10-20 international system, at a sampling frequency of 128 Hz. The EEGs were bandpass filtered with a third-order digital Butterworth filter (forward and reverse filtering) between 0.5 and 30 Hz.

2.1.3 Recording conditions common to both data sets

In both data sets, all recording sessions included in the analysis were conducted with the subjects in an awake but resting state with eyes closed, and the length of the EEG recording was about 5 minutes per subject. The EEG technicians prevented the subjects from falling asleep (vigilance control). After recording, the EEG data was carefully inspected. EEG recordings are prone to a various artifacts, for example, due to electronic noise, head movements, and muscular activity. For each patient, an EEG expert selected, by visual inspection, one 20-second segment of artifact-free EEG, blinded from the results of the present study. Only subjects whose EEG recordings contained at least 20 seconds of artifact-free data for all the channels were retained in the analysis. This selection was conducted blind from the outcomes of this study. Based on this requirement, the number of subjects in the MCI data set was further reduced to 22 MCI patients and 38 control subjects; in the Mild AD Data set no such reduction was required.

2.2. EEG measures

We consider two types of EEG measures: frequency power and synchrony. As a spectral measure, we selected Relative Power (RP). For synchrony, various measures were used: Correlation, Coherence, Granger Causality (including Granger coherence, Partial coherence (PC), Directed transfer function (DTF), Full frequency directed transfer function (ffDTF), Partial directed coherence (PDC), Direct directed transfer function (dDTF)), Omega Complexity, and Phase Synchrony. These synchrony measures have been reviewed earlier in [8]. A brief description is presented below.

2.2.1 Relative power

Relative Power (RP) measures the percentage of power within a specific frequency band compared to the power of the entire frequency range. RP is computed as:

$$RP_i\,(f_1, f_2) = \frac{P_i(f_1, f_2)}{P_i(f_{min}, f_{max})} \tag{1}$$

where $P_i(f_1, f_2)$ is the power spectrum in the frequency band of interest $[f_1, f_2]$ for channel i, and $P_i(f_{min}, f_{max})$ is the power spectrum in the entire frequency range $[f_{min}, f_{max}]$ for the same channel. The power spectrum is computed using Fast Fourier Transform [21,22].

2.2.2 Correlation coefficient

The correlation coefficient (r) measures the similarity between two time series; it is a measure of interdependence: $r = 1$ if the two time series are identical, $r = 0$ if the time series are completely different, and $r = -1$ if the two time series are identical and exactly out-of-phase. For two time series *x* and *y*, correlation is computed as [10]:

$$r = \frac{1}{N}\sum_{k=1}^{N}\frac{(x(k)-\tilde{x})}{\sigma_x}\frac{(y(k)-\tilde{y})}{\sigma_y} \quad (2)$$

where *N* is the length of the signals, $\tilde{x}$ and $\tilde{y}$ are the mean averages of the time series x and y, and σ_x and σ_y correspond to the variance of the time series x and y.

2.2.3 Coherence

Coherence estimates phase synchronization between two bivariate time series (*x* and *y*) in the frequency domain. Coherence is usually interpreted as an indicator of connectivity between two brain areas [23], and it is computed by dividing the time series in *M* segments of length *L*. Coherence $\left(c^2(f)\right)$ is defined as [10,24]:

$$c^2(f) = \frac{|\langle X(f)Y^*(f)\rangle|^2}{|\langle X(f)\rangle||\langle Y(f)\rangle|} \quad (3)$$

where $X(f)$ and $Y(f)$ denote the discrete Fourier transforms of the two signals *x* and *y*, Y^* is the complex conjugate of $Y \in \mathbb{C}$, $|Y(f)|$ is the magnitude of Y, and $\langle X(f)\rangle$ stands for the average of $X(f)$ computed over *M* segments, likewise $\langle Y(f)\rangle$ and $\langle X(f)Y^*(f)\rangle$.

2.2.4 Granger causality

Granger causality[1] refers to a family of synchrony measures derived from linear stochastic models of time series; they quantify the interdependence between multivariate signals.

Granger causality measures are derived from the multivariate autoregressive (MVAR) model of the multivariate time series. Let the set of *n* signals $x_1(k), x_2(k), \ldots, x_n(k)$, be the normalized

[1] The code for the Granger causality measures used in this study is implemented in the BioSig library: http://biosig.sourceforge.net

EEG time series of a recording (mean equal to zero and standard deviation equal to one). The MVAR of those times series is defined as:

$$x(n) = \sum_{l=1}^{p} \boldsymbol{A}(j)x(k-l) + e(k) \tag{4}$$

where $\boldsymbol{x}(k) = \left(x_1(k), x_2(k), \ldots, x_n(k)\right)^T$, p is the model order, the model coefficients $\boldsymbol{A}$(j) are $n \times n$ matrices, and $e(k)$ is a zero-mean Gaussian random vector of size *n*. Equation (4) can be rewritten as:

$$e(k) = \sum_{l=0}^{p} \widetilde{\boldsymbol{A}}(j)x(k-l) \tag{5}$$

where $\widetilde{\mathbf{A}}(0) = \mathrm{I}$ (identity matrix) and $\widetilde{\mathbf{A}}(\mathrm{j}) = -\mathbf{A}(\mathrm{j})$ for $j > 0$. Applying the z-transform and substituting $\mathrm{z} = \mathrm{e}^{-2\pi \mathrm{i} \Delta \mathrm{t}}$, with sampling rate $= 1/\Delta\,\mathrm{t}$, equation (5) can be written as:

$$\boldsymbol{E}(f) = \widetilde{\boldsymbol{A}}(f)\boldsymbol{X}(f) \tag{6}$$

$$\boldsymbol{X}(f) = \widetilde{\boldsymbol{A}}^{-1}(f)\boldsymbol{E}(f) = \boldsymbol{H}(f)\boldsymbol{E}(f) \tag{7}$$

If the variance of the noise $E(f)$ is represented by V, then the power spectrum matrix of the signal x(k) is defined as:

$$\boldsymbol{S}(f) = \boldsymbol{H}(f)\boldsymbol{V}\boldsymbol{H}^{\dagger}(f) \tag{8}$$

Granger measures are defined in terms of ***A***, ***H*** and ***S*** matrices.

Granger coherence

Granger Coherence describes the amount of in-phase components in signals *i* and *j* at the frequency *f*. With the matrix ***S***, Granger coherence $\left|K_{ij}(f)\right| \in [0,1]$ can be computed as [25]:

$$K_{ij}(f) = \frac{S_{ij}(f)}{\sqrt{S_{ii}(f)}\sqrt{S_{ij}(f)}} \tag{9}$$

Partial coherence

Partial coherence (PC) describes the amount of in-phase components in signals *i* and *j* at the frequency *f* when the influence of the other signals is statistically removed. PC can be written as [25]:

$$C_{ij}(f) = \frac{M_{ij}(f)}{\sqrt{M_{ii}(f)}\sqrt{M_{ij}(f)}} \tag{10}$$

where $M_{ij}(f)$ is the determinant of $\boldsymbol{S}$ with row *i* and column *j* removed. $\left|C_{ij}(f)\right| \in [0,1]$.

Directed transfer function

The Directed Transfer Function (DTF) quantifies the fraction of inflow to channel *i* stemming from channel *j*. It is computed in terms of the $\boldsymbol{H}$ transfer matrix [25]:

$$\gamma_{ij}^2(f) = \frac{\left|H_{ij}(f)\right|^2}{\sum_{j=1}^{m}\left|H_{ij}(f)\right|^2} \tag{11}$$

Full frequency directed transfer function

The Full Frequency Directed Transfer Function (ffDTF) is defined as [25]:

$$F_{ij}^2(f) = \frac{\left|H_{ij}(f)\right|^2}{\sum_f\sum_{j=1}^{m}\left|H_{ij}(f)\right|^2} \tag{12}$$

ffDTF can be seen as a variation of DTF, with a global normalization in frequency.

Partial directed coherence

Partial Directed Coherence (PDC) is an extension of PC. It represents the fraction of outflow from channel *j* to channel *i*. PDC $\left|P_{ij}(f)\right| \in [0,1]$ is described as [25]:

$$P_{ij}(f) = \frac{\tilde{A}_{ij}(f)}{\sqrt{\sum_{i=1}^{m}\left|\tilde{A}_{ij}(f)\right|^2}} \tag{13}$$

Direct directed transfer function

The Direct Directed Transfer Function (dDTF) is defined as [26]:

$$\chi_{ij}^2(f) = F_{ij}^2(f)C_{ij}^2(f) \tag{14}$$

dDTF is non-zero if the connection between channels *i* and *j* is causal. DTF, dDTF, PDC, and ffDTF are asymmetric measures, i.e.: $\gamma_{ij}^2 \neq \gamma_{ji}^2$, $F_{ij}^2 \neq F_{ji}^2$, $P_{ij} \neq P_{ji}$ and $\chi_{ij}^2 \neq \chi_{ji}^2$.

2.2.5 Omega complexity

Omega Complexity (Ω) is a synchrony measure for multichannel data sets. It quantifies the amount of spatial synchronization in a multivariate time series. Synchrony is evaluated with the Principal Component Analysis (PCA) of the obtained covariance of the data [27–29].

Given a data set of *n* signals, $x_1(k), x_2(k), \ldots, x_n(k)$; covariance matrix $\boldsymbol{R}_x \in \mathbb{R}^{n\times n}$ is computed. PCA is then used to obtain the eigenvectors and eigenvalues (λ_i) in descending order. Omega Complexity is defined in terms of those normalized eigenvalues [30]:

$$\Omega = \exp\left(-\sum_{i=1}^{n} \lambda_i \log \lambda_i\right) \tag{15}$$

The argument of the exponential in (15) is the entropy of the distribution obtained with the eigenvalues. Omega Complexity presents minimum value ($\Omega = 1$) for identical signals. The maximum value ($\Omega = \mathrm{n}$) is obtained for independent signals.

2.2.6 Phase synchrony

The Phase Synchrony index (γ) computes the synchronization between two time series $x(t)$ and $y(t)$. Phase Synchrony depends only on the phase between signals, even when the amplitudes of *x* and *y* are statistically independent. First, the instantaneous phase ϕ_x of a signal *x* is computed as [31]:

$$\phi_x(k) = \arg[x(k) + i\tilde{x}(k)] \tag{16}$$

where $\tilde{x}$ is the Hilbert transform of x. Then, the phase synchrony index for the two instantaneous phases of signals x and y (ϕ_x and ϕ_y) is defined as [31]:

$$\gamma = \left|\langle e^{i(n\phi_x - m\phi_y)}\rangle\right| \in [0,1] \quad (17)$$

where n and m are integers (usually $n = m = 1$), and $\langle\cdot\rangle$ is time averaging.

2.3 Bandpass filtering and computation of EEG measures

All the possible frequency ranges between 1 and 30 Hz were analysed with the set of measures presented in Section 2.2. To define the frequency ranges of study, the start frequency (F) was varied from 1 to 29 Hz, and the width (W) varied from 1 to 29 (e.g. 1-2 Hz, 1-3 Hz, 1-4 Hz… 1-30 Hz…, 29-30 Hz). The maximum frequency of analysis ($F + W$) was limited to 30 Hz. A total of 435 frequency ranges were studied, as detailed in figure 1:

$$W \in \mathbb{N}[1,29] \begin{bmatrix} 1-2 & & & & & \\ 1-3 & 2-3 & & & & \\ 1-4 & 2-4 & 3-4 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ 1-29 & 2-29 & 3-29 & & 28-29 & \\ 1-30 & 2-30 & 3-30 & \cdots & 28-30 & 29-30 \end{bmatrix}$$
$$F \in \mathbb{N}[1,29]$$

Figure 1. Range of frequency ranges considered in this study.

Before the measures were computed, the signals were bandpass filtered with Butterworth filters. These types of filters are characterized by a magnitude response that is maximally flat in the passband, and they offer good transition band characteristics at low coefficient orders, so they can be easily implemented [32].

In this study, we used third-order Butterworth filters as in [8], since such filters can handle narrow bands with a bandwidth of 1 Hz, such as 1 – 2 Hz, 2 – 3 Hz, …, 29 – 30 Hz. For frequencies ranges with $W = 1$, the frequencies F and $F + W$ have an attenuation of 3 dB, and the adjacent frequencies (i.e. $F - 1$ and $F + W + 1$) have an attenuation of at least 25 dB. Figure 2 illustrates the frequency response for a band-pass filter between 5 – 6 Hz. It can be seen that the adjacent frequencies 4 and 7 Hz have an attenuation of $-37\ dB$ and $-30\ dB$ respectively.

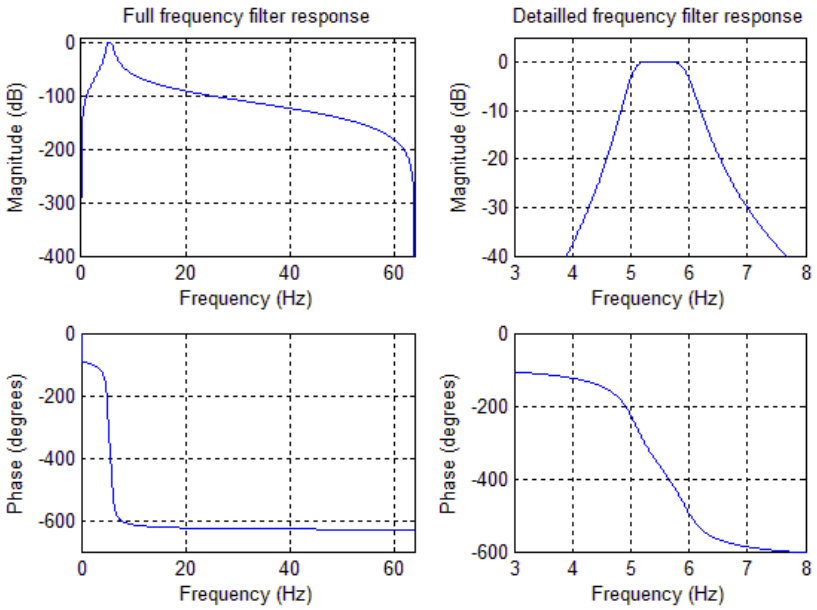

Figure 2. Frequency response of a third order Butterworth band-pass filter, filtering between 5 and 6 Hz with a sampling rate of 128 Hz. Left images show the frequency and phase response on the entire frequency range. Right images show in detail the frequency and phase response in the band pass.

Each measure was applied to the filtered signals. RP was computed for each channel independently. To obtain a global measure for each subject, the RP for all the channels was averaged. Since some of the synchrony measures were bivariate and others multivariate, we applied different approaches in evaluating them (for a review of those approaches see [8]).

For bivariate synchrony measures, we used the Local Approach 2 introduced in [8]. In this approach, the EEG signals are aggregated into five regions (frontal, left temporal, central, right temporal and occipital). Figure 3 presents the electrodes aggregated to each region. To compute the synchrony between two regions, one first computes the synchrony between each EEG signal from one region and each signal from the other. The next step is evaluating synchrony by computing the average synchrony values of these signal pairs. For example, synchrony between the left temporal and the occipital regions is evaluated by averaging the synchrony measures obtained from the 12 pairs of signals (F7,P3), (F7,P4), (F7,O1), (F7,O2), …., (T5,O2). Once the synchrony between each region is computed, the average of synchrony between regions (10 pairs) is calculated to obtain a global synchrony value for each subject. This approach was used for all the bivariate synchrony measures (Correlation, Coherence and Phase Synchrony).

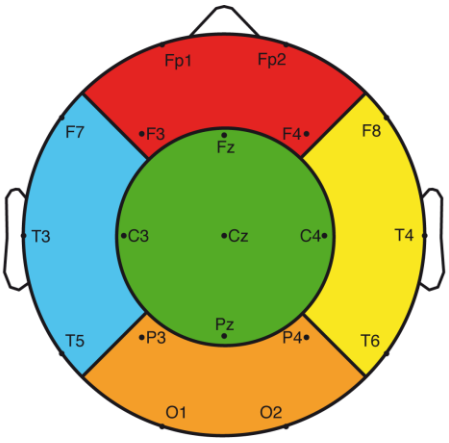

Figure 3. Layout and partition of the 21 electrodes used for the EEG recording. Regions defined by aggregation of electrodes: frontal area, left temporal area, parietal area, right temporal area and occipital area.

A different approach was used to compute the EEG synchrony for multivariate measures. Omega complexity was applied to all EEG signals of the data set. However, Granger measures require estimating a 21-dimensional MVAR model. To avoid this high-dimensional estimation, we calculated the time averaging between electrodes of the same region, obtaining averaged EEG time series for each region defined in figure 3. The Granger measures were then applied to these 5 averaged EEG signals. The Granger values between the regions were averaged (10 pairs) to obtain a global synchrony measure [8].

2.4 Statistical analysis

To evaluate the difference between populations, we calculated the statistical significance of the differences between MCI patients and control subjects as well as between Mild AD patients and control subjects using the Mann-Whitney test – a non-parametric test allowing us to investigate the statistical differences between two populations without assumptions of Gaussianity. Low p-values (close to zero, e.g., $p < 0.05$) indicate a large difference between the medians of the two populations.

EEG is highly non-stationary (see [8,10] for an extended review), and thus data characteristics may change over time; usually, therefore, time segmentation is used to compute synchrony measures. Exploring different parameters like window length (in all the synchrony measures) or

polynomial order (only in Granger measures) is important in selecting the parameters that are most effective in classifying the subjects as AD or healthy.

We used several time window length values: $L = 1\ s, 5\ s,$ and $20\ s$ for both data sets. We also evaluated several polynomial orders in Granger measures: p=1, 2, 3…. 9. Then, the Mann-Whitney test was computed along all the possible configurations in all the possible frequency ranges. The parameter configuration that presented the lowest p-value in any of the defined frequency ranges was defined as optimal and used in further analysis for that specific data set.

2.5 Separability criterion

We used a separability criterion $\left(J(F, F+W)\right)$ to represent the difference between the analysed classes in all the frequency ranges studied. $J(F, F+W)$ is a measure of distance between two normal distributions inspired by the z-score [9]. The distance $J(F, F+W)$ has large values when the mean difference between two populations is large, and the standard deviations of both distributions are small; the two populations can then be easily distinguished. On the other hand, if there is little difference between two populations, $J(F, F+W)$ presents a value close to 0.

We define the separability criterion as:

$$J(F, F+W) = \frac{|\mu_{Ctr}(F, F+W) - \mu_{Pat}(F, F+W)|}{\left(\sigma_{Ctr}(F, F+W) + \sigma_{Pat}(F, F+W)\right)} \quad (18)$$

where $(F, F+W)$ refers to the frequency range of study. F and $(F+W)$ refer to the start and end frequency of the study, respectively, as discussed in Section 2.3; μ_{Ctr} is the mean of the control population; and μ_{Pat} is the mean of the patient population (MCI or Mild AD depending on the data set). Similarly, σ_{Ctr} and σ_{Pat} refer to the standard deviations of the control and patient (MCI or Mild AD) groups. We computed this separability criterion for each proposed measure.

2.6 Classification

We investigated whether changes in EEG synchrony or RP allow us to distinguish between AD patients and healthy subjects. Using the proposed synchrony measures and the RP, we computed two different classifications. All measures were used as input features in a classifier, both

individually and in combinations determined to be optimal using the method described in the next section.

For both types of classification (individual and multiple features), Linear Discriminant Analysis (LDA) was used with leave-one-out (LOO) cross-validation methodology. The Classification Rate (CR), i.e. the percentage of subjects correctly classified, is obtained as a result. In addition, true positive rate, i.e. sensitivity (SE), and true negative rate, i.e. specificity (SP), were also computed.

LDA is a well-known scheme for feature extraction and dimension reduction [33]. Because LDA makes the assumption of Gaussian distribution for the input data, we confirmed the Gaussianity of the computed values by means of histograms and quantile-quantile plots.

2.6.1 Multiple feature classification

We performed multiple feature classification to determine which measures would be the most relevant for distinguishing MCI/Mild AD patients from healthy subjects. To control overfitting and to rank the input features by their significance, a multiple feature selection was used. This procedure is based on Gram-Schmidt Orthogonal Forward Regression (OFR).

The traditional OFR algorithm is the one presented in [34]. In that algorithm, initially the input features (z_i) are defined, after which the algorithm selects the feature that best correlates with the desired output, and projects the remaining features in the null space of the selected one. This procedure is repeated for all input features. The algorithm sorts the input features according to their correlation with the output. The traditional OFR algorithm is summarized as follows [34,35] :

i. Select the candidate feature (x_i) that is best correlated with the output (o) to be modeled: $x_i = \arg max_i \cos^2(z_i, o)$
ii. Project the output vector and all other candidate features on to the null space of the selected feature x_i, using Gram-Schmidt orthogonalization.
iii. Remove the selected feature (x_i) from the list of input features.
iv. Return to (i) until all features have been selected.

This OFR algorithm sorts the input features based on their relevance, but in order to control overfitting, does not select the optimal number of features. We applied the random probe method [36], which refers to random generations of data used to verify that the analysed data is more significant than random data. To compute the OFR with a random probe, one first creates a set of random probes. Then, one defines a risk level that corresponds to the risk that a feature might be kept despite being less relevant than the probe. The OFR algorithm is applied using all the different probes, one probe at a time. Finally, the cumulative distribution function of the position that the probe achieved in the OFR algorithm is computed to rank the probe. Selected features for each risk level are those that are ranked in a lower position than the probe.

In this study, a variation of OFR with a probe was used. The main difference between standard OFR with a probe and our variant lies in how the feature selection is performed. Whereas the standard variant sorts the given inputs and selects the optimal number of features to use, our method preselects the input features that are given to the OFR algorithm in each iteration. In our OFR implementation, there are multiple frequency ranges for each measure: the frequency range with the greatest difference between the two populations, as indicated by the Separability Criterion (*J*) value, is used. This process was repeated for all the measures. Feature normalization was not applied to the original selected features but to the values for all frequency ranges. Another difference is that random probes were not generated using random data. Instead, surrogate probes were generated with the same characteristics as the original data, with a different measure (or a synchrony measure or RP) used to generate each probe. The values of a specific measure for the two populations (AD patients and control subjects) were mixed together, and then labels for each class were assigned randomly. This process was repeated 500 times for each measure. The OFR algorithm used in this work can be summarized as follows:

i. Select the input features (z_i). For each z_i, select the frequency range that corresponds to the largest J. Repeat this procedure for all *i* measures.

ii. Select the candidate feature (x_i) that best correlates to the output (o) to be modelled: $x_i = \arg max_i \cos^2(z_i, o)$.

iii. Project the output vector on to the null space of the selected feature. Orthogonalize the rest of features using Gram-Schmidt orthogonalization in all the existing frequency ranges.

iv. Remove the selected feature (x_i) from the list of input measures.

v. Return to (i) until all features have been selected.

In our implementation, 5,500 probes were computed and added to the feature set to quantify the degree of overfitting.

3. Results

Table 1. Optimal time window lengths and orders used to compute the synchrony measures for the two data sets.

Measures	MCI data set		Mild AD data set	
	Window Length (s)	**Granger Order**	**Window Length (s)**	**Granger Order**
Correlation	20	-	20	-
Coherence	1	-	5	-
Granger Coherence	1	9	1	2
PC	1	4	20	4
DTF	20	8	1	7
ffDTF	20	8	1	2
PDC	1	3	1	2
dDTF	1	9	1	2
Omega Complexity	5	-	1	-
Phase Synchrony	20	-	20	-

Changes in the power and synchrony of EEG data of patients with MCI and Mild AD were evaluated. Each measure presented in Section 2.2 was computed in both data sets, after the computed values were used as input features for a classifier, first individually (results presented in Section 3.1), and next for a set of multiple features selected using the OFR algorithm (Section 3.2). The optimal configuration (window length and Granger Order) for each synchrony measure was used.

Table 1 presents the selected configuration for the synchrony measures, with the optimal time length of the window and the Granger order used to compute them.

3.1 Individual feature classification

Each measure was used independently as a classification feature. The results shown in table 2 show the CR for all measures for the MCI data set, while the results presented in table 3 show the results for the Mild AD data set. The results are presented only for the best set of parameters. The optimal frequency range for each measure is defined as the range in which the best CR is found; if a measure has several frequency ranges with the same CR, the one selected as the optimal frequency range is the one with the highest *J* for that measure. Figure 4 displays the CR obtained in each standard frequency band (δ, θ, α and β), in comparison with results obtained in the optimal frequency range for both data sets. Obtained CR in the optimal frequency range of each measure are always equal to or higher than the values in the standard frequency ranges.

The best CR was obtained in both data sets with RP, with a value of 78.33% for the MCI data set and 97.56% for the Mild AD data set. The results are obtained in a set of frequencies close to the θ frequency range, 2-9 Hz for the MCI data set and 4-7 Hz for Mild AD data set. The best result with a synchrony measure was 75.00% for dDTF in the frequency range of 14-16 Hz, for the MCI data set, and 95.12% for DTF measure in the frequency range of 5-6 Hz, for the Mild AD data set. Interestingly, for the Mild AD data set, most of the best classification results are in low frequency ranges, δ and θ. Our results for the MCI data set present only four measures (Correlation, Granger Coherence, PDC and Phase Synchrony) in which the best CR results for synchrony measures were obtained at low frequencies.

Table 2. CR for the MCI data set. CR is presented with the optimal frequency ranges and its corresponding p-value. The three best results for this data set are shown in bold.

Measures	Optimal frequency range (Hz)	CR optimal frequency range (%)	SE (%)	SP (%)	p-value
RP	**2 - 9**	**78.33**	**72.73**	**81.58**	**0.0001**
Correlation	**1 - 8**	**71.67**	**77.27**	**68.42**	**0.0012**
Coherence	8 - 13	68.33	63.64	71.05	0.0132
Granger Coherence	2 - 8	70.00	72.73	68.42	0.0021
PC	21 - 27	70.00	81.82	63.16	0.0157
DTF	6 - 27	65.00	63.64	65.79	0.2727
ffDTF	8 - 30	70.00	63.64	73.68	0.0013
PDC	1 - 2	66.67	68.18	65.79	0.0085
dDTF	**14 - 16**	**75.00**	**81.82**	**71.05**	**8.88×10^{-5}**
Omega Complexity	8 - 10	68.33	59.09	73.68	0.0138
Phase Synchrony	4 - 5	70.00	90.91	57.89	0.0201

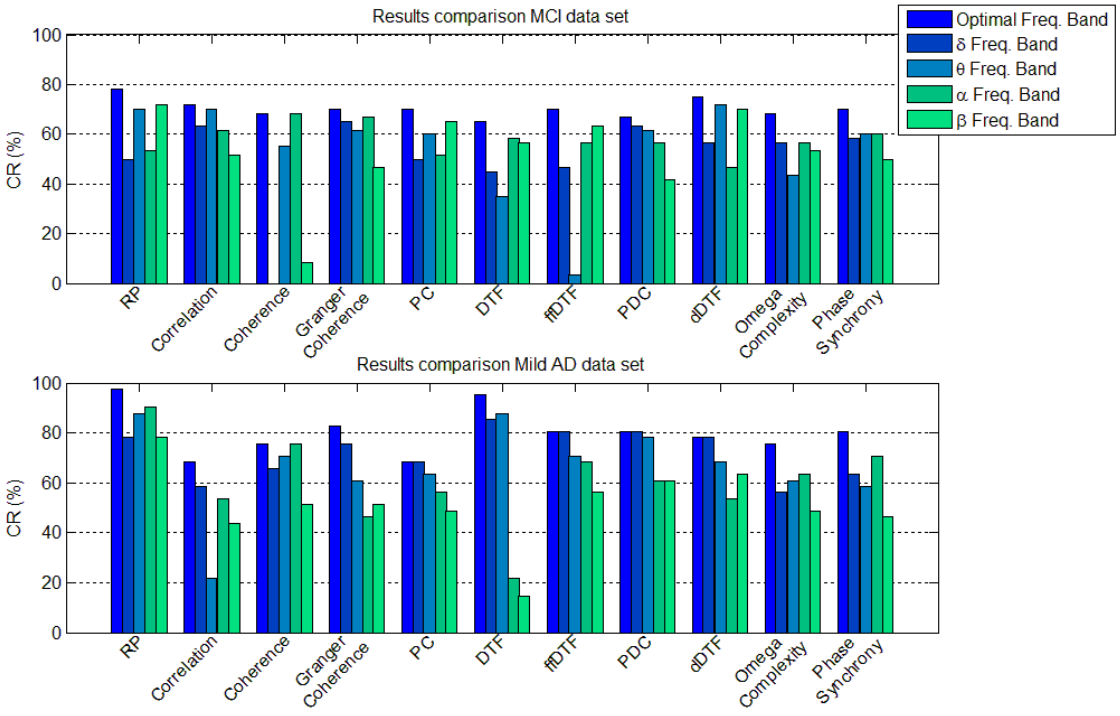

Figure 4. CR comparison between the optimal frequency range, δ, θ, α and β bands. Top image presents obtained results for the MCI data set. Bottom image presents obtained results for the Mild AD data set.

Table 3. CR obtained for the Mild AD data set. CR is presented with the optimal frequency ranges and its corresponding p-value. The three best results for this data set are shown in bold.

Measures	Optimal frequency range (Hz)	CR optimal frequency range (%)	SE (%)	SP (%)	p-value
RP	**4 - 7**	**97.56**	**94.12**	**100.00**	**8.38×10^{-8}**
Correlation	23 - 24	68.29	58.82	75.00	0.0699
Coherence	8 - 13	75.61	82.35	70.83	0.0003
Granger Coherence	**1 - 2**	**82.93**	**76.47**	**87.50**	**3.07×10^{-5}**
PC	3 - 4	68.29	70.59	66.67	0.0027
DTF	**5 - 6**	**95.12**	**100.00**	**91.67**	**2.64×10^{-6}**
ffDTF	1 - 2	80.49	88.24	75.00	3.00×10^{-6}
PDC	1 - 4	80.49	88.24	75.00	1.78×10^{-6}
dDTF	2 - 4	78.05	76.47	79.17	8.50×10^{-5}
Omega Complexity	7 - 8	75.61	82.35	70.83	0.0005
Phase Synchrony	9 - 10	80.49	88.24	75.00	5.45×10^{-5}

To evaluate the redundant information presented by the features, Pearson's linear correlation coefficient was computed between measures in their optimal frequency range. This methodology has been used in other publications [7,8] aiming to evaluate the same principle as we do. Nevertheless, we inspected scatterplots of pairs of features in order to verify that they have a linear relationship in our specific case. Figure 5 presents the correlation modulus of the obtained results. For each data set, high correlation ($|r| > 0.80$) was found between some measures. The MCI data set only presents a correlation value higher than $|r| > 0.80$. This value is found between Correlation and Granger Coherence. However, results found for Mild AD data set present six pairs of features with high correlation ($|r| > 0.80$). These values are found between pairs of Granger measures (DTF–PC, ffDTF–DTF, PDC–DTF, PDC–ffDTF, dDTF–ffDTF and dDTF–PDC). Therefore, obtained features achieve good classification rates when they are used as single feature but they introduce overfitting if we combine them into a

multifeature classifier. Overfitting appears because the system learns redundant information of the same aspects of the data. Therefore a feature selection method is needed in order to minimize it.

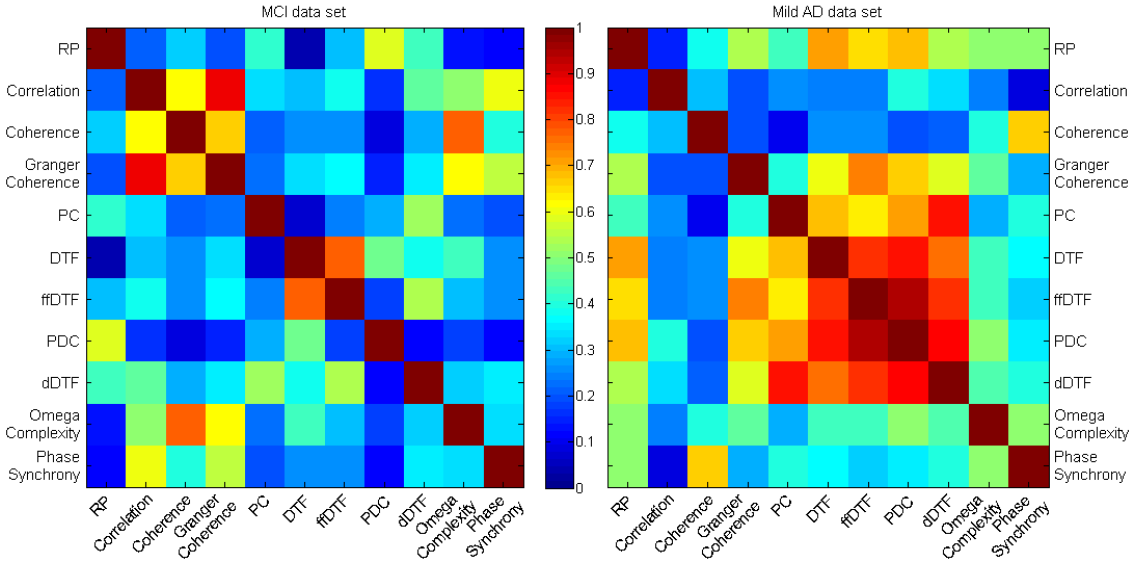

Figure 5. Modulus of the Pearson's linear correlation coefficient computed between all the measures. The optimal frequency range of each measure was used.

3.2 Multiple feature classification

Classification was also evaluated using multiple feature classification. The OFR algorithm defined in Section 2.6.1 was used in order to select the best parameters to perform a multiple feature classification. All computed measures in all frequency ranges were used as input features for the OFR algorithm.

The obtained features that best define each data set are presented in table 4 (MCI) and table 5 (Mild AD). The measures are presented in the order selected by the OFR algorithm along with selected OFR frequency ranges and the corresponding standard frequencies.

The results presented in table 4 demonstrate that the best feature for differentiating MCI patients from healthy subjects is RP in the frequency range 2-8 Hz. Table 5 again shows RP as the best feature, and the range from 4-7 Hz as the optimal range. These results seem to be consistent with the results obtained in Section 3.1, where RP obtained the higher CR in those frequency ranges in both data sets.

Table 4. Features and frequency ranges selected by the OFR algorithm for MCI data set. The last column presents the standard frequency bands corresponding to the measures.

Algorithm Order	Features	OFR Selected frequency ranges (Hz)	Standard frequency bands (Hz)
1	RP	2 - 8	4 - 8
2	Correlation	3 - 8	4 - 8
3	Coherence	1 - 6	1 - 4
4	PDC	1 - 3	1 - 4
5	ffDTF	9 - 29	13 - 30
6	Omega Complexity	24 - 25	13 - 30
7	Granger Coherence	1 - 30	1 - 4, 4 - 8, 8 – 13, 13 - 30
8	DTF	4 - 5	4 - 8
9	PC	1 - 10	1 - 4, 4-8
10	Phase Synchrony	28 - 30	13 - 30
11	dDTF	1 - 2	1 - 4

Table 5. Features and frequency ranges selected by the OFR algorithm for Mild AD data set. The last column presents the standard frequency bands corresponding to the measures.

Algorithm Order	Features	OFR Selected frequency ranges (Hz)	Standard frequency bands (Hz)
1	RP	4 – 7	4 - 8
2	Granger Coherence	1 – 2	1 - 4
3	Correlation	9 – 10	8 - 13
4	Phase Synchrony	25 – 26	13 - 30
5	PC	13 – 14	13 - 30
6	dDTF	2 – 6	4 - 8
7	Coherence	5 – 6	4 - 8
8	Omega Complexity	11 – 14	8 - 13
9	ffDTF	6 – 19	8 - 13
10	DTF	20 – 21	13 - 30
11	PDC	1 – 2	1 - 4

In order to study the improvement in the CR that results from including more of the selected features as input features to a classifier, we examined the evolution between the obtained CR and the number of features used as input. We also studied the improvement of performance by computing the CR using the features and frequency range selected by the OFR algorithm, and the same features in the standard frequency range. SE and SP were also evaluated for the different numbers of input features. Figure 6 presents this relationship for MCI subjects, and Figure 7 presents the same for Mild AD patients. In both images the top image represents the CR evolution, comparing the CR obtained with selected OFR frequency ranges and the CR computed with the standard frequency bands. Vertical lines indicate the percentage of noise introduced in the data set using the probe method. In the same figures, the bottom image presents the evolution of SE and SP.

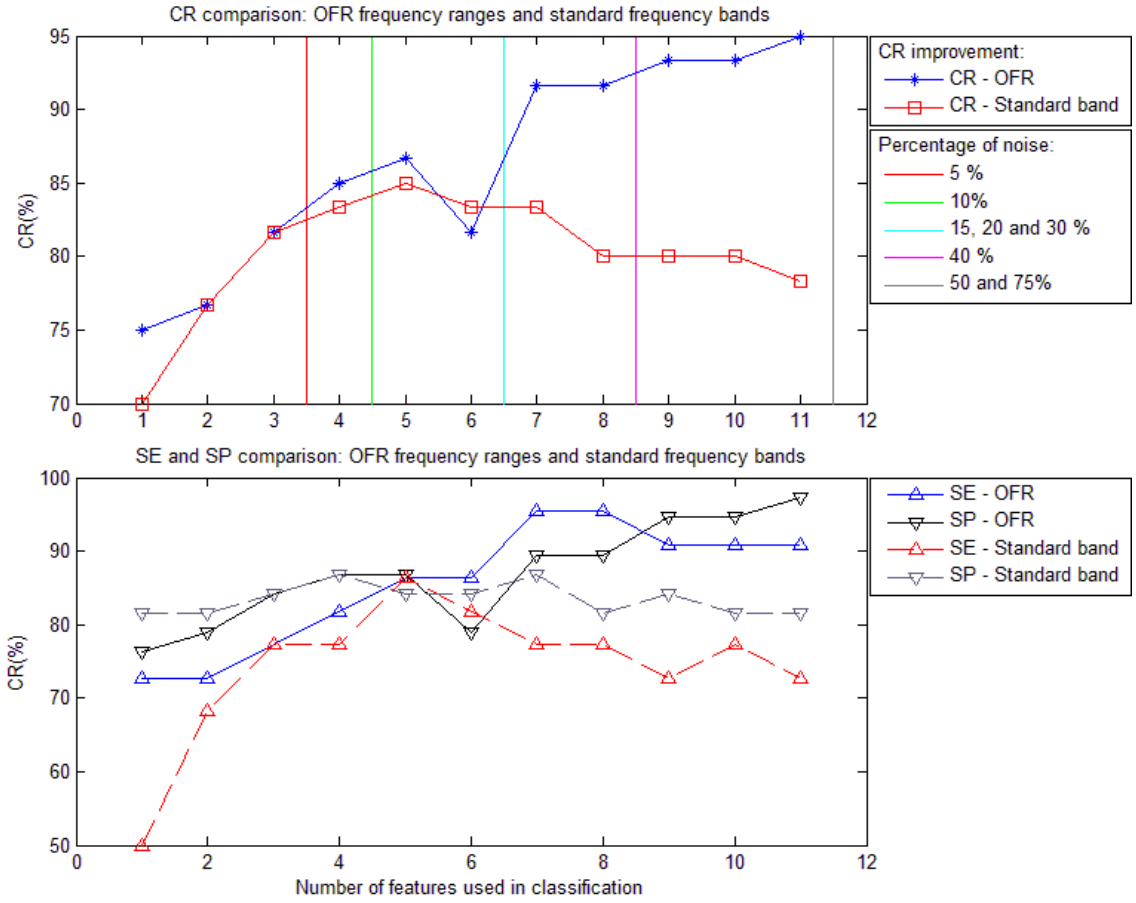

Figure 6. Relationship between the CR and the different number of features selected by the OFR algorithm. The top image presents the evolution of CR using OFR frequency ranges and the standard frequency bands. The bottom image presents the evolution of SE and SP. Results presented for the MCI data set.

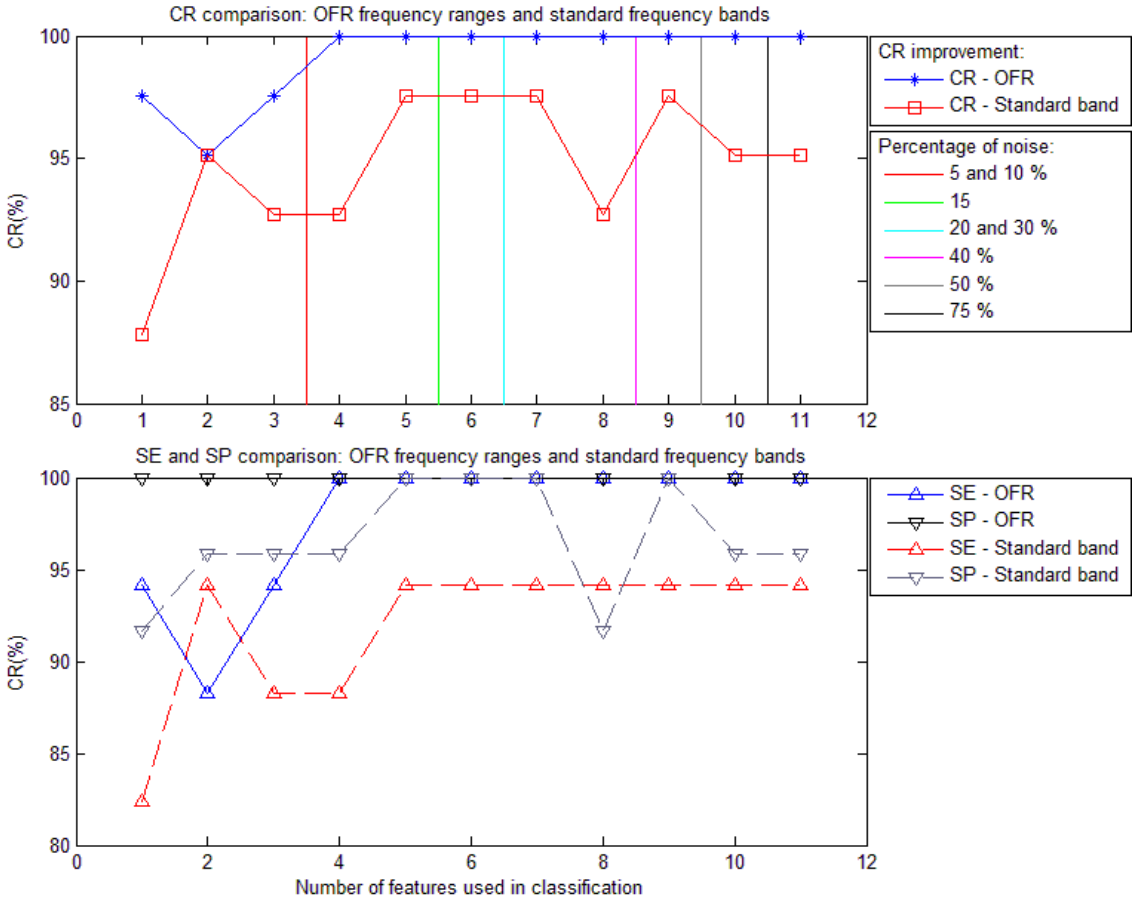

Figure 7. Relationship between the CR and the different number of features selected by the OFR algorithm. The top image presents the evolution of CR using OFR frequency ranges and the standard frequency bands. The bottom image presents the evolution of SE and SP. Results presented for the Mild AD data set.

Results presented in figure 6 (MCI data set) show that there is an improvement in the classification performance when we include more of the selected features as input parameters in a classifier. The same figure shows that using the OFR-selected frequency range achieves a better CR than using the standard frequency bands. The best value obtained using the OFR-selected frequency ranges is 95%, whereas using the classical frequency ranges the CR attains a maximum value of 85% and then decreases. SE and SP also present the same improvement with the increase of the number of features.

The results for Mild AD (figure 7) also show increases in performance when more input features are used. In this case, using only four features in the selected OFR frequency ranges, a CR of 100% is achieved; this value is stable when more features are added to the classifier. On

the other hand, with the standard frequency ranges the CR achieved is not as impressive. The maximum CR with those frequencies is 97.56%, using 5, 6, or 7 features. The use of 8, 9, or 10 features does not improve performance; instead the CR decreases to 92.68%. The evolution of SE and SP shows that SP is always equal to 1 for the OFR-selected frequencies, and it is SE that improves with the use of more features. This may indicate that adding more synchrony measures can characterize Mild AD patients better.

To check the redundant information provided by the selected features, Pearson's linear correlation between measures was computed as in Section 3.1. However, this time the frequency ranges selected by OFR were analysed. The features chosen by OFR are more salient for discriminating between patients and healthy subjects. The modulus of the results is presented in figure 8. We observe that correlation values are now lower than those presented in figure 5. For the MCI data set the highest correlation is obtained between PDC and DTF ($|r| = 0.72$). Results obtained for the Mild AD data set present an important decrease compared to the ones depicted in figure 5. Now, only the correlation between PDC and dDTF presents a high value ($|r| = 0.85$), in contrast with the six pairs obtained without using OFR.

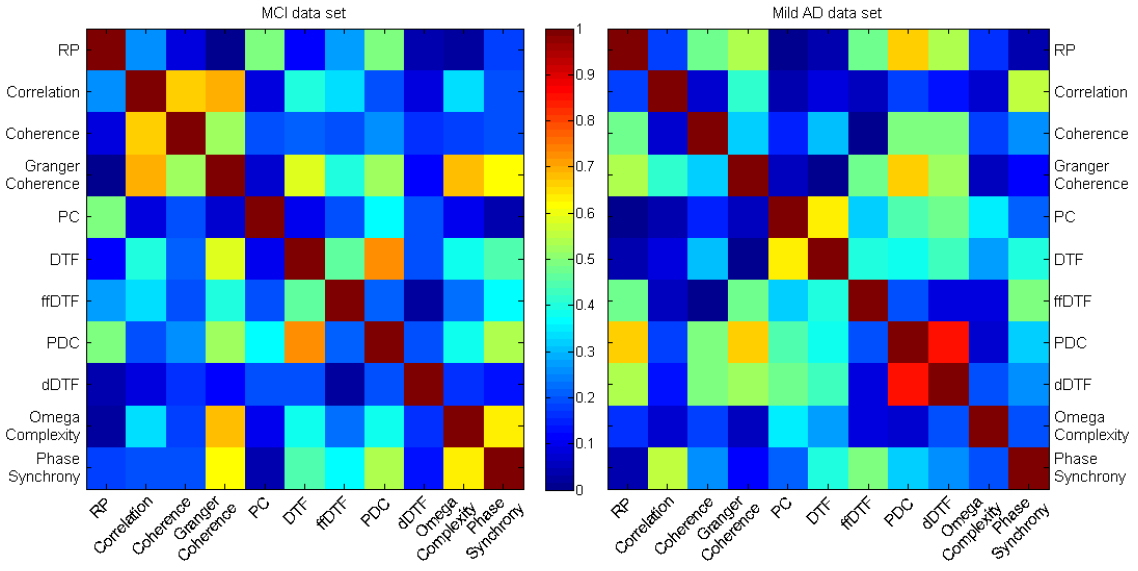

Figure 8. Modulus of the Pearson's linear correlation coefficient computed between all the measures. Frequency ranges defined by the OFR algorithm were used.

In order to standardize the obtained results, we carried out one more experiment. As the selected OFR features for each data set are different, both data sets were evaluated using the obtained

parameters from the other data set. Figure 9 presents the evolution of the CR using different numbers of features, where in this case the Mild AD data set was evaluated using the OFR-selected features for the MCI data set (top line) and the MCI data set was evaluated using the OFR-selected features for the Mild AD data set (bottom line). As can be seen, the change of parameters clearly reduces the CR obtained for the MCI data set but only presents a slight decrease for the Mild AD data set in comparison with the results obtained for each data set using its own OFR-selected measures and frequency ranges (see figure 6 for MCI and figure 7 for Mild AD). These results suggest that the results obtained for the MCI data set can be extended to the Mild AD data set to implement a system that could be used in hospitals.

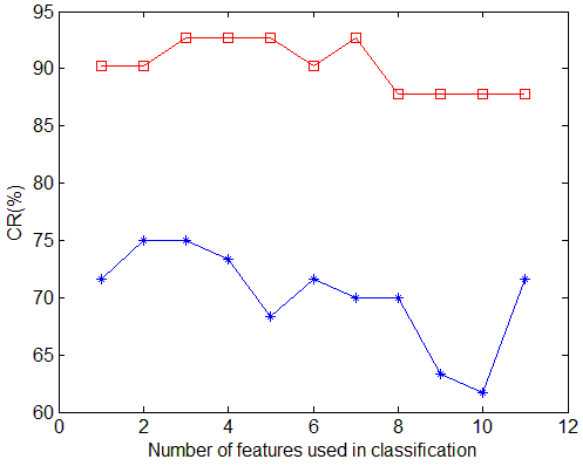

Figure 9. Evolution of the CR obtained, using the different numbers of features. The line with asterisks shows the MCI data set using OFR-selected features for Mild AD patients. The line with squares indicates the Mild AD data set using OFR-selected features for MCI patients.

In order to study stability across subjects who present the selected OFR features, we carried out a new experiment using feature selection through leave-one-out cross-validation. This cross-validation was performed by leaving a different subject out of the study in each iteration, with the aim of checking whether the selected features were stable all along the data set. Figure 10 presents the results obtained for the MCI data set, and figure 11 presents the same results for the Mild AD data set. In both cases, the features are listed in the same order as that obtained using the OFR algorithm. The results presented in figure 10 and figure 11 show that RP is the most stable feature selected for both data sets, for Mild AD data set RP was selected as first feature

for all the patients. In the MCI data set, Correlation and Coherence are stable across subjects, but PDC is not stable across subject variation. For the Mild AD data set, Granger Coherence, Correlation and Phase Synchrony are stable across the variation of subjects. For both data sets, there was variation in the percentage of times the latter features were selected in the same position as that obtained through the OFR algorithm. This may be due to the fact that in the OFR algorithm, each time that a feature is retained the remaining features are orthogonalized based on the selected feature. Consequently, if the first features present variability, this variability could be extended to the other features in the orthogonalization process. Taking into account these results, we can assume that the previously reported 95% of CR for MCI, the best result obtained with this data set so far, can be a generalizable value and not an overfitting effect.

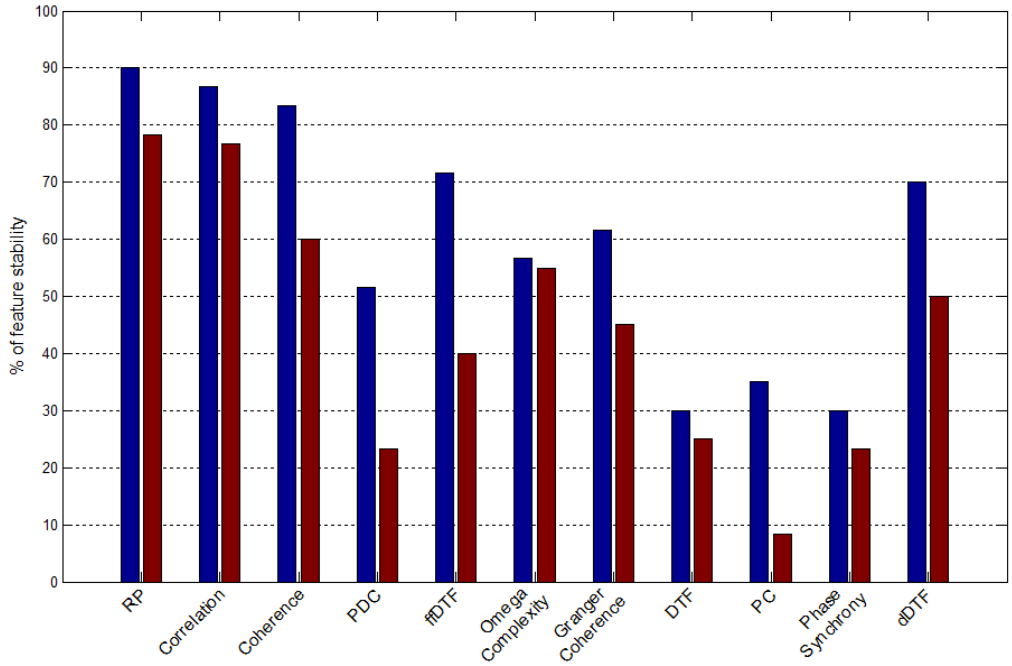

Figure 10. Results of computing the feature cross-validation in the OFR algorithm for the MCI data set. For each measure, first column stands for the % of times that each measure was selected in the same order as by the OFR algorithm and second column stands for the % of times that a feature was selected in the same order and with the same frequency range as by the OFR algorithm. The MCI data set contains 60 subjects.

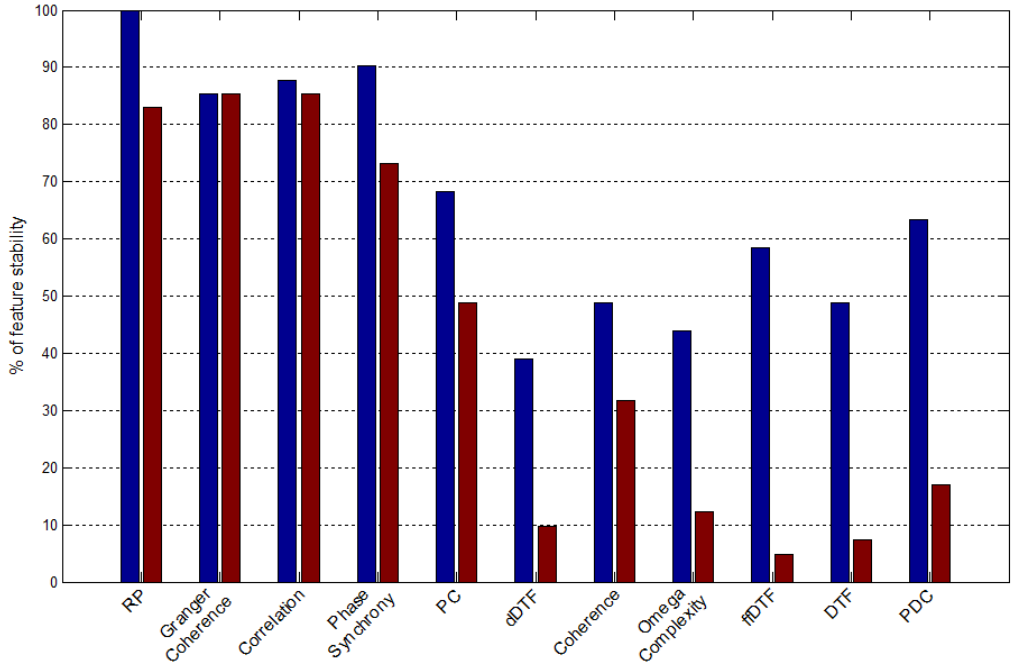

Figure 11. Results of computing the feature cross-validation in the OFR algorithm for the Mild AD data set. For each measure, first column stands for the % of times that each measure was selected in the same order as by the OFR algorithm and second column stands for the % of times that a feature was selected in the same order and with the same frequency range as by the OFR algorithm. The Mild AD data set contains 41 subjects.

4. Discussion

In this study we investigated the use of synchrony measures and a frequency power measure in the whole set of frequency ranges existing between 1 and 30 Hz. As we use OFR, each extracted feature is orthogonalized with respect to the previous extracted ones. Since RP was selected as the foremost discriminative feature, complexity measures were decorrelated on this basis. The results obtained in this study show that using a single measure, the classification is not as robust as can be with more attributes, and that a combination of RP and synchrony measures results in better classification performance. Furthermore, the use of specific frequency ranges for each measure improves the classification performance in comparison with the results obtained in the classical frequency range (δ, θ, α and β).

The presented results show that when using only one measure, RP is the best discriminating feature for the classification of AD patients versus healthy subjects. For the MCI data set, using RP we obtain a CR of 78.33%. On the other hand, the use of only a single synchrony measure

achieves the best CR, 75.00% for dDTF. Previous studies using this data set achieved similar results. For instance [15], using a completely different approach – Blind Source Separation (BSS) and RP in a different frequency range – achieved a CR of 80%. In our case, 78.33% is obtained without applying any decomposition technique. Our results using synchrony measures present some improvement over results presented in the literature. In [8,16], using only one measure as an input feature and LDA, the best obtained classification result was 70% using ffDTF. These studies evaluated the synchrony measures in the frequency range of 4-30 Hz. Our results show that analysing an optimal frequency range for each measure results in a better CR than using the whole frequency range. Using multiple feature classification the results improved to 78.33% in [7]. A number of other studies have also presented an improvement of CR, though only with using multiple features. In [18], 88.3% of CR is obtained by dividing the time series into small windows and computing the RP in each one. The value used as a discriminative feature in that study is the maximum value of RP, the best value obtained using the values of four electrodes. In [7], the best CR, using a combination of RP and a synchrony measure, is again 88.33%. The best CR obtained for the MCI data set was obtained in [17], achieving 93.3% using bump modelling [37,38], an approach completely different from the one presented here, which exploit time-frequency space information using a synchrony model, whereas our system only exploit the frequency information.

Using only one feature, a CR of 97.56% was obtained for the Mild AD data set. The use of only one single synchrony measure did not improve this result, because the best classification obtained was 95.12% for DTF (Mild AD). In [16] the best CR obtained was 82.9% using only one measure and in [7] the use of three measures as input features to a classifier achieved a CR of 95.12%. Our results are better in both cases. In [19], a CR of 97.6% was obtained using multiway array decomposition – in other words, the same value as is obtained in this study for RP used as individual feature classification, but using a more complex approach based on multiway array decomposition.

Using multiple features classification, the OFR algorithm selected RP as the most significant feature in both data sets. Interestingly, the frequency range obtained for RP is close, for both

data sets, to the standard θ range, which is the one that is usually analysed to study the slowing of EEG [5,39]. The results obtained also suggest that correlation may play an important role in the diagnosis of AD. For both data sets, correlation appears among the first three positions of the OFR-selected features, for MCI in the frequency range 3-8 Hz, and for Mild AD patients in the frequency range 9-10Hz. Another measure that appears to be significant is coherence in its different variants (Coherence and Granger Coherence).

For the MCI data set, using the eleven measures as input features for the classifier, a CR of 95% is achieved – the best result obtained with this data set. However, the level of significance using random probes is 50% (see figure 6), which indicates that those results may be overfitted. For the Mild AD data set, a CR of 100% was achieved using four features. The level of significance at which this value was obtained is less than 15% (see figure 7), which indicates that those measures were able to clearly identify AD patients in an advanced stage of the disease. This may indicate that MCI is a stage difficult to identify in comparison with the Mild AD stage. In the case of MCI, patients start to present some memory impairments but preserve other cognitive domains, whereas in the Mild AD stage subjects begin to display some cognitive deficits.

We have demonstrated that combining two of the well-known perturbations in AD EEG data (EEG slowing and changes in EEG synchrony) improve the ability to distinguish between AD patients and healthy subjects. The effect of the slowing of EEG, characterized by RP, appears to be more discriminative than the changes in synchrony. These results are in agreement with [40], where the effect of slowing of EEG was used to predict the progression from MCI to dementia. On the other hand, combining EEG slowing with the changes of EEG synchrony, quantified by coherence or correlation coefficient for instance, allows us to better differentiate between AD patients and healthy subjects. Changes of synchrony on EEG signals have been related with changes in functional connections between cortical regions [5] and brain cortical and subcortical atrophy [41].

Even though presented results achieve a good classification performance, i.e., 95% for MCI and 100% for Mild AD patients, several limitations of the present study should be emphasized. Due

to the limited number of subjects in each database, results may be prone to overfitting caused by parameters selection (time window and Granger order). On the other hand, the subjects in the Mild AD dataset are not age matched. Age differences have been related with changes in complex brain functional networks and cognitive decline [42]. Furthermore, studies comparing young and old AD patients have also described changes in the brain due to age [43]. Therefore, results obtained for the Mild AD patients have to be taken carefully because of the age differences between subjects.

In this study we limit ourselves to global values for each subject. Therefore, information specific to a particular pair of electrodes might get lost due to the average process. Regional analyses comparing the activity on different brain regions and different electrodes may further facilitate the differentiation of patients against healthy subjects.

Finally, we have to take into account that Mild AD is a stage in which the cognitive deficits are more pronounced than in MCI and, therefore, it is easier to classify the Mild AD subjects compared to MCI ones. Even if some shortcomings can be identified, the methodology described in this article opens an interesting line of research that could help to improve the diagnosis of the early stages of AD.

5. Conclusions

In this study, a group of synchrony measures and a frequency power measure were used to distinguish between healthy subjects and AD patients in different stages (MCI and Mild AD). Single features were used to compute CR in order to obtain the optimal frequency range that best discriminates between AD patients and healthy subjects. A multiple feature classification approach based on OFR was also presented, with the aim of obtaining a final CR that improves upon state of the art results, was described.

The two data sets analysed in this study (MCI and Mild AD) were obtained through different EEG recordings in two different hospitals, with different EEG systems and slightly different protocols. We can therefore expect significant variations in the experimental conditions. Consequently, comparing these two data sets is a challenging task. We chose to perform an independent study of each database separately. Interestingly, with both data sets high

classification rates were obtained: 95% for the MCI data set (using 11 features), and 100% for the Mild AD set (using 4 features). The results in frequency ranges differing from the standard bands were shown to be more discriminant. It seems that using a specific configuration and computing neural synchrony in a specific frequency range is more effective than standardizing all configurations. Furthermore, with the aim of obtaining a practical classification system, we explored the possibility of using the same features for both data sets. In this case, using features optimal for the MCI data set, we obtained promising results for the Mild AD data set, while, of course, maintaining the result for MCI. The standardization of features for the MCI and Mild AD data sets is worth future investigation.

Finally, it must be noted that the two data sets used are fairly small. A larger database is needed in order to generalize our results. Data sets containing different types of dementia, and optimally the evolution of MCI subjects to Mild AD, could significantly facilitate the early diagnosis of AD. In future work, we will analyse whether a better CR could be achieved by investigating and comparing the synchrony in each of the obtained regions, instead of computing a mean value for all the subjects. This methodology will allow us to identify which regions exhibit the most significant changes.

Acknowledgements

Esteve Gallego-Jutglà wishes to thank the University of Vic – Central University of Catalonia (UVic – UCC) for partially supporting this project under a PhD grant ("Amb el suport de l'ajut predoctoral de la Universitat de Vic") and under the mobility grants 2011MOB005 and 2013_1MOB/003. The authors would like to thank Drs. Sarah Umbrene Khan and Srinivasan Kannan for their valuable suggestions regarding the manuscript.

References

[1] Alzheimer's Association 2013 2013 Alzheimer's disease facts and figures *Alzheimer's Dement.* **9** 208–45

[2] Dauwels J, Vialatte F and Cichocki A 2010 Diagnosis of Alzheimer's Disease from EEG Signals: Where Are We Standing? *Curr. Alzheimer Res.* **7** 487–505

[3] Petersen R C, Roberts R O, Knopman D S, Boeve B F, Geda Y E, Ivnik R J, Smith G E and Jack Jr C R 2009 Mild cognitive impairment: ten years later *Arch. Neurol.* **66** 1447

[4] Weiner M W, Veitch D P, Aisen P S, Beckett L A, Cairns N J, Green R C, Harvey D, Jack C R, Jagust W, Liu E, Morris J C, Petersen R C, Saykin A J, Schmidt M E, Shaw L, Siuciak J A, Soares H, Toga A W and Trojanowski J Q 2012 The Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception *Alzheimer's Dement.* **8** S1–S68

[5] Jeong J 2004 EEG dynamics in patients with Alzheimer's disease *Clin. Neurophysiol.* **115** 1490–505

[6] Scinto L F M and Daffner K R 2000 *Early diagnosis of Alzheimer's disease* (Totowa, New Jersey: Humana Press Inc.)

[7] Dauwels J, Srinivasan K, Ramasubba Reddy M, Musha T, Vialatte F-B, Latchoumane C, Jeong J and Cichocki A 2011 Slowing and Loss of Complexity in Alzheimer's EEG: Two Sides of the Same Coin? *Int. J. Alzheimers. Dis.* **2011** 539621

[8] Dauwels J, Vialatte F B, Musha T and Cichocki A 2010 A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG *Neuroimage* **49** 668–93

[9] Kropotov J D 2009 *Quantitative EEG, Event-Related Potentials and Neurotherapy* (Elsevier Inc.)

[10] Nunez P and Srinivasan R 2006 *Electric fields of the brain: the neurophysics of EEG* (Oxford university press)

[11] Babiloni C, Ferri R, Binetti G, Cassarino A, Forno G D, Ercolani M, Ferreri F, Frisoni G B, Lanuzza B and Miniussi C 2006 Fronto-parietal coupling of brain rhythms in mild cognitive impairment: a multicentric EEG study *Brain Res. Bull.* **69** 63–73

[12] Koenig T, Prichep L, Dierks T, Hubl D, Wahlund L O, John E R and Jelic V 2005 Decreased EEG synchronization in Alzheimer's disease and mild cognitive impairment *Neurobiol. Aging* **26** 165–71

[13] Elgendi M, Vialatte F, Cichocki A, Latchoumane C, Jeong J and Dauwels J 2011 Optimization of EEG frequency bands for improved diagnosis of Alzheimer disease *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE* pp 6087–91

[14] Gallego-Jutglà E, Elgendi M, Vialatte F, Solé-Casals J, Cichocki A, Latchoumane C, Jeong J and Dauwels J 2012 Diagnosis of Alzheimer's disease from EEG by means of

synchrony measures in optimized frequency bands. *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE* vol 2012 pp 4266–70

[15] Cichocki A, Shishkin S L, Musha T, Leonowicz Z, Asada T and Kurachi T 2005 EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease *Clin. Neurophysiol.* **116** 729–37

[16] Dauwels J, Vialatte F, Latchoumane C, Jeong J and Cichocki A 2009 EEG synchrony analysis for early diagnosis of Alzheimer's disease: a study with several synchrony measures and EEG data sets. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2009** 2224–7

[17] Vialatte F, Cichocki A, Dreyfus G, Musha T, Rutkowski T M and Gervais R 2005 Blind source separation and sparse bump modelling of time frequency representation of EEG signals: New tools for early detection of Alzheimer's disease *2005 IEEE Workshop on Machine Learning for Signal Processing* pp 27–32

[18] Woon W L, Cichocki A, Vialatte F and Musha T 2007 Techniques for early detection of Alzheimer's disease using spontaneous EEG recordings *Physiol. Meas.* **28** 335

[19] Latchoumane C-F V., Vialatte F-B, Solé-Casals J, Maurice M, Wimalaratna S R, Hudson N, Jeong J and Cichocki A 2012 Multiway array decomposition analysis of EEGs in Alzheimer's disease *J. Neurosci. Methods* **207** 41–50

[20] Vialatte F-B, Solé-Casals J, Maurice M, Latchoumane C, Hudson N, Wimalaratna S, Jeong J and Cichocki A 2009 Improving the Quality of EEG Data in Patients with Alzheimer's Disease Using ICA *Adv. Neuro-Information Process.* 979–86

[21] Cooley J W and Tukey J W 1965 An algorithm for the machine calculation of complex Fourier series *Math. Comput.* 297–301

[22] Frigo M and Johnson S G 1998 FFTW: An adaptive software architecture for the FFT *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on* vol 3 (IEEE) pp 1381–4

[23] Sanei S and Chambers J A 2007 *EEG Signal Processing* (John Wiley & Sons)

[24] Pascual-Marqui R D 2007 Instantaneous and lagged measurements of linear and nonlinear dependence between groups of multivariate time series: frequency decomposition *arXiv Prepr. arXiv0711.1455*

[25] Kaminski M and Liang H 2005 Causal Influence: Advances in Neurosignal Analysis *Crit. Rev. Biomed. Eng.* **33** 347–430

[26] Korzeniewska A, Mańczak M, Kamiński M, Blinowska K J and Kasicki S 2003 Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method *J. Neurosci. Methods* **125** 195–207

[27] Bhattacharya J 2000 Complexity analysis of spontaneous EEG *Acta Neurobiol. Exp. (Wars).* **60** 495–501

[28] Saito N, Kuginuki T, Yagyu T, Kinoshita T, Koenig T, Pascual-Marqui R D, Kochi K, Wackermann J and Lehmann D 1998 Global, Regional, and Local Measures of

Complexity of Multichannel Electroencephalography in Acute, Neuroleptic-Naive, First-Break Schizophrenics *Biol. Psychiatry* **43** 794–802

[29] Wackermann J 1999 Towards a quantitative characterisation of functional states of the brain: from the non-linear methodology to the global linear description *Int. J. Psychophysiol.* **34** 65–80

[30] Wackermann J 1996 Beyond mapping: Estimating complexity of multichannel EEG recordings *Acta Neurobiol. Exp. (Wars).* **56** 197–208

[31] Lachaux J-P, Rodriguez E, Martinerie J and Varela F J 1999 Measuring phase synchrony in brain signals *Hum. Brain Mapp.* **8** 194–208

[32] Oppenheim A V and Schafer R W 1975 *Digital signal processing* (London etc.: Prentice-Hall International)

[33] Duda R O, Hart P E and Stork D G 2001 *Pattern Classification* (John Wiley \& Sons,)

[34] Chen S, Billings S A and Luo W 1989 Orthogonal least squares methods and their application to non-linear system identification *Int. J. Control* **50** 1873–96

[35] Dreyfus G, Martinez J-M, Samuelides M, Gordon M B, Badran F and Thiria S 2008 *Apprentissage statistique: Réseaux de neurones - Cartes topologiques - Machines à vecteurs supports* (Paris: Éditions Eyrolles)

[36] Stoppiglia H, Dreyfus G, Dubois R and Oussar Y 2003 Ranking a random feature for variable and feature selection *J. Mach. Learn. Res.* **3** 1399–414

[37] Vialatte F B, Dauwels J, Solé-Casals J, Maurice M and Cichocki A 2009 Improved sparse bump modeling for electrophysiological data *Lecture Notes in Computer Science* vol 5506 LNCS (Springer Berlin Heidelberg) pp 224–31

[38] Vialatte F B, Solé-Casals J, Dauwels J, Maurice M and Cichocki A 2009 Bump time-frequency toolbox: A toolbox for time-frequency oscillatory bursts extraction in electrophysiological signals *BMC Neurosci.* **10**

[39] Czigler B, Csikós D, Hidasi Z, Anna Gaál Z, Csibri É, Kiss É, Salacz P and Molnár M 2008 Quantitative EEG in early Alzheimer's disease patients—power spectrum and complexity features *Int. J. Psychophysiol.* **68** 75–80

[40] Vecchio F, Babiloni C, Lizio R, De Vico Fallani F, Blinowska K, Verrienti G, Frisoni G and Rossini P M 2013 Resting state cortical EEG rhythms in Alzheimer's disease: Toward EEG markers for clinical applications: A review *Suppl. Clin. Neurophysiol.* **62** 223–36

[41] Babiloni C, Vecchio F, Lizio R, Ferri R, Rodriguez G, Marzano N, Frisoni G B and Rossini P M 2011 Resting state cortical rhythms in mild cognitive impairment and Alzheimer's disease: electroencephalographic evidence. *J. Alzheimer's Dis.* **26 Suppl 3** 201–14

[42] Bamidis P D, Vivas A B, Styliadis C, Frantzidis C, Klados M, Schlee W, Siountas A and Papageorgiou S G 2014 A review of physical and cognitive interventions in aging. *Neurosci. Biobehav. Rev.* **44** 206–20

[43] De Waal H, Stam C J, de Haan W, van Straaten E C W, Scheltens P and van der Flier W M 2012 Young Alzheimer patients show distinct regional changes of oscillatory brain dynamics. *Neurobiol. Aging* **33** 1008.e25–31