A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia

A Novel Multi-modal Machine Learning based Approach for Automatic Classification of EEG recordings in Dementia

Cosimo Ieracitano[a,*], Nadia Mammone[b], Amir Hussain[c], Francesco C. Morabito[a]

[a] *DICEAM, University Mediterranea of Reggio Calabria, Via Graziella, Feo di Vito, 89060 Reggio Calabria (Italy)*
[b] *IRCCS Centro Neurolesi Bonino-Pulejo, Via Palermo c/da Casazza, SS. 113 98124, Messina (Italy)*
[c] *School of Computing, Edinburgh Napier University, Edinburgh EH10 5DT, Scotland, U.K.*

Abstract

Electroencephalografic (EEG) recordings generate an electrical map of the human brain that are useful for clinical inspection of patients and in biomedical smart Internet-of-Things (IoT) and Brain-Computer Interface (BCI) applications. From a signal processing perspective, EEGs yield a nonlinear and nonstationary, multivariate representation of the underlying neural circuitry interactions. In this paper, a novel multi-modal Machine Learning (ML) based approach is proposed to integrate EEG engineered features for automatic classification of brain states. EEGs are acquired from neurological patients with Mild Cognitive Impairment (MCI) or Alzheimer's disease (AD) and the aim is to discriminate Healthy Control (HC) subjects from patients. Specifically, in order to effectively cope with nonstationarities, 19-channels EEG signals are projected into the time-frequency (TF) domain by means of the Continuous Wavelet Transform (CWT) and a set of appropriate features (denoted as *CWT features*) are extracted from δ, θ, α_1, α_2, β EEG sub-bands. Furthermore, to exploit nonlinear phase-coupling information of EEG signals, higher order statistics (HOS) are extracted from the bispectrum (BiS) representation. BiS generates a second set of features (denoted as *BiS features*) which are also evaluated in the five EEG sub-bands. The CWT and BiS features are fed into a number of ML classifiers to perform both 2-way (AD vs. HC, AD vs. MCI, MCI vs. HC) and 3-way (AD vs. MCI vs. HC) classifications. As an experimental benchmark, a balanced EEG dataset that includes 63 AD, 63 MCI and 63 HC is analyzed. Comparative results show that when the concatenation of CWT and BiS features (denoted as *multi-modal (CWT+BiS) features*) is used as in-

*Corresponding author
Email addresses: cosimo.ieracitano@unirc.it (Cosimo Ieracitano), nadia.mammone@irccsme.it (Nadia Mammone), a.hussain@napier.ac.uk (Amir Hussain), morabito@unirc.it (Francesco C. Morabito)

put, the Multi-Layer Perceptron (MLP) classifier outperforms all other models, specifically, the Autoencoder (AE), Logistic Regression (LR) and Support Vector Machine (SVM). Consequently, our proposed multi-modal ML scheme can be considered a viable alternative to state-of-the-art computationally intensive deep learning approaches.

Keywords: Machine Learning, Continuous Wavelet Transform, Bispectrum, Alzheimer's disease, Mild Cognitive Impairment, Data Fusion

1. Introduction

Dementia due to Alzheimer's disease (AD) is a chronic degenerative neuropathology that occurs mainly in elderly individuals. AD accounts for about 60% of overall dementia cases and causes a progressive impairment of cognitive abilities, typically linked to behavioral problems such as restlessness, depression and language dysfunctions. The preclinical, intermediate stage between healthy aging and AD is referred to as amnestic Mild Cognitive Impairment (MCI), where deficits of cognitive abilities may start to be observable without interfering with the everyday life of the individual [1]. When a MCI subject is actually affected by AD, he/she will develop dementia in around 7 years [2]. This transition is known as "MCI to AD conversion". It has been estimated that MCI to AD conversion rate is 10%-15% per year [3]. The detection of the disorder onset is a difficult multidimensional task for clinicians. Even today, the definitive diagnosis of AD is possible only post-mortem, by autopsy. Early diagnosis would enhance the quality of life of both patients and caregivers, and motivate their active participation in the treatment program, potentially reducing the death rate of AD. This would also help the researchers to recruit more participants, especially people at the earliest stages of the disease. The Electroencephalogram (EEG) is a widely used diagnostic tool as it is a non-expensive, non-invasive method endowed with a high temporal resolution. EEG reflects, at a macroscopic level, the superposition of electromagnetic fields generated by the interaction between cortical neurons. The behaviour of the underlying neuronal populations, can be studied indirectly through EEG [4]. Some abnormalities in MCI/AD patients are observed in EEG recordings in the form of slowing of EEG rhythms, loss of synchronization between pairs of electrodes and loss of complexity [5], [6], [7], [8], [9]. In this context, there is a growing body of research focused on extracting relevant features from MCI/AD EEG signals. Specifically, considering the evidence that AD causes slowing in EEG waves, most studies focused on the analysis of EEG spectrum which is also typically available in clinical practice to capture possible correlations between the progression of the disease and alteration of the spectral profile. The use of spectral representations can enable significant computational advantages in neural networks learning. However, conventional power spectral analysis estimates only the spectral content of the signals without taking into account the EEG dynamics: some relevant additional information can be captured through

a time-frequency (TF) approach. In addition, Fourier analysis does not capture any nonlinear high order interaction among EEG signal components (which are known to be inherently nonlinear) [4]. Synchronization can also occur locally in frequency or between different frequencies, in a more complex, nonlinear way [10]. Further, nonlinear changes in cortical generating processes induce alterations in the electrical signals collected at the scalp. Bispectral analysis has been recently shown to represent an alternative methodology to detect such changes [11]. This is because nonlinear interactions that generate the multivariate time series can be estimated by higher-order phase coupling [12].

In this work, a set of features estimated from both the Bispectrum (BiS) and the time-frequency representation extracted via the Continuous Wavelet Transform (CWT) are used to automatically classify EEGs from AD, MCI and healthy elderly controls (HC). These are denoted as BiS and CWT features, respectively. Considering the inherent non-stationary nature of EEGs, the proposed approach involves partitioning EEG signals into segments (i.e. epochs) and processing them segment by segment. Subsequently, for each epoch of the EEG, CWT and BiS features are extracted.

Machine Learning (ML) techniques have been widely applied as successful approaches to develop computer-aided systems in healthcare applications (e.g.[13], [14], [15], [16], [17]). In this study, the extracted CWT and BiS features are vectorized and used as multi-modal input to a ML system to discriminate EEG epochs belonging to AD, MCI or HC subjects. To classify the signals' epochs, four different standard ML classifiers are used, specifically the Autoencoder (AE), Multilayer Perceptron (MLP), Logistic Regression (LR), Support Vector Machine (SVM). These are trained over three different sets: 1) using only CWT features; 2) using only BiS features; 3) using a concatenation of CWT and BiS features vector (denoted as *multi-modal* (CWT+BiS) features vector). The classifiers were trained to perform both 2-way (AD vs. HC, AD vs. MCI, HC vs. MCI) and 3-way (AD vs. MCI vs. HC) classification tasks. Experimental results show that the 1-hidden layer MLP classifier outperforms other approaches in both 2-way and 3-way classification, when the multi-modal input features vector is employed. It is worth mentioning that the reason for using standard neural network models reflects the potential for further clinical acceptance and practical deployment of these models. Deep learning (DL) schemes are powerful approaches utilised to extract relevant knowledge on their own, but require a proportionally large dataset to learn good representations, typically more difficult to interpret with respect to models based on engineered features.

The main contributions of this paper can be summarized as follows:

- development of an original multi-modal features extraction methodology based on CWT and BiS analysis of EEG recordings, that can effectively account for nonstationarity and nonlinearity effects;
- development of a novel data-driven multi-modal (CWT+BiS) ML system for classifying AD, MCI and HC EEG recordings;

- formulation of a generic multi-modal ML-framework potentially applicable in clinical applications.

The rest of this paper is organized as follows: Section 2 gives a short literature review of correlated works. Section 3 describes materials including the available EEG recordings and their pre-processing, the proposed multi-modal methodology, and the CWT/BiS features extraction procedure. Specific implementation details of the proposed ML classification models are also introduced. Section 4 reports comparative experimental results. These are then discussed in Section 5. Finally, Section 6 concludes the paper and outlines some future research directions.

2. Related works

Several recent studies discuss classification of brain states from EEG signals. Most of these rely on feature extraction procedures. Some works are focused on discriminating AD and/or MCI recordings from healthy aging controls. In some works the extraction of features is carried out in the spectral domain.

For example, Lehmann et al. [18] measured the relative and absolute spectral power, power spectral distribution and spatial synchronization from EEGs of 45 HC, 116 mild-AD and 81 moderate AD. Different classification techniques were used. Specifically, a sensitivity of 85% and specificity of 78% were achieved in mild-AD vs. HC classification using a random forest based classifier; whereas, a sensitivity of 89% and specificity of 88% were achieved in moderate-AD vs. HC classification by using SVM and MLP based classifiers. McBride et al. [19] evaluated regional spectral and complexity based features from EEGs of 15 HC, 16 early MCI and 17 early AD. The proposed 3-way SVM-based classifier achieved classification accuracies of 83.3%, 85.4%, 79.2% for the eyes-opened resting state, eyes-closed counting task and eyes-closed resting state, respectively. In [20] the authors applied eight different feature selection methods to discriminate 22 AD and 12 HC. Experimental results showed classification accuracies of up to 91.18%, when a reduced subset of spectral features (extracted from the EEG sub-bands) where used as input to a SVM classifier, previously proposed in [21]. Ruiz-Gomez et al. [22] selected an optimal set of spectral and nonlinear features, extracted from 111 EEG signals (37 AD, 37 MCI, 37 HC), to train a linear discriminant analysis (LDA), a quadratic-DA and a MLP based classifier. The MLP classifier achieved the best performance with an accuracy of 78.43% (in HC vs. all) and 76.47% (in AD vs. all). Neto et al. [23] analyzed spectral features (such as absolute δ power and amplitude, center and dispersion of α power) of the EEGs of 114 HC, 114 AD and 114 vascular-AD patients. A binary-based regularized-LDA classifier produced a best accuracy of 77% in classifying HC vs. dementia patients (AD or VAD), using a reduced set of features. Kashefpoor et al [24] estimated 19 spectral features from each channel of 16 HC and 11 MCI EEG signals. After selecting the best discriminative features by means of a correlation-based approach, a neuro-fuzzy-k-nearest neighbor classifier (NF-kNN) was developed, achieving accuracy rates up to 88.89% in HC

vs. MCI classification. In [25] Kulkarni used different complexity based features (such as spectral entropy, spectral centroid, spectral roll-off, zero crossing rate) extracted from 100 EEG signals (50 AD and 50 HC) as input to a 2-way KNN classifier, reporting accuracy of 94% in AD vs. HC classification.
There are only a limited number of works that have exploited time-frequency or bispectrum based features as discriminating coefficients. For example, with regard to the time-frequency approach, Fiscon et al. [26] extracted spectral and time-frequency (discrete wavelet) based features from 109 EEG signals (49 AD, 37 MCI, 23 HC) and then trained a decision tree classifier. Experimental results showed that wavelet coefficients evaluated by applying the discrete wavelet transform achieved the highest accuracy rates: 83.3% for AD vs. HC; 91.7% for MCI vs. HC; 79.1% for AD vs. MCI. Kulkarni and Bairagi in [27] decomposed EEG signals into five EEG sub-bands (δ, θ, α, β, γ) using the wavelet decomposition technique. The means and variances of wavelet coefficients were evaluated and used as input to a SVM classifier, achieving accuracy value of 88% in AD vs. HC classification. Vialatte et al. [28] transformed 60 EEG recordings (22 MCI and 38 HC) into time-frequency representations using the complex Morlet wavelet function. Bump modeling for features generation was then applied, and selected features used as input to a MLP classifier, achieving very good performance (93% accuracy, 91% sensitivity, 94.7% specificity). Recently, Morabito et al. [29] proposed a time-frequency based AD, MCI, HC classification system. Specifically, time-frequency maps of 69 EEG recordings (23 AD, 23 MCI, 23 HC) were extcolorredestimated using the CWT (with Mexican hat wavelet function) and statistical parameters (mean, standard deviation, skewness) were extracted from three EEG roughly defined sub-bands. The features were fed into a DL based Convolutional Neural Network (CNN) architecture, achieving accuracy rates up of 85%, 85%, 78% in AD vs. HC, MCI vs. HC and AD vs. MCI classification, respectively.

Wang et al. [12] explored the nonlinear high-order information of bispectrum, and estimated both spectral (i.e., median frequency, spectral entropy, normalized spectral squared entropy) and bispectrum (i.e., the normalized bispectral entropy, the normalized bispectral squared entropy) features. These were input to a SVM classifier to differentiate the activity in the five brain areas (frontal, central-parietal, occipital, left and right temporal) of 14 AD and 14 HC, reporting accuracy rate up to 90.2%. Gomez et al. [11] studied the bispectrum of 39 EEG recordings (17 AD, 19 HC) and extracted five bispectral features (such as normalized bispectral entropy, normalized bispectral squared entropy and sum of logarithmic amplitudes of the bispectrum). A LR classifier was used to perform the 2-way AD vs. HC classification, achieving an accuracy of 86.11%.

The drawbacks of the aforementioned studies lay mainly in the limited size of the dataset or in the missing comparison between AD vs. MCI classification, the most challenging task, that is also of practical interest in clinical frameworks. In this paper, a novel multi-modal approach based on both time-frequency and bispectrum features extraction is proposed. The selected features are used as input to ML classifiers to perform the 2-way (AD vs. HC, AD vs. MCI, MCI vs.

HC) and 3-way (AD vs. MCI vs. HC) classifications. In addition, a relatively high number of subjects is recruited (63 AD, 63 MCI, 63 HC) to construct the EEG dataset.

3. Materials and Methodology

3.1. EEG recording and preprocessing

Study population. A cohort of 189 subjects was recruited at IRCCS Centro Neurolesi Bonino-Pulejo of Messina (Italy): 63 affected by AD, 63 affected by MCI, and 63 healthy controls. A protocol for the EEG-based differential diagnosis of AD/MCI was designed and approved by the local Ethics Committee of the IRCCS Centro Neurolesi Bonino-Pulejo. The diagnosis of AD or MCI was formulated following the guidelines of the Diagnostic and Statistical Manual of Mental Disorders (fifth edition, DSM-5) [30]. The inclusion criteria was the diagnosis of AD or MCI, whereas the exclusion criteria were: presence of neurological or psychiatric disorders that may induce cognitive impairment, complex systemic disorders, presence of epileptiform patterns in the EEGs, hydrocephalus, stroke, traumatic brain injuries or other neurological conditions. The possible exclusion from the study was subjected to the evaluation of a neuroimaging examination, which helped highlight the possible presence of any of the aforementioned pathological conditions, that may cause symptoms similar to AD or MCI. The goal of the present study was explained in detail to patients and their caregivers, who signed an informed consent form. In relation to AD patients, the possible effects of medical treatment like cholinesterase inhibitors (ChEis), memantine, anti-depressants, anti-psychotics and anti-epileptic drugs, were considered. In particular, the target dose of memantine, such as cholinesterase inhibitors (ChEis), was 20 mg/day. The dose of anti-depressants (citalopram) was set at 30 mg/day. No anti-psychotics or anti-epileptic drugs were used. MCI subjects were not under any medical treatments.

EEG recording. Before starting the EEG recording session, patients and their caregivers were asked about the quality of their last sleep and last meal. EEG clinicians reassured subjects about the total non-invasiveness of the examination and helped them feel at ease to undertake the EEG recording session in a relaxed way. The participants sat on a comfortable chair and kept their eyes closed throughout the EEG acquisition. Thus, EEG was recorded in an eye-closed relaxed resting state. Subjects remained awake, as confirmed by EEG experts who monitored the EEG traces in real-time to detect any possible sleep pattern. After acquisition, EEG traces were manually reviewed offline and any segments corrupted by artifacts were marked by experts and discarded from the analysis. The average EEG length, after artifact rejection, was 4.1 min. EEGs were recorded using the standard 19-channels montage according to the 10-20 International System (channels Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz and Pz) with linked earlobe (A1-A2) referencing. Sampling frequency was 1024 Hz and a notch filter was applied at 50 Hz.

EEG preprocessing. The n-channel EEG recordings (with n=19) were band-pass filtered between 0.5 Hz and 32 Hz using the open source Matlab toolbox *EEGLab* ([31]). EEG signals were down-sampled to 256Hz and segmented into M non-overlapping epochs of 5s duration. With f_s=256 Hz, each epoch included N= 5 x 256 = 1280 samples. Thus, for each subject under consideration, M epochs EEG^{ε} ($\varepsilon = 1, 2, .., M$), sized n x N, were stored to be further processed one by one.

3.2. Methodology

Our methodology flowchart is illustrated in our proposed multi-modal ML-based EEG classification framework in Figure 1. It includes the following processing sub-modules:

- *EEG pre-processing.* The 19-channel EEG signal is recorded and stored on a computer. Next, the artefactual patterns are removed through visual inspection by an expert operator. Finally, the EEG recording is segmented into E non-overlapping epochs of 5 seconds and then analyzed on an epoch by epoch basis.

- *Time-Frequency analysis.* CWT is computed to project every EEG signal into the time-frequency domain: given the E^{th} epoch under analysis, for each EEG channel the corresponding time-frequency representation (TFR) is estimated for a grand total of 19 TFR maps, i.e., one per channel (Figure 2 shows three sample maps extracted from AD, MCI and HC subjects). Next, the 19 TFR are averaged, resulting in a single average TFR (ATFR).

- *Bispectrum analysis.* The HOS analysis is carried out using the Bispectrum (BiS) estimation: given the E^{th} epoch under analysis, for each EEG channel the corresponding bispectrum representation (BiSR) is estimated, for a grand total of 19 BiSR (one per channel). Next, the 19 BiSR are averaged, producing a single average BiSR (ABiSR).

- *Features extraction.* Given the E^{th} epoch under analysis, two sets of features are extracted, one from the ATFR map and one from the ABiSR:

    1. *CWT features extraction*: the E^{th} ATFR is partitioned into five time-frequency sub-maps corresponding to the five EEG rhythms: δ [0.5-4 Hz], θ [4-8 Hz], α_1 [8-10 Hz], α_2 [10-13 Hz], β [13-32 Hz] (Figure 2 shows how a TFR map is partitioned into 5 sub-bands). Next, the following five CWT features are extracted in each EEG sub-band: mean (CWT_1), standard deviation (CWT_2), skewness (CWT_3), kurtosis (CWT_4) and entropy (CWT_5). As shown in Figure 2, the time-frequency maps can reveal a different distribution of the EEG signal energy in the analyzed sub-bands. Furthermore, the number of relevant CWT coefficients is differently located in the representation. To reduce the high redundancy of the CWT coefficients, we consider average quantities extracted from the time-frequency sub-maps.

2. *BiS features extraction*. Similar to the ATFR, the E^{th} ABiSR is partitioned into five bispectrum maps corresponding to the five aforementioned EEG rhythms (δ, θ, α_1, α_2, β). The following BiS features are then extracted in each EEG sub-band:
   - normalized bispectral entropy (BiS_1);
   - normalized bispectral squared entropy (BiS_2);
   - sum of logarithmic amplitudes of the bispectrum (BiS_3);
   - sum of logarithmic amplitudes of diagonal elements of the bispectrum (BiS_4);
   - first-order spectral moment of the amplitudes of diagonal elements of the bispectrum (BiS_5);
   - second-order spectral moment of the amplitudes of diagonal elements of the bispectrum (BiS_6).

   It is worth mentioning that BiS features are extracted within the non-redundant triangular sub-region (Ω^*, see Section 3.3.3).

- *Features preparation*. Three feature vectors are generated and used as input to the proposed multi-modal ML classifiers:
  1. *only-CWT features* vector: only the features extracted from the ATFR are selected; with the corresponding length of the vector is given by:
     5 (# features) x 5 (# EEG rhythms)= 25 CWT features;
  2. *only-BiS features* vector: only the features extracted from ABiSR are selected; with a corresponding length of the vector is given by:
     6 (# features) x 5 (# EEG rhythms)= 30 BiS features;
  3. *multi-modal (CWT+BiS) features* vector: this comprises a concatenation of CWT and BiS features vectors. The corresponding length is thus given by:
     25 (# CWT features) + 30 (# BiS features)= 55 CWT+BiS features.

- *Classification*: four standard ML techniques are applied to discriminate EEG epochs of HC, MCI and AD subjects over the proposed feature set. To this end, multio-modal ML classifiers based on AE, MLP, LR and SVM architectures are developed to carry out 2-way and 3-way classifications tasks.

3.3. Features Extraction

3.3.1. Time-Frequency analysis

Each EEG recording is analyzed in the time-frequency domain by means of using the Continuous Wavelet Transform (CWT) [32]. If we let *s(t)* be the n-channel EEG recording, the CWT is defined as follows:

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int s(t)\psi^*(\frac{t-b}{a}) \tag{1}$$

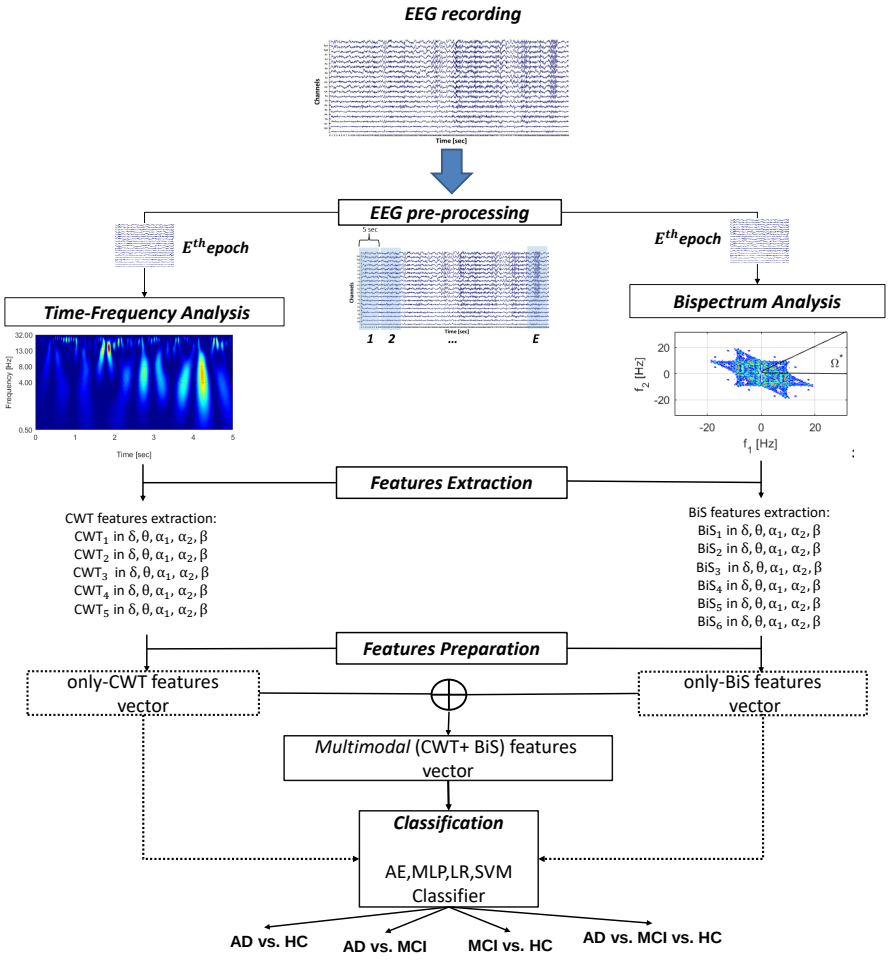

Figure 1: Flowchart of the proposed multi-modal ML-based EEG classification framework. The 19-channel EEG signal is filtered from artifacts and partitioned into N_E epochs. The E^{th} 19-channel EEG epoch is processed through the time-frequency and bispectrum analysis, resulting in the averaged TF and BiS representation, denoted here as ATFR and ABiSR, respectively. Initially, the ATFR and ABiSR are decomposed in five sub-maps (corresponding to the δ, θ, α_1, α_2, β EEG rhythms); next, five CWT features and six BiS features are extracted from each sub-map of ATFR and ABiSR, respectively. Three features vectors are generated: only-CWT features (sized 5 (# features) x 5 (# EEG rhythms)= 25), only-BiS features (sized 6 (# features) x 5 (# EEG rhythms)= 30), multi-modal (CWT+BiS) features (sized 25 (# CWT features) + 30 (# BiS features)= 55). Finally, the feature vectors are used as input to four multi-modal ML classifiers based on AE, MLP, LR and SVM architectures to perform the binary and multiple classifications.

where a and b denote the dilation (or scale) and shifting (or translation) variables, respectively, ψ represents the *mother wavelet* function, $CWT(a,b)$ represents the wavelet coefficients and * is the complex conjugate operator. There is an approximate relationship between scale and frequency where low scale values refer to high frequencies and vice versa ($f_a = \frac{f_c}{T_s a}$, with f_c central frequency of the mother wavelet, T_s sampling period, f_a *pseudo-frequency* corresponding to the scale a). The mother wavelet function chosen here was the *Mexican hat*, which has been proven to match well EEG signals dynamics [33].

3.3.2. CWT features extraction

Given the n^{th} channel of the EEG epoch, the CWT is estimated, resulting in a TFR sized f x S, where f represents the number of frequencies ranging between [0 - $f_{Nyquist}$ Hz] (with $f_{Nyquist} = \frac{f_s}{2} = 128$ Hz), and S is the number of samples in each epoch of 5s (S=1280). Hence, for each EEG epoch, 19 TFR maps are evaluated (one per channel). Next, the 19 TFR are averaged in order to produce a single time-frequency representation (ATFR) for the associated epoch. Assuming N_E epochs are available for a given EEG recording, N_E ATFR maps are generated from it (one per epoch). From the E^{th} ATFR, the following features (denoted as *CWT features*, [34]) are evaluated in each band: mean (CWT_1), standard deviation (CWT_2), skewness (CWT_3), kurtosis (CWT_4) and entropy (CWT_5). Overall, 5 (# features) x 5 (# sub-maps) = 25 CWT features are extracted for each EEG epoch.

3.3.3. Bispectrum analysis

The third-order spectrum is known as *bispectrum* [35]. It quantifies nonlinear interactions (i.e., quadratic phase coupling) and deviations from normality. The bispectrum is defined as the Fourier transform of the third-order cumulant sequence:

$$B(f_1, f_2) = E[S(f_1)S(f_2)S^*(f_1 + f_2)] \tag{2}$$

where $S(f)$ represents the Fourier transform of the n-channel EEG recording s(t), $*$ is the complex conjugate, E $[\cdot]$ is the expectation operation and $B(f_1, f_2)$ is the bispectrum, a complex-value function characterized by symmetric properties. Specifically, as shown in Figure 3(a), the bispectrum has 12 symmetric regions in the plane (f_1,f_2) and is uniquely defined in the non-redundant triangular region (Ω, $0 \leq f_2 \leq f_1 \leq (f_1+f_2) \leq 1$) where frequencies are normalized by the Nyquist frequency [35]. In this study, since we are interested in the EEG frequencies range [0.5-32 Hz], a sub-region of the non-redundant triangular area (denoted as Ω^*) is considered (red triangle, Figure 3). Hence, further observations and analyses of bispectrum are related to Ω^*.

3.3.4. BiS features extraction

Similar to CWT features extraction, given the n^{th} channel of the EEG epoch, the BiS is estimated, leading to a square BiSR matrix sized $\tilde{f}$ x $\tilde{f}$, where $\tilde{f}$ represents the number of frequencies ranging between [0-$f_{Nyquist}$]. Hence, for every EEG

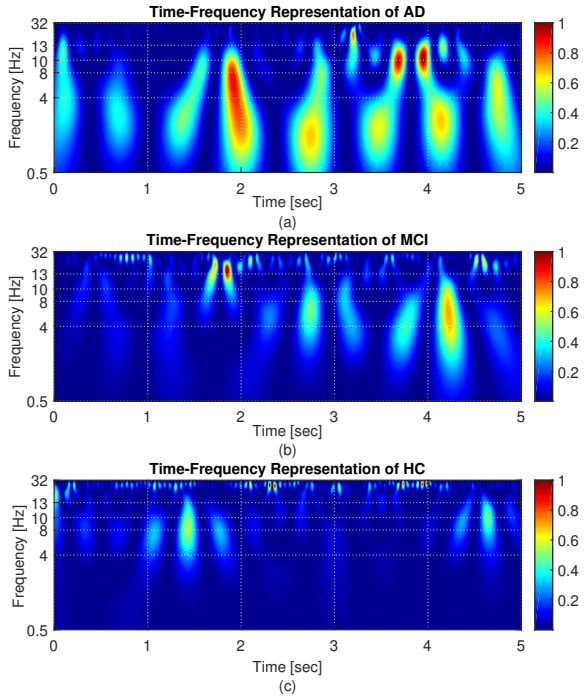

Figure 2: Average time-frequency representations of a 5s EEG epoch related to an AD (a), MCI (b) and HC (c) subject, respectively. X-axis represents time, Y-axis represents frequency and the energy of wavelet coefficients is encoded with coloration ranging from blue to red. The boundaries of the sub-bands taken into account (δ [0.5-4 Hz], θ [4-8 Hz], α_1 [8-10 Hz], α_2 [10-13 Hz], β [13-32 Hz]) are shown with horizontal dashed lines. The specific example is chosen to highlight the different distribution of wavelet coefficients energy in the different sub-bands in the three classes of subjects.

epoch, 19 BiSR are evaluated (one per channel). Subsequently, the 19 BiSR are averaged in order to produce a single average bispectrum representation (ABiSR). Finally, the non-redundant triangular sub-region (Ω^*) is selected from the related bispectrum map under analysis. Hence, given an EEG recording with N_E epochs, N_E average triangular bispectrum matrices, one per epoch (triangular-ABiSR), are produced.

Figure 4 shows a typical 2-dimensional contour map (left) and 3-dimensional surface (right) of the bispectrum. Dark blue and red colors indicate the relative changes in amplitude of bispectrum, that is the highest decrease and increase, respectively. Notably, an explanatory ABiSR related to a 5s epoch belonging to the three different categories of subjects (AD Figure 4 (a), MCI Figure 4 (b), HC Figure 4 (c)) is reported. As expected, due to *slowing effect* observable in EEGs of AD and MCI, the phase-coupled harmonics shift from higher to lower

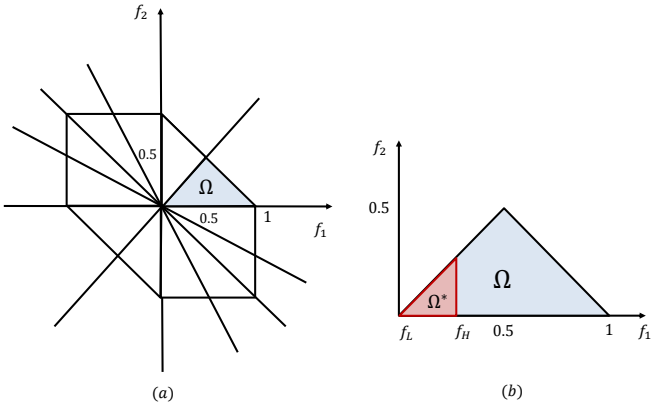

Figure 3: (a) Symmetric regions of the bispectrum, including the triangular non-redundant region (Ω). Frequencies are normalized by the Nyquist frequency (for this case: $f_{Nyquist}$=128 Hz). (b) Triangular non-redundant sub-region (Ω^*, red triangle) ranging between [f_L,f_H] where f_L and f_H correspond to 0.5 and 32 Hz, respectively.

frequencies [6]. The AD bispectrum shows sharp peaks in the δ sub-band ([0.5-4 Hz]), the MCI bispectrum shows dominant peaks in the δ and θ ([4-8 Hz]) sub-bands, whereas the HC bispectrum shows sharp peaks in the α band ([10-13 Hz]). Such findings suggest that the reduction of bispectral peaks correspond to a decrease in non-gaussianity and nonlinearity of EEG signals in AD.

From the E^{th} triangular-ABiS, five sub-regions related to the five main EEG rhythms are taken into account: δ [0.5-4 Hz], θ [4-8 Hz], α_1 [8-10 Hz], α_2 [10-13 Hz], β [13-32 Hz]. Finally, the following most commonly used bispectral features (denoted as *BiS features*), are evaluated in each band [11], [12], [36]:

- Normalized bispectral entropy:

$$BiS_1 = -\sum_i u_i log(u_i) \tag{3}$$

  where $u_i = \frac{|B(f_1,f_2)|}{\sum_{f_1,f_2 \in Sb} |B(f_1,f_2)|}$ and $i = 1, 2, ..I$ is the number of points within the sub-band (Sb) under analysis.

- Normalized bispectral squared entropy:

$$BiS_2 = -\sum_i v_i log(v_i) \tag{4}$$

  where $v_i = \frac{|B(f_1,f_2)|^2}{\sum_{f_1,f_2 \in Sb} |B(f_1,f_2)|^2}$.

- Sum of logarithmic amplitudes of the bispectrum:

$$BiS_3 = \sum_{f_1, f_2 \in Sb} log(|B(f_1, f_2)|) \tag{5}$$

where $B(f_1, f_2)$ is the bispectral content at the point (f_1, f_2) in the Sb under analysis.

- Sum of logarithmic amplitudes of diagonal elements of the bispectrum:

$$BiS_4 = \sum_{f_d, f_d \in Sb} log(|B(f_d, f_d)|) \tag{6}$$

where $B(f_d, f_d)$ is the bispectral value at the diagonal element (f_d, f_d) and $d = 1, ..., D$ is the number of diagonal elements in the Sb under analysis.

- First-order spectral moment of the amplitudes of diagonal elements of the bispectrum:

$$BiS_5 = \sum_{f_d, f_d \in Sb} d * log(|B(f_d, f_d)|) \tag{7}$$

- Second-order spectral moment of the amplitudes of diagonal elements of the bispectrum:

$$BiS_6 = \sum_{f_d, f_d \in Sb} (d - BiS_5)^2 * log(|B(f_d, f_d)|) \tag{8}$$

Overall, 6 (# features) x 5 (# sub-maps) = 30 BiS features are calculated for each EEG epoch.

3.4. Machine Learning Systems

The vectors of the extracted features for the E^{th} EEG epoch under analysis (only-CWT, only-BiS, multi-modal (CWT+BiS) features) are used as input to different ML based systems for binary and multi-classifications: AD vs. HC, AD vs. MCI, MCI vs. HC and AD vs. MCI vs. HC. Specifically, AE, MLP, LR and SVM based multi-modal ML classifiers are developed.

3.4.1. AE classifier

The autoencoder (AE) consists of an encoder ($e : X \rightarrow C$) and a decoder ($d : C \rightarrow X$) module (Figure 5 (a)). Firstly, the input vector $X = [x_1, x_2, ..., x_N]$ is mapped into the compressed representation $C = [c_1, c_2, ..., c_N]$ in the encoding stage:

$$c_i = e(x_i) = \rho_1(W_1 x + b_1) \tag{9}$$

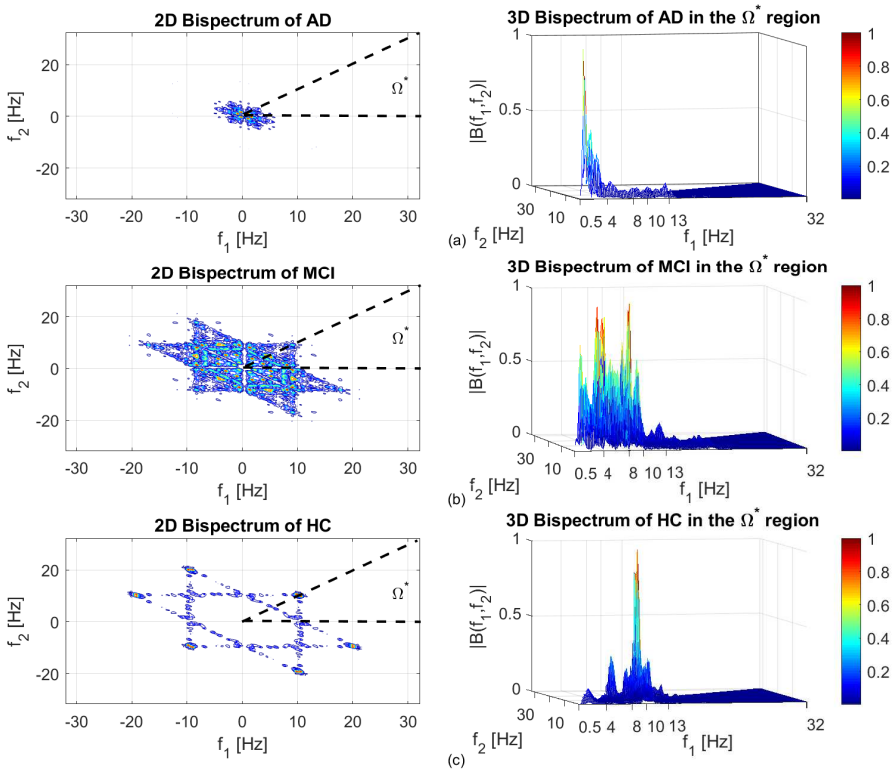

Figure 4: Typical 2D contour map (left) and 3D surface (right) representation of the bispectrum of a 5s EEG epoch related to an AD (a), MCI (b) and HC (c) subject, respectively. Both axes represent frequency, the bispectrum maps show phase coupled harmonics for frequencies (f_1,f_2) in the range under consideration ($0.5-32Hz$). The phase coupling reflects bispectral peaks in the 3D surfaces.

where, W_1 is the weight matrix, b_1 is the bias between the input and the hidden layer, ρ_1 represents the activation function (here, *saturating linear activation function*). Next, c_i is mapped back to x_i in the decoding stage:

$$\hat{x_i} = d(e(x_i)) = \rho_2(W_2 c_i + b_2) \tag{10}$$

where, W_2 is the weight matrix, b_2 is the bias between the hidden and the output layer, ρ_2 represents the activation function (here, *linear transfer function*). The aim of the AE is to minimize the error between the original input data and its reconstruction by minimizing a loss function, that is the mean square error (MSE).

Four AE based architectures are developed (whose parameters are experimentally optimized, using a trial and error approach), specifically: AE_1 with a

single hidden layer comprising 18 units; AE_2 with two hidden layers of 18 and 10 units; AE_3 with a single hidden layer of 30 units; AE_4 with two hidden layers of 30 and 18 units. The AE_1 and AE_2 classifiers are employed when the only-CWT and the only-BiS features are used as input; whereas, the AE_3 and AE_4 classifiers exploit multi-modal (CWT+BiS) features vector as input. The features extracted from each AE architecture are used as input to a fully connected layer with softmax activation function trained with supervised learning to perform the classification tasks. Next, the AE_j (with j=1,2,3,4) followed by the SF layer is fine-tuned (with supervised training) to improve the discrimination performance. As an example, Figure 5 (a-b) show the best performing AE architectures when the multi-modal (CWT+BiS) features are used as input. The compressed representations are then employed to develop the AE_3 and AE_4 classifiers (Figure 5 (d) and (e), respectively). Note that the size of the hidden layers of AE_1, AE_2, AE_3 and AE_4 were chosen by evaluating the minimum reconstruction errors.

3.4.2. MLP classifier

The Multi-Layer Perceptron (MLP) is the most popular feed-forward neural network, typically composed of an input layer, one or more hidden layers, an output layer. It is trained with supervising learning procedure through standard backpropagation technique [37]. In this study, similarly to the case of AE classifiers, four MLP based architectures are developed: MLP_1, MLP_2, MLP_3, MLP_4. For a fair comparison, the MLP_j classifier has the same structure as the AE_j classifier (with j=1,2,3,4). For example, Figure 6 (a) and (b) reports the best performing MLP_3 and MLP_4 classifiers, respectively, when the multi-modal CWT+BiS features vector is used as input. Specifically, MLP_3 consists of 1-hidden layer with 10 units; whereas, MLP_4 consists of 2-hidden layers with 30 and 18 units. All the MLP_j outputs with a softmax layer. Each hidden layer of the MLPs has a saturating ReLu activation function. The MLPs are trained over 10^3 epochs until the convergence of the cross-entropy loss function is observed.

3.4.3. LR classifier

The Logistic Regression (LR) is a statistical technique used to model the probability of occurrence of a specific event (here, the presence of disease (MCI/AD) or healthy condition) by using a linear combination of predictors (indipendent variables, η) [38]. The sigmoid function ($\sigma(\eta) = \frac{1}{1+e^{\eta}}$) represents the logistic function. Hence, the LR classifier provides the probability score value of a given input (i.e EEG epoch) as belonging to the G group (i.e. AD/MCI/HC).

3.4.4. SVM classifier

The Support Vector Machine (SVM) is a well-known statistical technique, that finds the best hyperplane providing the maximum separation among classes [39]. In this study, for comparative evaluation purposes, a linear kernel has been used to perform the classification tasks. A detailed mathematical formulation of the SVM technique can be found in [40].

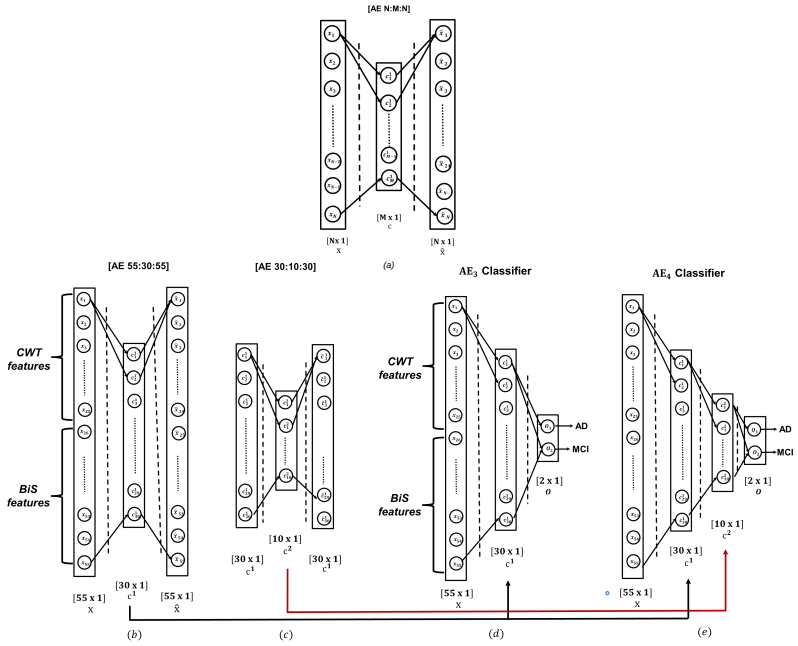

Figure 5: (a) Standard AE architecture, where c is the compressed representation of the N-dimensional input vector x and $\hat{x}$ is the vector reconstructed from c, so that $\hat{x} \approx x$. (b) AE_3 architecture [AE 55:30:55]. The 55-dimensional multi-modal features (x) are compressed into a 30-dimensional features vector (c^1), that is used to produce the same input representation $\hat{x}$. (c) AE_4 architecture [AE 30:10:30]. The features vector c^1 extracted through the AE_3 forms the input to a second AE that compresses the input space into a 10-dimensional features representation (c^2) (d) AE_3 classifier. This includes the 30-dimensional features vector (c^1) and a softmax layer (o) to perform the 2-way and 3-way classifications. (e) AE_4 classifier. This includes the 30-dimensional features vector (c^1), 10-dimensional features vector (c^2) and a softmax layer (o) for classification purposes. The AE_3, AE_4 classifiers are then fine-tuned using the standard back propagation method to improve classification performance. The figure illustrates examples classifiers for the binary configuration case: AD vs. MCI.

4. Experimental results

In this research, an EEG database of 63 AD, 63 MCI and 63 HC, has been used as benchmark to perform the 2-way and 3-way classification tasks. Specifically, each n-channel EEG recording (n=19), was divided into N_E epochs. Next, for all the epochs, the time-frequency transform and the bispectrum were estimated, resulting in N_E ATFR and ABiSR, in accordance with the procedure described in Section 3.3. Given the E^{th} ATFR and ABiSR under analysis, 25 CWT features and 30 BiS features were evaluated, respectively. Four ML based systems were developed to discriminate EEG epochs as belonging to AD, MCI or HC class: AE_j, MLP_j (with j=1,2,3,4), LR and SVM. Next, the performance of the proposed classifiers was tested firstly with the only-CWT features vector, then with the only-BiS features vector and finally with the concatenation of CWT

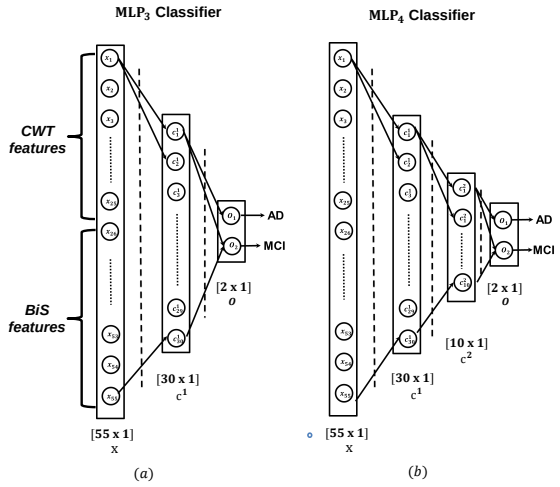

Figure 6: (a) MLP_3 classifier. It comprises 1-hidden layer with 30 hidden units followed by a softmax output layer for performing 2-way and 3-way classifications. (b) MLP_4 classifier. It comprises of 2-hidden layers with 30 and 10 hidden units, respectively, and a softmax layer for classification purposes. The figure illustrates examples classifiers for binary classification: AD vs. MCI.

and BiS features vectors, the latter here denoted as *multi-modal (CWT+BiS) features*.
The following measures were used to quantify the effectiveness of our developed classification systems:

$$PRECISION(prc) = \frac{TP}{TP + FP} \tag{11}$$

$$RECALL(rec) = \frac{TP}{TP + FN} \tag{12}$$

$$F - SCORE(fsc) = 2 * \frac{PC * RC}{PC + RC} \tag{13}$$

$$ACCURACY(acc) = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

where TP, TN, FP, FN represent the true positive, true negative, false positive and false negative, respectively [41]. Furthermore, the k-fold cross validation procedure (with k=10) was used: in each fold, 70% of data was used as the training set and the remaining 30% as testing set. Thus, since the overall classification performances were evaluated by estimating the average (avg) and standard deviation (std) over the 10 folds, all results are expressed in terms of average value $\pm$ standard deviation (i.e. avg(acc) $\pm$ std (acc)).

4.1. Only-CWT features based classification

Table 1 reports the AD vs. HC, AD vs. MCI, MCI vs. HC, AD vs. MCI vs. HC classification performance when the only-CWT features vector was used as input to the proposed classifiers.

1) *AD vs. HC*: in this scenario, the AE_1 and AE_2 classifiers reported very good performance, achieving accuracy values of 91.84 ± 0.5% and 92.35 ± 0.5%, respectively, and F-scores of 85.93 ± 0.9% and 86.90 ± 0.8%, respectively. The LR and SVM based classifiers, on the other hand, reported lower values in terms of F-score (81,19 ± 1.0% and 79.60 ± 0.8%, respectively) and accuracy (89.25 ± 0.5% and 88.48 ± 0.5%, respectively). However, the MLP architectures reported the highest performance. Specifically, MLP_1 and MLP_2 achieved accuracy values up to 95.76 ± 0.5% and 95.4 ± 0.6%, respectively.

2) *AD vs. MCI*: in this scenario, the AE_1 and AE_2 classifiers achieved accuracy values of 76.60 ± 0.8% and 77.45 ± 0.8%; whereas, the LR and SVM classifiers achieved values of 71.87 ± 0.7% and 70.48 ± 0.7%, respectively. However, the MLP_1 and MLP_2 classifiers outperformed all the other networks, reporting F-score of 81.70 ± 1.3% and 78.73 ± 1.4%, respectively, and accuracy values of 86.84± 1.0% and 84.70 ± 1.1%, respectively.

3) *MCI vs. HC*: in this case, the AE_1 and AE_2 classifiers reported good accuracy performances of 83.13 ± 1.8% and 83.93 ± 1.1%, respectively. The LR and SVM classifiers achieved acceptable values in terms of the F-score (72.76 ± 0.6% and 73.03 ± 0.6%, respectively) and accuracy (78.07 ± 0.5% and 78.61 ± 0.5%, respectively). However, the MLP_1 and MLP_2 based classifiers reported the best performance: delivering accuracies of 91.80 ± 0.9% and 91.53 ± 0.5%, respectively; F-score of 89.99 ± 0.6% and 72.76 ± 0.6%, respectively;

4) *AD vs. MCI vs. HC*: in this case, low accuracy values were achieved by AE_1, AE_2, LR, SVM classifiers: 69.96 ± 1.5%, 70.61 ± 0.8%, 60.08 ± 0.6% and 66.95 ± 0.6%, respectively. The MLP_1 and MLP_2 classifiers, on the other hand, reported acceptable accuracy (82.13 ± 0.5% and 80.93 ± 0.6%, respectively) and F-scores (70.25 ± 1.3% and 68.19 ± 1.4%, respectively).

4.2. Only-BiS features based classification

Table 2 reports the AD vs. HC, AD vs. MCI, MCI vs. HC classification performance when the only-BiS features vector was used as input to the proposed classifiers.

1) AD vs. HC: the AE_1 classifier achieved accuracy and F-score values of 82.82 ± 0.9% and 70.56 ± 1.4%, respectively. The AE_2 classifier reported similar results (accuracy of 82.59 ± 1.0% and F-score of 70.46 ± 1.6%). Comparable performances were observed with the LR classifier (accuracy of 83.03 ± 0.9% and F-score of 70.67 ± 1.5%); whereas, lower values were observed with the SVM model. However, the MLP classifiers reported the highest classification performance. Specifically, MLP_1 achieved an accuracy of 87.59 ± 0.9% and F-score of 79.15 ± 1.6%; whereas, MLP_2 achieved an accuracy of 87.28 ± 0.8% and F-score of 78.69 ± 1.3% .

2) AD vs. MCI: the AE_1 classifier achieved accuracy and F-score values of 72.12 ± 1.0% and 57.18 ± 2.0%, respectively. The AE_2 classifier reported comparable

values (accuracy of 71.76 ± 0.9% and F-score of 57.53 ± 1.4%). Similarly, the LR and SVM classifiers reported accuracies of 72.76 ± 1.0% and 71.03 ± 0.8%, respectively. The MLP_1 and MLP_2 classifiers reported the highest classification accuracies of 79.38 ± 0.8% 79.20 ± 1.2%, respectively.
3) MCI vs. HC: the AE_1 and AE_2 classifiers achieved similar accuracy values, that is: 79.25 ± 0.6% and 79.56 ± 1.0%, respectively. The LR classifier reported an accuracy and F-score values of 76.43 ± 0.6% and 71.88 ± 0.9%, respectively; whereas, the SVM classifier reported accuracy of 76.20 ± 0.8% and F-score of 71.87 ± 1.0%, respectively. However, the MLP classifiers reported the highest classification performance with an accuracy of 87.14 ± 0.7% .
4) *AD vs. MCI vs. HC*: low accuracy values were observed in the AE_1, AE_2, LR, SVM classifiers, reporting accuracy rate up to 65.75 ± 1.2% (AE_2). Slightly better performance were achieved with the MLP based classifiers, with accuracy rate up to 74.75 ± 0.9% (for MLP_1)

4.3. multi-modal (CWT+BiS) features based classification

Table 3 reports the AD vs. HC, AD vs. MCI, MCI vs. HC classification performance when the multi-modal (CWT+BiS) features vector was used as input to the proposed classifiers.
1) AD vs. HC: in this case, high discrimination performances were observed with all the classifiers. Specifically, the AE_3 classifier reported accuracy and F-score values of 93.85 ± 0.3% and 89.56 ± 0.5%, respectively; the AE_4 classifier achieved an accuracy of 94.00 ± 0.5% and F-score of of 89.86 ± 0.8%, respectively. The LR and SVM reported accuracies of 93.21 ± 0.6% and 91.88 ± 0.5%; whereas the MLP architectures reported the highest performance with accuracy values of 96.95 ± 0.5% (MLP_3) and 96.71 ± 1.1% (MLP_4).
2) AD vs. MCI: in this case, the AE_3 and AE_4 classifiers reported acceptable accuracy values of 78.79 ± 1.0% and 79.26 ± 0.9%; whereas, the LR and SVM classifiers achieved values of 76.64 ± 0.8% and 75.85 ± 0.5%, respectively. However, the MLP_3 and MLP_4 outperformed all the other classifiers, achieving accuracy rates up to 90.24 ± 0.7% and 89.81 ± 1.0%, respectively.
3) MCI vs. HC: in this scenario, higher classification performances were observed with the AE_3 (accuracy of 90.53 ± 0.8%, F-score of 88.83 ± 0.9%) and AE_4 (accuracy of 90.57 ± 0.9%, F-score of 88.83 ± 1.0%) compared to LR (accuracy of 84.89 ± 0.5%, F-score of 82.03 ± 0.5%) and SVM (accuracy of 84.41 ± 0.7%, F-score of 81.48 ± 0.5%). On the other hand, the MLP models reported the highest performances with accuracy values of 96.24 ± 0.5% (MLP_3) and 95.91 ± 0.3% (MLP_4).
4) *AD vs. MCI vs. HC*: acceptable accuracies were reported with the AE_3 (76,91 ± 2.0%), AE_4 (77,30 ± 1.4%) and SVM (74,54 ± 0.5%) classifiers; whereas, low accuracy performance were observed with the LR technique (65.35 ± 0.6%). Finally, in this scenario, the MLP_3 and MLP_4 classifiers reported the highest accuracy values: 89.22 ± 0.7% and 88.56 ± 0.7%, respectively.

5. Discussion

This paper exploited the potential of both CWT and BiS features extracted from time-frequency and bispectrum representations able to better discriminate among AD/MCI/HC EEG epochs. Specifically, three sets of features were generated: only-CWT features, only-BiS features and multi-modal (CWT+BiS) features. The extracted features were fed to the proposed ML classifiers (AE_j, MLP_j, LR, SVM with j=1,2,3,4) to perform binary and multiple EEG epoch classifications: AD vs. HC, AD vs. MCI, MCI vs. HC and AD vs. MCI vs. HC. Comparative results showed that our multi-modal (CWT+BiS) approach outperformed the only-CWT and only-BiS approaches, reporting the best classification performance in all of the analysed scenarios. This endorses the hypothesis that high order features extracted from the bispectrum and TF maps, can provide enhanced AD/MCI/HC classification. We conjecture that the improved classification performance is due to the ability of bispectrum to capture nonlinear couplings between channels in the frequency domain (since it takes into account the phase component of the Fourier Transform). Indeed, EEGs are known to be inherently generated by non-linear interactions between neuronal circuitry. However, the bispectrum is not sufficient for representing all required information from the EEGs, particularly on account of their non-stationary nature. Thus augmenting time-frequency domain representations can help improve the performance of multi-modal classifiers. Time-frequency approaches, such as the CWT, can provide information on how the frequency content changes over time. Comparative performance evaluation demonstrated the superior classification ability of MLP_3 as compared to other proposed models. In particular, the MLP_3 was found to achieve the highest accuracy of 96.95 $\pm$ 0.5% in AD vs. HC, 90.24 $\pm$ 0.7% in AD vs. MCI; 96.24 $\pm$ 0.5% in MCI vs. HC and 89.22 $\pm$ 0.67% in AD vs. MCI vs. HC classification. It is worth noting that the developed MLP_3 had a very simple architecture (1-hidden layer with 30 neurons), thus requiring significantly reduced number of learning parameters and training time.

It is also interesting to note that the MLPs outperformed AE networks. AEs are typically used to extract relevant features from high dimensional representations. In this study, the maximum input size was only 55 features (for the case of the multi-modal approach) thus it could be argued that the compression ability of AE was not fully exploited in this case. The possible over compression of features potentially caused loss of relevant information, and consequently a misclassification of the EEG epoch under analysis. In particular, Figure 7 shows outputs of a pair of nodes (i.e., 6 vs. 10) of the second hidden layer of the AE[55:30:10:2], when the multi-modal (CWT+BiS) features were used as input in the AD vs. HC classification scenario. As can be seen, the compressed representation allows to approximately reconstruct the input vector by also separating the two classes (i.e. AD, HC), notwithstanding the absence of labels since the learning is unsupervised. In case of the MLP, in contrast, the hidden nodes are forced to learn the dual classification problem. Availability of labels improves discrimination performance with respect to the scheme based on

AEs, as the compression procedure invariably causes some information loss. In the future, a wider set of features will be introduced and extracted from EEGs in order to capture more diagnostically relevant information as possible. In such a case, a multi-modal AE model might improve classification performance by reducing redundancy in the relatively large input space.

To the best of our knowledge, this is the first work that uses a multi-modal approach based on a concatenation of time-frequency and bispectrum features to classify EEGs segments of AD, MCI and HC.

In a previous work, Morabito et al. [29] proposed CNN based classification of EEG recordings of AD/MCI/HC subjects by extracting statistical parameters mean (μ), standard deviation (σ) and skewness (κ) from time-frequency representations of the EEG time series. The proposed system achieved good classification performance, specifically: 85% accuracy in AD vs. HC, 78% accuracy in AD vs. MCI, 85% accuracy in MCI vs. HC and 82% accuracy in AD vs. MCI vs. HC classification. However, the classifier was tested on a limited dataset (69 individuals, 23 per class) and the CWT features were extracted from empirically defined time-frequency sub-maps that did not correspond to the standard EEG rhythms, that are known to be associated with different brain states, and can thus provide relevant information when investigated separately. In [42], instead, Ieracitano et al. used the same EEG database that was used in this paper. The authors proposed and trained a CNN_1 classifier with 2D-power spectral density (PSD) images of AD, MCI and HC. Experimental results reported classification accuracies of 92.95% in AD vs. HC, 84.62% AD vs. MCI 91.88% in MCI vs. HC and 83.33% in AD vs. MCI vs. HC. However, the methodology proposed in [42] was heavily dependent on the position of power peaks in the frequency domain. In particular, an unexpected shifting of the dominant α peak (for example, due to artefactual components) can cause a misclassification of the epoch under analysis.

In contrast, in this study, we extracted the CWT features and BiS features from standard EEG sub-bands and differently from the aforementioned studies, the 10-fold cross validation method was also used. Moreover, it has been proven in our current study, that the concatenation of CWT and BiS features can provide a more effective AD/MCI/HC classification performance. Specifically, experimental results show that, the proposed MLP classifier based on the multi-modal feature set outperformed other schemes. Nevertheless, it is worth mentioning that, the idea to mix CWT and BiS features is not merely a combination of two different techniques (time-frequency and bispectrum), but, comes from a more accurate analysis of their properties in each sub-band (within each category). Indeed, in order to further emphasize the effectiveness of extracted features in the discrimination of HC, MCI and AD subjects, the probability density functions (pdf) of the only-CWT, only-BiS and multi-modal (CWT+BiS) features were also evaluated in each sub-band (i.e. δ, θ, α_1, α_2, β). Comparative experimental results show that the pdf distributions highlight the different features of AD, MCI and HC. For example, Figure 8 reports the δ band distribution of three set of features vectors for each class category (AD/MCI/HC). As can be seen, the distributions of multi-modal (CWT+BiS) features (Figure

8 (g), (h), (i)) take into account multi-modal information gathered from both TF (Figure 8 (a), (b), (c)) and bispectrum (Figure 8 (d), (e), (f)) analysis. Notably, as regards the HC histogram of the CWT features (Figure 8 (c)), it reports a bimodal distribution with local maxima around 0.1 and 0.4; similarly, as regards the HC histogram of the BiS features (Figure 8 (f)) a bimodal distribution is also observed but with local maxima around 0.4 and 0.9. Finally, the HC histogram of the multi-modal (CWT+BiS) features, combine the CWT and BiS properties by showing a trimodal distribution with local maxima around 0.1, 0.4 and 0.9. Similar behaviors can be observed in MCI and AD distributions. Furthermore, AD distributions report lower pdf values than HC, while MCI shows values in between those in AD and HC. This is rather expected as the MCI is a prodromal stage of AD. The discrimination performance was also confirmed by the analysis of the Receiver Operating Characteristics (ROC) curve and in particular by measuring the areas under ROC curve (AUC). As an example, Figure 9 reports the ROC curves and the corresponding AUC values related to the AD vs. MCI scenario when only-CWT, only-BiS and multi-modal (CWT+BiS) features are used as input to the proposed classifiers. As seen from the comparative experiment results, MLP architectures with multi-modal input features vector achieved the best performance (AUC rate up to 96.5%).

Nevertheless, our proposed methodology suffers from some drawbacks. The EEG database includes AD and MCI patients at different stages of the disorders. AD is a neurodegenerative disease, which progressively damages the brain causing neuronal death, synapse loss and atrophy; it develops gradually and reflects differently upon the EEG waves, depending on its stage. The group of MCI subjects includes all those who show a mild cognitive impairment, however, not all are inherently affected by AD, hence not all of them will develop dementia due to AD. Some of them may remain stable whilst some others may improve. This means that some MCI subjects, inherently affected by AD and at a stage of the pathology close to degeneration into dementia, may show EEG patterns closer to the AD category, than to MCI, which may cause a misclassification in the EEG of that subject.

6. Conclusions

In this paper, a novel multi-modal ML system for classifying EEG recordings of AD, MCI and HC individuals is proposed. The originality of the present research lies in combining higher order statistics extracted from the bispectrum and from the time-frequency representation of EEG signals in order to enhance discrimination performance between the AD, MCI and HC categories. EEG is generated by both linear and non-linear processes related to the interaction between neurons at the cortical level. The bispectrum is able to capture such non-linear interactions in the frequency domain. Such kind of interactions show differently in the three classes of subjects. For this purpose, a feature set comprising higher order statistics extracted from the bispectrum ("only-BiS" dataset) was defined. Bispectrum analysis is an attractive technique to monitor brain states in AD subjects as it is capable of detecting EEG nonlinearities via

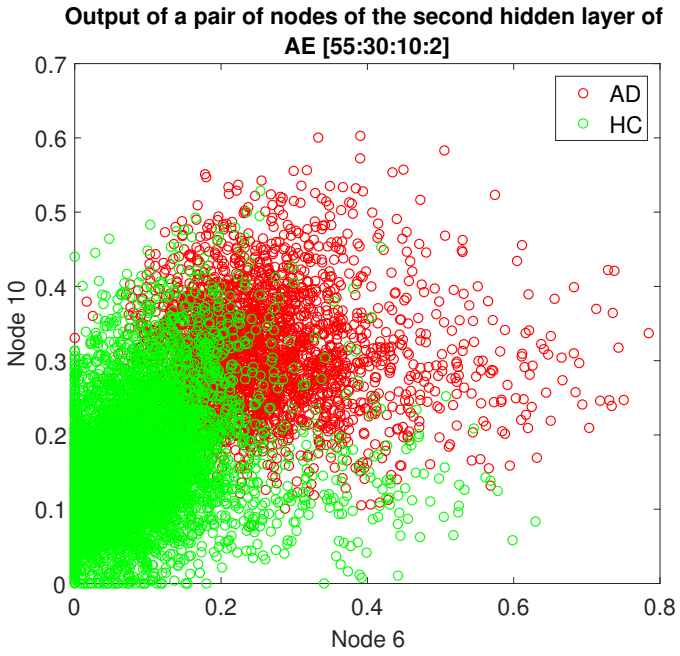

Figure 7: Output of the pair of nodes (6 vs. 10) of the second hidden layer of the autoencoder sized [55:30:10:2], when the multi-modal (CWT+BiS) features are used as input. It is here evident that the compressed representation allows to approximately reconstructing the input vector by also separating the two classes although the learning is unsupervised, thus the label information is not available. In the case of MLP, in contrast, the hidden nodes are forced to learn the binary nature of the input. This improves discrimination performance with respect to the scheme based on the AEs as the compression procedure causes invariably some information loss.

the information garthered from higher order phase coupling. Encouraged by previous findings from time-frequency representations of EEG, a set of features was extracted from the time frequency maps of EEG signals (termed "only-CWT" dataset). Four different architectures (MLP, SVM, AE, LR) were designed and tested over three different feature sets: only-CWT, only-BiS and a combination of CWT and BiS. Comparative results demonstrate the potential of our proposed multi-modal (CWT+BiS) approach. The best performance was achieved when the multi-modal features vector was used, thus endorsing our hypothesis that the joint use of TF and bispectrum can more effectively empower EEG-based differential diagnosis of AD, MCI and healthy elderly individuals. To the best of our knowledge, this is the first study that exploits and uses both time-frequency and bispectral coefficients as discriminating features to improve the diagnosis of MCI/AD. Notably, we propose a computationally-efficient MLP architecture with only one hidden layer and 30 hidden units, which is shown to outperform other state-of-the-art learning models, delivering accuracy rates of up to: 96.95 $\pm$ 0.5% in AD vs. HC; 90.24 $\pm$ 0.7% in AD vs. MCI; 96.24

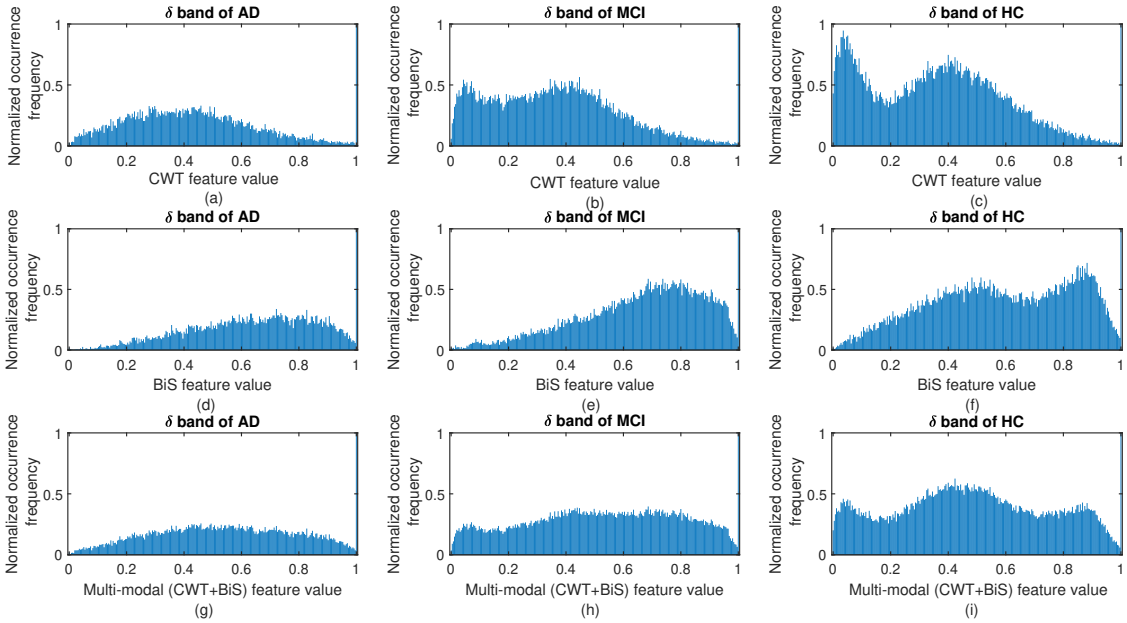

Figure 8: Probability density function (pdf) distribution of the three set of features vectors (only-CWT, only-BiS and multi-modal (CWT+BiS) features) in the δ band of each class category (AD, MCI and HC). As can be seen, the multi-modal (CWT+BiS) features distribution takes into account both infomation of CWT and BiS features distributions. For example, the HC distribution of the multi-modal (CWT+BiS) features (i) somehow collects the HC properties of the CWT (c) and BiS (f) features vectors.

$\pm$ 0.5% in MCI vs. HC and 89.22 $\pm$ 0.7% in AD vs. MCI vs. HC. It is worth noting that this study focused on the AD/MCI/HC classification task and did not address prediction of early onset of AD, which is an open diagnostic and clinical issue. In order to enhance the predictive power, an analysis is required on the evolution of predictive models using longitudinal data. Imaging diagnostics are mostly used for longitudinal studies e.g. Alzheimer's Disease Neuroimaging Initiative (ADNI). Nevertheless, the use of EEG is rather useful in early phases, on account of its low-cost, ease of multiple scanning and potential acceptance by prospective patients. In this context, we previously carried out a longitudinal study on EEG-AD data [43] that showed the alterations of brain connectivity during the course of the disease. Hence, motivated by the excellent AD/MCI/HC classification performance achieved here, through our proposed multi-modal ML system, we intend to carry out future longitudinal (or multiple-time follow-up) studies on MCI patients and HC individuals in order to extract potentially relevant biomarkers for monitoring disease progression. This could prove useful for designing an early protocol for patients eligible for prescription of anti-dementia drugs. It is also worth mentioning that the

classification power refers to discrimination of classes of subjects, whereas predictive power relates to early interpretation of disease progression in individuals. Monitoring more subjects will generate the required database to enable possible extraction of "general" rules for conversion from MCI to AD, and HC to MCI, based on the discovered features. Further, the follow-up of healthy individuals could yield relevant information for predicting discrimination in healthy and pathological aging. The main goal of future research will be to extract efficient features showing different longitudinal paths.

The identification of AD state can also be carried out using recently developed architectures such as the Spiking Neural Network (SNN) based NeuCube, which was shown to deliver very promising results [44]. For future work, it will be interesting to integrate our multi-modal ML techniques with the SNN in order to better understand changes in brain activity [45] [46]. Additionally, motivated by the promising results of our multi-modal approach, a larger cohort of patients will be taken into account, and performance benchmarked against a range of other state-of-the-art deep and reinforcement learning approaches (including those recently reported in [47]). Further, we intend to apply our multi-modal ML system to investigate epileptic seizure activity, and compare with the recently reported approach in [48], which exploited high order spectral features of bispectrum and the conventional MLP neural network to enhance seizure prediction (in particular, classification of interictal and preictal samples). Finally, novel saliency based approaches [49] may also be explored, in an attempt to further enrich the multi-modal representations.

Acknowledgment

This work was partially funded by the Italian Ministry of Health grant Ref. GR-2011-02351397. A. Hussain conceived the original idea of exploiting higher order statistics (bispectrum features) as part of a novel multi-modal machine learning approach. He would like to acknowledge the UK Engineering and Physical Sciences Research Council (EPSRC) grant Ref. EP/M026981/1 (AVCOGHEAR). The authors would also like to thank the doctors of IRCCS Centro Neurolesi Bonino Pulejo Messina (Italy) for providing the EEG dataset used in the present work.

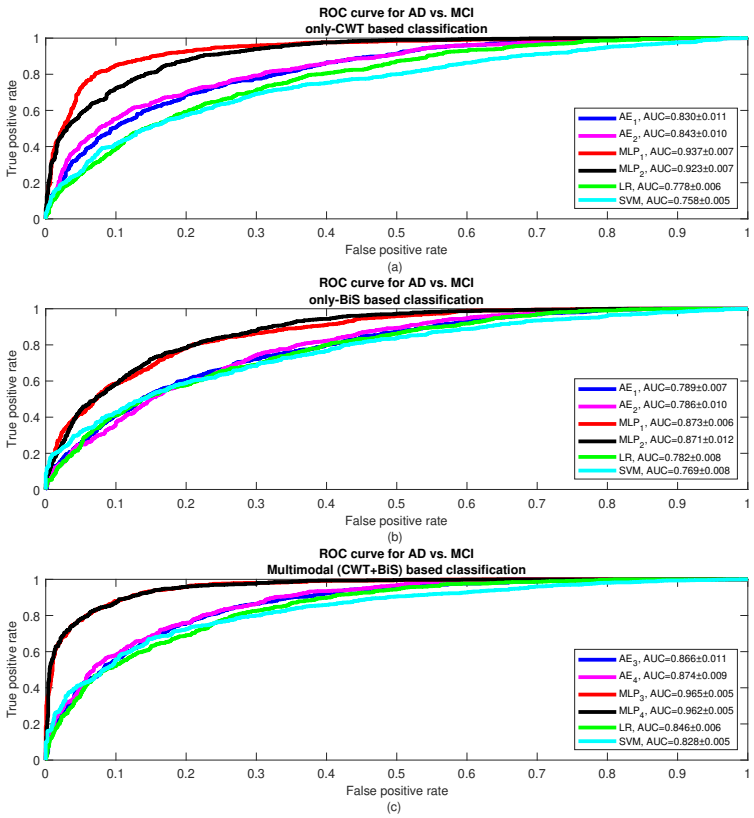

Figure 9: ROC curves and AUC values of the proposed AEs, MLPs, LR and SVM classifiers for the AD vs. MCI when the only-CWT (a), only-BiS (b), multi-modal (CWT+BiS) features vector (b) is used as input.

References

[1] R. C. Petersen, G. E. Smith, S. C. Waring, R. J. Ivnik, E. G. Tangalos, E. Kokmen, Mild cognitive impairment: clinical characterization and outcome, Archives of Neurology 56 (3) (1999) 303–308.

[2] R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack Jr, J. Kaye, T. J. Montine, et al., Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease, Alzheimer's & dementia 7 (3) (2011) 280–292.

[3] 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).

[4] P. L. Nunez, R. Srinivasan, Electric fields of the brain: the neurophysics of EEG, Oxford University Press, USA, 2006.

[5] J. Jeong, EEG dynamics in patients with Alzheimer's disease, Clinical neurophysiology 115 (7) (2004) 1490–1505.

[6] J. Dauwels, K. Srinivasan, M. Ramasubba Reddy, T. Musha, F.-B. Vialatte, C. Latchoumane, J. Jeong, A. Cichocki, Slowing and loss of complexity in Alzheimer's EEG: two sides of the same coin?, International Journal of Alzheimer's Disease 2011.

[7] T. König, L. Prichep, T. Dierks, D. Hubl, L. Wahlund, E. John, V. Jelic, Decreased EEG synchronization in Alzheimers disease and Mild Cognitive Impairment, Neurobiology of aging 26 (2) (2005) 165–171.

[8] N. Mammone, C. Ieracitano, H. Adeli, A. Bramanti, 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 (99) (2018) 1–14.

[9] N. Mammone, S. De Salvo, L. Bonanno, C. Ieracitano, S. Marino, A. Marra, A. Bramanti, F. C. Morabito, Brain network analysis of compressive sensed high-density EEG signals in AD and MCI subjects, IEEE Transactions on Industrial Informatics 15 (1) (2019) 527–536.

[10] F. Chella, A. D'Andrea, A. Basti, V. Pizzella, L. Marzetti, Non-linear analysis of scalp EEG by using bispectra: the effect of the reference choice, Frontiers in Neuroscience 11 (2017) 262.

[11] C. Gomez, F. Vaquerizo-Villar, J. Poza, S. J. Ruiz, M. A. Tola-Arribas, M. Cano, R. Hornero, Bispectral analysis of spontaneous eeg activity from patients with moderate dementia due to alzheimer's disease, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017, pp. 422–425.

[12] R. Wang, J. Wang, S. Li, H. Yu, B. Deng, X. Wei, Multiple feature extraction and classification of electroencephalograph signal for alzheimers' with spectrum and bispectrum, Chaos: An Interdisciplinary Journal of Nonlinear Science 25 (1) (2015) 013110.

[13] J. Li, Z. Zhang, H. He, Hierarchical convolutional neural networks for EEG-based emotion recognition, Cognitive Computation 10 (2) (2018) 368–380.

[14] M. Badarna, I. Shimshoni, G. Luria, S. Rosenblum, The importance of pen motion pattern groups for semi-automatic classification of handwriting into mental workload classes, Cognitive Computation 10 (2) (2018) 215–227.

[15] J. Mekyska, Z. Galaz, T. Kiska, V. Zvoncak, J. Mucha, Z. Smekal, I. Eliasova, M. Kostalova, M. Mrackova, D. Fiedorova, et al., Quantitative analysis of relationship between hypokinetic dysarthria and the freezing of gait in parkinson's disease, Cognitive Computation 10 (6) (2018) 1006–1018.

[16] V. Sachnev, S. Suresh, N. Sundararajan, B. S. Mahanand, M. W. Azeem, S. Saraswathi, Multi-region risk-sensitive cognitive ensembler for accurate detection of attention-deficit/hyperactivity disorder, Cognitive Computation (2019) 1–15.

[17] A. Kalantari, A. Kamsin, S. Shamshirband, A. Gani, H. Alinejad-Rokny, A. T. Chronopoulos, Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions, Neurocomputing 276 (2018) 2–22.

[18] C. Lehmann, T. Koenig, V. Jelic, L. Prichep, R. E. John, L.-O. Wahlund, Y. Dodge, T. Dierks, Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG), Journal of neuroscience methods 161 (2) (2007) 342–350.

[19] J. C. McBride, X. Zhao, N. B. Munro, C. D. Smith, G. A. Jicha, L. Hively, L. S. Broster, F. A. Schmitt, R. J. Kryscio, Y. Jiang, Spectral and complexity analysis of scalp EEG characteristics for Mild Cognitive Impairment and early Alzheimer's disease, Computer methods and programs in biomedicine 114 (2) (2014) 153–163.

[20] L. Trambaiolli, N. Spolaôr, A. Lorena, R. Anghinah, J. Sato, Feature selection before EEG classification supports the diagnosis of Alzheimer's disease, Clinical Neurophysiology 128 (10) (2017) 2058–2067.

[21] L. R. Trambaiolli, A. C. Lorena, F. J. Fraga, P. A. Kanda, R. Anghinah, R. Nitrini, Improving Alzheimer's disease diagnosis with machine learning techniques, Clinical EEG and neuroscience 42 (3) (2011) 160–165.

[22] S. J. Ruiz-Gómez, C. Gómez, J. Poza, G. C. Gutiérrez-Tobal, M. A. Tola-Arribas, M. Cano, R. Hornero, Automated multiclass classification of spontaneous EEG activity in Alzheimer's disease and Mild Cognitive Impairment, Entropy 20 (1) (2018) 35.

[23] E. Neto, F. Biessmann, H. Aurlien, H. Nordby, T. Eichele, Regularized linear discriminant analysis of EEG features in dementia patients, Frontiers in Aging Neuroscience 8 (2016) 273.

[24] M. Kashefpoor, H. Rabbani, M. Barekatain, Automatic diagnosis of mild cognitive impairment using electroencephalogram spectral features, Journal of Medical Signals and Sensors 6 (1) (2016) 25.

[25] N. Kulkarni, Use of complexity based features in diagnosis of mild alzheimer disease using EEG signals, International Journal of Information Technology 10 (1) (2018) 59–64.

[26] G. Fiscon, E. Weitschek, A. Cialini, G. Felici, P. Bertolazzi, S. De Salvo, A. Bramanti, P. Bramanti, M. C. De Cola, Combining EEG signal processing with supervised methods for Alzheimer's patients classification, BMC medical informatics and decision making 18 (1) (2018) 35.

[27] N. Kulkarni, V. Bairagi, Extracting salient features for EEG-based diagnosis of Alzheimer's disease using support vector machine classifier, IETE Journal of Research 63 (1) (2017) 11–22.

[28] F. Vialatte, A. Cichocki, G. Dreyfus, T. Musha, S. L. Shishkin, R. Gervais, Early detection of Alzheimer's disease by blind source separation, time frequency representation, and bump modeling of EEG signals, in: International Conference on Artificial Neural Networks, Springer, 2005, pp. 683–692.

[29] F. C. Morabito, M. Campolo, C. Ieracitano, J. M. Ebadi, L. Bonanno, A. Bramanti, S. Desalvo, N. Mammone, 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, IEEE, 2016, pp. 1–6.

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

[31] A. Delorme, S. Makeig, EEGlab: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, Journal of neuroscience methods 134 (1) (2004) 9–21.

[32] C. Torrence, G. P. Compo, A practical guide to wavelet analysis, Bulletin of the American Meteorological Society 79 (1) (1998) 61–78.

[33] O. Faust, U. R. Acharya, H. Adeli, A. Adeli, Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis, Seizure 26 (2015) 56–64.

[34] C. Ieracitano, N. Mammone, A. Bramanti, S. Marino, A. Hussain, F. C. Morabito, A time-frequency based machine learning system for brain states classification via EEG signal processing, in: 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, 2019.

[35] C. L. Nikias, M. R. Raghuveer, Bispectrum estimation: A digital signal processing framework, Proceedings of the IEEE 75 (7) (1987) 869–891.

[36] N. Kumar, K. Khaund, S. M. Hazarika, Bispectral analysis of EEG for emotion recognition, Procedia Computer Science 84 (2016) 31–35.

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

[38] D. W. Hosmer Jr, S. Lemeshow, R. X. Sturdivant, Applied logistic regression, Vol. 398, John Wiley & Sons, 2013.

[39] O. Bousquet, S. Boucheron, G. Lugosi, Introduction to statistical learning theory, in: Advanced Lectures on Machine Learning, Springer, 2004, pp. 169–207.

[40] J. A. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Processing Letters 9 (3) (1999) 293–300.

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

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

[43] F. C. Morabito, M. Campolo, D. Labate, G. Morabito, L. Bonanno, A. Bramanti, S. De Salvo, A. Marra, P. Bramanti, A longitudinal EEG study of Alzheimer's disease progression based on a complex network approach, International Journal of Neural Systems 25 (02) (2015) 1550005.

[44] E. Capecci, F. C. Morabito, M. Campolo, N. Mammone, D. Labate, N. Kasabov, A feasibility study of using the neucube spiking neural network architecture for modelling Alzheimer's disease EEG data, in: Advances in Neural Networks: Computational and Theoretical Issues, Springer, 2015, pp. 159–172.

[45] N. K. Kasabov, Spiking neural networks for deep learning and knowledge representation, Neural Networks: The Official Journal of the International Neural Network Society 119 (2019) 341.

[46] N. K. Kasabov, Time-space, spiking neural networks and brain-inspired artificial intelligence, Vol. 7, Springer, 2018.

[47] M. Mahmud, M. S. Kaiser, A. Hussain, S. Vassanelli, Applications of deep learning and reinforcement learning to biological data, IEEE Transactions on Neural Networks and Learning Systems 29 (6) (2018) 2063–2079.

[48] E. B. Assi, L. Gagliano, S. Rihana, D. K. Nguyen, M. Sawan, Bispectrum features and multilayer perceptron classifier to enhance seizure prediction, Scientific Reports 8 (1) (2018) 15491.

[49] L. Wang, B. Jiang, Z. Tu, A. Hussain, J. Tang, Robust pixelwise saliency detection via progressive graph rankings, Neurocomputing 329 (2019) 433–446.

Table 1: Classification performance of the proposed classifiers evaluated on the test sets, when the only-CWT features vector is used. All the results are quantified as average value ± standard deviation.

only-CWT based classsification				
AD vs HC				
Method	Precision	Recall	F-score	Accuracy
AE_1	87.87 ± 1.5 %	84.07 ± 1.5 %	85.93 ± 0.9 %	91.84 ± 0.5 %
AE_2	88.19 ± 0.9 %	85.64 ± 1.2 %	86.90 ± 0.8 %	92.35 ± 0.5 %
MLP_1	92.96 ± 1.1 %	92.72 ± 1.7 %	92.84 ± 1.0 %	95.76 ± 0.6 %
MLP_2	92.24 ± 1.1 %	92.40 ± 0.9 %	92.32 ± 0.7 %	95.45 ± 0.4 %
LR	88.19 ± 1.2 %	78.29 ± 1.5 %	81.19 ± 1.0 %	89.25 ± 0.5 %
SVM	83.75 ± 1.3 %	75.85 ± 1.1 %	79.60 ± 0.8 %	88.48 ± 0.5 %
AD vs. MCI				
Method	Precision	Recall	F-score	Accuracy
AE_1	71.28 ± 1.1 %	59.53 ± 2.0 %	64.88 ± 1.4 %	76.60 ± 0.8 %
AE_2	72.47 ± 1.5 %	61.10 ± 2.0 %	66.30 ± 1.4 %	77.45 ± 0.8 %
MLP_1	82.52 ± 1.9 %	80.89 ± 1.7 %	81.70 ± 1.3 %	86.84 ± 1.0 %
MLP_2	79.51 ± 2.0 %	77.95 ± 1.6 %	78.73 ± 1.4 %	84.70 ± 1.1 %
LR	72.47 ± 1.2 %	50.30 ± 1.5 %	56.50 ± 1.2 %	71.87 ± 0.7 %
SVM	64.20 ± 1.6 %	42.25 ± 1.6 %	50.97 ± 1.3 %	70.48 ± 0.7 %
MCI vs. HC				
Method	Precision	Recall	F-score	Accuracy
AE_1	83.54 ± 2.0 %	75.07 ± 2.8 %	79.08 ± 2.4 %	83.13 ± 1.8 %
AE_2	84.17 ± 1.6 %	76.56 ± 1.7 %	80.18 ± 1.4 %	83.93 ± 1.1 %
MLP_1	90.62 ± 1.2 %	90.03 ± 1.2 %	90.32 ± 1.1 %	91.80 ± 0.9 %
MLP_2	90.37 ± 1.0 %	89.62 ± 0.9 %	89.99 ± 0.6 %	91.53 ± 0.5 %
LR	84.17 ± 1.0 %	68.98 ± 0.7 %	72.76 ± 0.6 %	78.07 ± 0.5 %
SVM	78.61 ± 0.9 %	68.19 ± 0.7 %	73.03 ± 0.6 %	78.61 ± 0.5 %
AD vs. MCI vs. HC				
Method	Precision	Recall	F-score	Accuracy
AE_1	58.88 ± 3.6 %	42.72 ± 4.4 %	49.52 ± 3.9 %	69.96 ± 1.5 %
AE_2	58.11 ± 1.9 %	43.04 ± 4.5 %	49.45 ± 3.3 %	70.61 ± 0.8 %
MLP_1	72.54 ± 1.7 %	68.10 ± 1.2 %	70.25 ± 1.3 %	82.13 ± 0.5 %
MLP_2	70.32 ± 1.7 %	66.19 ± 2.1 %	68.19 ± 1.4 %	80.93 ± 0.6 %
LR	58.11 ± 1.1 %	36.34 ± 0.9 %	42.99 ± 0.8 %	60.08 ± 0.6 %
SVM	57.06 ± 2.7 %	38.61 ± 0.9 %	46.06 ± 1.1 %	66.95 ± 0.6 %

Table 2: Classification performance of the proposed classifiers evaluated on the test sets, when the only-BiS features vector is used. All the results are quantified as average value ± standard deviation.

only-BiS based classsification				
AD vs. HC				
Method	Precision	Recall	F-score	Accuracy
AE_1	71.66 ± 2.1 %	69.48 ± 2.0 %	70.56 ± 1.4 %	82.82 ± 0.9 %
AE_2	70.86 ± 2.1 %	70.06 ± 1.5 %	70.46 ± 1.5 %	82.59 ± 1.0 %
MLP_1	78.81 ± 1.7 %	79.41 ± 1.9 %	79.15 ± 1.6 %	87.59 ± 0.9 %
MLP_2	78.14 ± 1.7 %	79.24 ± 1.3 %	78.69 ± 1.3 %	87.28 ± 0.8 %
LR	70.86 ± 1.8 %	69.03 ± 1.9 %	70.67 ± 1.5 %	83.03 ± 0.9 %
SVM	69.00 ± 1.7 %	69.64 ± 1.9 %	69.32 ± 1.4 %	81.74 ± 0.9 %
AD vs. MCI				
Method	Precision	Recall	F-score	Accuracy
AE_1	64.64 ± 1.6 %	51.26 ± 2.5 %	57.18 ± 2.0 %	72.12 ± 1.0 %
AE_2	63.38 ± 1.6 %	52.67 ± 1.6 %	57.53 ± 1.4 %	71.76 ± 0.9 %
MLP_1	73.26 ± 1.5 %	68.04 ± 1.7 %	70.55 ± 1.2 %	79.38 ± 0.8 %
MLP_2	72.31 ± 1.9 %	69.24 ± 1.8 %	70.74 ± 1.7 %	79.20 ± 1.2 %
LR	63.38 ± 1.8 %	53.91 ± 1.0 %	58.97 ± 1.3 %	72.76 ± 1.0 %
SVM	63.84 ± 1.8 %	46.63 ± 1.3 %	53.89 ± 1.2 %	71.03 ± 0.8 %
MCI vs. HC				
Method	Precision	Recall	F-score	Accuracy
AE_1	75.58 ± 1.0 %	75.56 ± 1.5 %	75.57 ± 0.8 %	79.25 ± 0.6 %
AE_2	75.53 ± 1.3 %	76.74 ± 1.2 %	76.13 ± 1.1 %	79.56 ± 1.0 %
MLP_1	83.88 ± 0.8 %	86.32 ± 1.2 %	85.08 ± 0.8 %	87.14 ± 0.7 %
MLP_2	83.42 ± 0.9 %	86.24 ± 0.9 %	84.81 ± 0.8 %	86.88 ± 0.7 %
LR	75.53 ± 0.9 %	70.90 ± 1.7 %	71.88 ± 0.9 %	76.43 ± 0.6 %
SVM	72.15 ± 1.3 %	71.60 ± 1.6 %	71.87 ± 1.0 %	76.20 ± 0.8 %
AD vs. MCI vs. HC				
Method	Precision	Recall	F-score	Accuracy
AE_1	48.58 ± 1.9 %	35.77 ± 2.8 %	41.20 ± 1.9 %	65.54 ± 0.6 %
AE_2	48.58 ± 2.0 %	33.77 ± 4.4 %	39.85 ± 3.5 %	65.75 ± 1.2 %
MLP_1	58.75 ± 2.3 %	51.99 ± 2.3 %	55.17 ± 2.2 %	74.75 ± 0.9 %
MLP_2	56.33 ± 2.6 %	49.18 ± 2.7 %	52.51 ± 2.4 %	73.59 ± 1.2 %
LR	48.58 ± 1.9 %	10.07 ± 0.6 %	16.77 ± 0.8 %	56.49 ± 0.8 %
SVM	48.85 ± 1.3 %	38.03 ± 1.5 %	42.76 ± 1.4 %	64.20 ± 0.7 %

Table 3: Classification performance of the proposed classifiers evaluated on the test sets, when the multi-modal (CWT+BiS) features vector is used. All the results are quantified as average value ± standard deviation.

multi-modal (CWT+BiS) based classsification				
AD vs. HC				
Method	Precision	Recall	F-score	Accuracy
AE_3	90.02 ± 0.9 %	89.10 ± 1.1 %	89.56 ± 0.5 %	93.85 ± 0.3 %
AE_4	89.97 ± 1.1 %	89.74 ± 1.2 %	89.86 ± 0.8 %	94.00 ± 0.5 %
MLP_3	94.72 ± 1.0 %	94.91 ± 1.1 %	94.85 ± 0.9 %	96.95 ± 0.5 %
MLP_4	94.38 ± 2.1 %	94.56 ± 1.8 %	94.46 ± 1.8 %	96.71 ± 1.1 %
LR	89.97 ± 1.3 %	87.51 ± 1.4 %	88.43 ± 1.0 %	93.21 ± 0.6 %
SVM	86.82 ± 1.0 %	85.60 ± 0.7 %	86.21 ± 0.8 %	91.88 ± 0.5 %
AD vs. MCI				
Method	Precision	Recall	F-score	Accuracy
AE_3	73.22 ± 1.2 %	65.56 ± 2.7 %	69.18 ± 1.8 %	78.79 ± 1.0 %
AE_4	73.05 ± 1.7 %	67.97 ± 2.1 %	70.42 ± 1.4 %	79.26 ± 0.9 %
MLP_3	87.14 ± 1.0 %	85.71 ± 1.9 %	86.46 ± 1.1 %	90.24 ± 0.7 %
MLP_4	86.08 ± 1.6 %	85.83 ± 1.5 %	85.95 ± 1.3 %	89.81 ± 1.0 %
LR	73.05 ± 1.2 %	64.23 ± 1.5 %	66.63 ± 1.2 %	76.64 ± 0.8 %
SVM	68.97 ± 0.9 %	60.89 ± 1.3 %	64.68 ± 0.9 %	75.85 ± 0.5 %
MCI vs. HC				
Method	Precision	Recall	F-score	Accuracy
AE_3	89.00 ± 1.3 %	88.65 ± 1.2 %	88.83 ± 0.9 %	90.53 ± 0.8 %
AE_4	89.45 ± 1.6 %	88.21 ± 1.4 %	88.83 ± 1.0 %	90.57 ± 0.9 %
MLP_3	95.31 ± 0.8 %	95.86 ± 0.6 %	95.58 ± 0.6 %	96.24 ± 0.5 %
MLP_4	94.95 ± 0.8 %	95.44 ± 0.5 %	95.19 ± 0.3 %	95.91 ± 0.3 %
LR	89.45 ± 0.9 %	81.22 ± 0.6 %	82.03 ± 0.5 %	84.89 ± 0.5 %
SVM	82.24 ± 0.8 %	80.75 ± 1.0 %	81.48 ± 0.8 %	84.41 ± 0.7 %
AD vs. MCI vs. HC				
Method	Precision	Recall	F-score	Accuracy
AE_3	64.87 ± 2.0 %	55.85 ± 1.4 %	62.00 ± 0.7 %	76.91 ± 2.1 %
AE_4	63.66 ± 1.8 %	55.89 ± 2.7 %	59.52 ± 2.1 %	77.30 ± 1.4 %
MLP_3	80.74 ± 1.5 %	80.99 ± 2.0 %	80.87 ± 1.4 %	89.22 ± 0.7 %
MLP_4	79.92 ± 1.4 %	78.21 ± 2.0 %	79.10 ± 1.5 %	88.56 ± 0.7 %
LR	63.66 ± 1.2 %	47.52 ± 1.4 %	49.96 ± 1.0 %	65.35 ± 0.6 %
SVM	63.86 ± 1.4 %	56.74 ± 1.3 %	60.09 ± 1.0 %	74.54 ± 0.5 %