

An explainable Artificial Intelligence approach to study MCI to AD conversion via HD-EEG processing

Clinical EEG and Neuroscience
2023, Vol. 54(1) 51–60
© EEG and Clinical Neuroscience
Society (ECNS) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15500594211063662
journals.sagepub.com/home/eeg



Francesco Carlo Morabito¹ , Cosimo Ieracitano¹,
and Nadia Mammone¹

Abstract

An explainable Artificial Intelligence (xAI) approach is proposed to longitudinally monitor subjects affected by Mild Cognitive Impairment (MCI) by using high-density electroencephalography (HD-EEG). To this end, a group of MCI patients was enrolled at IRCCS Centro Neurolesi Bonino Pulejo of Messina (Italy) within a follow-up protocol that included two evaluations steps: T0 (first evaluation) and T1 (three months later). At T1, four MCI patients converted to Alzheimer's Disease (AD) and were included in the analysis as the goal of this work was to use xAI to detect individual changes in EEGs possibly related to the degeneration from MCI to AD. The proposed methodology consists in mapping segments of HD-EEG into channel-frequency maps by means of the power spectral density. Such maps are used as input to a Convolutional Neural Network (CNN), trained to label the maps as "T0" (MCI state) or "T1" (AD state). Experimental results reported high intra-subject classification performance (accuracy rate up to 98.97% (95% confidence interval: 98.68–99.26)). Subsequently, the explainability of the proposed CNN is explored via a Grad-CAM approach. The procedure detected which EEG-channels (i.e., head region) and range of frequencies (i.e., sub-bands) were more active in the progression to AD. The xAI analysis showed that the main information is included in the delta sub-band and that, limited to the analyzed dataset, the highest relevant areas are: the left-temporal and central-frontal lobe for Sb01, the parietal lobe for Sb02, the left-frontal lobe for Sb03 and the left-frontotemporal region for Sb04.

Keywords

High-Density Electroencephalography, Mild Cognitive Impairment, Alzheimer's Disease, Convolutional Neural Network, explainable Artificial Intelligence

Received June 21, 2021; revised October 14, 2021; accepted October 26, 2021.

Introduction

Dementia affects more than 40 million people worldwide, with a serious increasing tendency to more than 60 million by 2030. Nearly 60% of dementia cases are due to Alzheimer's Disease (AD). Since AD is still an incurable disorder, commonly diagnosed in elderly, after the diagnosis the patient has, on average, only 4 to 8 more years to live. As also postulated by the National Institute on Aging and the Alzheimer's Association, AD is preceded by a pre-dementia phase, known as Mild Cognitive Impairment (MCI). This prodromal stage passes often unnoticed as it affects cognitive abilities in subtle ways. An MCI subject is still able to live his/her daily life autonomously, however, especially in the amnesic subtype, he/she has a high risk to evolve to dementia due to AD with the aging process. MCI subjects can also remain stable or recovery¹, depending on the inherent causes of the disorder. Hence, longitudinal follow-up programs on MCI are extremely important to monitor the course of AD and early diagnose it as soon as neural deficits occur. In this context, electroencephalography

(EEG), considered one of the main technology to study the electrical brain activity, provided promising results in the diagnosis of AD and MCI. EEG is a non-invasive tool consisting in a set of electrodes (i.e., channels) located on the scalp to record the electric potentials generated by cortical neurons. Anomalies between inter-neuron communication may reflect on abnormalities in the EEG recordings. Indeed, the so-called "slowing effect" (i.e., an increase of the relative EEG power content at low frequencies), loss of complexity and synchronization among channels are common neurodegeneration phenomena related to AD². Conventional low-density EEG (LD-EEG) with less than 64 channels is widely used as diagnostic tool. This offers a high

¹DICEAM, University Mediterranea of Reggio Calabria, Via Graziella Feo di Vito, 89124, Reggio Calabria, Italy

Corresponding Author:

Francesco Carlo Morabito, DICEAM, University Mediterranea of Reggio Calabria, Via Graziella Feo di Vito, 89124, Reggio Calabria, Italy.

Email: morabito@unirc.it

Full-color figures are available online at journals.sagepub.com/home/eeg

temporal resolution but suffers from poor spatial resolution because of large inter-electrode distance³. High-Density EEG (HD-EEG) overcomes this issue. Indeed, it provides superior spatial resolution, allowing to detect more relevant features in longitudinal MCI/AD studies.

However, a few number of HD-EEG-based works on AD exists in the literature. In⁴ authors compared HD-EEG signals of AD and healthy controls (HC) observing a reduction of activity in all frequency bands in the right cerebral hemisphere of AD; while, in⁵ HD-EEG signals of AD and HC were analysed, showing a dysfunction in the parietal and medial temporal regions and also an adaptive reorganization in AD. Recently, compressive sensing technique was applied⁶ to reconstruct HD-EEG of MCI and AD patients aiming to demonstrate that the compression has no measurable effect on the complex brain network analysis. A few EEG-based longitudinal studies on MCI-to-AD progression exist in the current state-of-the-art, furthermore, they are based on LD-EEG signals. Authors in⁷, carried out a longitudinal study on MCI, observing a decreased alpha global field power and a source localization of alpha, theta and beta frequency in a more anterior area, in MCI subjects who progressed to AD. Longitudinal changes were also studied in terms of relative power⁸, reporting higher theta and lower beta relative power, especially at the temporal and temporo-occipital regions, in AD patients. The theta/gamma and alpha3/alpha2 ratios⁹ were investigated for prognosticating the dementia due to AD progression, observing that the increase of alpha3/alpha2 ratio was related to the AD conversion. A quantitative spectral analysis on EEG signals of MCI subjects was carried out in¹⁰, revealing a decreased alpha activity in follow-up MCI converted to AD, especially over posterior leads; whereas, in¹¹ a complex network based strategy was proposed, showing increasing characteristic path length and decreasing efficiency along with AD progression. In¹² and¹³ two novel coupling strength metrics between time series, namely, the Permutation Disalignment Index, (PDI) and the Permutation Jaccard Distance (PJD), respectively, were introduced. Experimental results reported an increase of PDI and PJD, namely a decrease of coupling strength in delta and theta sub-band, in the converted patients. Finally, an EEG-based eLORETA longitudinal analysis on MCI subjects¹⁴ was carried out, reporting an increased power in delta and theta bands for subject converted to AD. Several EEG-based Artificial Intelligence (AI) models have been instead emerging to classify MCI and AD subjects, achieving impressive results^{15–20}. However, most of these perform an inter-subject classification and are based on LD-EEG. Hence, the potential advantages of HD-EEG in AD research is still widely unexplored. In order to fill this gap and exploration of the potential of xAI in the dementia field, here, an intra-subject explainable artificial intelligence (xAI) approach to longitudinal HD-EEG classification is proposed. The objective of the study is to investigate the capability of xAI to offer new perspectives in follow-up (longitudinal) studies on subjects at

risk of developing dementia due to Alzheimer's, hence the attention was focused on patients that experienced the conversion from MCI to AD. To this end, a small group of MCI subjects was enrolled at IRCCS Centro Neurolesi Bonino Pulejo of Messina (Italy) within a follow-up program: all patients were diagnosed MCI at time T0 and AD at time T1 (three months later). The study was designed to include a stage of intra-subject classification with the aim of training an artificial neural network to detect subject-specific characteristics potentially associated with the process of degeneration from MCI to AD. Specifically, the characteristics of EEG signals in the frequency domain were used as input to the neural network so that it learned to discriminate between the EEGs recorded from a given subject at time T0 (MCI condition) from the EEGs recorded from the same subject at time T1 (AD condition) and, on the basis of such classification, xAI explained what characteristics were more relevant to classification thus are more likely to be involved in the disease development process. For each EEG channel of the HD-EEG epoch under analysis, the Power Spectral Density (PSD) is estimated, resulting in a channel-frequency (CF) map, here denoted as *HD-CF epoch*, used as input to a custom Convolutional Neural Network (CNN) meant to perform the binary epoch-classification task (AD vs. MCI) of the patient taken into account. It is worth mentioning that AI-based systems typically work as a *black-box* and no explanation of the results are generally provided. Here, xAI is applied to the trained CNN. In particular, a Grad-CAM-based analysis is carried out by using the HD-CF maps as input to the trained model to investigate which set of channels and which range of frequencies were mostly relevant to label the input as "AD" rather than "MCI". This allowed exploration of the longitudinal changes in the brain electrical activity that each subject experienced from time T0 to time T1.

Materials

Patient description

A groups of 15 MCI subjects were initially recruited in the study at IRCCS Centro Neurolesi Bonino Pulejo of Messina (Italy). The study consisted in a follow-up program: each participant was evaluated at a baseline time, T0, and after three months (time T1). Unfortunately, only 11 of them could precisely meet the protocol schedule (3 months between time T0 and time T1). By the time T1, 4 of these 11 patients were diagnosed as converted to dementia due to AD and were therefore included in the present study. Three of them were female. From the very beginning, the objective of the study was to investigate the capability of xAI to monitor, longitudinally, subjects at risk of developing dementia due to AD. The attention was therefore focused on patients converted from MCI to AD and a system was designed to learn how to detect subject-specific characteristics potentially associated with the conversion. Each patient agreed to participate in the study by

signing an informed consent document. The local Ethical Committee approved the clinical protocol and an expert team of neurologists and neuropsychologists carried out all the cognitive and medical examinations according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-V²¹). All participants were subjected to a neuroradiological examination to exclude any other possible pathology such as strokes, tumors or other neural deficit. None of patients was undergoing any medical treatment.

Methodology

HD-EEG recording

EEG signals were recorded by using a high-density (HD) EGI 256-channels Geodesic Sensor Net (Figure 1). This is a wet-electrodes system, it must be kept immersed in a saline solution for 10 minutes before the application on the patient's head. The central electrode (Cz) was the reference location and, as recommended by the EGI guidelines, the impedance of each electrode was kept lower than 50 kΩ, possibly with the help of additional saline solution, and the sampling frequency f_s was set at 250 Hz. EEG recordings took place in the morning. Before starting the EEG acquisition, participants were interviewed about their last meal and about the quality of their last night sleep. Throughout the recording, patients remained in an eye closed resting state condition, while an expert operator continuously monitored EEG signals in order to detect any possible signal trend related to drowsiness. To make the recording conditions of sessions T0 and T1 as similar as possible, the same patient position and room conditions (noise, temperature, etc) were reproduced, as well as the quality and quantity of meal and sleep preceding the recording.

HD-EEG pre-processing

Every EEG recording was pre-processed by applying a band-pass filter between 1 and 40 Hz in order to select the main EEG rhythms: δ (1–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz) and γ (30–40 Hz). The filtering operation was carried out by the *Net Station EEG software* of the Electrical Geodesics EEG system. Artifactual segments in the EEG traces were manually removed by an expert neurophysiologist. In this study, the signals recorded by electrodes located on the cheeks and on the neck were excluded, as they are likely to be corrupted by muscle artifacts and by bad skin-electrode artifacts, providing weak information about the cortical activity. As result, a sub-set of 173 scalp electrodes, those enclosed in the blue region of the montage representation shown in Figure 1b, were taken into account. Then, each HD-EEG recording was partitioned into non-overlapping temporal epochs of 1s and processed epoch by epoch. Since $f_s = 250$ Hz, every epoch consisted of $M = 250$ samples. Finally, every HD-EEG epoch was spatially filtered by means of surface Laplacian, with the aim of reducing the

effects of volume conduction and therefore improving EEG spatial resolution²².

Channel-frequency representation of HD-EEG

Given a HD-EEG epoch, each of the 173 EEG channels was mapped into the spectral domain by means of the Power Spectral Density (PSD²³). A common estimate of PSD is the *periodogram*, defined as follows:

$$P(f) = \frac{T_s}{M} \left| \sum_{m=0}^{M-1} e_m(t) e^{-2\pi i f m} \right|^2 \quad (1)$$

where T_s is the sampling period, $e(t)$ is the HD-EEG recorded by one of the electrodes of length M and $-\frac{1}{2T_s} < f \leq \frac{1}{2T_s}$. It is to be noted that when the input signal (i.e., EEG recording) is multiplied by a window function w_m , the *modified periodogram* is achieved:

$$\bar{P}(f) = \frac{T_s}{M} \left| \sum_{m=0}^{M-1} w_m e_m(t) e^{-2\pi i f m} \right|^2 \quad (2)$$

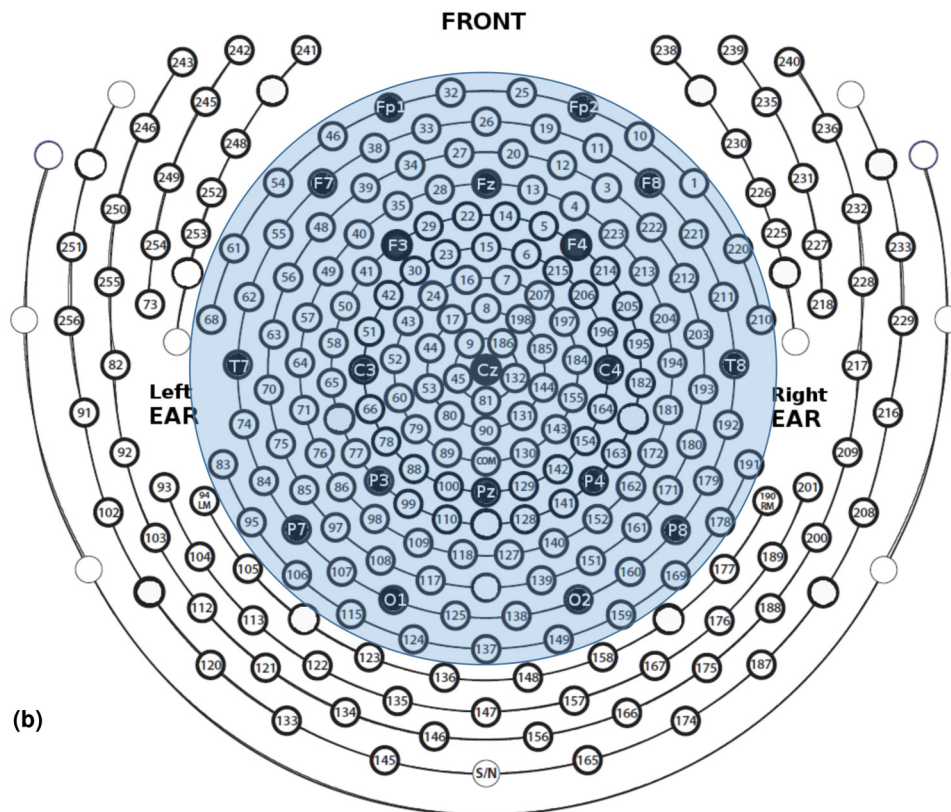
The modified periodogram reduces the spectral leakage of the standard periodogram and smoothes the edges of the signal. In this study, the PSD of the e^{th} EEG recording was estimated by applying the modified periodogram with basic rectangular window function. In particular, a vector of $F = 40$ frequencies in the range (1–40 Hz) was determined in order to cover the five main EEG sub-bands. For every EEG channel of the epoch under analysis, a PSD vector sized $1 \times F = 1 \times 40$ was evaluated, for an overall of 173 spectral profiles (one per channel). The result is a Channel-Frequency (CF) map sized $N \times F = 173 \times 40$. Hence, every HD-EEG epoch was associated to a CF map (herein denoted to as *HD-CF epoch*) meant to be later used as input to the proposed classification system.

Longitudinal epoch-based classification system

In this study, we propose a longitudinal HD-EEG epoch-based classification system composed of a custom Convolutional Neural Network (CNN). This is a class of Deep Learning (DL²⁴) architectures, typically applied to image recognition. It performs feature extraction automatically by processing the input data through a set of convolution, activation and pooling layers. Further theoretical details on CNN are reported in^{18,25} Figure 2 shows the proposed CNN that consists of one convolutional layer (Cv_1 , followed by a ReLU activation function), one max pooling layer (Mp_1) and a 2-hidden layer NN with softmax output function. The convolutional layer has a bank of 8 filters sized 3×2 with stride 2; while, the filters of the pooling layer have dimension 2×2 with step size 2. Cv_1 outputs 8 features maps sized 86×20 , whereas Mp_1 produces 8 features maps sized 43×10 . The latter are reshaped into a single vector sized $1 \times (8 \times 43 \times 10) = 1 \times 3440$ used as input to a 2-hidden-layer NN with 1500 and 500 neurons.



(a)



(b)

Figure 1. (a) The High-Density (HD) EGI 256-channel Geodesic Sensor Net EEG system. (b) Electrodes layout of the 256-channel HD EEG³⁰, where black sensors correspond to the conventional 19-channel montage and the blue zone refers to the 173 sensors on the scalp used in this study. Note that the 256-channels of the Geodesic Sensor Net EEG system are labelled as E1, E2,...,E256. In the Figure, for sake of readability, the letter "E" is omitted.

The network ends with a softmax output function for the binary epoch-classification only related to a subject under analysis: MCI vs. AD. The optimized Adaptive Moment algorithm was

applied with exponential decay values β_1 , β_2 of 0.9 and 0.999, respectively and learning rate equal to 0.001. Note that the topology and set-up of the developed CNN was designed

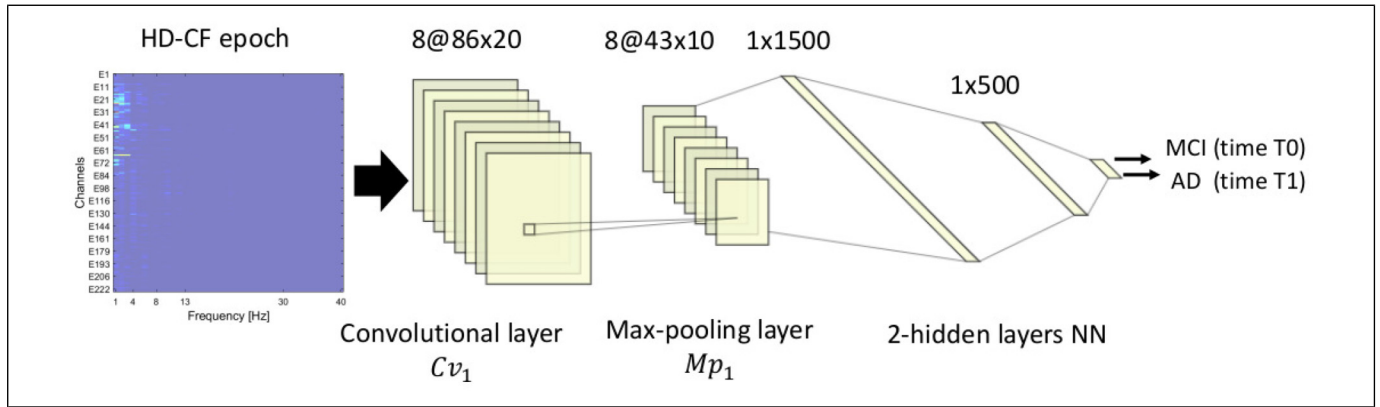


Figure 2. Architecture of the proposed longitudinal epoch-based classification system.

according to a *trial and error* procedure. Finally, the network was implemented in Matlab R2021a and experiments were carried out on a workstation equipped with one NVIDIA GeForce RTX 2080 Ti GPU and a RAM of 64 GB installed.

Grad-CAM-based analysis of channel-frequency maps

The main challenge of AI is to explain the predictive decisions achieved by the developed networks in order to provide more trustworthy and reliable systems. In this context, xAI research is emerging. This refers to the methodologies able to “open” the *AI-black box* and discover which part of it contributed to achieve a specific result²⁶. In this study, one of the most widely employed xAI technique to understand CNN-based models, namely, the Gradient-weighted Class Activation Mapping (Grad-CAM²⁷) is exploited. Specifically, Grad-CAM visualizes the input areas that are relevant for predictions. Let o^c the score of a specific category c (i.e., MCI or AD). First, the gradient of o^c with respect to the features representations R^n (with n number of features maps) of a Cv layer is estimated: $\frac{\partial o^c}{\partial R^n}$. Here, features maps of the last convolutional layer are used. Next, the global average pooling is computed to calculate the neuron importance weights \hat{w}_n^c , defined as follows:

$$\hat{w}_n^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial o^c}{\partial R_{i,j}^n} \quad (3)$$

where Z is the total number of pixels in the feature map; whereas i, j identifies the pixels. Finally, a weighted combination of the features maps R^n is performed:

$$\hat{w}_n^c = \text{ReLU} \left(\sum_n \hat{w}_n^c R^n \right) \quad (4)$$

where w_n^c are used as weights and the *ReLU* set all the negative values to zero. The result is the so called *Grad-CAM map* or *importance map* where relevant input parts for the classification are identified with coloration grading from blue (low importance) to red (high importance). Such relevance map can be overlapped to the original input (i.e. the HD-CF channel x frequency map) to infer information about the relevance of the input regions. Deriving

information about the relevance of the input regions (CF maps associated to a HD-EEG epoch) means deriving information about how relevance changes across channels (thus across the brain regions) and across frequencies. In this study, for each subject, the the Grad-CAM technique was performed by using the HD-CF epochs, correctly classified as MCI/AD as input to the pre-trained CNN. The output is a visual attention map able to provide information on both the brain regions (i.e., EEG-channels) mostly activated in the progression towards AD (at time T1) and the frequency ranges mainly involved. As an example, Figure 3 reports a Grad-CAM representation of a HD-CF epoch related to Sb01 extracted at time T1 (AD diagnosis). As can be seen, the area associated to the EEG-channels ranged E51-E61 (refer to Figure 1b for the corresponding scalp position) in the δ sub-band (1-4 Hz) resulted the most relevant to classifying the corresponding epoch as “AD”.

Classification metrics

Conventional classification metrics are used to measure the performance of the proposed epoch-based classification system; they are:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (6)$$

$$\text{Positive Predicted Value (PPV)} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Negative Predicted Value (NPV)} = \frac{TN}{TN + FN} \quad (8)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

where TP= True Positive, TN= True Negative, FP= False Positive, FN= False Negative. It is worth noting that the k -fold cross

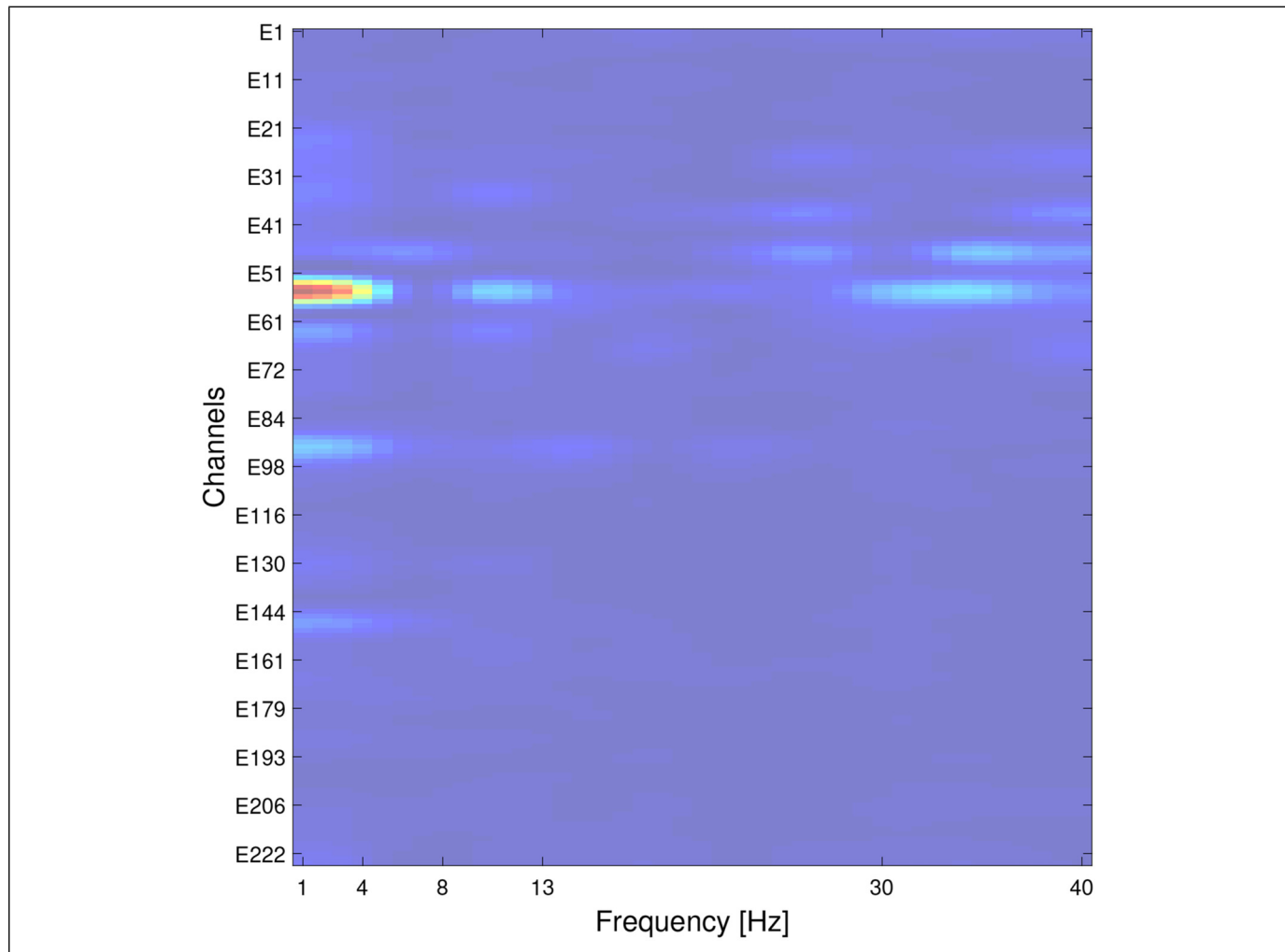


Figure 3. Grad-CAM visualization map of a high-density channel-frequency epoch. Red areas correspond to the input regions highly relevant for the classification; vice-versa, blue areas are the input regions less relevant.

validation technique (with $k=7$) was applied. Hence, classification statistics are expressed in terms of mean value and the corresponding 95% confidence interval²⁸ (95%CI).

Results

Longitudinal epoch-based classification performance

Table 1 reports the longitudinal epoch-based intra-subject classification results achieved by the proposed CNN. As can be

observed, high performance was observed in all the participants. Indeed, it is worth mentioning that the maximum classification performance in terms of accuracy was reported by Sb02 with a score of 98.97% (other statistics were: sensitivity of 98.38% (95% CI: 98.68-99.26) specificity of 99.45% (95% CI: 99.31-99.58), PPV of 99.31% (95% CI:99.14-99.48), NPV of 98.69% (95% CI:98.30-99.09)); whereas, the minimum result was achieved by Sb04 with an accuracy rate up even to 97.79% (other statistics were: sensitivity of 99.80% (95% CI: 99.60-100) specificity of 97.42% (95% CI:

Table 1. Performance of the proposed longitudinal epoch-based classification approach for each subject. Results are showed in terms of mean value and the corresponding 95% confidence interval (CI) reported in the square brackets.

Subject	Sensitivity [%]	Specificity [%]	PPV [%]	NPV [%]	Accuracy [%]
Sb01	98.30, (97.37–99.22)	98.97, (98.65–99.30)	97.54, (96.78–98.29)	99.30, (98.92–99.68)	98.78, (98.48–99.07)
Sb02	98.38, (97.89–98.88)	99.45, (99.31–99.58)	99.31, (99.14–99.48)	98.69, (98.30–99.09)	98.97, (98.68–99.26)
Sb03	98.39, (97.81–98.96)	98.72, (98.13–99.31)	96.74, (95.29–98.20)	99.38, (99.16–99.60)	98.63, (98.22–99.04)
Sb04	99.80, (99.60–100)	97.42, (95.39–99.45)	88.71, (81.17–96.24)	99.96, (99.94–100)	97.79, (96.08–99.49)

95.39-99.45), PPV of 88.71% (95% CI:81.17-96.24), NPV of 99.96% (95% CI:99.94-100)). Overall, the proposed CNN reported an average accuracy among subjects of about 98.54%.

Explainability of channel-frequency maps

Grad-CAM analysis was performed in order to find out which area of the input HD-CF map most contributed in the classification process and, consequently, to find out which channels (i.e. head areas) were mostly involved in the evolution from MCI to AD and in which sub-band. In particular, for every subject, an averaged Grad-CAM map was evaluated by calculating the mean across the Grad-CAM maps extracted from the HD-CF epochs (i.e., HD-EEG epochs) correctly classified as MCI or AD. Figure 4 shows the average Grad-CAM maps at time T0 (MCI state) and at time T1 (AD state) for each subject Sb01, Sb02, Sb03, Sb04. For example, in Figure 4a, the average representation of Sb01 at time T0 (MCI state), showed that the region located around 1–4 Hz and related especially to EEG locations roughly from E11 to E21 (refer to Figure 1b for the corresponding scalp position) appears the most significant in the decision process. At time T1, when the subject converted to AD, an alteration, of the most relevant areas, was observed especially in terms of set of EEG-channels and range of frequencies. Such line of reasoning can be applied also to the other patients. A similar behavior was observed in all the four subjects: overall, different group of channels activated especially in the δ sub-band at time T0 and T1. No difference among channels could be detected in higher frequency sub-bands, which indeed showed low relevance in some subjects (blue color in Sb01 T0-T1, Sb02 T0, Sb03 T0-T1, Sb04 T1) and high relevance in other subjects (red color in Sb02 T1, Sb04 T0). The importance of δ band was also endorsed by the estimation of common relative powers (RP²⁹) reported in Table 2. As can be seen, the highest contribution was detected in the δ sub-band at T0 and T1, with an increase at T1 in Sb02 and Sb03, likely due to the slowing process that characterizes the evolution of AD. Hence, in order to pinpoint the brain activity alterations possibly linked to AD evolution, Figure 5 reports the topographic map of the scalp, only related to the δ sub-band, for each subject at time T0 and time T1. In particular, Sb01 exhibited high relevance in the central area at T0 (MCI diagnosis) and an alteration of the left-temporal and central-frontal regions at T1 (AD diagnosis). Sb02 reported high relevance in the left-frontal zone at time T0 and in the parietal area at T1, respectively. The left-frontal lobe was the most significant area at time T0 and T1 in Sb03. Sb04 reported that the right-frontal area was involved at time T0, whereas the left-frontotemporal area was the most relevant at T1.

Discussion

The presented research is a proof of concept aimed at investigating the capability of xAI to offer new personalized diagnostic opportunities in follow-up studies on subjects at risk of developing dementia due to AD. The study was designed to include a stage of intra-subject classification with the aim of training an artificial neural network (i.e., CNN) to detect

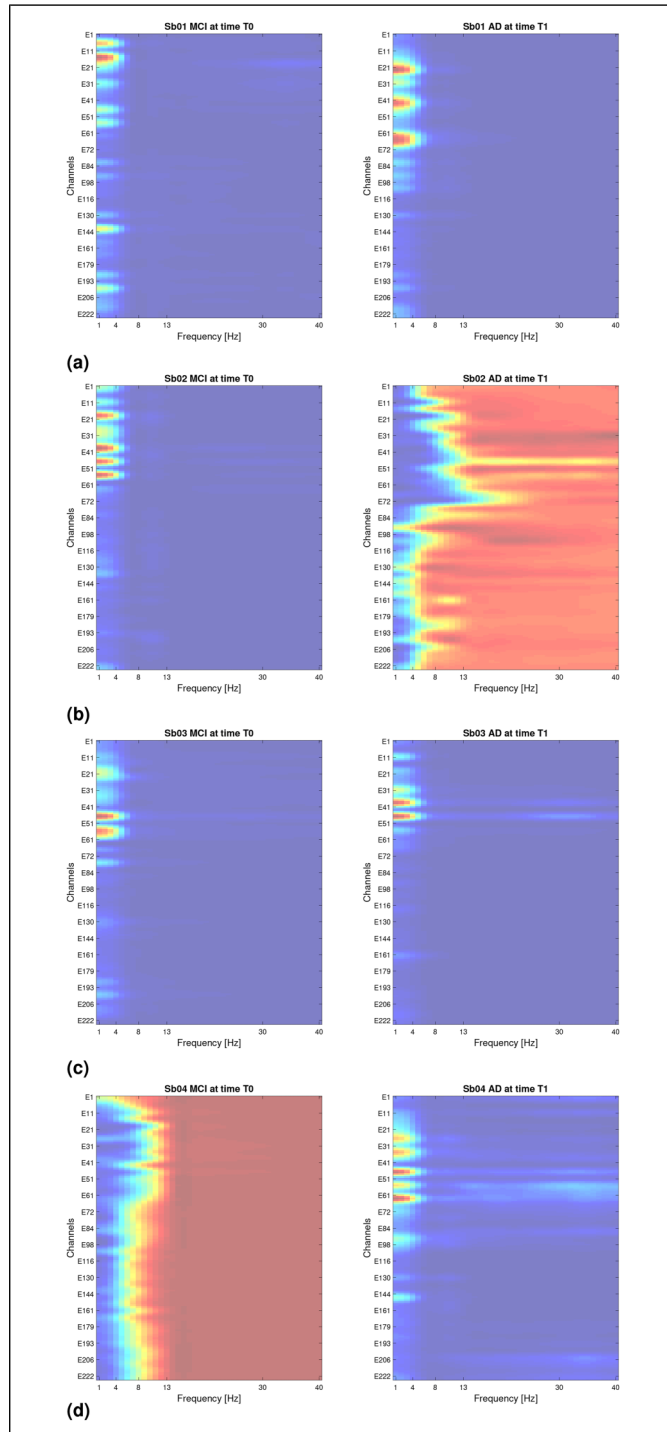


Figure 4. Grad-CAM maps of Subject 01 (a), Subject 02 (b), Subject 03 (c), Subject 04 (d) at time T0 (MCI state) and at time T1 (AD state). Red colour denotes regions with the highest relevance; vice-versa, blue colour denotes regions with the lowest relevance.

Table 2. Relative Power (RP) calculated at time T0 and T1 for each subject.

Subject	Relative Power (RP) at time T0 [%]					Relative Power (RP) at time T1 [%]				
	δ	θ	α	β	γ	δ	θ	α	β	γ
Sb01	97.55	3.38	0.47	0.36	0.23	97.46	8.53	0.44	0.01	0.08
Sb02	66.15	41.23	9.19	0.50	0.02	97.65	2.67	0.55	0.38	0.32
Sb03	65.23	42.36	9.37	0.51	0.02	98.36	2.96	0.21	0.13	0.07
Sb04	78.78	33.17	2.82	0.08	0.01	66.19	41.15	9.13	0.53	0.03

subject-specific characteristics potentially associated with the process of degeneration from MCI to AD.

Experimental results showed that the proposed CNN was able to successfully detect MCI from AD epochs of a same patient with high classification performance in all the participants. In order to find out which EEG channels (scalp electrodes), thus which head regions, were more likely to have been involved in the evolution of the disease from the MCI condition to AD condition, a Grad-CAM-based analysis was carried out. In particular, the inspection of the highly important regions (depicted in red color) allowed to detect not only the corresponding most significant EEG-channel (i.e., head region) but also the frequency range (i.e., sub-band) involved. It is worth mentioning that subjects experiencing a worsening from MCI to AD are not expected to exhibit the same involvement in terms of head regions deterioration. Some general trend can be found in the literature about MCI conversion to AD (i. e. slowing effect, loss of functional connectivity, etc), but individual and subject-dependent characteristics should be expected. To the best of our knowledge, this is the first work that attempts to explore the longitudinal changes from MCI to AD by means of xAI. Nonetheless, the proposed approach has some drawbacks. The major limitation lies in the small number of patients involved. In the future, a significant cohort of MCI subjects will be recruited and monitored longitudinally, considering also more steps in the follow-up program (i.e, T2, T3 etc.). This is indeed strictly necessary and of extreme importance to objectively estimate the evolution of MCI to AD and also to improve the reliability and validity of the proposed AI-based model. Furthermore, since AD causes a gradual decline of the brain, MCI and AD patients may experience different states of the disease. This means that an EEG epoch of a severe MCI can show similar patterns to an EEG epoch belonging to a mild AD patient, in this way causing a misclassification of that epoch and consequently a decrease of the classification performance. Despite some rather universal characteristics of AD development (slowing in EEG rhythms, functional connectivity decrease, etc), the way the disease affects a given area of the brain can vary significantly across patients as the brain deteriorates in an individual way. This makes clear why follow-up studies are crucial in AD development monitoring and why we did not expect to observe similar results in all the patients. Intra-subject, follow-up, studies are prone to a bias induced by an inherent correlation between samples recorded within the same session. This is a challenging

issue in all follow-up studies which, however, are strictly necessary for monitoring neurodegenerative disorders longitudinally. In order to reduce such bias and to make the recording conditions as similar as possible across T0 and T1 sessions, the same patient's position and room conditions (noise, temperature, etc) were reproduced, as well as the quality and quantity of meal and sleep preceding the recording. Furthermore, each EEG recording was pre-processed conveniently to delete noise/artifacts and was inspected by an expert EEG technician in order to exclude any defective electrode. The impedance of each electrode was set lower than 50 k Ω . Notwithstanding, the identifiability of the session cannot be hypothesized to have been totally suppressed. In the future, longitudinal changes in EEGs could be compared with longitudinal changes detected by means of other diagnostic techniques like, for example, functional Magnetic Resonance Imaging (fMRI) or functional Near-Infrared Spectroscopy (fNIRS), even though it is worth to keep in mind that both techniques are sensitive to blood oxygenation rather than to the electrical activity of the brain thus cannot offer a different perspective on the very same phenomenon.

Conclusion

In this work a longitudinal epoch-based classification approach able to discriminate CF maps of HD-EEG signals recorded at time T0 (MCI diagnosis) and at time T1 (AD diagnosis) was proposed. A custom CNN was developed to perform the intra-subject binary epoch-based classification: MCI vs. AD, reporting accuracy rate up to 98.97% (95% CI: 98.68-99.26) (Sb02). Furthermore, in order to identify the input channel-frequency regions that were mostly involved in the classification process, xAI was employed. Specifically, Grad-CAM-based analysis allowed to detect which head region activated during the progression of AD. However, there is still a long way to go in the longitudinal monitoring of AD. In the future, a large number of participants will be enrolled in a follow-up program and high-density EEG recordings will be collected. In addition, motivated by the encouraging results, we intend to use the proposed framework also to investigate the evolution of other forms of neural disorders.

Authorship

Francesco Carlo Morabito: Conceptualization, Methodology, interpretation of data, Supervision, Writing- Reviewing and Editing;

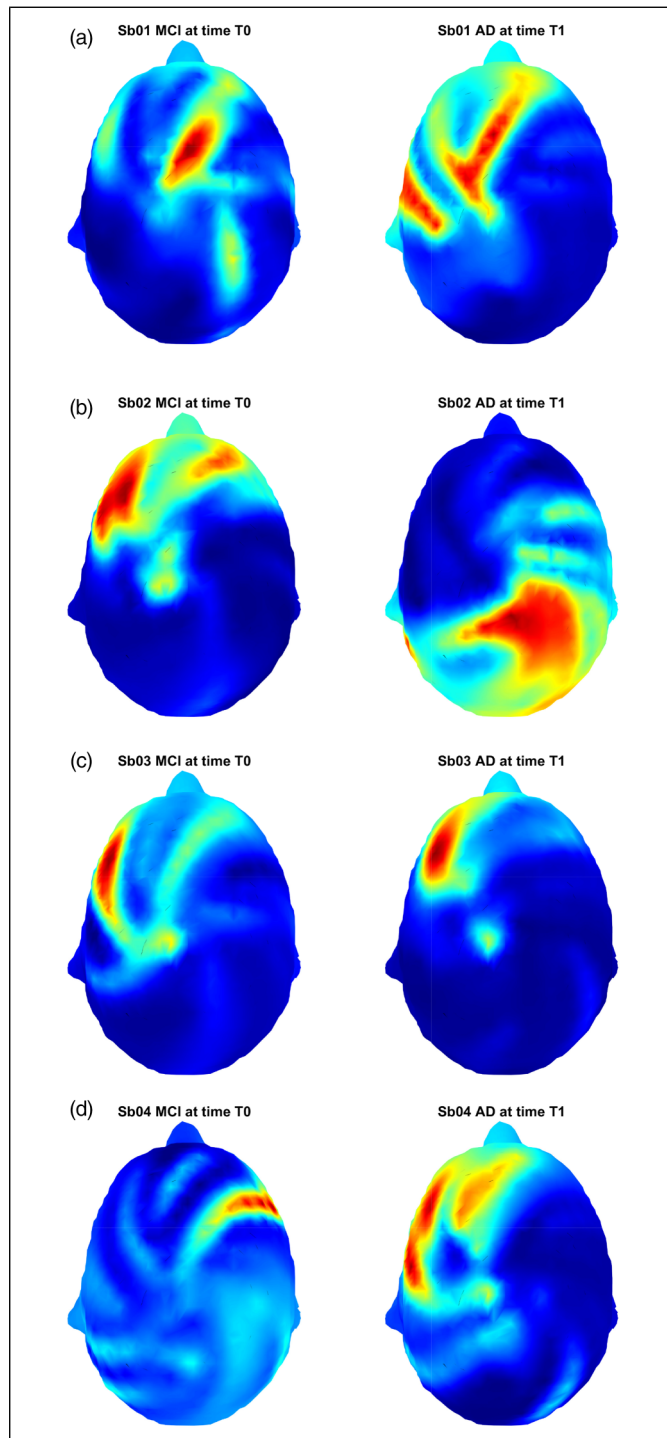


Figure 5. Topographic representations of the Grad-CAM maps, only related to the δ sub-band, for Subject 01 (a), Subject 02 (b), Subject 03 (c), Subject 04 (d) at time T0 (MCI state) and at time T1 (AD state). Red colour denotes regions with the highest relevance; vice-versa, blue colour denotes regions with the lowest relevance.


Cosimo Ieracitano: Conceptualization, Methodology, Investigation, Implementation, Validation, Visualization, Formal analysis, Interpretation of data, Writing- Original draft preparation, Reviewing

and Editing; **Nadia Mammone:** Conceptualization, Methodology, interpretation of data, Writing- Reviewing and Editing;

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

ORCID iD

Francesco Carlo Morabito  <https://orcid.org/0000-0003-0734-9136>

References

1. Sperling RA, Aisen PS, Beckett LA et al. Toward Defining the Preclinical Stages of Alzheimer's Disease: Ecommendations From the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease. *Alzheimer's Dementia* 2011;7(3):280-292.
2. Jeong J. EEG Dynamics in Patients with Alzheimer's Disease. *Clinical Neurophysiol* 2004;115(7):1490-1505.
3. Nunez PL, Srinivasan R. *Electric fields of the brain: the neuro-physics of EEG*. Oxford University Press, USA, 2006.
4. Aghajani H, Zahedi E, Jalili M et al. . Diagnosis of Early Alzheimer's Disease Based on EEG Source Localization and a Standardized Realistic Head Model. *IEEE J Biomed Health Inform* 2013;17(6):1039-1045.
5. Dubovik S, Bouzerda-Wahlen A, Nahum L et al. Adaptive Reorganization of Cortical Networks in Alzheimer's Disease. *Clin Neurophysiol* 2013;123(1):35-43.
6. Mammone N, De Salvo S, Bonanno L et al. Brain Network Analysis of Compressive Sensed High-density EEG Signals in AD and MCI Subjects. *IEEE T Ind Inform* 2018;15(1):527-536.
7. Huang C, Wahlund LO, Dierks T et al. Discrimination of Alzheimer's Disease and Mild Cognitive Impairment by Equivalent EEG Sources: a Cross-sectional and Longitudinal Study. *Clin Neurophysiol* 2000;111(11):1961-1967.
8. Jelic V, Johansson S, Almkvist O et al. Quantitative Electroencephalography in Mild Cognitive Impairment: Longitudinal Changes and Possible Prediction of Alzheimer's Disease. *Neurobiol Aging* 2000;21(4):533-540.
9. Moretti D, Frisoni GB, Fracassi C et al. MCI Patients' EEGs Show Group Differences Between those Who Progress and those Who Do Not Progress to AD. *Neurobiol Aging* 2011;32(4):563-571.
10. Luckhaus C, Grass-Kapanke B, Blaaser I et al. Quantitative EEG in Progressing Vs Stable Mild Cognitive Impairment (MCI): Results of a 1-year Follow-up Study. *Inter J Geriatr Psychiat: A J Psychiat Late Life Allied Sci* 2008;23(11):1148-115.
11. Morabito FC, Campolo M, Labate D et al. A Longitudinal EEG Study of Alzheimer's Disease Progression Based on a Complex Network Approach. *Intern J Neural Syste* 2015;25(02):1550005.
12. Mammone N, De Salvo S, Ieracitano C et al. A Permutation Disalignment Index-based Complex Network Approach to Evaluate Longitudinal Changes in Brain-electrical Connectivity. *Entropy* 2017;19(10):548.
13. Mammone N, Ieracitano C, Adeli H et al. . Permutation Jaccard Distance-based Hierarchical Clustering to Estimate EEG Network Density Modifications in MCI Subjects. *IEEE T Neural Networks Learning Syst* 2018;29(10):5122-5134.

14. Dattola S, La Foresta F. An ELORETA Longitudinal Analysis of Resting State EEG Rhythms in Alzheimer's Disease. *Appl Sci* 2020;10(16):5666.
15. Şeker M, Özbek Y, Yener G, Özerdem MS. Complexity of EEG Dynamics for Early Diagnosis of Alzheimer's Disease Using Permutation Entropy Neuromarker. *Compu Meth Progra Bio* 2021;206:106116.
16. Huggins CJ, Escudero J, Parra MA et al. Deep Learning of Resting-state Electroencephalogram Signals for 3-class Classification of Alzheimer's Disease, Mild Cognitive Impairment and Healthy Ageing. *J Neural Eng* 2021; (0): –.
17. Toural JES, Pedrón AM, Reyes EJM. A New Method for Classification of Subjects with Major Cognitive Disorder, Alzheimer Type, Based on Electroencephalographic Biomarkers. *Inform Med Unlocked* 2021;23:100537.
18. Ieracitano C, Mammone N, Bramanti A et al. . A Convolutional Neural Network Approach for Classification of Dementia Stages Based on 2D-spectral Representation of EEG Recordings. *Neurocomputing* 2019;323:96-107.
19. Morabito FC, Campolo M, Ieracitano C et al. Deep convolutional neural networks for classification of mild cognitive impaired and Alzheimer's disease patients from scalp EEG recordings. in *2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)*:1–6IEEE, 2016.
20. Ieracitano C, Mammone N, Hussain A et al. . A Novel Multi-modal Machine Learning Based Approach for Automatic Classification of EEG Recordings in Dementia. *Neural Networks* 2020;123:176-190.
21. Edition F. Diagnostic and Statistical Manual of Mental Disorders. *Am Psychiatric Assoc* 2013;21.
22. Kayser J, Tenke CE. On the Benefits of Using Surface Laplacian (current Source Density) Methodology in Electrophysiology. *Intern J Psychophysiol* 2015;97(3):171.
23. Heinzel G, Rüdiger A, Schilling R, Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new at-top windows, 2002.
24. Yan LC, Yoshua B, Geoffrey H. Deep Learning. *Nature* 2015;521(7553):436-444.
25. Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification with Deep Convolutional Neural Networks. *Adv Neural Inform Proc Syst* 2012;25:1097-1105.
26. Holzinger A, Malle B, Saranti A et al. . Towards Multi-modal Causability with Graph Neural Networks Enabling Information Fusion for Explainable AI. *Inform Fusion* 2021;71:28-37.
27. Selvaraju RR, Cogswell M, Das A et al. . Grad-cam: Visual explanations from deep networks via gradient-based localization. in *Proceedings of the IEEE international conference on computer vision*:618–626, 2017.
28. Hazra A. Using the Confidence Interval Confidently. *Thoracic Disease* 2017;9(10):4125.
29. Bian Z, Li Q, Wang L et al. . Relative Power and Coherence of EEG Series are Related to Amnesic Mild Cognitive Impairment in Diabetes. *Frontiers Aging Neurosci* 2014;6:11.
30. Mammone N, De Salvo S, Ieracitano C et al. Compressibility of High-Density EEG Signals in Stroke Patients. *Sensors* 2018;18(12):4107.