A Convolutional Neural Network approach for Classification of Dementia Stages based on 2D-Spectral Representation of EEG recordings

Cosimo Ieracitano[a,*], Nadia Mammone[b], Alessia Bramanti[c], Amir Hussain[d], 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] *Institute of Applied Sciences and Intelligent Systems Eduardo Caianiello (ISASI), National Research Council (CNR), Messina (Italy)*
[d] *Division of Computing Science & Maths, University of Stirling, FK9 4LA Stirling (UK)*

Abstract

A data-driven machine deep learning approach is proposed for differentiating subjects with Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI) and Healthy Control (HC), by only analyzing noninvasive scalp EEG recordings. The methodology here proposed consists of evaluating the power spectral density (PSD) of the 19-channels EEG traces and representing the related spectral profiles into 2-d gray scale images (PSD-images). A customized Convolutional Neural Network with one processing module of convolution, Rectified Linear Units (ReLu) and pooling layer (CNN_1) is designed to extract from PSD-images some suitable features and to perform the corresponding two and three-ways classification tasks. The resulting CNN is shown to provide better classification performance when compared to more conventional learning machines; indeed, it achieves an average accuracy of 89.8% in binary classification and of 83.3% in three-ways classification. These results encourage the use of deep processing systems (here, an engineered first stage, namely the PSD-image extraction, and a second or multiple CNN stage) in challenging clinical frameworks.

Keywords: Deep Learning, Convolutional Neural Network, Power Spectral Density, Alzheimer's disease, Mild Cognitive Impairment

*Corresponding author
Email addresses: cosimo.ieracitano@unirc.it (Cosimo Ieracitano), nadia.mammone@irccsme.it (Nadia Mammone), alessia.bramanti@gmail.com (Alessia Bramanti), ahussain@ieee.org (Amir Hussain), morabito@unirc.it (Francesco C. Morabito)

1. Introduction

Alzheimer's disease (AD) is the most common form of dementia that occurs in older individuals, causing cognitive and functional deficits. The prodromal stage of AD is known as Mild Cognitive Impairment (MCI) due to AD [1], a neurological disorder characterized by a mild decline in mental abilities that does not interfere with the autonomy of the subject. Actually, MCI patients may remain stable, regress to a normal condition, whereas 10-15 % (per year) of them may develop dementia due to AD, which is a heavy progressive neurodegenerative disorder. The number of subjects affected by AD has been drastically increasing over years and it is expected to reach 15 million by 2050 [2]. In the last two decades, Electroencephalography (EEG) has been employed as a promising tool for early screening and to assist diagnosing AD. To this end, several EEG-based classification algorithms have been developed [5-11]. EEG is a low-cost, noninvasive network of measurement, which allows for the recording of the brain's electrical activity. The EEG is commonly decomposed into four, neurophysiological sub-bands: δ (0.5-4 Hz), θ (4-8 Hz), α (8-12 Hz), β (12-32 Hz). It has been shown [2], [3], that abnormalities in EEG recordings, such as slowing of the rhythms, loss of complexity and altered synchronization between channels, may be indicative of brain degeneration due to AD, although not specific of the disease. A significant amount of research has been focused on the early detection of AD but it is still a challenging open issue. The EEG of AD and MCI patients shows slowing effects in the brain EEG rhythms, as the AD progression is related to the relative prevalence of low frequencies (delta and theta bands) with respect to high frequencies (alpha and beta) [2]. This evaluation is carried out in a clinical setting by visually inspecting the EEG recordings, sometimes also by computing the relative power in the EEG sub-bands, by using proper functionality provided by the most common EEG review software. In particular, the Power Spectral Density (PSD), which describes the power frequency distribution of a signal, is routinely used. [4]. Empirically defined features based on the PSD have been proposed in the literature and machine learning (ML) techniques have been widely employed to classify EEGs of cognitively impaired subjects [5]. However, standard ML methods are based on the explicit definition of features and are not suitable to process high-dimensional volumes of data. Deep learning (DL) is an advanced ML technique able to overcome the aforementioned limitations; indeed, it extracts the most relevant features directly from raw input data. Since DL has been turning out to be the key solution for many real-world applications (i.e. biochemistry [6],[7], biomedicine [8], [9], image detection [10], cybersecurity [11]), recently, DL driven AD detection systems have been emerging. However, such systems are mostly based on neuroimaging analysis (structural and functional magnetic resonance imaging (sMRI, fMRI)) especially gathered from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database and only a few DL studies address the discrimination of AD referring to EEG recordings.
In this context, a dataset of 189 EEGs (63 AD, 63 MCI, 63 HC) were collected at IRCSS Centro Neurolesi Bonino-Pulejo of Messina (Italy) and here used to

develop a DL based system. Specifically, in this study, we propose a data-driven Convolutional Neural Network (CNN) framework capable of differentiating the PSD spectrograms extracted from the EEGs of AD, MCI and HC subjects. The proposed system includes an EEG signal transform into the frequency domain through the PSD estimation and a CNN architecture with a single stage of convolutional, ReLu and pooling layer (CNN_1), able to extract significant features after mapping the spectral profiles into 2-d gray scale images (PSD-images). It is to be noted that ReLu activation function has been utilized in this study, because it alleviates the vanishing gradient problem and introduces sparse representations (as it sets to zeros negative values) that consequently lead to better prevent the over-fitting and to a significant reduction of the free parameters of the network [12],[13],[14]. The proposed CNN_1 is compared with standard classifiers (Multi Layer Perceptron (MLP), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA)), all of them trained over two different kind of input: 1) the raw PSD spectrograms; 2) handcrafted features manually extracted from the PSD spectrograms. Experimental results showed that the proposed CNN_1 outperformed all other approaches achieving average accuracy rate up to 89.8% in binary (AD vs HC, MCI vs HC, AD vs MCI) and 83.3% in 3-ways (AD vs MCI vs HC) classification.
The list of contributions of the present paper can be summarized as follows:

- development of an innovative method based on 2-d spectral representations (PSD images) of EEG data;
- development of a data-driven DL approach based on a computationally-efficient Convolutional Neural Network (CNN_1) for classifying AD, MCI and HC subjects by only analyzing noninvasive scalp EEG recordings;
- development of a system with potential for clinical deployment in real-world applications.

The paper is organized as follows: in Section 2 the related literature review is discussed; in Section 3 the available EEG recordings are described and the proposed methodology is introduced. Section 4 illustrates the architecture of the proposed CNN_1. In Section 5 the achieved experimental results are presented and discussed. Section 6 and Section 7 address the discussion and conclusions, respectively.

2. Related works

Features-based approaches have been employed to aid the diagnosis of Alzheimer by using state-of-the-art ML algorithms (i.e. LDA, SVM, ANNs). Anderer et al. [15] and Pritchard et al. [16] reported significant classification performance (up to 90%) between AD patients and normal individuals, by using EEG markers as input to an ANN. Trambaiolli et al. [17] evaluated different set of features based on coherence and classified AD patients and healthy control

subjects (HC) through SVM, with a 79.9% accuracy. Huang et al. [18] analyzed, in the frequency domain, 30 sec eye-closed resting state EEG epochs recorded from 38 mild AD, 31 MCI and 24 HC. They observed that alpha and theta global frequency power (GFP) were the best discriminating indicators in AD/HC and AD/MCI classification, reporting 84% and 78% correct classification, respectively. Buscema et al. [19] developed a novel system based on ANNs, referred to as IFAST (Implicit Function As Squashing Time), to discriminate eye-closed/eye-open resting state EEG segments recorded from AD and MCI patients. They used as input of the model the connections weights of a nonlinear auto-associative ANN trained to reproduce the recorded EEG tracks, reaching 92.33% accuracy. Rossini et al. [20] tested IFAST procedure for the automatic classification of individual normal elderly (Nold) and MCI patients, achieving 93.46 % accuracy. McBride et al.[21] used regional spectral and complexity features in a SVM model to discriminate the EEG of 15 normal controls, 16 early MCI and 17 early stage AD, recorded according to a specific protocol, by including different cognitive and movement tasks. The three-way classifier provided accuracies of 83.3%, 85.4%, and 79.2% for the eye-open resting, eye-closed counting and eye-closed resting states, respectively.

Although the discussed models provided high classification performance, the literature lacks a clear and standardized EEG recording protocol for AD/MCI classification. Often, the analyzed cohorts of subjects are either small or the length of the EEG recordings is not sufficient (just a few seconds), which makes it difficult to compare the performance of the different methods and to replicate the experiments. One of the main drawbacks of the methods proposed in the literature is also the requirement for handcrafted features extraction, yielding a set of features not easily interpretable by clinicians. This issue can be overcome by using mixed DL techniques, that are able to introduce data-driven methods on well accepted data representations, like spectral data [22]. CNN is a DL-architecture which avoids the explicit definition of features and learns significant representations hierarchically from the input data. As CNNs provided impressive results, especially in image classification problems [12], [23], recently, several CNN models based on neuroimages (structural and functional magnetic resonance imaging (sMRI, fMRI)) have been emerging also in the early detection of Alzheimer's disease. In this context, Payan et al. [24] combined sparse autoencoder (AE) and 3D-CNN to classify MRI scans of HC, MCI, AD subjects gathered from ADNI dataset. Experimental results showed that the 3D-CNN outperformed others methods reporting accuracies of 95.39%, 86.64%,92.11%, 89.47% in AD vs HC, AD vs MCI, MCI vs HC and AD vs MCI vs HC classification, respectively. Similarly, Hosseini et al. [25] proposed a 3D-CNN based on a 3D convolutional AE (CAE). In this work, the authors pretrained the 3D-CAE on CADDementia Dataset to extract generic and transferable AD feature from MRI data. Then, the performance of the 3D-CNN proposed were evaluated on the ADNI dataset, reporting accuracies of 97.6%, 95.1%,90.8%, 89.1% in AD vs HC, AD vs MCI, MCI vs HC and AD vs MCI vs HC classification, respectively. Aderghal et al. [26] proposed a multi-projection fusion approach by using 2D-CNNs trained on a subset of the ADNI database. The authors studied the

binary classifications AD vs HC, AD vs MCI, MCI vs HC achieving accuracies of 91.4%, 69.5% and 65.6% respectively. Sarraf et al. [27] focused only on AD vs HC discrimination of fMRI data and achieved 96.85% accuracy by using the famous CNN based LeNet-5 architecture; whereas Billones et al. [28] used a modified 16-layers VGG network trained on 900 MRI images of ADNI, reaching accuracy rates up to 98.3%, 93.9%, 91.7% in AD vs HC, AD vs MCI and MCI vs HC discrimination.
Although the discussed CNN based solutions trained on MRI/fMRI images produced good classification performance, they require highly complex algorithms and expensive infrastructures. A few deep learning studies based on EEGs of AD patients are instead reported in literature. In [29] Morabito et al. proposed a stacked AE architecture able to extract the most significant features from the EEG time-frequency representation and discriminate four different groups: 26 normal subjects, 13 AD, 20 Creutzfeldt-Jakob Disease (CJD) and 17 rapidly progressive dementia (RPD) patients. The proposed AE model achievied average accuracy, sensitivity, specificity of 88%, 94% and 85% in CJD vs AD classification. Comparable results were obtained for CJD vs RPD and CJD vs HC discriminations. In [30] Morabito et al. proposed a data-driven CNN architecture based on mean, standard deviation, skewness variables of the EEG time-frequency representation to distinguish AD from MCI and HC. Experimental results showed appreciable performances (binary average accuracy of 82.66% and 3-ways accuracy of 82%) but a small subset of EEG recordings (23 AD, 23 MCI, 23 HC) was used.
In this paper, 189 EEG recordings (63 AD, 63 MCI, 63 HC) have been analyzed and a data-driven CNN has been developed to extract automatically relevant features from gray scale PSD images for two-ways (AD vs HC, AD vs MCI, MCI vs HC) and three-ways (AD vs MCI vs HC) classification purposes.

3. Method

The flowchart of the proposed method is pictorially represented in Figure 1: (1) acquisition of the n-channels EEG recording; (2) segmentation of the EEG signals into M non-overlapping epochs; (3) given the $\varepsilon - th\ EEG^{\varepsilon}$ epoch under analysis (with $\varepsilon = 1, 2, .., M$), the PSD is estimated in the range 0.5-32 for each EEG channel, ending up with a matrix PSD^{ε}, sized 19 x F (where F=159, is the number of discrete frequency values in the range [0.5-32] Hz); (4) the PSD^{ε} matrix of the epoch under analysis is converted into a 2-d gray scale representation sized 19 x F, denoted as $PSD\ image$ of epoch ε; (5) the 2-d PSD-images are used as input to a customized Convolutional Neural Network characterized by 1 convolution layer, 1 max pooling layer, 1 fully connected layer followed by a softmax layer which performs the 2-ways (AD vs HC, AD vs MCI, MCI vs HC) or 3-ways (AD vs MCI vs HC) classification. First, the CNN proposed is trained with the PSD-images of the subjects. At this stage, the ability of the network to correctly classify the PSD-images (i.e. the corresponding EEG epochs) as belonging to AD, MCI or HC subjects, is tested (*PSD epoch-based classification*). Successively, epochs of the same patient are

used as input to the trained CNN, performing an overall classification per patient (*patient-based classification*). In this case, taking into account all the epochs of a subject, if the number of PSD-images (i. e. epochs) classified by the network as class c (where c can be AD, MCI or HC) is larger than 50%, then the subject is assigned to class c.

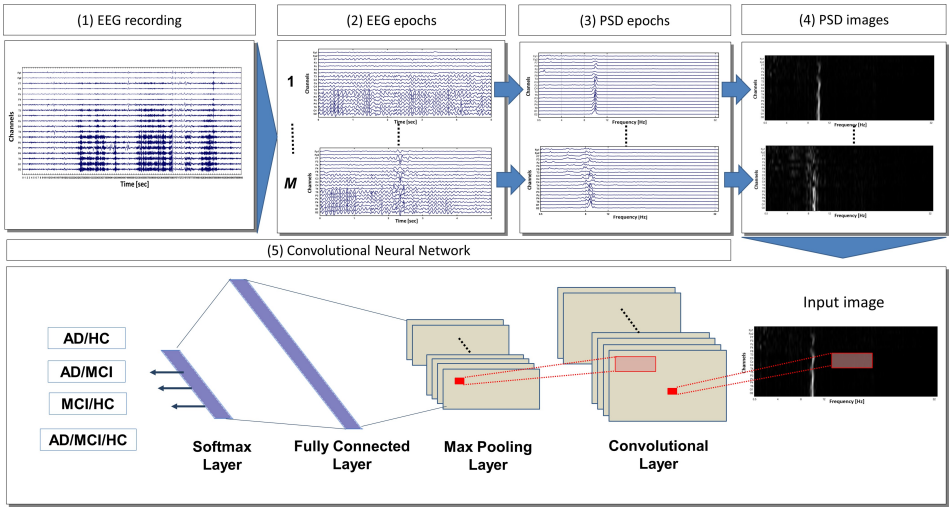

Figure 1: The EEG is recorded and stored on a computer. Then, the EEG is divided into M non-overlapping epochs of 5 sec each. For every EEG^{ε} ($\varepsilon = 1, 2, .., M$) epoch, the PSD over each channel is estimated, coming up with the corresponding PSD^{ε} matrix of spectral density profiles for epoch ε, PSD^{ε}. All PSD^{ε} matrices are then converted into 2-d gray scale images (PSD^{ε} *images*) by mapping each element of the matrix into an intensity value ranged between 0 (black pixel) and 1 (white pixel). The PSD-image is the input of a Convolutional Neural Networkthat includes 1 convolutional layer (+ ReLu), 1 max pooling layer, 1 fully connected layer and a softmax layer which performs the classification tasks (two ways: AD vs HC, AD vs MCI, MCI vs HC or three ways: AD vs MCI vs HC).

3.1. EEG recording and preprocessing

Study population. 189 subjects were enrolled at IRCCS Centro Neurolesi Bonino-Pulejo of Messina (Italy): 63 affected by Alzheimer's Disease, 63 affected by Mild Cognitive Impairment, 63 healthy controls. A clinical protocol has been approved by the local Ethics Committee of the IRCCS Centro Neurolesi Bonino-Pulejo. The inclusion criteria was diagnosis of AD or MCI, according to the guidelines of the Diagnostic and Statistical Manual of Mental Disorders (fifth edition, DSM-5) [31]. The Exclusion criteria were: evidence of other psychiatric or neurological disorders that might cause cognitive impairment, uncontrolled or complex systemic disorders, EEG epileptiform patterns, traumatic brain injuries. The protocol, the goals, the risks and the benefits of the present study were described to the patients and their caregivers, who signed an informed consent form. Every subject underwent a neuroimaging examination, in order to exclude other possible pathological conditions (stroke, traumatic brain injury,

hydrocephalus or other neurological conditions) which may cause symptoms similar to AD. The possible effects of any medication like cholinesterase inhibitors (ChEis), Memantine, anti-depressants, anti-psychotics and anti-epileptic drugs, were considered in AD patients. MCI subjects were under no medical treatment.

EEG recording. The EEG was recorded in the morning. The patients and their caregivers were interviewed about the quality and length of the last sleep as well as about the last meal. During the EEG recording, subjects sat in a comfortable, eye-closed resting state, but did not fall asleep, as confirmed by the EEG experts who continuously monitored the EEG traces so that any sleep pattern could be promptly detected. The EEG segments affected by artifacts were marked by the EEG experts and later discarded. The EEG was recorded according to the standard 10-20 International System with the following set up: 19-channels montage (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz and Pz), sampling frequency of 1024 Hz, notch filter at 50 Hz and linked earlobe (A1-A2) reference.

EEG preprocessing. The n-channel EEG recording was band-pass filtered at 0.5 - 32 Hz through the *eegfiltfft* function implemented in the Matlab toolbox *EEGLab* ([32]) which is based on the inverse Fast Fourier Transform (FFT). Every EEG signal was downsampled to 256 Hz and then partitioned into M non-overlapping epochs (of 5 sec length), since f_s\$=256 Hz, each epoch included N=1280 samples. Therefore, for every subject under consideration, the recorded epochs EEG^{ε} ($\varepsilon = 1, 2, .., M$), sized n x N, were stored on a computer and processed one at a time. The average EEG length, after artefactual segments rejection, was 4.1 min.

3.2. Power Spectral Density

Power spectral density (PSD) measures the frequency distribution of the informative content of a signal [33]. The conventional estimate of PSD is known as *periodogram*. It is a nonparametric evaluation of PSD that represents the Fourier transform of the autocorrelation function. Specifically, the periodogram of a signal x_l of length L, is defined as follow:

$$P(f) = \frac{T_s}{L} |\sum_{l=0}^{L-1} x_l e^{-2\pi i f l}|^2 \tag{1}$$

where $-\frac{1}{2T_s} < f \leq \frac{1}{2T_s}$ and T_s is the sampling interval. When the input time-series is multiplied by a window function w_l, the *modified periodogram* is obtained:

$$\hat{P}(f) = \frac{T_s}{L} |\sum_{l=0}^{L-1} w_l x_l e^{-2\pi i f l}|^2 Subjectto: \tag{2}$$

The modified periodogram smooths the edges of the signal under analysis and minimize the spectral leakage in the standard periodogram. Here, the PSD of the x^{th} EEG time series, was computed by using the modified periodogram with

rectangular windowing function, as it achieves good resolution and is well-suited for biomedical signals analysis [34] .

3.2.1. EEG Spectral profiles

PSD is commonly used in clinical practice to analyze spectral changes in EEG recordings of HC subjects and MCI/AD patients. Indeed, it is employed to roughly estimate the power of EEG signals within every sub-band, in order to evaluate the predominance of relatively lower (delta and theta) or higher (alpha and beta) frequencies. The main properties of typical HC, MCI, AD spectral profiles are listed below:

- HC-spectrum shape (of normal subjects in resting state): commonly characterized by a power peak in the alpha sub-band (particularly in the parieto-occipital electrodes) and low powers in slower and faster frequencies;
- MCI-spectrum shape: commonly characterized by a shift to the left of the alpha peak and an exponential decreasing of power toward the higher frequencies;
- AD-spectrum shape: commonly characterized by a substantial decreasing of alpha power and a significant increasing of delta power.

Such spectral profiles have been widely used to extract empirically defined features and classify dementia stages through conventional classification techniques [40,42].

3.3. 2-d Spectral Representation (PSD-image)

In this study, the EEG spectral profiles are instead represented as images and advanced ML techniques (CNN) are employed to carry out the automatic features extraction and classification. Given the PSD^{ε} matrix under analysis, the $p_{i,j}$ (with i=1,..19; j=1,..,159) element was mapped into an intensity value ranged between 0 - 1 (where 0 (black pixel) and 1 (white pixel) correspond to the minimum and maximum of the PSD^{ε} matrix, respectively) coming up with a 2-d gray scale image. Such PSD-images were firstly evaluated by visual inspection: they looked generally dark with bright bars at the frequencies corresponding to PSD peaks. This property was due to the inherent sparse nature of PSD^{ε} matrix. Fig. 2 shows an example of the conversion of a matrix PSD^{ε} (every row represents the PSD vs frequency profile of a given channel) into 2-d spectral representation. The epoch belongs to a HC. As can be observed, it exhibits a white vertical bar at 8-12 Hz (alpha band), typically observed in the eye-closed resting state EEG of healthy subjects. However, it is to be noted that the PSD-images of the epochs of the same patient can vary, due to the non-stationarity of the EEG signals, the possible presence of artifacts or unexpected electrical patterns. Such abnormalities affect the PSD-images, which may cause errors in the classification process.

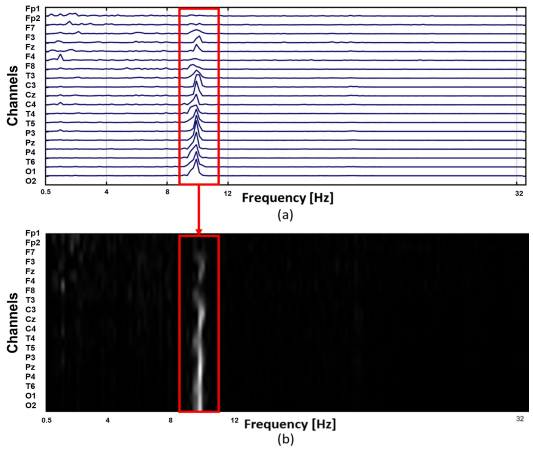

Figure 2: PSD-image conversion of an EEG epoch of an healthy control individual. (a) Power spectral profiles of each EEG channel. (b) 2-d gray scale PSD-image. Dark pixels correspond to low power values; brighter pixels correspond to high power values. Since a typical EEG epoch of a healthy subject is characterized by a dominant alpha activity, the PSD-image presents brighter pixels between 8-12 Hz. Alpha peak is boxed in red.

4. Convolutional Neural Network

A standard CNN is composed of two main building blocks: a feature extractor, which extracts features from raw data automatically through one or multiple stages of convolution, activation and pooling layers; and a fully connected multi-layer neural network which performs the classification task by using the features learned previously. The details of each level are now described briefly.
Convolution Layer. Let $X_i \in R^{hxw}$ be the $2d$-input maps and let $K_j \in R^{k_1 x k_2}$ be the weights matrices called *kernels* (or filters). The convolution operation is expressed as:

$$Y_j = \sum X_i * K_j + B_j \tag{3}$$

where B_j is the bias term and * denotes the convolution operator. The j^{th} filter convolves with a local region of X_i (called *receptive field*) and then it slides over the whole input map with a stride s by sharing the same set of weights. The output Y_j of each filter is referred as *features map* and its size is y_1 x y_2 where:

$$y_1 = \frac{h - k_1 + 2 \times p}{s} + 1 \tag{4}$$

and

$$y_2 = \frac{w - k_2 + 2 \times p}{s} + 1 \tag{5}$$

p is the zero padding parameter and it is typically used to control the output size by padding the input edges with zeros. In this study the zero padding was

applied and p was set to 1 so that the input and output spatial resolution are the same.

Activation Layer. It consists of a nonlinear transfer function which usually follows the convolutional layer. In standard Neural Networks, *sigmoid* ($f(x) = \frac{1}{1+e^{-x}}$) or *hyperbolic tangent* ($f(x) = tanh(x)$) are widely employed as activation functions. Since recent studies have proved that a novel function called "*Rectified Linear Units*" (ReLu, $f(x) = max(0, x)$) provides better performance in terms of generalization and learning time [12],[13],[14] especially in CNN applications, also in the architecture proposed it was applied to Y_j feature maps extracted from the previous layer.

Pooling Layer. It reduces the resolution of the input feature maps (Y_j) by taking the average (*average pooling*) or maximum (*max pooling*) of neighbour elements selected with a filter sized $\tilde{k}_1$ x $\tilde{k}_2$. The filter scans the features map with stride $\tilde{s}$ producing an sub-sampled representation of Y_j sized $\tilde{y}_1$ x $\tilde{y}_2$, where:

$$\tilde{y}_1 = \frac{y_1 - \tilde{k}}{\tilde{s}} + 1 \tag{6}$$

and

$$\tilde{y}_2 = \frac{y_2 - \tilde{k}}{\tilde{s}} + 1 \tag{7}$$

*w*ith $\tilde{k} = \tilde{k}_1 = \tilde{k}_2$. Although the average pooling reduces the spatial dimension, similarly to max pooling, it propagates the mean of all the input values selected by the filter to the subsequent layer. As the main objective is to detect the PSD peak (that is the maximum spectral value) in the image, the average pooling operator could also take into account those features that may not be important. Furthermore, in [35] Scherer et al. showed empirically that the max pooling operation was significantly superior because it captures better invariant features and improves generalization performance. For these reasons, in this study, max pooling was chosen as sub-sampling operator.

Fully Connected Layer. The convolution (+ ReLu) and pooling layers are followed by one or more fully connected layers (FC). The neurons of this layer are connected to every unit of the previous layer as a standard NN. The output size of the last FC is equal to the number of possible classes.

4.1. Architecture Proposed

The architecture of the CNN proposed is shown in Figure 3. The network is designed to accept fixed-size images n x F (where n=19, number of EEG channels; F=159, number of frequencies ranged 0.5-32 Hz). The number of filters (K) and its size (k_1 x k_2), have been chosen empirically after several experimental tests. Here, the convolutional layer (Conv_1) has K=16 learnable filters sized 3 x 3. Each filter convolves with the input image with a stride s=1, producing 16 features maps with the same input size 19 x 159. Indeed, according to Eq. 4, 5:

$$y_1 = \frac{h - k_1 + 2 \times p}{s} + 1 = \frac{19 - 3 + 2 \times 1}{1} + 1 = 19 \tag{8}$$

and

$$y_2 = \frac{w - k_2 + 2 \times p}{s} + 1 = \frac{159 - 3 + 2 \times 1}{1} + 1 = 159 \quad (9)$$

At this stage the CNN has W_{conv}= K x k_1 x k_2 = 16 x 3 x 3 = 144 weights and B_{conv} = 16 biases, for a grand total of 144 + 16 = 160 learnable parameters. The convolutional layer is followed firstly by the ReLu and then by the max pooling layer (MaxPool_1) which reduces the features maps size from 19 x 159 to 9 x 79 by using 3 x 3 filters with a stride $\tilde{s}$=2. Indeed, according to Eq. 6, 7:

$$\tilde{y}_1 = \frac{y_1 - \tilde{k}}{\tilde{s}} + 1 = \frac{19 - 3}{2} + 1 = 9 \quad (10)$$

and

$$\tilde{y}_2 = \frac{y_2 - \tilde{k}}{\tilde{s}} + 1 = \frac{159 - 3}{2} + 1 = 79 \quad (11)$$

The features extracted are the input of a FC layer with U_{FC} =300 hidden units. The number of weights $W_{conv-FC}$ depends on the output size of the previous convolutional layer ($\tilde{y}_1$ x $\tilde{y}_2$), the number of filters (K) and the number of hidden units in the FC (U_{FC}); indeed, $W_{conv-FC}$= $\tilde{y}_1$ x $\tilde{y}_2$ x K x U_{FC} = 9 x 79 x 16 x 300 = 3412800 weights. In this case, the number of parameters is 3412800 + 300 (biases)= 3413100. The network ends with a soft-max (SF) layer to estimate the class predictions in binary and three way classification. In this layer, the number of weights (W_{SF}) depends on number of neurons of the previous FC layer (U_{FC}) and number of neurons of the SF layer (U_{SF}); consequently, W_{SF}= U_{FC} x U_{SF} = 300 x 2 = 600 weights for the 2-ways classification and W_{SF}= U_{FC} x U_{SF} = 300 x 3 = 900 weights for the 3-ways classification. As an example, Table 1 reports the overall number of learnable parameters of the 2-ways classifier.

Table 1: Total number of learnable parameters of CNN_1, for the binary classification.

Layer Name	**Size**	**Weights**	**Biases**	**Parameters**
Input	19x159	-	-	-
Conv_1	19x159x16	144	16	160
MaxPool_1	9x79x16	-	-	-
FC	300x1	3412800	300	3413100
SF	2x1	600	2	602
Total				**3413862**

4.2. Learning set up

Learning parameters were set up following practical recommendations of [36], [37]. The network was trained by the stochastic gradient descent (SGD) algorithm with $mini - batch$ learning method and mini-batch size of 91 [38]. Let $x_{1,2,..,k}$ be the samples (PSD-images) of the training dataset, a subset $x_{1,2,..,b}$, called mini-batch, was processed at each iteration. Momentum coefficient α=0.9 and weight decay rate λ=10^{-4} were included in order to speed up learning and reduce over fitting (α and λ were chosen through experimental tests). The

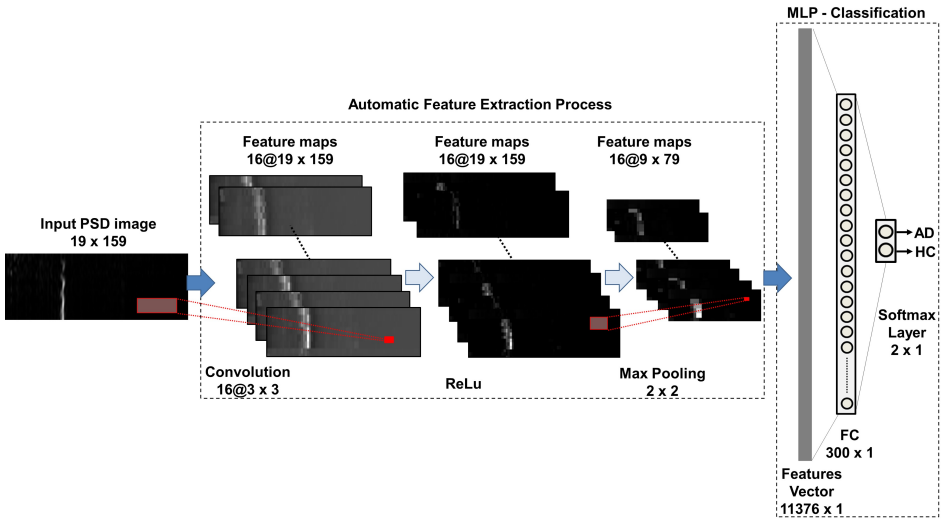

Figure 3: Convolutional Neural Network Architecture. It consists of an automatic feature extraction process and a standard MLP. First, the PSD-image convolves with 16 kernels sized 3 x 3, producing 16 features maps sized 19 x 159. Then, the ReLu function is applied to each feature map. The max pooling operation reduces the input image resolution, providing 16 features maps sized 9 x 79. At this stage, the feature maps are reshaped into a vector sized 16 x 9 x 79 = 11376 and processed by a standard Multi Layer Perceptron (MLP). It consists of a fully connected layer with 300 hidden neurons and a softmax layer which performs the binary or 3-way classification. In the figure, an example of the proposed architecture for binary classification is shown.

weights (w) were initialized from a zero-mean Gaussian distribution with standard deviation 0.01 and updated according to:

$$w_{t+1} = w_t - \eta \frac{1}{b} \sum_{i=1}^{b} \nabla J(x_i, w) + \alpha w_t - \lambda \eta w_t$$

where x_i is an example of the mini-batch under analysis, t is the iteration index, b is the mini-batch size, J is the cross-entropy loss function, η denotes the learning rate, ∇ the gradient operator and $\frac{1}{b} \sum_{i=1}^{b} \nabla J(x_i, w)$ is the gradient of the loss function averaged over a mini-batch. It is a good practice to start with high value of learning rate and then reduce it by a factor 3, until no divergence is observed [37]. Here, the optimal learning rate was reached at 0.0001 and the convergence was observed after about 2500 iterations. The network was implemented using MATLAB R2017b (The MathWorks, Inc., Natick, MA, USA) and executed on Intel(R) Xeon(R) CPU ES-2650 v3 processor with 128 GB RAM installed. The average processing time was 2024 secs (34 minutes).

5. Experimental results

As described in 3, the first goal was to assess the ability of the network to correctly classify PSD-images (i.e. EEG epochs) as belonging to class c (where

c can be AD, MCI or HC) (*epoch-based classification*). The second goal was to perform an overall subject's classification considering how his/her own epochs were labelled by the trained network (*patient-based classification*). Given a subject, if the number of PSD-images (i. e. epochs) labelled by the network as class c is larger than 50%, then the subject is assigned to class c.
The database included 189 EEG recordings. In the first stage, 117 EEG recordings (39 belonging to patients affected by AD, 39 belonging to patients affected by MCI and 39 belonging to healthy subjects) out of 189 were analyzed. EEG epochs of 5 sec were extracted and preprocessed to construct the dataset of PSD-images that was used to train and test the network. Successively, in order to carry out the overall patient-based classification, the trained CNN classifier was tested over the epochs of the remaining 72 EEG recordings (24 belonging to patients affected by AD, 24 belonging to patients affected by MCI and 24 belonging to healthy subjects).

5.1. Epoch-based Classification

Given the i^{th} (i=1,2..,117) EEG recording, M=20 EEG epochs of 5 sec were randomly extracted and converted into M PSD-images, for a grand total of 20 x 117 = 2340 PSD-images: 780 related to AD, 780 related to MCI and 780 related to HC. 70% of epochs was used for training and the remaining 30% was used for testing. Four classifiers were trained to classify the PSD-images (i.e. EEG epochs) of: 1) AD vs HC; 2) AD vs MCI 3) MCI vs HC 4) AD vs MCI and HC. The ability to correctly classify AD, MCI or HC epochs was evaluated through standard metrics: precision, recall, F-measure and accuracy:

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{14}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

where TP, TN, FP, FN correspond to true positive, true negative, false positive and false negative, respectively [39] (for example, in AD vs HC classification: TP represents the number of PSD-images correctly identified as images related to people with dementia (AD); TN represents the number of PSD-images correctly identified as images of healthy people (HC); FP is the number of PSD-images erroneously identified as images related to subjects with AD; FN is the number of PSD-images misclassified as images related to HC. Table 2 summarizes the results of the epoch-based classification achieved with CNN_1. Specifically, the best classification performances were observed in the binary classification AD vs HC and MCI vs HC, achieving an accuracy of 92.95% and 91.88%, respectively. The AD vs HC classifier provided the best performances in terms of

recall (95.30%); whereas, the MCI vs HC classifier provided the best performances in terms of precision (91.53%). Very good results were achieved also in AD vs MCI discrimination with an accuracy of 84.62% and F-measure of 84.68%. As regards the 3-way classifier, an accuracy of 83.33% accuracy was obtained. However, in order to exploit higher features levels and to find out the optimal CNN configuration, additional processing layers were tested. Specifically, a second convolutional layer (Conv_2) followed by ReLu and max poling layer (MaxPool_2) was added in the CNN_1 model described in Subsection 4.1, coming up with a CNN_2 model. In the second convolution, 32 kernels of size 3 x 3, stride s=1 and padding parameter p=1 were used, producing 32 features maps of size 9 x 79. Then, the max pooling layer with window size 2 x 2 and $\tilde{s}$=2 produced 32 features maps sized 4 x 39. Finally, the features extracted 32 x 4 x 39=4992 were the input of a fully connected layer with 300 hidden units followed by a soft-max layer employed to estimate the class predictions in the 2-ways and 3-ways classification. Table 3 reports the epoch-based test results of CNN_2. The best classification performances were observed in the binary classification AD vs HC and MCI vs HC with accuracies of 91.88% and 90.17%, respectively. However, the CNN_2 remained deficient in the AD vs MCI and AD vs MCI vs HC classification, reporting accuracy rate of 78.63% and 78.49%, respectively. Comparative results have proved that the CNN_1 architecture outperformed CNN_2 in all comparisons, therefore, the CNN_1 model was adopted. The reason why additional layers did not impact well on the performance may be due to the sparsity of the PSD spectrograms, described in Section 3.2. PSD-images were indeed characterized a widespread darkness with sharp bright bars related to PSD peaks, an example is shown in Fig. 2. Furthermore, PSD-images are inherently simple, the dataset of three groups essentially differ in the number, positioning and width of the bright bars. An increased complexity in the structure of the CNN may therefore cause overfitting, taking into account the limited size of the database here analyzed. It is worth to be noted that the small feature maps (sized 4 x 39) extracted from MaxPool_2 did not allow to add further convolutional layers and test deeper convolutional networks.

Table 2: Epoch-based classification performance of the proposed CNN_1 evaluated on test sets.

	CNN_1			
Classification	Precision [%]	Recall [%]	F-measure [%]	Accuracy [%]
AD vs HC	91.02	95.30	93.11	92.95
AD vs MCI	84.32	85.04	84.68	84.62
MCI vs HC	91.53	92.31	91.91	91.88
AD vs MCI vs HC	79.51	82.91	81.17	83.33

5.1.1. Comparison with standard classifiers (MLP - SVM - LDA) based on raw PSD

In order to evaluate the efficiency of the proposed approach, the architecture (CNN_1) was compared with three baseline machine learning techniques: Multi Layer Percerptron (MLP), Support Vector Machine with linear kernel

Table 3: PSD epoch-based classification performance of the proposed CNN_2 evaluated on test sets.

	CNN_2			
Classification	Precision [%]	Recall [%]	F-measure [%]	Accuracy [%]
AD vs HC	90.16	94.02	92.05	91.88
AD vs MCI	76.59	82.48	79.42	78.63
MCI vs HC	88.52	92.31	90.38	90.17
AD vs MCI vs HC	74.49	78.63	76.51	78.49

(l-SVM) and Linear Discriminant Analysis (LDA). Given the i^{th} (i=1,2..,117) EEG recording, for every epoch ε, the ε^{th} PSD-image, sized 19 x 159, was flattened into a 1-d vector sized 1 x 3021 (1-d PSD input). Afterwards, the 1-d PSD inputs were fed into the MLP, l-SVM and LDA classifiers to identify EEG patterns of AD, MCI and HC. Three MLP classifiers were trained and tested: MLP_1 classifier with one hidden layer of 1510 neurons; MLP_2 classifier with two hidden layers of 1510 and 755 neurons, respectively; MLP_3 classifier with three hidden layers of 1510, 755 and 300 neurons, respectively. All MLP classifiers were followed by a softmax output layer to perform the binary or 3-way classification. The topology of the classifiers was chosen empirically after several experimental tests. Table 4 reports the epoch-based test results of CNN_1, MLP_1, MLP_2, MLP_3, l-SVM and LDA classifiers in terms of F-measure (which includes also precision and recall information) and accuracy. Experimental results show that the proposed CNN_1 outperformed the other approaches in all classifications (AD vs HC, AD vs MCI, MCI vs HC and AD vs MCI vs HC), achieving an accuracy up to 89.8% in 2-ways classifier and 83.3% in 3-ways classifier. This result was confirmed also by the analysis of the area under the curve (AUC) for the Receiver Operating Curve (ROC). As can be observed in Figure 4, the CNN_1 classifier shows the highest AUC value in each scenario (AD vs HC, AD vs MCI, MCI vs HC and AD vs MCI vs HC). Specifically, it is to be noted that the CNN_1 achieved better AUC values in the most challenging classifications, AD vs MCI and AD vs MCI vs HC with AUC of 0.93 and 0.94, respectively.

Table 4: F-measure and accuracy performance of the proposed CNN_1 and conventional machine learning techniques (MLP_1,MLP_2, MLP_3, l-SVM, LDA), evaluated on test sets, when PSD-images are used as input.

	AD vs HC		AD vs MCI		MCI vs HC		AD vs MCI vs HC	
Method	F measure [%]	Accuracy [%]	F-measure [%]	Accuracy [%]	F measure [%]	Accuracy [%]	F measure [%]	Accuracy [%]
CNN_1	**93.11**	**92.95**	**84.68**	**84.62**	**91.91**	**91.88**	**81.17**	**83.33**
MLP_1	90.36	90.38	78.24	77.78	88.43	88.03	75.50	76.78
MLP_2	89.36	89.32	79.33	78.85	88.84	88.46	75.40	77.64
MLP_3	90.68	90.60	77.45	77.35	88.19	88.03	74.95	77.64
l-SVM	89.13	89.32	77.06	77.35	88.52	88.25	72.56	76.50
LDA	81.05	81.41	62.60	59.40	70.51	71.58	46.76	56.55

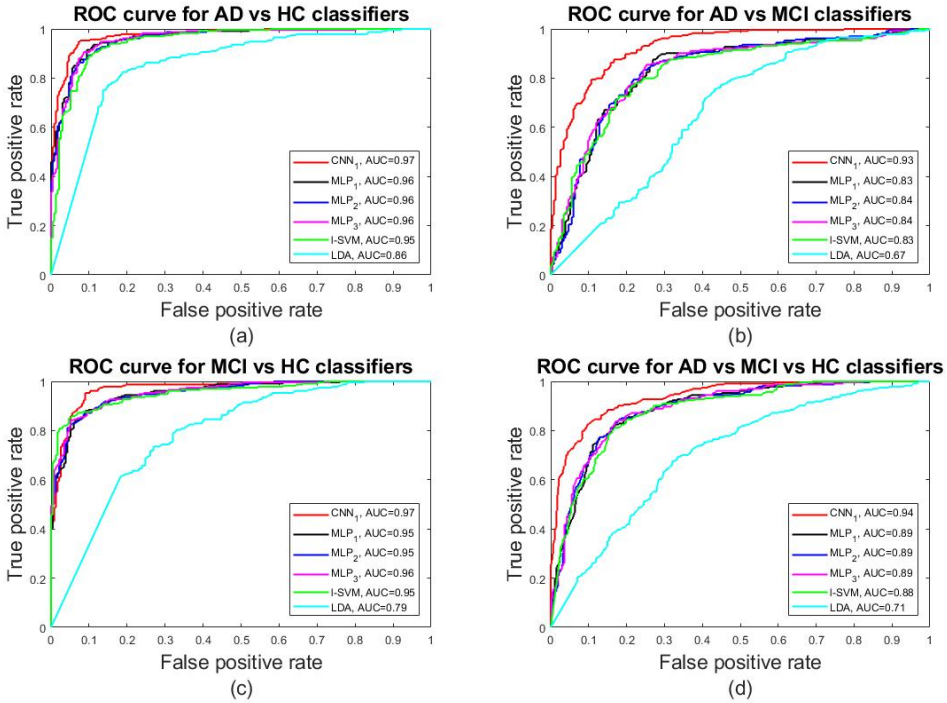

Figure 4: ROC curves of CNN_1, MLP_1, MLP_2, MLP_3, l-SVM, LDA classifiers for AD vs HC (a), AD vs MCI (b), MCI vs HC (c) and AD vs MCI vs HC (d) classification, when PSD-images are used as input.

5.1.2. Comparison with standard classifiers (MLP - SVM - LDA) trained on handcrafted features extracted from PSD

The proposed CNN_1 was also compared with standard classifiers trained on quantitative features manually extracted from the PSD spectrograms. Spectral features, relevant to the analysis of AD/MCI EEGs, were computed to be used as input to standard MLP, l-SVM and LDA classifiers. Specifically, given the i^{th} (i=1,2..,117) EEG recording, for each ε^{th} epoch under analysis, nine relevant discriminating spectral features of the EEG were computed:

- The ratio α_3/α_2 was introduced by Moretti et al [40] as a prognotic value for MCI to AD conversion;
- The ratios $\theta/(\alpha+\beta_1)$ and $(\delta+\theta)/(\alpha+\beta_1+\beta_2)$ were introduced by Bennys et al [41] to discriminate between AD and healthy controls;
- Transition Frequency (TF) and Individual alpha frequency (IAF), defined as the minimun power in the theta band and the maximum peak in the alpha frequency range, respectively, were introduced to discriminate mild AD from healthy elderly scontrols [42].

Table 5 reports the epoch-based test results of the proposed CNN_1 and MLP_1,

l-SVM and LDA classifiers in terms of F-measure and accuracy. Among the conventional classifiers the MLP_1 (with one hidden layer and five hidden neurons) outperformed the l-SVM and LDA classifiers, achieving average accuracy rate up to 75.42% in binary classification and only 59.40% in AD vs MCI vs HC classification. However, as can be observed, the proposed CNN_1 outperformed the MLP_1 and subsequently the SVM and LDA classifiers in all comparisons. Similar outcomes were achieved by the analysis of AUC reported in Figure 5, where the CNN_1 classifier showed the highest AUC value in AD vs HC (0.97), AD vs MCI (0.93), MCI vs HC (0.97) and AD vs MCI vs HC (0.94) classifications. This result has confirmed the effectiveness of the proposed approach over conventional classifiers based on the handcrafted extraction of spectral features from the PSD of EEG signals. The PSD spectrogram, taken as a whole, seems to carry more information than a set of features extracted from it, even though relevant features. PSD-images indeed display the behaviour of EEG over the frequency and over the channels (the i-th row of the PSD-image represents the PSD vs freq profile of the i-th EEG signal) in this way describing the behaviour of EEGs over frequency and over space.

Table 5: F-measure and accuracy performance of the proposed CNN_1 and conventional machine learning techniques (MLP_1, l-SVM, LDA), evaluated on test sets. In this case the inputs of MLP, SVM and LDA classifiers are handcrafted features manually extracted from spectral profiles.

	AD vs HC		AD vs MCI		MCI vs HC		AD vs MCI vs HC	
Method	F measure [%]	Accuracy [%]	F-measure [%]	Accuracy [%]	F measure [%]	Accuracy [%]	F measure [%]	Accuracy [%]
CNN_1	**93.11**	**92.95**	**84.68**	**84.62**	**91.91**	**91.88**	**81.17**	**83.33**
MLP_1	84.08	83.33	67.64	66.88	75.86	76.07	58.73	59.40
l-SVM	82.35	82.69	58.05	60.47	69.96	71.37	56.06	56.84
LDA	79.65	80.13	62.27	64.53	68.72	69.66	57.76	55.70

5.2. Patient-based Classification

Since CNN_1 classifier showed the best performances in the epoch-based classification (Table 4, 5), it was employed to perform the patient-based classification that estimates the probability that patient j^{th} belongs to class AD, MCI or CNT, depending on how his/her epochs were labelled by the classifier. Specifically, given the j^{th} (j=1,2..,72) EEG recording, the PSD-images (i.e. EEG epochs) were extracted and fed into the CNN_1 classifier. If the number of images classified as c (for example, AD) was larger than 50%, then the j^{th} EEG recording was labeled as belonging to c class (i.e. AD). Table 6 shows the results of the overall patient classification analysis. The AD vs HC classifier was able to correctly classify 23 AD patients out of 24 and 15 HC subjects out of 24, respectively. In the AD vs MCI classification, only 11 out of 24 AD patients were correctly discriminated as AD, whereas 13 were misclassified as MCI. This is probably related to the different stages of AD and MCI included in the present database. Similarly, 11 out of 24 MCI patients were correctly classified. Good performances were obtained also in the MCI vs HC discrimination: 12 out of 24 HC subjects and 18 out of 24 MCI patients were correctly classified, respectively. Finally, the 3-way classifier reported acceptable results: 13 AD, 5 MCI and 11 HC were properly classified.

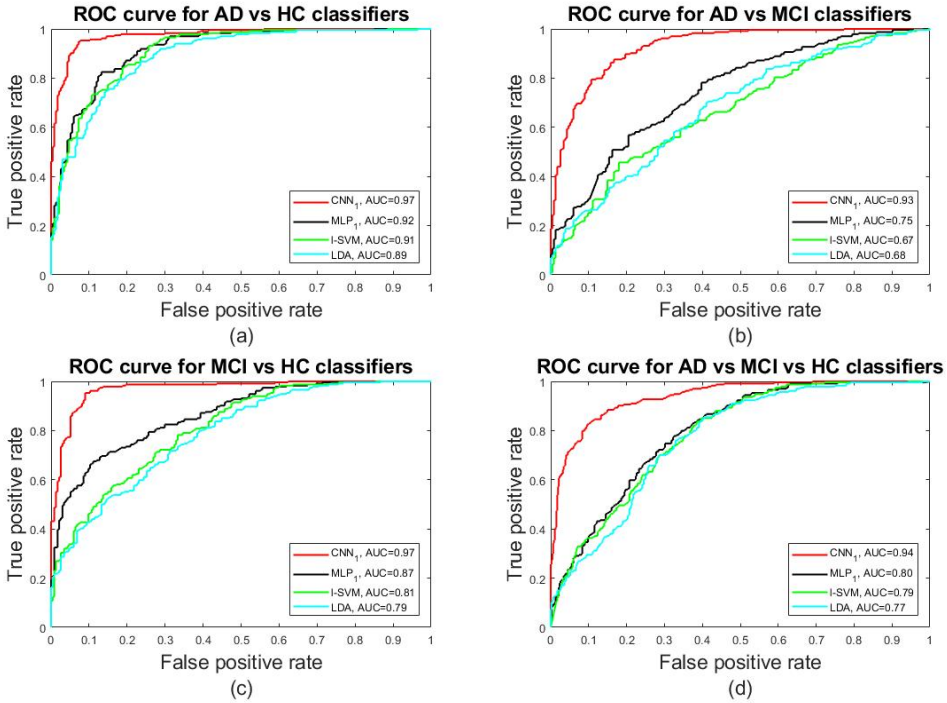

Figure 5: ROC curves of CNN_1, MLP_1, l-SVM, LDA classifiers for AD vs HC (a), AD vs MCI (b), MCI vs HC (c) and AD vs MCI vs HC (d) classification, when handcrafted features manually extracted from spectral profiles are used as input of MLP_1, l-SVM, LDA.

6. Discussion

The present research addressed the challenging issue of classifying EEG recordings of patients affected by Alzheimer's disease (AD), Mild Cognitive Impairment (MCI) and Healthy Control (HC) subjects, through advanced artificial intelligence methodologies. Deep Learning (DL) is an emerging machine learning technique able to automatically learn latent discriminating variables ($features$) from the input data automatically. DL overcomes the limitations that come with standard learning algorithms: in principle, it allows to bypass the handcrafted feature extraction process.

The visual inspection of PSD of the EEG recordings is a common practice to investigate anomalies in the spectrum of patients affected by AD. Accordingly one of the main contributions of this study was developing an innovative method based on spectral representation of EEG data. Specifically, we proposed a PSD-based CNN able to extract latent features from 2-d gray scale representations of PSD spectra (PSD-images) and subsequently differentiate AD, MCI, HC by only analyzing noninvasive scalp EEG recordings. To our best knowledge, this is the first work that has dealt with classification of AD, MCI and HC, through

Table 6: Patient-based classification performances of the proposed CNN_1 classifier.

Patient Class	AD vs HC	AD vs MCI	MCI vs HC	AD vs MCI vs HC
AD	23/24	11/24	-	13/24
MCI	-	11/24	18/24	5/24
HC	15/24	-	12/24	11/24

a data-driven deep learning framework based on PSD-images. In a previous work, Morabito et al. [30] introduced a CNN classifier to discriminate among EEG recordings of AD, MCI and HC subjects. The system was based on the engineering feature extraction of the statistical quantities mean (μ), standard deviation (σ) and skewness (υ) from the time-frequency representations of the EEG time series. Good classification performances were achieved: 85% accuracy in AD vs HC classification, 78% accuracy in AD vs MCI classification, 85% accuracy in MCI vs HC classification, 82% accuracy in AD vs MCI vs HC classification; however, the features were manually extracted and the classifier was tested on a limited dataset (only 23 subjects per group). Here, 3 cohorts of 63 subjects were analyzed. The proposed DL procedure, allowed to generate a high-level set of features automatically, from the PSD-images of the EEG recordings, yielding a very good performance in the epoch-based classification of AD vs HC (92.95% accuracy, 0.97 AUC), AD vs MCI (84.62% accuracy, 0.93 AUC), MCI vs HC (91.88% accuracy, 0.97 AUC) and AD vs MCI vs HC (83.33% accuracy, 0.94 AUC). Specifically, the simple spectral shape and the sparsity properties of the PSD-images have allowed to develop a CNN with only one module of convolutional, ReLu and max pooling layer (CNN_1) saving time and cost of elaboration. Indeed, the proposed CNN_1 performed very well both in binary and three ways classification, even when additional hidden layers are taken into account. Specifically, the number of learnable parameters in CNN_1 was more than 3 millions with an average computational cost of 2024 secs. When the second convolutional layer was added (CNN_2) the learnable parameters were about 1.5 millions but the computational cost was of 5434 secs. It is to be noted that, in this study, the increasing of computational complexity (and consequently of computational cost) led to a reduction of performances in the dementia stages classification. Hence, CNN_1 was experimentally determined as the optimal trade-off in terms of processing and classification performance. Moreover, experimental results have shown that the proposed CNN_1 classifier outperforms conventional MLP, SVM and LDA classifiers not only when raw PSD-images are used as input but also when handcrafted features (manually extracted from spectral profiles) are used. This result has confirmed the effectiveness of the proposed PSD-based CNN rather than conventional machine learning techniques and might represent a remarkable breakthrough in real-world clinical applications. However, the proposed method suffers from some limitations. The major limitation lies in the nonstationarity of the EEG epochs of each patient under analysis, which, of course, reflects upon the variability of the PSD-images. In fact, when the CNN attempts to discriminate epochs of the same patient, some of them are misclassified mainly because of the position

of power peaks in the frequency domain. A normal subject in a resting eyes-closed condition is characterized by a dominant activity in the alpha rhythm, especially in the parieto-occipital area. This is clearly visible in the spectral domain, where a peak in the alpha band, which corresponds to a white vertical line in the 2-d gray scale image, can be observed. The presence of artifacts, the unexpected lack of the dominant peak in alpha or a shifting towards low-frequencies (typically observed in the spectra of AD or MCI patients), may lead to the misclassification of that epoch. Similarly, epochs of an AD (or MCI) patient may not have steady PSD characteristics, resulting in a high variability of the PSD-images over the time. Moreover, the differences between the PSD of AD and MCI patients can be very mild, due to the fact that AD develops gradually thus, for example, a mild AD could have frequency characteristics more similar to MCI than to severe AD. This clarifies why CNN failed in classifying correctly all of the epochs of the same subject (AD, MCI or HC) causing a performance deterioration in the overall patient classification (Table 6).

7. Conclusions

In this paper, the authors introduced a novel DL method based on PSD for the brain states classification of patients affected by AD, MCI and HC individuals. The originality of the method proposed lies in mapping the power spectrum of each subject into a 2-d gray scale image and developing a Convolutional Neural Network (CNN) able to automatically extract latent features automatically from simple and sparse images representative of the power spectra. The proposed CNN_1 consisted of: a convolutional layer (+ ReLu), a max pooling layer, and a fully connected layer followed by a softmax layer.
The experimental results indicate that the CNN provided high performances in epochs classification also when compared with standard learning algorithms (MLP, SVM, LDA, Table 4, 5). Nevertheless, the proposed classification system remains deficient in the patient-based classification. (Table 6). The findings suggest that the power spectral analysis is not enough to detect the anomalies of EEG recording, but we believe that DL techniques can identify a significant set of variables that may support clinicians in the diagnosis of AD.
In the future the proposed PSD-based CNN architecture will be trained by using more powerful graphic processing units (GPU) in order to optimize the training times. A larger cohort of AD/MCI/HC subjects will be taken into account to further demonstrate and fully exploit the generalization potential of deep learning techniques in clinical applications. Classifications over the average PSD-image of whole EEG recordings will also be analyzed, since such as approach is likely to reduce the effects of epoch variability. In addition, different methods to map the spectral characteristics of the EEG, that may be more sensitive to the specific characteristics of AD or MCI patients, will be investigated. Finally, motivated by the preliminary, though promising results obtained in this paper, using limited available clinical data, future works will explore integrating the EEG-based CNN model proposed in this paper, with contextual processing of volumetric medical images [43] of the patient's brain.

This will lead to development of a more personalized, context-aware and multi-modal diagnostic system.

Acknowledgment

This work was funded by the Italian Ministry of Health, project code: GR-2011-02351397. Professor A. Hussain was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant no. EP/M026981/1. The authors also wish to thank the anonymous reviewers who helped improve the quality of the paper.

References

References

[1] 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 367 (9518) (2006) 1262–1270.

[2] J. Jeong, Eeg dynamics in patients with alzheimer's disease, Clinical neurophysiology 115 (7) (2004) 1490–1505.

[3] C. Stam, T. Montez, B. Jones, S. Rombouts, Y. Van Der Made, Y. Pijnenburg, P. Scheltens, Disturbed fluctuations of resting state eeg synchronization in alzheimer's disease, Clinical Neurophysiology 116 (3) (2005) 708–715.

[4] J. S. Bendat, A. G. Piersol, Engineering applications of correlation and spectral analysis, New York, Wiley-Interscience, 1980. 315 p.

[5] K. Van der Hiele, A. Vein, R. Reijntjes, R. Westendorp, E. Bollen, M. Van Buchem, J. Van Dijk, H. Middelkoop, Eeg correlates in the spectrum of cognitive decline, Clinical neurophysiology 118 (9) (2007) 1931–1939.

[6] N. Zeng, Z. Wang, H. Zhang, F. E. Alsaadi, A novel switching delayed pso algorithm for estimating unknown parameters of lateral flow immunoassay, Cognitive Computation 8 (2) (2016) 143–152.

[7] N. Zeng, Z. Wang, H. Zhang, W. Liu, F. E. Alsaadi, Deep belief networks for quantitative analysis of a gold immunochromatographic strip, Cognitive Computation 8 (4) (2016) 684–692.

[8] M. Mahmud, M. S. Kaiser, A. Hussain, S. Vassanelli, Applications of deep learning and reinforcement learning to biological data, arXiv preprint arXiv:1711.03985.

[9] S. Gasparini, M. Campolo, C. Ieracitano, N. Mammone, E. Ferlazzo, C. Sueri, G. G. Tripodi, U. Aguglia, F. C. Morabito, Information theoretic-based interpretation of a deep neural network approach in diagnosing psychogenic non-epileptic seizures, Entropy 20 (2) (2018) 43.

[10] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, A. M. Dobaie, Facial expression recognition via learning deep sparse autoencoders, Neurocomputing 273 (2018) 643–649.

[11] C. Ieracitano, A. Adeel, M. Gogate, K. Dashtipour, F. C. Morabito, H. Larijani, A. Raza, A. Hussain, Statistical analysis driven optimized deep learning system for intrusion detection, arXiv preprint arXiv:1808.05633.

[12] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[13] M. D. Zeiler, M. Ranzato, R. Monga, M. Mao, K. Yang, Q. V. Le, P. Nguyen, A. Senior, V. Vanhoucke, J. Dean, et al., On rectified linear units for speech processing, in: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, IEEE, 2013, pp. 3517–3521.

[14] V. Nair, G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp. 807–814.

[15] P. Anderer, B. Saletu, B. Klöppel, H. Semlitsch, H. Werner, Discrimination between demented patients and normals based on topographic eeg slow wave activity: comparison between z statistics, discriminant analysis and artificial neural network classifiers, Electroencephalography and clinical neurophysiology 91 (2) (1994) 108–117.

[16] W. S. Pritchard, D. W. Duke, K. L. Coburn, N. C. Moore, K. A. Tucker, M. W. Jann, R. M. Hostetler, Eeg-based, neural-net predictive classification of alzheimer's disease versus control subjects is augmented by non-linear eeg measures, Electroencephalography and clinical Neurophysiology 91 (2) (1994) 118–130.

[17] 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.

[18] C. Huang, L.-O. Wahlund, T. Dierks, P. Julin, B. Winblad, V. Jelic, Discrimination of alzheimer's disease and mild cognitive impairment by equivalent eeg sources: a cross-sectional and longitudinal study, Clinical Neurophysiology 111 (11) (2000) 1961–1967.

[19] M. Buscema, P. Rossini, C. Babiloni, E. Grossi, The ifast model, a novel parallel nonlinear eeg analysis technique, distinguishes mild cognitive impairment and alzheimer's disease patients with high degree of accuracy, Artificial intelligence in medicine 40 (2) (2007) 127–141.

[20] P. M. Rossini, M. Buscema, M. Capriotti, E. Grossi, G. Rodriguez, C. Del Percio, C. Babiloni, Is it possible to automatically distinguish resting eeg data of normal elderly vs. mild cognitive impairment subjects with high degree of accuracy?, Clinical Neurophysiology 119 (7) (2008) 1534–1545.

[21] 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.

[22] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[23] S. Lawrence, C. L. Giles, A. C. Tsoi, A. D. Back, Face recognition: A convolutional neural-network approach, IEEE transactions on neural networks 8 (1) (1997) 98–113.

[24] A. Payan, G. Montana, Predicting alzheimer's disease: a neuroimaging study with 3d convolutional neural networks, arXiv preprint arXiv:1502.02506.

[25] E. Hosseini-Asl, R. Keynton, A. El-Baz, Alzheimer's disease diagnostics by adaptation of 3d convolutional network, in: Image Processing (ICIP), 2016 IEEE International Conference on, IEEE, 2016, pp. 126–130.

[26] K. Aderghal, J. Benois-Pineau, K. Afdel, C. Gwenaëlle, Fuseme: Classification of smri images by fusion of deep cnns in 2d+ projections, in: Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing, ACM, 2017, p. 34.

[27] S. Sarraf, G. Tofighi, Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks, arXiv preprint arXiv:1603.08631.

[28] C. D. Billones, O. J. L. D. Demetria, D. E. D. Hostallero, P. C. Naval, Demnet: A convolutional neural network for the detection of alzheimer's disease and mild cognitive impairment, in: Region 10 Conference (TENCON), 2016 IEEE, IEEE, 2016, pp. 3724–3727.

[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] 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 27 (02) (2017) 1650039.

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

[32] 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.

[33] G. Heinzel, A. Rüdiger, R. Schilling, Spectrum and spectral density estimation by the discrete fourier transform (dft), including a comprehensive list of window functions and some new at-top windows.

[34] J. G. Proakis, Digital signal processing: principles algorithms and applications, Pearson Education India, 2001.

[35] D. Scherer, A. Müller, S. Behnke, Evaluation of pooling operations in convolutional architectures for object recognition, Artificial Neural Networks–ICANN 2010 (2010) 92–101.

[36] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, K. He, Accurate, large minibatch sgd: Training imagenet in 1 hour, arXiv preprint arXiv:1706.02677.

[37] Y. Bengio, Practical recommendations for gradient-based training of deep architectures, in: Neural networks: Tricks of the trade, Springer, 2012, pp. 437–478.

[38] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, 2015, pp. 448–456.

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

[40] D. Moretti, Conversion of mild cognitive impairment patients in alzheimer's disease: prognostic value of alpha3/alpha2 electroencephalographic rhythms power ratio, Alzheimer's research & therapy 7 (1) (2015) 80.

[41] K. Bennys, G. Rondouin, C. Vergnes, J. Touchon, Diagnostic value of quantitative eeg in alzheimer's disease, Neurophysiologie Clinique/Clinical Neurophysiology 31 (3) (2001) 153–160.

[42] D. V. Moretti, C. Babiloni, G. Binetti, E. Cassetta, G. Dal Forno, F. Ferreric, R. Ferri, B. Lanuzza, C. Miniussi, F. Nobili, et al., Individual analysis of eeg frequency and band power in mild alzheimer's disease, Clinical Neurophysiology 115 (2) (2004) 299–308.

[43] S. K. Wajid, A. Hussain, K. Huang, Three-dimensional local energy-based shape histogram (3d-lesh)-based feature extraction–a novel technique, Expert Systems with Applications.