A Time-Frequency based Machine Learning System for Brain States Classification via EEG Signal Processing

Cosimo Ieracitano
DICEAM Department
Mediterranean University of Reggio Calabria
Reggio Calabria, Italy
cosimo.ieracitano@unirc.it

Nadia Mammone
IRCCS
Centro Neurolesi Bonino-Pulejo
Messina, Italy
nadia.mammone@irccsme.it

Alessia Bramanti
ISASI
National Research Council
Messina, Italy
a.bramanti@isasi.cnr.it

Silvia Marino
IRCCS
Centro Neurolesi Bonino-Pulejo
Messina, Italy
silvia.marino@irccsme.it

Amir Hussain
School of Computing
Edinburgh Napier University
Edinburgh, United Kingdom
a.hussain@napier.ac.uk

Francesco Carlo Morabito
DICEAM Department
Mediterranean University of Reggio Calabria
Reggio Calabria, Italy
morabito@unirc.it

***Abstract*—In the last decades, the use of Machine Learning (ML) algorithms have been widely employed to aid clinicians in the difficult diagnosis of neurological disorders, such as Alzheimer's disease (AD). In this context, here, a data-driven ML system for classifying Electroencephalographic (EEG) segments (i.e. epochs) of patients affected by AD, Mild Cognitive Impairment (MCI) and Healthy Control (HC) individuals, is introduced. Specifically, the proposed ML system consists of evaluating the average Time-Frequency Map (aTFM) related to a 19-channels EEG epoch and extracting some statistical coefficients (i.e. mean, standard deviation, skewness, kurtosis and entropy) from the main five conventional EEG sub-bands (or EEG-rhythms: delta, theta, alpha$_1$, alpha$_2$, beta). Afterwards, the time-frequency features vector is fed into an Autoeconder (AE), a Multi-Layer Perceptron (MLP), a Logistic Regression (LR) and a Support Vector Machine (SVM) based classifier to perform the 2-ways EEG epoch-classification tasks: AD vs HC and AD vs MCI. The performances of the proposed approach have been evaluated on a dataset of 189 EEG signals (63 AD, 63 MCI and 63 HC), recorded during an eye-closed resting condition at IRCCS Centro Neurolesi Bonino Pulejo of Messina (Italy). Experimental results reported that the 1-hidden layer MLP (MLP$_1$) outperformed all the other developed learning systems as well as recently proposed state-of-the-art methods, achieving accuracy rate up to 95.76 % $\pm$ 0.0045 and 86.84% $\pm$ 0.0098 in AD vs HC and AD vs MCI classification, respectively.**

***Index Terms*—Machine learning, Time-Frequency features, EEG recording, Alzheimer's disease, Mild Cognitive Impairment.**

I. Introduction

Nowadays, 50 million of individuals suffer from mental disorders worldwide [1]. This number is expected to rise to 82 million by 2030 and even to 116 million by 2050. Furthermore, 60% of dementia cases are due to Alzheimer's disease (AD) [2]. AD causes irreversible cognitive damages such as thinking, memory, and significantly undermines patients' autonomy. However, AD is a progressive degenerative neurological disease. The neurological state prodromal to AD is known as Mild Cognitive Impairment (MCI) [3]. A person with MCI suffers from mild memory problems and is able to carry out daily activities. Nevertheless, MCI patients' symptoms typically aggravate along with the disease's progression. Indeed, the rate conversion from MCI to AD is of 10%-15% per year. The rest remains stable, develops other forms of dementia and rarely regresses to a healthy condition. Moreover, it is to be noted that the early detection of AD is still a challenging open issue. A timely diagnosis would improve the patients' and caregivers' quality of life and reduce the mortality rate due to AD. The most common method to monitor abnormalities of the electrical brain activity is analyzing the low-cost, noninvasive scalp Electroencephalographic (EEG) recordings, with a particular attention to the following specific rhythms (or sub-bands): delta [0.5-4 Hz], theta [4-8 Hz], alpha$_1$ [8-10 Hz], alpha$_2$ [10-13 Hz], beta [13-32 Hz]. Indeed, several studies showed that EEG may reflect anomalies (such as slowing of the rhythms, loss of complexity, loss of connectivity between channels) due to the cortical degeneration caused by AD [4], [5], [6], [7], [8]. In this context, EEG-based machine learning (ML) techniques have been widely used to aid the diagnosis of cognitively impaired patients. In the last years, ML-classification systems have been developed to classify EEG recordings of AD/MCI subjects especially, by using empirically defined EEG spectral features. Indeed, it is a common clinical practice to investigate anomalies in the spectral domain of AD as it is possible to observe potential correlations between the stage of disease and

978-1-7281-2009-6/$31.00 paper N-20207.pdf

the EEG power frequency distribution [9]. For example, Huang et al. [10] analyzed EEG signals of 38 mild AD, 31 MCI, 24 HC and used alpha and theta global frequency power to perform AD vs HC and AD vs MCI classifications, achieving accuracy values of 84% and 78%, respectively, with linear discriminant analysis (LDA). Lehmann et al. [11] focused on AD vs HC discrimination using absolute and relative spectral power, distribution of spectral power, spatial synchronization measurements as input of conventional ML architectures (such as LDA, Support Vector Machine (SVM), Artificial Neural Network (ANN)), reporting sensitivity and specificity values up to 89% and 88%, respectively. Trambaiolli et al. [12] extracted a set of spectral measurements to discriminate EEG recordings of 22 AD and 12 HC subjects. They developed a SVM classifier, achieving accuracy rate up to 91.18%. Ruiz-Gmez et al. [13] analyzed a dataset of 111 EEG recordings (37 AD , 37 MCI, 37 HC) and proposed a multiclass classification system via binary classifiers (HC vs. all and AD vs. all). After selecting a set of optimal features (individual alpha frequency, relative power in delta frequency sub-band, sample entropy), the proposed 1-hidden layer neural network reported better classification performances (78.43% and 76.47% accuracy in HC vs. all and AD vs. all, respectively) as compared with standard discriminate analysis techniques. In contrast, recently, Fiscon et al. [14] performed both spectral and time-frequency (TF) analysis on 109 EEG traces belonging to AD, MCI, and HC, for training a decision tree classifier. They found out that TF features achieved best accuracy performances: 83.3% (AD vs HC), 91.7% (MCI vs HC), 79.1% (AD vs MCI). In this context, advanced machine learning techniques (known as *deep learning* (DL) [15]) have been also employed in developing EEG-based classification systems. In a recent work, Ieracitano et. [16] developed a data-driven Convolutional Neural Network (CNN) to differentiate AD, MCI and HC subjects through 2D-spectral representations (PSD-images) of EEG recordings, achieving accuracy rate up to 93.11% (in AD vs HC classification). In [17] Morabito et al. extracted some statistical parameters (mean (m), standard deviation (d), skewness (v)) from pre-processed time-frequency maps of sub-bands EEG signals. Then, they developed a stacked Autoecoder (AE) architecture to capture latent features and discriminate EEG of four categories of subjects: CJD (Creutzfeldt-Jakob Disease), RPD (Rapidly Progressive Dementia), AD, HC. Experimental results reported good classification performance (i.e. 88% in CJD vs AD). Similarly, in [18] Morabito et al. adopted the same time-frequency features extraction approach to classify EEG recordings of psychogenic non-epileptic seizures (PNES) from HC; whereas in [19], the time-frequency features representation was used as input of a CNN architecture to classify EEG of AD, MCI and HC, achieving accuracy values of 85% in AD vs HC and MCI vs HC discrimination and 78% in AD vs MCI classification. However, in all the aforementioned TF approaches a small cohort of subjects was analyzed and the TF features were extracted from three bands roughly corresponding to the rhythms of interest. Motivated by the promising results achieved with TF analysis of EEG signals ([18], [17], [19]) we propose a Time-Frequency based Machine Learning System for Classifying EEG recordings of AD, MCI and HC subjects. Here, a dataset of 189 EEG signals (63 AD, 63 MCI, 63 HC, collected at IRCSS Centro Neurolesi Bonino-Pulejo of Messina, Italy) is used and a larger vector of statistical coefficients is evaluated from the main five conventional EEG sub-bands (delta, theta, alpha$_1$, alpha$_2$, beta). Specifically, as the TF features proposed in [18], [17], [19], provided very good classification performance, in this study, we decided not only to estimate mean (m), standard deviation (d), skewness (v), but also to evaluate higher order statistics (such as kurtosis (k) and entropy (h)) in order to find out more correlated discriminating features. Hence, for each EEG sub-band, mean (m), standard deviation (d), skewness (v), kurtosis (k) and entropy (h), are measured. The TF features vector is the input of conventional ML (Multi-Layer Perceptron, Support Vector Machine, Logistic Regression) as well as DL (Autoencoder) based architectures to perform EEG-epoch classifications tasks: AD vs HC and AD vs MCI. The rest of the paper is organized as follows: in Section II the proposed procedure is introduced. In Section III and Section IV the EEG data pre-processing and features extraction are described, respectively. Section V presents the developed classification architectures. Section VI reports the simulation results. Section VII addresses the discussion and conclusions.

II. Methodology

The procedure of the proposed method is shown in Figure 1 and summarized as follows:

1) *n-channels EEG recording and pre-processing* (n=19 electrodes): the e^{th} EEG signal is filtered from artifactual components through clinical (visual) inspection and partitioned into s non-overlapping 5-seconds segments (i.e. *epochs*);
2) *Time-Frequency analysis*: given the EEG segment *s*, time-frequency analysis is performed by Continuous Wavelet Function (CWT) estimation (with Mexican hat function as wavelet mother function). Specifically, a time-frequency map (TFM) is estimated for each EEG channel, coming up with 19 TFM per epoch. Then, the average TFM (aTFM) is evaluated;
3) *features extraction*: the aTFM (of the epoch *s*) is split into five sub-matrices corresponding to the five main EEG sub-bands: delta [0.5-4 Hz], theta [4-8 Hz], alpha$_1$ [8-10 Hz], alpha$_2$ [10-13 Hz], beta [13-32 Hz]. Afterwards, given the aTFM sub-matrix of the EEG band under analysis, five CWT features are extracted: mean (m), standard deviation (d), skewness (v), kurtosis (k), entropy (h). Hence, 5 (# sub-bands) x 5 (# features) = 25 CWT features are extracted for each EEG epoch and used as input to the proposed classifiers;
4) *classification*: four machine learning techniques are employed to classify EEG epochs as belonging to AD, MCI or HC subject. Specifically, an Autoencoder (AE), a Muli-Layer Perceptron (MLP), a Logistic Regression

(LR) and a Support Vector Machine (SVM) based classifiers are proposed to perform the following binaries classifications: AD vs HC and AD vs MCI.

III. EEG RECORDING AND PREPROCESSING

Study population. A cohort of 189 individuals were recruited at IRCCS Centro Neurolesi Bonino-Pulejo of Messina (Italy). Specifically, it consists of 63 patients diagnosed with AD, 63 patients diagnosed with MCI and 63 HC subjects. The study was developed in fulfillment of a clinical protocol approved by the local Ethical Committee of IRCCS Centro Neurolesi Bonino-Pulejo. The guidelines of the Diagnostic and Statistical Manual of Mental Disorders (fifth edition, DSM-5) [20] were followed in order to formulate the diagnosis of AD or MCI and and the following exclusion criteria were defined: existence of other psychiatric or neural disorders that may lead to cognitive impairment, complex or unstable systemic disorders, epileptic EEG activity, traumatic brain damages. All the patients and caregivers signed a consent agreement and were informed about the aims, benefits and possible risks of this study. Furthermore, each subject was evaluated through neuroradiological examinations, in order to rule out other possible pathological conditions that may cause similar clinical features to AD, like lesions due to stroke, tumors or traumatic brain damage. The possible effects due to any medical therapy like cholinesterase inhibitors (ChEis), Memantine, anti-depressants, anti-psychotics and anti-epileptic drugs, were taken into account in AD patients; whereas, MCI patients were not under any medical treatment.

EEG recording. The EEG recordings were acquired in the morning. All the necessary information about the last meal and the quality and duration of the last sleep were collected by interviewing the patients and her/his caregivers. The EEG was recorded in accordance with the conventional 10-20 International System: the standard 19-channels montage (Figure 2) is adopted (which means that 19 EEG signals are simultaneously recorded). The sampling frequency was set at 1024 Hz, a notch filter was applied at 50 Hz. Signals were referenced to linked earlobe (A1-A2). During EEG acquisition, subjects were sitting comfortably and keeping their eyes closed. The subjects remained awake, as confirmed by the EEG experts who continuously monitored the EEG traces in order to promptly detect any possible pattern of sleep activity. The acquired EEG recordings were later manually reviewed by the experts in order to cancel the segments affected by artifacts.

EEG preprocessing. Every EEG signal was firstly band-pass-filtered between 0.5 and 32 Hz by using the *eegfiltfft* Matlab function of the *EEGLab* toolobox [21]. Afterwards, it was downsampled to a sampling frequency f_s=256 Hz and segmented into s non-overlapping EEG epochs sized l=5 seconds. Hence, every EEG epoch consisted of L = 256 x 5 = 1280 samples. For each patient under examination, n x L EEG epochs were stored on a computer and processed one by one.

IV. FEATURES EXTRACTION

Time-Frequency analysis. A time-frequency analysis was carried out on every EEG signal using the Continuous Wavelet Transform (CWT) [23]. A wavelet is a zero-mean function of finite duration which is localized both in time and frequency. The CWT of a given EEG time series x(t) is defined as:

$$CWT(\alpha,\beta) = \frac{1}{\sqrt{\alpha}} \int x(t)\psi^*(\frac{t-\beta}{\alpha}) \tag{1}$$

where α is the scale parameter (or dilatation/compression coefficient), β is the translation parameter (or shifting coefficient), ψ is the wavelet basis function known as *mother wavelet*, * denotes the complex conjugate operator and $CWT(\alpha,\beta)$ are the wavelet coefficients. The square modulus of the CWT coefficients ($|CWT(\alpha,\beta)|^2$) represents the energy distribution in the time-scale (or time-frequency) plane. Indeed, an approximate relationship between scale and frequency exists: high scale values correspond to low frequencies and vice-versa:

$$f_\alpha = \frac{f_c}{T_s\alpha} \tag{2}$$

where f_c is the center frequency of the chosen mother wavelet, T_s is the sampling period and f_α is called *pseudo-frequency* associated with the scale α. In this study, the Mexican hat mother wavelet function was used, because it is suitable for EEG signals analysis [24].

CWT features. Given an EEG epoch (n x L, with n=19 channels, L=1280 samples), the CWT is evaluated for each channel, according to eq. 1, coming up with 19 TFM (each TFM is sized f x L, where f = 253, is the number of frequencies in the range [0.5-32 Hz]). Then, the mean over the 19 f x L TFM is calculated, coming up with a single averaged TFM (aTFM) of the same size. Each aTFM is partitioned into five sub-matrices corresponding to the five sub-bands analyzed (delta [0.5-4 Hz], theta [4-8 Hz], alpha$_1$ [8-10 Hz], alpha$_2$ [10-13 Hz], beta [13-32 Hz]). For each sub-matrix (i.e. sub-band) the following CWT features are estimated: mean (m), standard deviation (d), skewness (v), kurtosis (k), entropy (h). The resulting CWT vector has size 5 (# sub-bands) x 5 (# features) = 25 features. Hence, for each EEG epoch, the corresponding CWT feature vector (sized 1 x 25) is extracted and used as input to the proposed classification architectures.

V. CLASSIFICATION

Given a s^{th} EEG epoch, in order to classify it as AD, MCI or HC, different ML classifiers are proposed. Specifically, the following most common ML architectures are developed to perform the 2-ways classifications tasks: AD vs HC and AD vs MCI.

1) AE classifier: Figure 3 (a) illustrates a typical AE architecture. It compresses the input representation (τ) into a lower dimensional space (*encoder* operation):

$$\xi = \rho_1(\tau\Sigma + b) \tag{3}$$

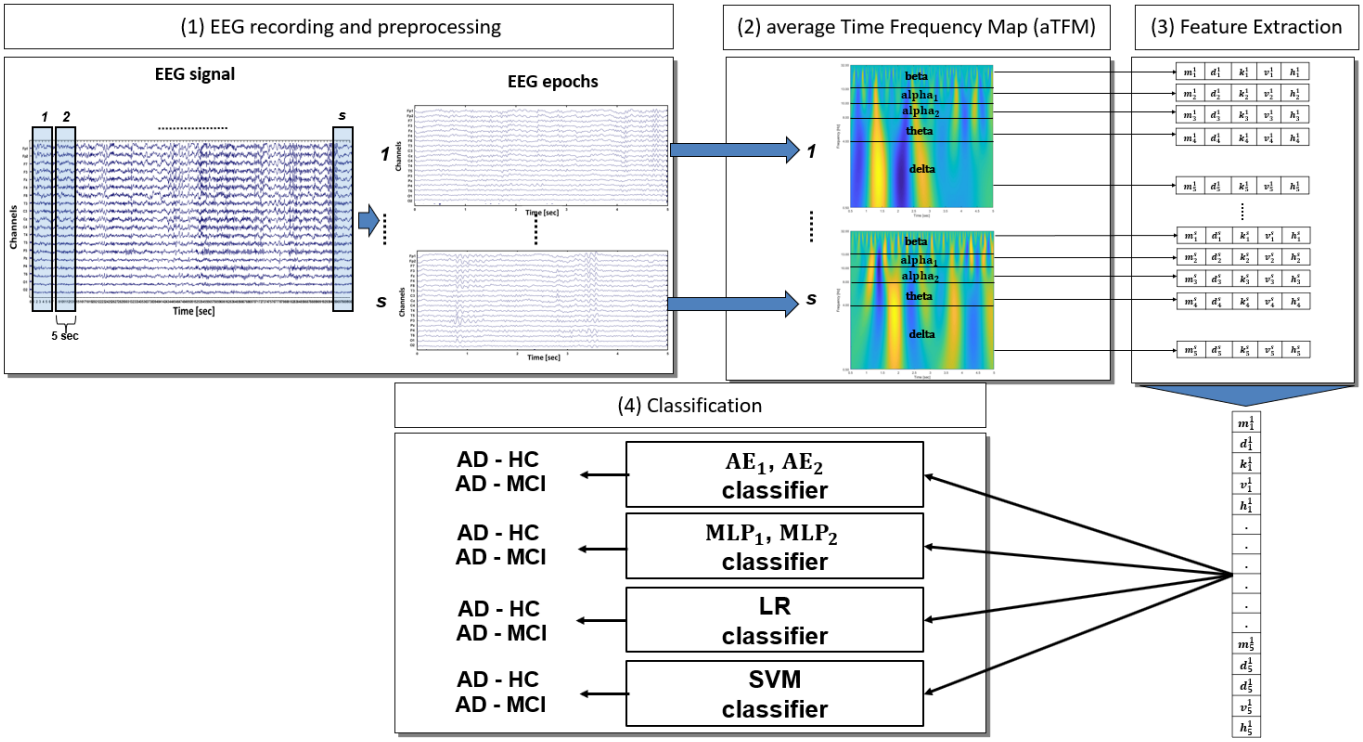

Fig. 1. Flowchart of the proposed method.

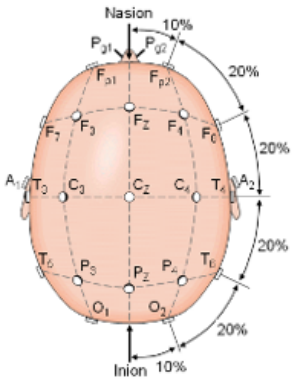

Fig. 2. 19-EEG channels according to the 10- 20 International System: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz and Pz. Odd and even numbers are related to the left and right hemispheres, respectively [22].

where Σ, b and ρ_1 are the weight matrix, the bias vector and the transfer function for the encoder stage. Then, it attempts to reproduce the input representation τ from the compressed vector ξ (*decoder* operation):

$$\hat{\tau} = \rho_2(\xi\Sigma^T + b) \tag{4}$$

where ρ_2 denotes the transfer function for the decoder and $\hat{\tau}$ is the reconstructed vector so that $\hat{\tau} \approx \tau$ [25]. Here, the *saturating linear transfer function* ($\rho_1(q)$ = 0 when q $\leq$ 0, $\rho_1(q) = q$ when $0 < q < 1$ and $\rho_1(q)$ = 0 when $\rho_1(q) \geq 1$) is used for the encoder and the *linear transfer function* ($\rho_2(q) = q$) for the decoder. Furthermore, the reconstruction error between τ and $\hat{\tau}$ is minimized through the stochastic gradient descent (SGD) algorithm and measured with the mean square error (MSE). In this study, the extracted 25 CWT features are the input of 1-hidden layer AE (AE_1, with 18 hidden units) and 2-hidden layers AE (AE_2, with 18 and 10 hidden units, respectively). The AE_1 compresses the input from 25 into 18-dimensional significant features used as input of a softmax layer (SF) trained in supervised fashion to perform the binary classifications tasks. The whole network (AE_1[25:10] + SF) is then trained with supervised learning to increase the classification performance (*fine tuning* technique). In AE_2, instead, the 25-dimensional input vector is firstly compressed into 18-dimensional features (as AE_1) and then into 10-dimensional significant parameters. At this stage, the 10-dimensional features vector is the input of SF. Then, the whole network (AE_2 [25:18:10] + SF) is fine tuned. It is worth to be noted that the AE_1 and AE_2 classifiers are trained until the cross-entropy loss function [26] converges, that is for $4*10^2$ epochs. Moreover, the topology of the developed AE_1 and AE_2, including hidden layer size and activation function, were set empirically after several tests and more specifically by measuring the minimum reconstruction error, that was $3.2*10^{-3}$ and $8.4*10^{-4}$, respectively.

2) MLP classifier: MLP is a widely used feed-forward neural network, trained through the standard backpropagation algorithm [27]. It consists of one input layer, one or more hidden layers and one output layer. In this study, two MLP classifiers are developed: MLP_1 and MLP_2. Specifically, for

fair comparisons, MLP_1 and MLP_2 have the same architecture of AE_1 and AE_2 classifiers. Indeed, MLP_1 is composed of 1-hidden layer with 18 hidden units; whereas MLP_2 is composed of 2-hidden layers with 18 and 10 hidden units, respectively. Both MLP_1 and MLP_2 are followed by SF output layer to perform the 2-ways classifications. Similarly to AEs, the hidden neurons of MLP networks use the *saturating linear transfer function* and are trained for $4*10^2$ epochs.

3) LR classifier: LR is a statistical method used to predict the likelihood of an event (i.e. presence of disease (AD) or healthy state) using a linear combination of independent variables (κ) [28]. The logistic function is represented by the sigmoid function defined as follow: $\sigma(\kappa) = \frac{1}{1+e^{-\kappa}}$. Hence, the LR classifier outputs the probability value P (ranged between 0 and 1) as belonging to the C class (i.e. AD, MCI or HC). In this study, two binaries classification problems are addressed: AD vs HC and AD vs MCI classification.

4) SVM classifier: SVM technique belongs to the statistical learning theory [29]. SVM finds the best hyperplane that provides the maximum separation between classes. Different kernel can be used in a SVM algorithm; however, simple linear kernel is applied in this study. Detailed mathematical formulation of SVM technique is reported in [30].

VI. Results

In this study, a dataset of 189 EEG signals (63 EEG of people with AD, 63 EEG of people with MCI, 63 EEG of people with no neural deficits (HC)) was used. Given the e^{th} EEG signal, it was firstly partitioned into *s* EEG epochs and then time-frequency transformed into *s* aTFM (where *s* depends on the EEG length). An overall of 13950 aTFM was obtained. Afterwards, given an aTFM under analysis, 25 CWT features were extracted (according to the procedure described in Section IV) and used as input of the proposed AE, MLP, LR, SVM classifiers to classify the s^{th} EEG epoch as belonging to the AD, MCI or HC group (*epoch-classification*). In this study, the binary epoch-classifications AD vs HC and AD vs MCI was carried out. The classification performances were evaluated using the following standard metrics (precision (*PC*), recall (*RC*), F-measure (*FM*) and accuracy (*ACC*):

$$PC = \frac{TP}{TP + FP} \tag{5}$$

$$RC = \frac{TP}{TP + FN} \tag{6}$$

$$FM = 2 * \frac{PC * RC}{PC + RC} \tag{7}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

where TP, TN, FP, FN stand for true positive, true negative, false positive and false negative, respectively [31]. The *10*-fold cross validation technique was applied: for each class, 70% of epochs were used as train set and, 30% as test set. Hence, all the results are expressed in terms of mean value $\pm$ standard deviation (i.e. mean(ACC) $\pm$ std(ACC)).

1) AD vs HC: Table I reports the AD vs HC classification performances in terms of PR, RC, FM and ACC. As can be seen, MLP_1 classifier outperformed all the other networks. Indeed, MLP_1 reported PR of 92.96% $\pm$ 0.0105, RC of 92.72% $\pm$ 0.0166 and consequently FM 92.84% $\pm$ 0.0099; MLP_2 reported FM value of 92,32% $\pm$ 0.0074; the AE_1 and AE_2 classifiers achieved values of 85.93% $\pm$ 0.0087 and 86.9% $\pm$ 0,0084, respectively; whereas, LR and SVM, FM scores of 81,19% $\pm$ 0.010 and 79,6% $\pm$ 0.0077. The MLP_1 achieved also the highest ACC of 95.76 % $\pm$ 0.0045, as compared with MLP_2 (95.45% $\pm$ 0,0057 ACC), AE_1 (91.84% $\pm$ 0.0051 ACC), AE_2 (92.35% $\pm$ 0.0047 ACC), LR (89.25% $\pm$ 0.0054 ACC) and SVM (88.48% $\pm$ 0.0046 ACC) classifiers. Furthermore, in order to assess and confirm the ability of the network to accurately detect AD and HC EEG epochs, the area under the ROC curve (AUC) was also estimated. Specifically, Figure 5 (a) reports the ROC curves and AUC values of each discrimination technique. The best performance was observed with the MLP_1 classifier (AUC score of 0.990 $\pm$ 0.002). However, comparable results was also achieved with MLP_2 classifier (AUC score of 0.989 $\pm$ 0.002).

2) AD vs MCI: Table II reports the AD vs MCI classification performances in terms of PR, RC, FM and ACC. The AE_1 and AE_2 classifiers achieved FM values of 64.88% $\pm$ 0.0145 and 66.30% $\pm$ 0.0141, respectively; whereas, the MLP_1 and MLP_2 classifiers reported values of 81.70% $\pm$ 0.0130 and 78.73% $\pm$ 0.0142, respectively. As regards LR and SVM classifiers, they achieved acceptable performance in terms of PR (64.43% $\pm$ 0.0121 and 64.21% $\pm$ 0.0158, respectively) but low values of RC (and consequently of FM). Hence, comparative results showed that the MLP_1 classifier outperformed all the others architectures developed. Such outcome was observed also in terms of ACC. Indeed, the MLP_1 achieved the highest accuracy of 86.84% $\pm$ 0.0098, as compared with MLP_2 (84.70% $\pm$ 0.0109 ACC), AE_1 (76.60% $\pm$ 0.0078 ACC), AE_2 (77.45% $\pm$ 0.0080 ACC), LR (71.87% $\pm$ 0.0069 ACC) and SVM (70.48% $\pm$ 0.0069 ACC) classifiers. Similarly to AD vs HC classifcation, the ROC curves and AUC values ware estimated. As can be seen in Figure 5 (b) , also in this scenario, MLP_1 and MLP_2 classifiers achieved the optimal AUC scores: 0.937 $\pm$ 0.007 and 0.923 $\pm$ 0.007, respectively.

TABLE I
AD vs HC epoch-classification performances (Recall, Precision, F-measure, Accuracy) of the proposed classifiers, measured on the test sets. Results are expressed as mean value $\pm$ standard deviation.

Method	**AD vs HC**			
	Recall	Precision	F-measure	Accuracy
AE_1	84.07% $\pm$ 0.0147	87.87% $\pm$ 0.0149	85.93% $\pm$ 0.0087	91.84% $\pm$ 0.0051
AE_2	85.64% $\pm$ 0.0123	88.20% $\pm$ 0.0091	86.90% $\pm$ 0.0084	92.35% $\pm$ 0.0047
MLP_1	92.72% $\pm$ 0.0166	92.96% $\pm$ 0.0105	92.84% $\pm$ 0.0099	95.76% $\pm$ 0.0057
MLP_2	92.40% $\pm$ 0.0089	92.24% $\pm$ 0.0111	92.32% $\pm$ 0.0074	95.45% $\pm$ 0.0045
LR	78.30% $\pm$ 0.0155	84.31% $\pm$ 0.0118	81.19% $\pm$ 0.0010	89.25% $\pm$ 0.0054
SVM	75.85% $\pm$ 0.0109	83.75% $\pm$ 0.0133	79.60% $\pm$ 0.0077	88.48% $\pm$ 0.0046

VII. Discussion and Conclusion

In this research we exploited the potential of machine learning (ML) algorithms to differentiate EEG segments (i.e.

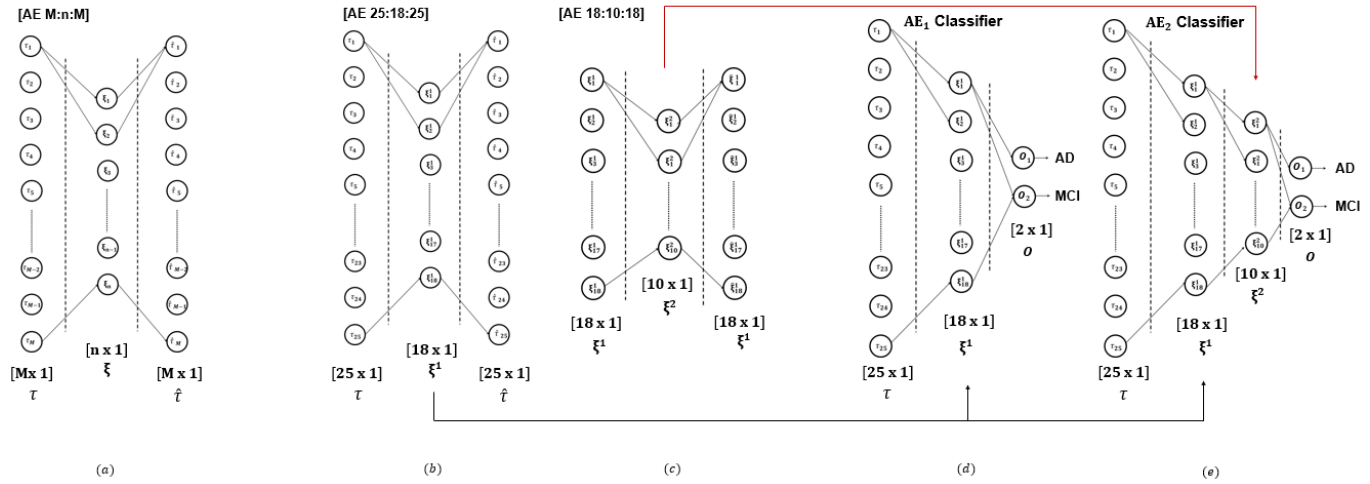

Fig. 3. (a) Standard AE configuration. The input τ is typically compressed into a lower dimensional features vector ξ. Then, ξ is used to reconstruct the original input vector, so that $\hat{\tau} \approx \tau$. (b) AE_1 architecture [AE 25:18:25]. It compresses the 25-dimensional CWT features (τ) into a 18-dimensional features vector (ξ_1). Then, it attempts to produce the same input representation $\hat{\tau}$. (c) AE_2 architecture [AE 18:10:18]. The compressed features vector ξ_1 (of AE_1) is used as input of a second AE model that reduces the input size into a 10-dimensional features vector (ξ_2) (d) AE_1 classifier architecture. It uses the 18-dimensional features vector (ξ_1) followed by a softmax layer (o) to perform the binaries classification tasks. (e) AE_2 classifier architecture. It uses the 18-dimensional features vector (ξ_1) and 10-dimensional features vector (ξ_2). The last layer is a softmax layer (o) for binaries classification tasks. Finally, the whole architectures (AE_1, AE_2 classifiers) are fine-tuned using conventional back propagation algorithm. In the figure, the classifiers refer to AD vs MCI classification task.

TABLE II
AD VS MCI EPOCH-CLASSIFICATION PERFORMANCES (RECALL, PRECISION, F-MEASURE, ACCURACY) OF THE PROPOSED CLASSIFIERS, MEASURED ON THE TEST SETS . RESULTS ARE EXPRESSED AS MEAN VALUE ± STANDARD DEVIATION.

Method	AD vs MCI			
	Recall	Precision	F-measure	Accuracy
AE_1	59.53% ± 0.0199	71.28% ± 0.0110	64.88% ± 0.0145	76.59% ± 0.0078
AE_2	61.10% ± 0.0200	72.47% ± 0.0146	66.30% ± 0.0141	77.44% ± 0.008
MLP_1	80.89% ± 0.0166	82.52% ± 0.0188	81.70% ± 0.0130	86.84% ± 0.0098
MLP_2	77.95% ± 0.0158	79.51% ± 0.0197	78.73% ± 0.0142	84.69% ± 0.0109
LR	50.31% ± 0.0147	64.43% ± 0.0121	56.50% ± 0.0129	71.86% ± 0.0069
SVM	42.25% ± 0.0134	64.21% ± 0.0158	50.97% ± 0.0134	70.47% ± 0.0069

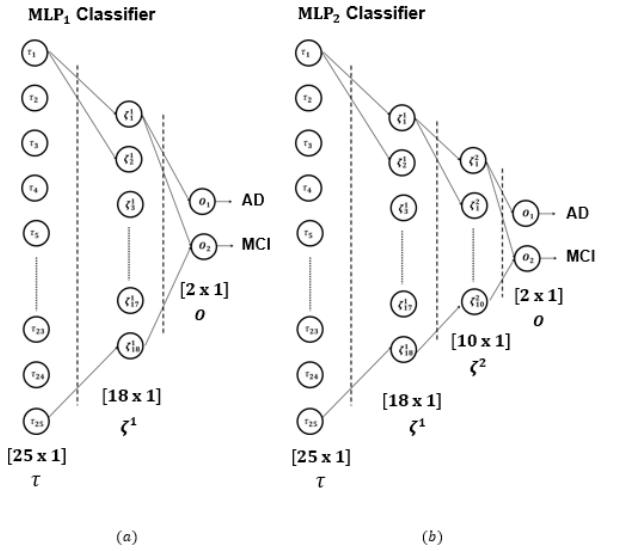

Fig. 4. (a) MLP_1 classifier architecture. It has 1-hidden layer of 18 hidden neurons followed by softmax layer for binaries classification tasks. (b) MLP_2 classifier architecture. It has 2-hidden layers of 18 and 10 hidden neurons respecitvely. The last layer is a sofmax output layer to perform the 2-way classifications. In the figures, the MLP_1 and MLP_2 classifiers refer to AD vs MCI classification task.

epochs) of patients affected by neurological disorders (i.e. AD, MCI) and healthy people (HC). Specifically, a data driven ML framework based on time-frequency representation of EEG recording was proposed. Here, the Continuous Wavelet Transform (CWT) was used and 25 CWT features were evaluated for each EEG epoch. The extracted CWT features vector was the input of the proposed deep (AE_1, AE_2) and shallow (MLP_1, MLP_2, LR, SVM) classifiers to perform the 2-ways EEG-epoch-classifications: AD vs HC and AD vs MCI. Experimental results showed that the MLP with a single hidden layer classifier (MLP_1) achieved better classification performance as compared with MLP_2, LR, SVM and AE_1, AE_2 classfiers. Indeed, MLP_1 reported the highest accuracy values of 95.76 % ± 0.0045 and 86.84% ± 0.0098 in AD vs HC and AD vs MCI classification, respectively. result was also statistically confirmed by the analysis of ROC curves. As shown in Figure 5, MLP_1 achieved the highest AUC of 0.99±0.002 (AD vs HC) and 0.937±0.007 (AD vs MCI). It is worth mentioning that, as discussed in Section V, even if the AEs and MLPs had similar topologies and training/learning parameters (such as number of epochs, 10^2), it was observed that the averaged training times (performed on a high performance NVIDIA GPU Tesla K40c) were 66.2 s and 30.5 s with AE and MLP, respectively. Furthermore, deep neural networks usually outperform conventional machine learning models. In this study, instead, standard MLPs reported better performance

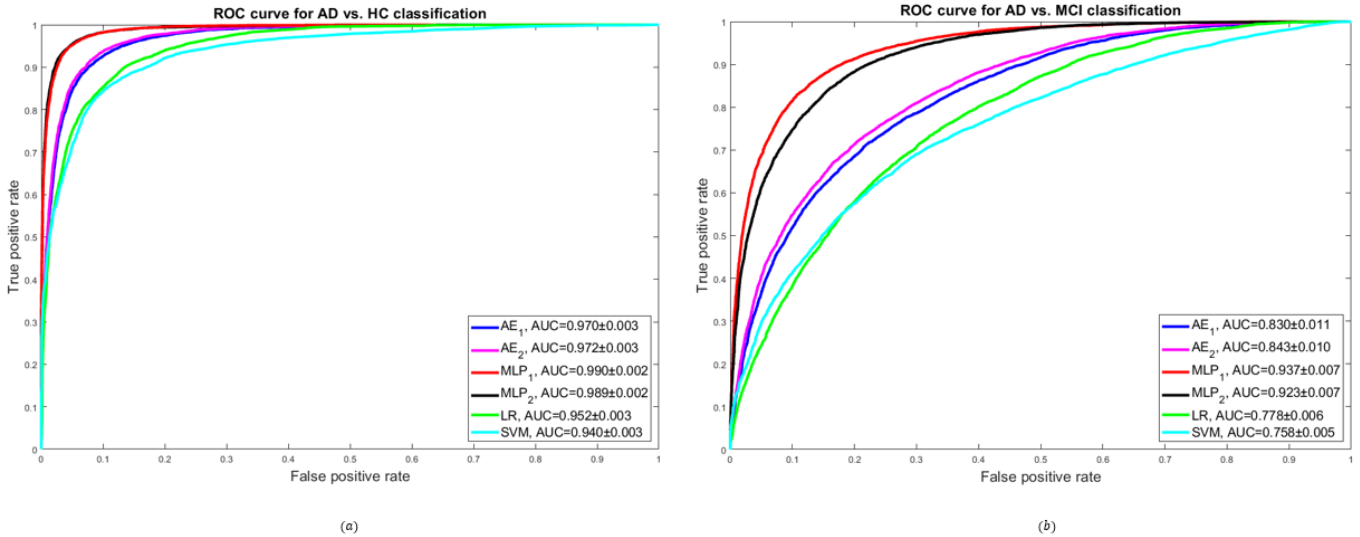

Fig. 5. ROC curves and AUC values of the proposed machine learning classification systems (AE_1, AE_2, MLP_1, MLP_2, LR, SVM) for AD vs HC classification (a) and AD vs MCI classification (b).

in detecting AD, MCI or HC EEG epochs, than AEs that are topologically shallow but fall into the class of deep learning techniques. This is possible due to the fact that AE architectures are generally employed to reduce redundancy of high dimensional input data. Here, an input vector of only 25 features was used. This means that the developed autoencoders caused over compression of features and consequently loss of significant information. In addition, it is to be noted that it is difficult to compare our results with other studies that used different databases. Hence, for fair comparison, we compared the results of the proposed classifier with a recent previous work where the EEG epoch-based classification of AD, MCI, HC subjects was performed and the same dataset was used. Specifically, in [16], the proposed CNN_1 classifier (trained with spectral images of EEG segments) achieved epoch-classification performances of 92.95% (ACC) and 84.62% (ACC) in AD vs HC and AD vs MCI, respectively; whereas, here, better ACC values (95.76 % $\pm$ 0.0045 (AD vs HC) and 86.84% $\pm$ 0.0098 (AD vs MCI)) have been reported even using the 10-fold cross validation technique. Although, the CNNs avoid hand-designed features extraction and let the procedure faster than traditional approaches, we focused on the ability of networks to correctly classify AD/MCI. Indeed, the proposed feature extraction procedure and the developed ML system achieved very good classification accuracy.

Nevertheless, the proposed processor has some limitations. One of the main limitation is related to the different stages of AD and MCI included in the dataset analyzed. For example, a severe MCI might have time-frequency features more similar to AD than to mild MCI. This causes a misclassifation of the EEG epoch under analysis and consequently a reduction of classification performances. A second limitation is that different epochs of the same patient may exist both in train and test sets. For this reason we decided not to perform the overall patient-classification. A third limitation is that we only developed the binaries classifications (AD vs HC, AD vs MCI). However, this work has to be considered as a preliminary study. Indeed, in the future, the multiple classification (AD vs MCI vs HC) will be also taken into account. In addition, since the time-frequency analysis showed promising results in AD/MCI/HC EEG epoch detection, in the future, an overall patient's classification will be performed. Specifically, all the epochs belonging to the AD/MCI/HC subject (not used as train examples) will be the input of the trained classifier and the maximum number of epochs classified as class *C* (i.e. AD, MCI or HC), will indicate the class to which the subject belongs.

Acknowledgment

This work was supported by the Italian Ministry of Health, grant no.: GR-2011-02351397.

References

[1] W. H. Organization *et al.*, "Meeting on the implementation of the global action plan of the public health response on dementia 2017-2025: meeting report: 11-12 december 2017, world health organization, geneva, switzerland," tech. rep., World Health Organization, 2018.

[2] Alzheimer's Association *et al.*, "2018 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 14, no. 3, pp. 367–429, 2018.

[3] S. Gauthier, B. Reisberg, M. Zaudig, R. C. Petersen, K. Ritchie, K. Broich, S. Belleville, H. Brodaty, D. Bennett, H. Chertkow, *et al.*, "Mild Cognitive Impairment," *The Lancet*, vol. 367, no. 9518, pp. 1262–1270, 2006.

[4] J. Jeong, "EEG dynamics in patients with Alzheimer's disease," *Clinical neurophysiology*, vol. 115, no. 7, pp. 1490–1505, 2004.

[5] 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.

[6] T. König, L. Prichep, T. Dierks, D. Hubl, L. Wahlund, E. John, and V. Jelic, "Decreased EEG synchronization in Alzheimers disease and Mild Cognitive Impairment," *Neurobiology of aging*, vol. 26, no. 2, pp. 165–171, 2005.

[7] N. Mammone, C. Ieracitano, H. Adeli, A. Bramanti, and F. C. Morabito, "Permutation jaccard distance-based hierarchical clustering to estimate EEG network density modifications in mci subjects," *IEEE Transactions on Neural Networks and Learning Systems*, no. 99, pp. 1–14, 2018.

[8] N. Mammone, S. De Salvo, L. Bonanno, C. Ieracitano, S. Marino, A. Marra, A. Bramanti, and F. C. Morabito, "Brain network analysis of compressive sensed high-density EEG signals in AD and MCI subjects," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 1, pp. 527–536, 2019.

[9] H. Soininen, J. Partanen, V. Laulumaa, E.-L. Helkala, M. Laakso, and P. Riekkinen, "Longitudinal EEG spectral analysis in early stage of Alzheimer's disease," *Clinical Neurophysiology*, vol. 72, no. 4, pp. 290–297, 1989.

[10] 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.

[11] 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.

[12] L. Trambaiolli, N. Spolaôr, A. Lorena, R. Anghinah, and J. Sato, "Feature selection before EEG classification supports the diagnosis of Alzheimer's disease," *Clinical Neurophysiology*, vol. 128, no. 10, pp. 2058–2067, 2017.

[13] S. J. Ruiz-Gómez, C. Gómez, J. Poza, G. C. Gutiérrez-Tobal, M. A. Tola-Arribas, M. Cano, and R. Hornero, "Automated multiclass classification of spontaneous EEG activity in Alzheimer's disease and Mild Cognitive Impairment," *Entropy*, vol. 20, no. 1, p. 35, 2018.

[14] 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 Alzheimer's patients classification," *BMC medical informatics and decision making*, vol. 18, no. 1, p. 35, 2018.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[16] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, and F. C. Morabito, "A convolutional neural network approach for classification of dementia stages based on 2d-spectral representation of EEG recordings," *Neurocomputing*, vol. 323, pp. 96–107, 2019.

[17] F. C. Morabito, M. Campolo, N. Mammone, M. Versaci, S. Franceschetti, F. Tagliavini, V. Sofia, D. Fatuzzo, A. Gambardella, A. Labate, *et al.*, "Deep learning representation from electroencephalography of early-stage creutzfeldt-jakob disease and features for differentiation from rapidly progressive dementia," *International journal of neural systems*, vol. 27, no. 02, p. 1650039, 2017.

[18] F. C. Morabito, M. Campolo, C. Ieracitano, J. M. Ebadi, L. Bonanno, A. Bramanti, S. Desalvo, N. Mammone, and P. Bramanti, "Deep convolutional neural networks for classification of Mild Cognitive Impaired and Alzheimer's disease patients from scalp EEG recordings," in *Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), 2016 IEEE 2nd International Forum on*, pp. 1–6, IEEE, 2016.

[19] S. Gasparini, M. Campolo, C. Ieracitano, N. Mammone, E. Ferlazzo, C. Sueri, G. G. Tripodi, U. Aguglia, and F. C. Morabito, "Information theoretic-based interpretation of a deep neural network approach in diagnosing psychogenic non-epileptic seizures," *Entropy*, vol. 20, no. 2, p. 43, 2018.

[20] A. P. Association, ed., *Diagnostic and statistical manual of mental disorders (5th ed.)*. 2013.

[21] 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.

[22] J. Malmivuo and R. Plonsey, *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*. Oxford University Press, USA, 1995.

[23] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.

[24] O. Faust, U. R. Acharya, H. Adeli, and A. Adeli, "Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis," *Seizure*, vol. 26, pp. 56–64, 2015.

[25] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[26] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Annals of operations research*, vol. 134, no. 1, pp. 19–67, 2005.

[27] D. W. Patterson, *Artificial neural networks: theory and applications*. Prentice Hall PTR, 1998.

[28] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*, vol. 398. John Wiley & Sons, 2013.

[29] O. Bousquet, S. Boucheron, and G. Lugosi, "Introduction to statistical learning theory," in *Advanced lectures on machine learning*, pp. 169–207, Springer, 2004.

[30] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[31] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.