An integrated approach based on EEG signals processing combined with supervised methods to classify Alzheimer's disease patients

1st Giulia Fiscon
[1]*ACT Operations Research, Rome, Italy*
[2]*Institute of Systems Analysis and Computer Science, National Research Council, Rome, Italy*
giulia.fiscon@iasi.cnr.it

2nd Emanuel Weitschek
Department of Engineering, Uninettuno International University, Rome, Italy
emanuel.weitschek@iasi.cnr.it

3th Maria Cristina De Cola
Centro Neurolesi "Bonino-Pulejo", Messina, Italy
cristina.decola@gmail.com

4rd Giovanni Felici
Institute of Systems Analysis and Computer Science, National Research Council, Rome, Italy
giovanni.felici@iasi.cnr.it

5th Paola Bertolazzi
[1]*ACT Operations Research, Rome, Italy*
[2]*Institute of Systems Analysis and Computer Science, National Research Council, Rome, Italy*
paola.bertolazzi@iasi.cnr.it

***Abstract*—Alzheimer's Disease (AD) is the most widespread and incurable neurodegenerative disorder, and together with its preliminary stage - Mild Cognitive Impairment (MCI) - its detection still remains a challenging issue. Electroencephalography (EEG) is a non-invasive and repeatable technique to diagnose brain abnormalities. However, the analysis of EEG spectra is still carried out manually by experts and effective computer science methods to extract relevant information from these signals become a necessity. Through a data mining approach, which guides the automated knowledge discovery process, we aim to achieve an automatic patients classification from the EEG biomedical signals of AD and MCI, in order to support medical doctors in the diagnosis formulation. Specifically, we design an integrated procedure that encompasses the following steps: (1) data collection; (2) data preprocessing of EEG-signals data; (3) features extraction by applying time-frequency transforms on EEG-signals (Fourier and Wavelet analysis); and (3) a supervised learning approach to classify samples in patients suffering from AD, patients affected by MCI, and healthy control (HC) subjects. By applying our procedure, we are able to extract human-interpretable classification models that allow to automatically assign the patients into their belonging class. In particular, by exploiting a Wavelet feature extraction we achieve 83%, 92%, and 79% of accuracy when dealing with HC vs AD, HC vs MCI, and MCI vs AD classification problems, respectively. By comparing the classification performances with both feature extraction methods, we find out that Wavelets analysis outperforms Fourier. Thus, we suggest it in combination with supervised methods for automatic patients classification based on their EEG signals for aiding the medical diagnosis of dementia. We provided processed data from our study at ftp://bioinformatics.iasi.cnr.it/public/EEG/.**

***Index Terms*—EEG signals analysis, supervised learning, Alzheimer's Disease**

I. Introduction

Alzheimer's disease (AD) is the most widespread form of dementia, for which actually no cure is known [1]. 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 [2]. 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. In particular, Electroencephalography (EEG) appears as non-invasive and repeatable technique to diagnose brain abnormalities [3]. Different studies have shown that AD has (at least) three major effects on EEG signals: enhanced complexity, slowing of signals, and perturbations in EEG synchrony [4], [5]. The open challenge is to perform clinical studies in order to shed light on biological and medical questions related to AD and MCI [6]. 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 [7].

In [8], we presented an integrated classification procedure that combined EEG-signal preprocessing with supervised machine learning methods in order distinguish different dementia related pathology (e.g., AD and MCI) and help also to discriminate the healthy control subjects (HC).

978-1-5386-5488-0/18/$31.00 ©2018 IEEE

Processed data from our study are freely available at ftp://bioinformatics.iasi.cnr.it/public/EEG/.

II. Materials and Methods

The classification procedure was designed with the following steps (Fig. 1): (1) data collection; (2) data preprocessing of EEG signals; (3) features extraction by means of the Discrete Fourier and Wavelet Transforms; and (3) classification with tree-based supervised methods.

Fig. 1. Flowchart of the integrated procedure for EEG signals analysis and classification.

A. Data collection

We enrolled a total of 109 subjects in 2012 and 2013 at The IRCCS Centro Neurolesi "Bonino-Pulejo": 23 healthy controls samples (HC) and 86 patients affected by dementia (AD, MCI). According to the World Health Organization definition, an expert neurologist classified the patients in affected by AD or MCI, including 37 men and 49 women with the mean aged of 78.4 $\pm$ 6.4 for AD and 74.1 $\pm$ 9.4 years for MCI. HC subjects includes 13 men and 10 women with the mean age of 65.6 $\pm$ 7.9 years. Subjects under pharmacological treatment that could change the activity of the brain have been excluded from the study, whereas subjects capable of undergoing an electroencephalogram and with a negative anamnesis for neurological comorbid disease have been included. The enrolled subjects can be divided in three main etiological classes: (1) patients with Alzheimer's disease (AD), (2) patients with Mild Cognitive Impairment (MCI), and (3) healthy control samples (HC).

B. Data preprocessing

We acquired multi-channel EEG signals by using 19 electrodes (E), by setting their placement according to the International 10-20 System [9], and by exploiting monopolar connections with earlobe electrode landmark [10]. The brain activity of the subjects in resting condition and closed eyes was measured in terms of electrical potential (μV). We recorded the EEG signals by capturing 300 seconds with 256 or 1024 sampling frequency (Hz). For each signal we selected the central 180 seconds (i.e., from 60 to 240 seconds) to avoid initial and final EEG recording artifacts. Additionally, to normalize the sampling frequency we converted each signal to 256 Hz.

C. Features Extraction

We extracted features from EEG signals in frequency domain by applying the Fourier and the Wavelet Transform for estimating their spectrum [7], [11], [12]. In particular, the Fast Fourier Transform (FFT) was applied to each EEG signal of 180 seconds and we obtained $M = 16$ Fourier Coefficients for each electrode. Thus, for each sample we extracted 304 features (i.e., $M \cdot E = 16$ coefficients $\cdot$ 19 electrodes) and we arranged them in a matrix with 109 rows (i.e., samples) and 305 columns (304 referring to the features, and one referring to the sample class).

Then, the Discrete Wavelet Transform (DWT) was applied to each EEG signal of 180 seconds and we obtained $M = 48$ Wavelet Coefficients for each electrode. Once again, for each sample we extracted 912 features (i.e., $M \cdot E = 48$ coefficients $\cdot$ 19 electrodes) and we arranged them in a matrix with 109 rows (i.e., samples) and 913 columns (912 referring to the features, and one referring to the sample class).

Table I reports a schematic representation of these feature matrices, where N is the number of samples, M is the number of coefficients (Fourier or Wavelet), E is the number of EEG electrodes, $M \cdot E$ is the number of features, and $x_{i,j}$ refers to a generic matrix element.

TABLE I
Example of output matrix from feature extraction step.

Sample	Coefficient$_{(1,1)}$	$\cdots$	Coefficient$_{(E,E\cdot M)}$	Sample Type
sample$_1$	$x_{(1,1)}$	$\cdots$	$x_{(1,E\cdot M)}$	MCI
sample$_2$	$x_{(2,1)}$	$\cdots$	$x_{(2,E\cdot M)}$	AD
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
sample$_N$	$x_{(N,1)}$	$\cdots$	$x_{(N,E\cdot M)}$	HC

The Wavelet and Fourier spectral analysis of the EEG signals were performed by using MATLAB R2014a [13].

D. Classification

We performed a supervised learning analysis in order to automatically classify the samples to their types (MCI, AD, HC) by processing their associated features [14], [15]. We aimed to extract a human readable model specific for each type of sample (MCI, AD, HC) based on a small subset of features (e.g., 'if $Wavelet_{15} > 0.4$ and $Wavelet_{13} < 0.7$ then the sample can be classified as AD") that could aid clinicians to identify key features related to the understudy neurodegenerative disease. Specifically, we applied tree-based classifier (J48 that implements the C4.5 algorithm) from Weka 3.6.9 environment [16] to address the following classification tasks: (1) HC *vs* AD; (2) HC *vs* MCI; (3) MCI *vs* AD; (4) HC *vs* CASE (MCI+AD), where the CASE class is composed of AD joint to MCI samples for testing diseased subjects recognition with respect to the healthy ones.

III. Results and Discussion

In Table II, we reported the results of the Decision Tree C4.5 classifier considering the EEG signals analysis performed both by using the Fourier Transform and Wavelet Transform.

We used $M = 16$ Fourier Coefficients and $M = 48$ Wavelet Coefficients as features and a leave-one-out sampling with 72, 60, 86, 109 folds for HC *vs* AD, HC *vs* MCI, MCI *vs* AD, HC *vs* CASE, respectively.

For what concerns Fourier analysis, we obtained 72%, 72%, 75%, 80% of accuracy when dealing with HC *vs* MCI, HC *vs* AD, HC *vs* CASE, and MCI *vs* AD classification problems, respectively.

On the other hand, the features extraction based on the Wavelet Transform achieved high classification performance in all metrics and for all classification problems. In particular, we obtained 92%, 83%, 73%, and 79% of accuracy when dealing with HC *vs* MCI, HC *vs* AD, HC *vs* CASE, , and MCI *vs* AD classification problems, respectively. Furthermore, the Wavelet spectral analysis outperformed the Fourier analysis when dealing with EEG signals classification of HC *vs* MCI, HC *vs* AD, and HC *vs* CASE. Conversely, for MCI *vs* AD both signal processing methods lead to comparable classification performances.

TABLE II
CLASSIFICATION RESULTS [%] FOR FOURIER ANALYSIS AND WAVELET ANALYSIS ON EEG SIGNALS.

	HC *vs* AD	HC *vs* MCI	MCI *vs* AD	HC *vs* CASE
Fourier				
Accuracy	72.2	71.7	80.2	74.7
F-measure	71.4	71.8	80.1	74.7
Sensitivity	72.2	71.7	80.2	74.7
Precision	71.1	78.9	80.2	74.0
Specificity	59.0	79.0	78.5	46.3
Wavelet				
Accuracy	72.2	71.7	80.2	74.7
F-measure	71.4	71.8	80.1	74.7
Sensitivity	72.2	71.7	80.2	74.7
Precision	71.1	78.9	80.2	74.0
Specificity	59.0	79.0	78.5	46.3

For validating our results, we applied our procedure by randomly permuting the class labels of data (i.e., hundred random permutations for each classification problem and for each EEG signal processing technique) and we obtained an overall average classification accuracy of 50.6%.

IV. CONCLUSIONS

We presented an integrated procedure for EEG signals classification of subjects affected by Alzheimer Disease (AD) and Mild Cognitive Impairment (MCI) with respect to Healthy Control samples (HC). The analysis relies on a data preprocessing and feature extraction phase followed by a classification procedure based on the widespread supervised learning approach to distinguish AD, MCI, and HC samples. We tested our procedure on EEG signals recorded on 109 human samples (23 HC, 37 MCI, 49 AD). By combining the Wavelet-based signal analysis and the tree-based classifier C4.5, we identified HC, MCI, and AD experimental samples with better accuracy than the spectral analysis carried out with Fast Fourier Transform.

REFERENCES

[1] T. Bird, "Alzheimer's disease and other primary dementias," *Harrisons principles of internal medicine*, vol. 2, pp. 2391–2398, 2001.

[2] R. C. Petersen, "Early diagnosis of alzheimers disease: is mci too late?" *Current Alzheimer Research*, vol. 6, no. 4, p. 324, 2009.

[3] H. Braak and E. Braak, "Neuropathological stageing of alzheimer-related changes," *Acta neuropathologica*, vol. 82, no. 4, pp. 239–259, 1991.

[4] 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.

[5] 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.

[6] 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.

[7] 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*. IEEE, 2006, pp. 4626–4629.

[8] G. Fiscon, E. Weitschek, A. Cialini, G. Felici, P. Bertolazzi, S. De Salvo, A. Bramanti, P. Bramanti, and M. C. De Cola, "Combining eeg signal processing with supervised methods for alzheimers patients classification," *BMC medical informatics and decision making*, vol. 18, no. 1, p. 35, 2018.

[9] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10–20 system electrode placement," *Electroencephalography and clinical neurophysiology*, vol. 66, no. 4, pp. 376–382, 1987.

[10] H. H. Jasper, "The ten twenty electrode system of the international federation," *Electroencephalography and clinical neurophysiology*, vol. 10, pp. 371–375, 1958.

[11] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of eeg records in an epileptic patient using wavelet transform," *Journal of neuroscience methods*, vol. 123, no. 1, pp. 69–87, 2003.

[12] 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.

[13] MATLAB, *version 8.3.0 (R2014a)*. Natick, Massachusetts: The MathWorks Inc., 2014.

[14] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Addison Wesley, 2005.

[15] E. Weitschek, G. Felici, and P. Bertolazzi, "Clinical data mining: Problems, pitfalls and solutions." Los Alamitos: IEEE Computer Society, 2013, pp. 90–94.

[16] 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, no. 1, pp. 10–18, Nov. 2009.