Computer Methods and Programs in Biomedicine 275 (2026) 109159

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.sciencedirect.com/journal/computer-methods-and-programs-in-biomedicine

Revealing EEG signatures of intervention in disorder of consciousness using artificial intelligence: methodology and feasibility

Davide Borra [a,*], Valentina Bonsangue [b], Sofia Straudi [c], Elisa Magosso [a]

[a] Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi" (DEI), University of Bologna, Cesena, Italy
[b] Severe Brain Injury Unit, Ferrara University Hospital, Ferrara, Italy
[c] Neuroscience and Rehabilitation Department, Ferrara University, Ferrara, Italy

ARTICLE INFO

Keywords:
Minimally conscious state
Therapeutic intervention
EEG
Transcranial direct current stimulation
Explainable artificial intelligence
Frequency-domain EEG signatures

ABSTRACT

Background and Objective: Electroencephalography (EEG) is a crucial tool for monitoring recovery in patients with disorders of consciousness (DOC) after therapeutic interventions. It helps in identifying the neural correlates and in guiding the development of personalized treatments. Spectrum power measures are widely employed. However, these measures are manually handcrafted, not patient-specific, and not tailored to the specific intervention.
Methods: To address these limitations, we propose an explainable artificial intelligence (XAI) framework designed to automatically uncover the most salient frequency-domain EEG signatures in an intervention- and patient-specific manner. The framework integrates an interpretable convolutional neural network, which is capable of learning interpretable frequency-domain EEG features, with an explanation technique, which quantifies the relevance of the learned spectral features. This approach enables the automatic tracking of patient-specific spectral EEG changes and refines the analysis toward neural features that are more closely associated with key clinical variables.
Results: We showcase the potential of our approach by applying it to EEG signals collected from patients in a minimally conscious state following an intervention based on transcranial direct current stimulation. The XAI results reveal a prominent role of alpha-band EEG oscillations in DOC intervention, supporting evidence that functional improvements due to intervention are associated with an increase in alpha-band spectral content.
Conclusions: Our XAI-driven analysis offers a robust, individualized, and transparent alternative (or complement) to conventional EEG analyses, thereby enhancing the EEG characterization of DOC patients.

1. Introduction

Disorders of consciousness (DOC) are conditions characterized by impaired wakefulness and awareness. Patients surviving severe brain injuries can develop DOC, such as coma, minimally conscious state (MCS), and unresponsive wakefulness syndrome [1]. Coma patients are unarousable, and unaware of themselves and their environment (unwakefulness, reflex behaviors only). In unresponsive wakefulness syndrome, patients show signs of wakefulness but no signs of awareness of self and the environment (reflex behaviours only). In contrast, in MCS there is behavioral evidence of self or environmental awareness. Patients recovering from DOC (e.g., recovering functional communication or functional use of objects) emerge from MCS [2]. The clinical management of DOC patients is challenging, primarily because they are unable to communicate and rely entirely on others for their care [3]. Despite this, only a few studies have investigated the treatment of DOC patients [3]. Therapeutic strategies to improve consciousness include pharmacological and non-pharmacological interventions [4]. Pharmacological interventions involve agents such as amantadine, midazolam, or zolpidem [5–7]. Non-pharmacological interventions involve both invasive brain stimulation, such as deep-brain stimulation, and non-invasive brain stimulation, such as transcranial direct current stimulation (tDCS) [8–11] or repetitive transcranial magnetic stimulation [12–15]. DOC patients exhibit heterogeneous responses to therapy, reflecting diverse etiologies and types of brain injury. Generally, MCS patients show greater responsiveness to treatments but the outcomes vary considerably across individuals [4]. These factors underscore the potential benefit of patient-specific intervention protocols tailored to each

* Corresponding author at: Department of Electrical, Electronic and Information Engineering "Guglielmo Marconi", University of Bologna – Cesena Campus, Via dell'Università 50, 47521, Cesena (FC), Italy.
E-mail address: davide.borra2@unibo.it (D. Borra).

https://doi.org/10.1016/j.cmpb.2025.109159
Received 11 July 2025; Received in revised form 7 November 2025; Accepted 8 November 2025
Available online 9 November 2025

individual's clinical and neurophysiological profile.

Assessing the cerebral response during therapy is critical for understanding the mechanisms in action and guiding the development of personalized treatment strategies. Combining interventions with brain activity assessments can improve our knowledge of the neural correlates underlying clinical responses and the possible neuroplastic mechanisms occurring after brain injury [3,16]. In clinical practice, non-invasive and cost-effective techniques such as electroencephalography (EEG) are widely used to assess DOC patients [4]. EEG supports clinicians in the diagnosis (e.g., distinguishing MCS from unresponsive wakefulness syndrome), prognosis (e.g., identifying prognostic indicators), and monitoring therapeutic interventions, typically via resting-state recordings [4]. Measures derived from EEG analyses have been applied to track the brain response to both pharmacological and non-pharmacological DOC interventions [5,6,8,9,12,13,15]. Spectrum power measures, which quantify the strength of brain oscillations [17], are frequently used. Indeed, a common EEG modulation in DOC patients involves increased power at low frequencies (delta and theta) and decreased power at higher frequencies (alpha) [4]. This slowing of EEG oscillations correlates with the severity of the traumatic brain injury [18,19]. The recovery from DOC is associated with spectral changes, typically showing a shift of the dominant EEG band from lower (delta and/or theta) to higher (alpha) frequencies. This represents the most common EEG spectral change in patients who recover consciousness [4, 20]. Notably, alpha-band EEG power positively correlates with functional improvements of consciousness, as measured by clinical recovery scales [9]. Taken together, these findings highlight the key role of frequency-domain changes, particularly within the alpha band, in assessing brain responses to DOC interventions.

Building on this evidence, previous studies investigated EEG signals in DOC and assessed EEG spectral modulations during the intervention [5,6,8,9,12,13,15], reporting promising findings. However, these analyses present some limitations that could be addressed via artificial intelligence. First, the measures used to characterize DOC patients and interventions are *manually handcrafted*. They are generally derived by computing the power within EEG bands, either by aggregating signals across predefined groups of channels (e.g., all frontal channels) or by focusing only on a small, preselected subset of channels from the full multivariate EEG recording (e.g., mainly frontal, central, or parietal channels). Second, and more importantly, the measures are *patient-* and *intervention-aspecific*, as they are defined in the same way across patients and fail to capture the neural aspects uniquely associated with DOC and the therapeutic intervention. These limitations can be resolved by adopting analysis frameworks able to automatically process the full neural activity, preserving potentially relevant information, and to define measures that inherently reflect DOC-related and intervention-related characteristics on a patient-by-patient basis. Such approaches have the potential to boost the analysis and understanding of the neural substrates underlying DOC, and to foster the development of innovative biomarkers for personalized therapeutic interventions.

Recent advancements in EEG analysis have introduced deep learning-based approaches that enable a data-driven characterization of neural signals [21–27]. These methods rely on interpretable neural networks that, after a training process, automatically extract the most informative neural features to map the input EEG onto specific output states (e.g., cognitive, or behavioral states). A key advantage of interpretable neural networks is their ability to learn features that are inherently interpretable in domains relevant for EEG signals, such as the frequency domain. Importantly, these approaches can process high dimensional data, thus, can exploit the full multi-variate brain activity as input, without discarding potentially relevant information. Moreover, by properly setting network training, it is possible to reveal the neural features specific to individual patients (e.g., by using patient-specific data during training), and to specific therapeutic conditions (e.g., by considering different moments of DOC intervention as output). To further enhance interpretability, the models can be combined with an explanation technique [28] to quantify the importance of the learned EEG features for predicting the output states [23–27]. This combination realizes an explainable artificial intelligence (XAI) algorithm. This algorithm operates directly on the full multi-variate neural activity, enabling the *automatic* definition of EEG measures that are *subject-* and *intervention-specific*. Moreover, XAI approaches proved to reveal the neural correlates underlying cognitive tasks better than traditional EEG analysis pipelines [24,25]. Nevertheless, the potential of XAI for the analysis of patients' signals and for revealing the neural traits due to intervention are still underexplored.

In this study, we aim at overcoming the current limitations in the characterization of EEG signatures associated with DOC and DOC intervention (i.e., the use of manually handcrafted, patient-aspecific, and intervention-aspecific measures), by showcasing the potential of XAI to uncover EEG traits in MCS patients following a tDCS-based intervention in a longitudinal pilot clinical study [9]. Our XAI approach integrates an interpretable neural network (Sinc-ShallowNet [23]) with an explanation technique (DeepLIFT [29]). The neural network automatically learns frequency-domain EEG features that are directly interpretable, and it is applied to discriminate resting-state EEG acquired before vs. after tDCS, separately for each patient. The explanation technique quantifies the relevance of the frequency-domain features used for discrimination (between before and after tDCS), thereby identifying the spectral components most affected by the brain stimulation. Then, the spectral relevance emerging from the XAI analysis was related to clinical scores assessing behavioral improvement after the intervention compared to before. With this approach, we primarily aspire to explore, for the first time, the feasibility of using artificial intelligence to reveal the neural changes due to intervention in DOC, in an automatic, patient-specific, and intervention-specific manner. The main contribution of this work lies in developing and testing the feasibility of a methodological pipeline capable of capturing clinically relevant EEG patterns of DOC at the individual level. This individual level analysis is of extreme importance when dealing with patient data, due to the almost uniqueness of neuropathological alterations and of neural mechanisms underlying recovery. As a secondary contribution, this study extends the validation of XAI approaches to clinical EEG signals collected from patients, while the application of these approaches in the literature remains primarily limited to healthy participants. Ultimately, our approach could provide important clinical implications, as it can be exploited, in prospective, to derive novel spectral EEG biomarkers specific of each DOC patient and to monitor the neural effects of DOC interventions.

2. Methods

2.1. Data description and data pre-processing

In this study, we analyzed the data acquired by Straudi et al. [9] in a longitudinal pilot EEG-tDCS study involving patients in a minimally conscious state following traumatic brain injury who underwent a tDCS-based intervention. In this section, first a brief description of the longitudinal pilot study is provided; additional details about the pilot study, tDCS stimulation, and data recordings can be found in the reference publication [9]. Then, the pre-processing of EEG signals is described.

2.1.1. Participants

Ten patients were enrolled in the study (35.5 $\pm$ 12.6 years, 7 males, 5.5 $\pm$ 5.4 years post trauma) at the Severe Brain Injury Unit (Neuroscience and Rehabilitation Department, Ferrara University Hospital, Italy). Inclusion criteria were: i) age between 18 and 60 years; ii) MCS diagnosis; iii) traumatic etiology; iv) >12 months after the injury. The clinical study was approved by the ethics committee of the Ferrara University Hospital and registered on the Clinicaltrial.gov database (NCT02288533). For each patient, the legal representative and the

medical doctor were informed regarding the study objectives and procedures, and a written informed consent was signed by the legal representative.

2.1.2. Experimental protocol

The experimental protocol consisted of 10 sessions of anodal tDCS (five sessions per week, 40 minutes per session) using two electrodes (anode) placed in correspondence of the primary motor cortex bilaterally, and one electrode (cathode) positioned at the nasion.

The clinical assessment was conducted by two experienced clinicians using the Italian version of the Coma Recovery Scale-Revised (CRS-R) [30], the gold standard for the assessment of DOC with adequate reliability and validity [31]. This standardized neurobehavioral scale evaluates the residual functions of patients across multiple domains, including auditory, visual, motor, verbal, communication, and arousal functions. The instrumental assessment consisted of a 15-minute resting-state EEG recording. Electrodes were placed according to the 10-20 international system (ground at AFz) and were: Fp1, Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, and O2 (Fig. 1a). Signals were acquired at a sampling frequency of 1000 Hz.

Both clinical and instrumental assessments were performed after 5 sessions (T1) and at the end of the 10 sessions (T2) of tDCS, as well as two weeks before (T-1) and one day before (T0) the start of the experimental protocol (Fig. 1b). Patients no. 6 and 7 lacked the EEG recording in one timeline before the tDCS (at T0 and at T-1 respectively), due to technical issues. Patient no. 8 did not complete the stimulation protocol due to a septic shock, unrelated to the tDCS application, that affected his clinical stability (see Straudi et al. [9] for additional details); therefore, this patient was excluded from the analyses, and the data from 9 patients were used. The total CRS-R scores (i.e., the sum of the scores across the different subscales) obtained from the clinical assessment are summarized in Table 1.

Table 1
Clinical assessment. The Coma Recovery Scale-Revised (CRS-R) total scores collected before and after the tDCS intervention are reported in the second and third columns, respectively. Each value represents the average CRS-R score across the two timelines acquired before (T-1 and T0) and after (T1 and T2) the tDCS. CRS-R values corresponding to timelines without EEG recordings (i.e., one timeline before tDCS in patients no. 6 and 7, see also Section 2.1.2) were excluded from the averages. Individual CRS-R scores for each timeline are reported within brackets, i.e., at (T-1, T0) and at (T1, T2), respectively for before- and after-tDCS conditions. The CRS-R improvement was computed as the difference of the CRS-R between after- vs. before-tDCS when the difference is positive, and zero otherwise; negative or zero differences were treated as no improvement.

Patient ID	CRS-R before-tDCS $= (CRSR_{T-1} + CRSR_{T0})/2$	CRS-R after-tDCS $= (CRSR_{T1} + CRSR_{T2})/2$	CRS-R improvement $= max(CRSR_{after} - CRSR_{before}, 0)$
1	15 (15, 15)	16 (16, 16)	1
2	10 (11, 9)	11 (12, 10)	1
3	13 (12, 14)	15 (18, 12)	2
4	13 (13, 13)	13 (13, 13)	0
5	12 (12, 12)	14.5 (15, 14)	2.5
6	9 (None, 9)	10.5 (11,10)	1.5
7	9 (9, None)	7 (7, 7)	0
9	7 (8, 6)	9.5 (8, 11)	2.5
10	9.5 (9, 10)	10.5 (10, 11)	1

2.1.3. EEG pre-processing

For each patient, EEG signals acquired at different timelines were pre-processed as follows, adopting a pre-processing pipeline similar to that used in previous EEG studies [32,33]:

i. Resampling to 500 Hz.
ii. High-pass filtering at 0.5 Hz (elliptic infinite-impulse-response filter) and notch filtering at 50 Hz.
iii. Identification of bad channels within signals of each timeline via random sample consensus method [34] (3 bad channels, on average across patients).
iv. Concatenation of electrode signals across the four timelines.
v. Removal of channels labeled as bad at least in one timeline.
vi. Removal of the main artifacts (ocular, muscular) via independent component analysis (3 removed components, on average across patients).
vii. Spherical spline interpolation of the bad channels removed.
viii. Extraction of 2s non-overlapped epochs with amplitude <100 μV in all channels, separately from each timeline.
ix. Downsampling to 125 Hz, to reduce computational cost in the XAI approach.

Following this pre-processing, 316, 414, 352, and 318 EEG epochs were extracted on average across patients, for the T-1, T0, T1, and T2 timelines, respectively. Each epoch had a shape of $(C, T) = (27, 250)$, where C and T denote the number of EEG channels and time samples,

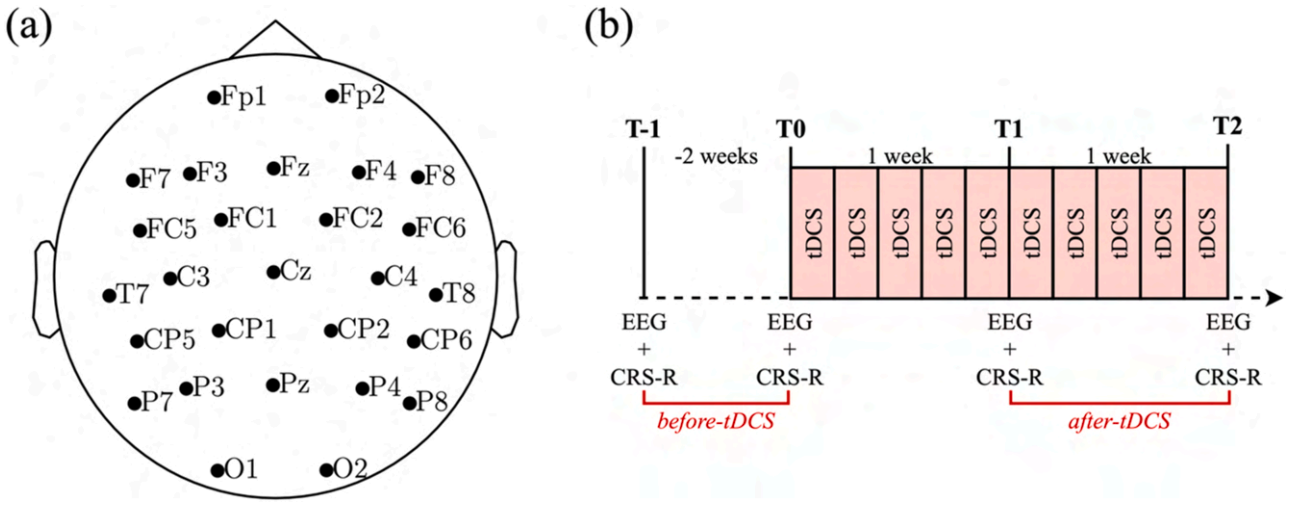

Fig. 1. EEG electrode location and study timeline. Panel a – Electrode layout. Twenty-seven electrodes were placed according to the 10-20 international system (ground at AFz). Panel b – Study timeline. Clinical (CRS-R scores) and instrumental (EEG) assessments were performed at four timelines during the pilot study (T-1 and T0: before applying tDCS; T1 and T2: after applying tDCS).

respectively. As described below, individual resting-state EEG epochs were used as input to a neural decoder trained to discriminate before-tDCS vs. after-tDCS epochs. The knowledge learned by the decoder to solve this classification problem was then exploited to analyze the EEG signatures associated with the intervention at the single-patient level, implementing the XAI-based EEG analysis.

2.2. Analysis of the spectral EEG signatures of DOC intervention via artificial intelligence

The XAI approach is illustrated in Fig. 2. It comprises an interpretable neural network (Sinc-ShallowNet [23]) and an explanation technique (DeepLIFT [29]). The neural network is designed to learn interpretable EEG spectral features (light-red box in Fig. 2) and is trained to distinguish between before and after tDCS from the resting-state EEG epochs collected from DOC patients. The explanation technique is then applied to quantify the relevance of the EEG spectral features learned for the discrimination (yellow box in Fig. 2). Together, these algorithms form a powerful framework capable of automatically identifying the frequency-domain EEG signatures that exhibit the most prominent changes following the therapeutic intervention, in a patient-specific manner.

2.2.1. Network architecture

We employed Sinc-ShallowNet [23], an interpretable neural network that learns EEG spectral features in an interpretable way. The network processes the input EEG epoch $X \in \mathbb{R}^{C\times T}$ by first applying a temporal convolution followed by a spatial convolution, and finally produces the decision $y \in L = \{l_0, l_1\} = \{\text{before} - \text{tDCS}, \text{after} - \text{tDCS}\}$. The key element of this network is the adoption of an interpretable temporal convolution, named temporal sinc-based convolution, realized by constraining the filters to be ideal bandpass finite-impulse-response filters.

Denoting with $x[n]$ a 1D time series (e.g., one EEG channel) and $h[n]$ a convolutional kernel, the output $y[n]$ of the convolution is given by $y[n] = (x*h)[n]$. Unlike traditional convolutional layers, where each kernel sample is learned ($h[n], \forall n$), the sinc-based convolution assumes a known analytical expression for the impulse response of the filter (ideal bandpass filter, denoted by $g(n;\theta)$) and exposes only a limited set of parameters θ to the training. These parameters are easily interpretable in the frequency domain, such as the filter cutoff frequencies.

$$\begin{cases} y[n] = (x*h)[n;\theta] = \sum_l x[n-l]h[l;\theta] \\ h[n;\theta] = g(n;\theta) \end{cases} . \tag{1}$$

In case of a trainable bank of ideal bandpass filters, the analytical expression of the impulse response is given by:

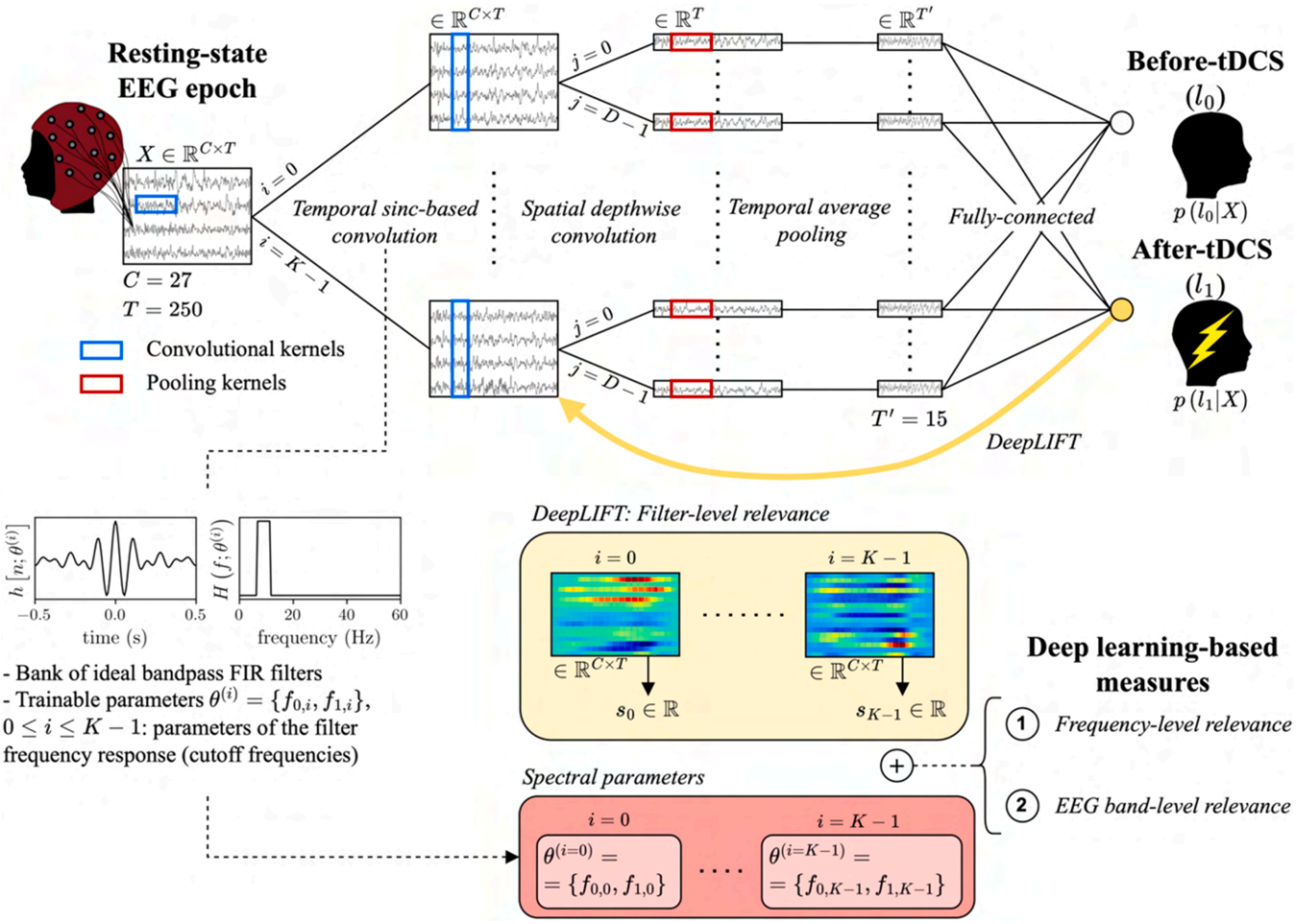

Fig. 2. Explainable artificial intelligence approach for analyzing tDCS-based intervention in disorder of consciousness (DOC). An interpretable neural network classifies before-tDCS vs. after-tDCS resting-state EEG epochs recorded from DOC patients with a minimally conscious state diagnosis. The network architecture is illustrated at a high-level on top. The model learns interpretable frequency-domain features, estimating the spectral parameters of the optimal bank of bandpass filters (cutoff frequencies, light-red box) used for classification. An explanation technique (DeepLIFT, yellow arrow and box) is then applied to quantify the relevance of each bandpass filter (filter-level relevance) for predicting the after-tDCS condition. The derived relevance values, together with the interpretable spectral parameters of the network, are combined to derive deep learning-based measures sensitive to DOC intervention, revealing the frequencies (frequency-level relevance) and the EEG bands (band-level relevance) most modulated after the intervention.

$$h[n;\theta=\{f_1,f_2\}]=g(n;\theta)=2f_2 sinc(2\pi f_2 n)-2f_1 sinc(2\pi f_1 n), \quad (2)$$

where f_1 and f_2 denote the low and high cutoff frequencies of the ideal bandpass filter, respectively, and constitute the interpretable parameters. Indeed, the frequency response is given by the difference of two heavy-side step functions:

$$H(f;\theta=\{f_1,f_2\})=rect\left(\frac{f}{2f_2}\right)-rect\left(\frac{f}{2f_1}\right). \quad (3)$$

To mitigate the effects of windowing the impulse response, the filter values $h[n;\theta=\{f_1,f_2\}]$ are multiplied by a Hamming window. Of course, in practice K convolutional kernels are trained and used (rather than one, as shown previously for brevity). Thus, the set of interpretable parameters is $\theta^{(i)}=\left\{f_{1,i},f_{2,i}\right\}, 0\le i\le K-1$.

The network architecture consists of three main blocks (Table 2 and Fig. 2), and it is described in the following; the main network hyper-parameters were set as in Borra et al. [23], where Sinc-ShallowNet was originally introduced.

i. *Spectral and spatial feature extractor (block no. 1)*. This block includes the interpretable temporal sinc-based convolution, which filters the input through a bank of $K=32$ trainable ideal bandpass filters, each with a kernel size $F=(1, 63)$. Subsequently, a spatial depthwise convolution optimally recombines the information across all electrode sites, by learning $D=2$ spatial filters for each bandpass-filtered version of the input (for a total of 64 spatial filters), each with a kernel size $F=(C, 1)=(27, 1)$. Neurons are activated via exponential-linear functions (ELUs) [35], and batch normalization [36] is included after each convolutional layer.
ii. *Feature aggregator (block no. 2)*. To reduce the computational cost, temporal average pooling summarizes the information over time, with a pooling size $F=(1, 63)$ and stride $S=(1, 13)$. This corresponds to averaging the activity of approximately 500 ms of data with a stride of 100 ms, reducing the number of time samples to process (T) from 250 to 15.
iii. *Classifier (block no. 3)*. The feature maps produced by the feature aggregator block are flattened and passed to a fully-connected layer with $N=2$ neurons, one per class (before-tDCS and after-tDCS). Output neurons are activated using a softmax activation function to produce the conditional probabilities $p(l|X),\ l\in L=\{l_0,\ l_1\}=\{\text{before}-\text{tDCS},\text{after}-\text{tDCS}\}$.

In addition to batch normalization, acting as a regularizer, dropout [37] (dropout probability $p=0.5$) is applied immediately before the block no. 3. The parameters of the spatial depthwise convolutional layer and of the fully-connected layer are constrained to have a maximum norm of 1 and 0.25, respectively. Furthermore, the parameters of the interpretable temporal convolutional layer (i.e., the 32 pairs of cutoff frequencies) are constrained to have the central frequency $\in[1,13]$ Hz and the bandwidth $\in[2,5]$ Hz. These ranges reflect the spectral modulations typically associated with DOC recovery, which involves changes of brain oscillations in delta, theta, and alpha band [20].

Table 2
Interpretable neural network architecture: layer details. The table reports the name, main hyper-parameters, number of trainable parameters, and output shape for each layer. Unless specified, unit stride (S) and no padding (P) are applied. The total number of trainable parameters is 3.9k.

Block ID	Layer name	Main hyper-parameters	Trainable param.	Output shape
	Input	-	0	(1, 27, 250)
1	*SincConv2D*	$K=32$, $F=(1, 63)$	64	(32, 27, 250)
	BatchNorm2D	-	64	(32, 27, 250)
	DepthConv2D	$D=2$, $F=(27, 1)$	1728	(64, 1, 250)
	BatchNorm2D	-	128	(64, 1, 250)
	ELU	-	0	(64, 1, 250)
2	*AvgPool2D*	$F=(1, 63)$, $S=(1, 13)$	0	(64, 1, 15)
	Dropout	$p=0.5$	0	(64, 1, 15)
3	*Flatten*	-	0	(960)
	Fully-connected	$N=2$	1922	(2)
	Softmax	-	0	(2)
			3906	

2.2.2. Network training and evaluation

The cross-entropy was used as loss function, and parameters optimized using adaptive moment estimation (150 training epochs, learning rate set at 1e-4, mini-batch size of 32). Parameter updates were weighted depending on the class frequencies, assigning more weight to the under-represented class ([1.0, 1.1] on average across patients for before- and after-tDCS classes, respectively). The neural network was trained on patient-specific data adopting a within-patient training strategy. Specifically, the patient's 2s-epochs at T-1 and T0 were pooled together and labeled as 'before-tDCS', while epochs at T1 and T2 were pooled together and labeled as 'after-tDCS'. This resulted in an overall dataset of 1400 EEG epochs per patient, on average across patients (730 epochs for before-tDCS, and 670 epochs for after-tDCS). Pooling data across different timelines reduced the influence of session-specific effects on the results, such as day-to-day variability or differences in EEG cap placement [38]. Thus, this way, we promoted the learning of network features more specifically related to the intervention itself, minimizing the influence of confounding factors related to natural inter-session variability. Notably, the number of trainable parameters introduced by our neural network (3906) and the size of each patient-specific dataset (1400 examples, on average across patients) fell within ranges commonly employed in state-of-the-art deep learning methods applied for EEG decoding (generally 1k-100k trainable parameters and 0.5k-2k examples in subject-specific datasets) [39,40]. A 5-fold cross-validation procedure was applied to train and test the network. Within each fold, 20 % of the training examples were used as validation set, and the model with the lowest validation loss was selected for performance evaluation and EEG analysis (see Section 2.2.3). For each patient (9) and each cross-validation fold (5), the confusion matrix, area under the ROC curve (AUC), balanced accuracy, sensitivity and specificity were computed on the held-out test set.

2.2.3. Network decision explanation

The interpretable neural network was combined with an explanation technique [28] to explain the model decision in the frequency domain. Specifically, this technique was used to quantify the relevance of each bandpass filtered version of the input for predicting the after-tDCS condition. DeepLIFT [29] was employed as explanation technique, as it was identified as the most promising method in a recent benchmark study comparing different explanation techniques applied to deep neural networks trained on EEG signals [41]. DeepLIFT is a backward propagation technique that propagates the contribution of a neuron of interest (e.g., the neuron corresponding to the after-tDCS class) back to a target layer (e.g., the input layer or a hidden layer), thereby quantifying the relevance of neurons in the target layer for the neuron under analysis.

For each trained interpretable model (45 in total, across 9 patients and 5 cross-validation folds), we computed the importance of each bandpass filter for predicting the after-tDCS condition. Using DeepLIFT, we extracted the relevance of each feature map from the first convolutional layer (i.e., temporal sinc-convolution) for predicting the after-tDCS condition, by considering all the test set examples belonging to that class. For each example, $K=32$ relevance maps of shape ($C=27$, $T=250$) were obtained (yellow box in Fig. 2), quantifying the importance of each EEG channel (27) and time sample (250) in the filtered input. Since the input was filtered in a specific frequency range defined by the cutoff frequencies (f_1, f_2), we obtained one relevance map for each frequency range (32 ranges overall). The relevance maps were then averaged across EEG channels, time samples and examples, to

obtain the *filter-level relevance scores* $s \in \mathbb{R}^{32}$ (yellow box in Fig. 2). Each score $s_i, 0 \leq i \leq K-1$, quantifies how much a specific frequency range (given by $\theta^{(i)} = \{f_{1,i}, f_{2,i}\}$) contributed to the classification, i.e., to the output probability $p(\text{after} - tDCS|X)$. A positive (negative) score indicates that the frequency components in that range increase (decrease) the probability of the after-tDCS class. For each patient, based on the scores s_i, we derived two deep learning-based measures to identify the frequency components and EEG bands most relevant to the after-tDCS class (i.e., the frequencies and bands that changed more from before to after the intervention).

i. *Frequency-level relevance*. For the i-th bandpass filter – identified by the learned cutoff frequencies $\theta^{(i)} = \{f_{1,i}, f_{2,i}\}$ – the frequencies between $f_{1,i}$ and $f_{2,i}$ were weighted by the associated score s_i, while the frequencies outside this range were assigned a value of zero, that is:

$$r_i[f] = s_i\left(rect\left(\frac{f}{2f_{2,i}}\right) - rect\left(\frac{f}{2f_{1,i}}\right)\right), 0 \leq i \leq K-1, \quad (4)$$

where f represents the frequency values obtained by discretizing the frequency axis from 1 to 15 Hz with 500 points. These representations were averaged across all bandpass filters obtaining the *frequency-level relevance* $\bar{r}[f] \in \mathbb{R}^{500}$ ($\bar{r}[f] = \frac{1}{K}\sum_i r_i[f]$), which quantifies the importance of each frequency component in distinguishing resting-state EEG after tDCS from before.

ii. *EEG band-level relevance*. The frequency-level relevance $\bar{r}[f]$ was further aggregated to reveal the importance of the different EEG bands in predicting the after-tDCS condition. Specifically, $\bar{r}[f]$ was averaged across the frequency bins belonging to delta (1-4 Hz), theta (4-8 Hz), low-alpha (8-11 Hz) and high-alpha (11-15 Hz) bands, producing a *band-level relevance score* ($\in \mathbb{R}$) for each considered EEG band.

Following this procedure, 45 frequency-level relevance patterns and 45 EEG band-level relevance scores (for each of the four EEG bands) were obtained, across the 9 patients and 5 cross-validation folds.

2.3. Analysis of the spectral EEG signatures of DOC intervention with a traditional approach

We performed a traditional analysis based on the scalp-level power spectral density (PSD) and EEG band power, in order to compare the results obtained with the XAI approach to those derived from conventional methods. For each patient (9) and each EEG channel (27), the PSD was estimated using the Welch method (2s-length windows, 25 % overlap). PSD estimates were then averaged across epochs, separately for the epochs collected before and after the tDCS. The percentage variation of the PSD between before- and after-tDCS conditions was computed as $\Delta_{PSD} = 100 \cdot (PSD_{\text{after - tDCS}} - PSD_{\text{before - tDCS}})/PSD_{\text{before - tDCS}}$. The resulting Δ_{PSD} values were then averaged across EEG channels within four groups: parietal (7 channels: parietal and occipital sites), central (13 channels: central, fronto-central, and centro-parietal sites), frontal (7 channels: frontal and fronto-polar sites) and all electrode sites (27 channels). Finally, for each patient (9) and each group of channels (4), we integrated the Δ_{PSD} values within delta, theta, low-alpha, high-alpha band, obtaining the percentage variation of the power (Δ_P) for each EEG band.

2.4. Statistical analysis

The EEG band-level relevance scores (from point ii. of Section 2.2.3) were subjected to the following statistical analyses.

i. Comparison vs. null relevance. EEG band-level relevance scores were aggregated separately across non-improving patients (10 relevance observations per band, from 2 patients across 5 cross-validation folds) and improving patients (35 relevance observations per band, from 7 patients across 5 cross-validation folds). For each EEG band (delta, theta, low-alpha, and high-alpha), the relevance scores were compared against the null relevance value (0) to identify significant contributions, separately for non-improving and improving patients. Pairwise comparisons were performed using two-tailed Wilcoxon signed-rank tests [42] (4 tests in total, separately for non-improving and improving patients). False discovery rate correction at $\alpha = 0.05$ using the Benjamini-Hochberg procedure [43] was employed to correct for multiple tests.

ii. Comparison between EEG bands. To assess differences among the relevance scores of the four EEG bands (delta, theta, low-alpha, and high-alpha), we conducted a Friedman test [44] separately for non-improving and improving patients, using the same patient aggregation strategy described in point i. Significant differences were observed for both non-improving patients ($p = 3.14\text{e-}05$) and improving patients ($p = 0.002$). For each group of patients, post-hoc pairwise comparisons were then conducted using two-tailed Wilcoxon signed-rank tests [42] for all possible EEG band combinations (6 tests in total, separately for non-improving and improving patients). False discovery rate correction at $\alpha = 0.05$ using the Benjamini-Hochberg procedure [43] was employed to correct for multiple tests.

iii. Correlation analysis. We examined whether EEG band-relevance scores were related to the CRS-R improvement by computing the Pearson correlation coefficient between these measures, separately for each EEG band (delta, theta, low-alpha, high-alpha). The CRS-R improvement was defined as the difference between after- and before-tDCS CRS-R scores when positive, and zero otherwise (Table 1). P-values were corrected for multiple comparisons (4 tests in total) via the Benjamini-Hochberg procedure [43].

Finally, the percentage variation of EEG band power (Δ_P) was analyzed at the individual patient level (i.e., using patient-specific resting-state EEG epochs) to determine whether the observed patient-level changes in Δ_P were statistically significant. For each patient, each electrode group (4) and each EEG band (4), a two-tailed permutation test (1000 permutations) [45] was performed on the Δ_P values. In each permutation, EEG epochs from the before- and after-tDCS conditions were pooled, randomly shuffled, and reassigned to two groups with the same number of epochs recorded before and after tDCS. Then, the percentage variation Δ_P between the two groups was computed as specified in Section 2.3. This process yielded a null distribution of Δ_P values, against which the observed variation was tested.

3. Results

3.1. EEG spectral signatures of DOC intervention: artificial intelligence approach

We first report the decoding performance achieved while classifying before- vs. after-tDCS from the resting-state EEG recorded in DOC patients. All performance metrics refer to the held-out test set. Table 3 summarizes the AUC, accuracy, sensitivity, and specificity achieved by the patient-specific models on the test set, separately for each patient. Fig. 3 additionally displays the model accuracy (left panel), and the confusion matrix (right panel). Across patients, neural decoders achieved an AUC of 95.0 ± 4.3 % (mean and standard deviation), an accuracy of 95.6 ± 4.4 %, a sensitivity of 95.5 ± 5.0 %, and a specificity of 94.4 ± 4.8 %. Notably, the high decoding performance of the patient-specific models on unseen examples indicates strong generalization, with no evidence of overfitting.

The high decoding performance of the neural network in

Table 3
Neural decoding performance: AUC, accuracy, sensitivity, and specificity. Performance metrics are reported per patient (mean ± std. across cross-validation folds), and across patients (mean ± std., last row).

Patient ID	AUC	Accuracy	Sensitivity	Specificity
1	0.932 ± 0.050	0.942 ± 0.045	0.909 ± 0.082	0.954 ± 0.053
2	0.969 ± 0.028	0.968 ± 0.029	0.988 ± 0.012	0.950 ± 0.054
3	0.966 ± 0.029	0.966 ± 0.029	0.949 ± 0.042	0.982 ± 0.027
4	0.841 ± 0.133	0.839 ± 0.146	0.834 ± 0.256	0.849 ± 0.215
5	0.996 ± 0.004	0.996 ± 0.004	0.995 ± 0.004	0.996 ± 0.005
6	0.960 ± 0.039	0.960 ± 0.046	0.959 ± 0.074	0.961 ± 0.041
7	0.934 ± 0.099	0.975 ± 0.032	0.990 ± 0.010	0.878 ± 0.191
9	0.987 ± 0.013	0.988 ± 0.013	0.984 ± 0.019	0.991 ± 0.011
10	0.964 ± 0.061	0.969 ± 0.050	0.991 ± 0.008	0.938 ± 0.120
	0.950 ± 0.043	0.956 ± 0.044	0.955 ± 0.050	0.944 ± 0.048

discriminating EEG recorded before- vs. after-tDCS suggests that the resting-state EEG in these two conditions contained distinct characteristics and patterns. Moreover, it emerges that the trained networks have successfully captured EEG features relevant for the discrimination, likely reflecting modulations induced by the tDCS intervention.

Fig. 4 shows representative examples of the bandpass filter banks associated with models trained on data from a non-improving patient (no. 4, Fig. 4a), a highly improving patient (no. 9, Fig. 4b), and a mildly improving patient (no. 6, Fig. 4c), along with the corresponding filter-level relevance scores. The filter banks are visualized using bars, where each bar spans the passband (between the two learned cutoff frequencies), and its thickness and color encode the filter-level relevance score (thickness: magnitude of s_i; color: sign of s_i).

In the highly improving patient, the low frequency components in delta and theta bands contributed negatively to predict after-tDCS resting-state EEG epochs, whereas the higher frequencies in high-alpha band contributed positively. The opposite pattern was observed in the non-improving patient. The mildly improving patient exhibited an intermediate, more dispersed pattern along the frequency axis, falling between those of the non-improving and highly improving patients. These representative examples suggest that the network processes frequency components differently depending on the neurobehavioral improvement of patients (in this case, mediated by tDCS intervention). This distinctive aspect was further analyzed using the deep learning-based spectral measures provided by our XAI approach (frequency-level relevance and EEG band-level relevance).

Fig. 5 shows the frequency-level relevance for all patients as an heatmap (patients by rows, sorted by CRS-R improvement). As the CRS-R improvement increased, the relevance associated to high frequencies progressively shifted from negative values (non-improving patients) to larger positive values (improving patients). In contrast, the relevance associated with low-frequency components exhibited an approximately opposite trend. An exception was patient no. 3, who showed an improvement in CRS-R with larger positive relevance at low frequencies. This might be related to the fact that patient no. 3 was the only patient with no change in the minimum CRS-R score between the two before-tDCS timelines (T-1, T0) and the two after-tDCS timelines (T1, T2), see Table 1. In all the other improving patients, the minimum CRS-R score after tDCS was higher than before tDCS; this could indicate a more robust improvement in these patients compared to patient no. 3.

To better visualize the dynamics of the relevance, Fig. 6 displays the frequency-level relevance for a non-improving patient (no. 4, Fig. 6a), for a highly improving patient (no. 9, Fig. 6b), and for a mildly improving patient (no. 6, Fig. 6c). These are the same patients considered in Fig. 4. Finally, Fig. 6d shows the pattern of the frequency-level relevance aggregated across all patients (9 patients), across non-improving patients (2 patients), and across improving patients (7 patients). On average, in improving patients, the relevance was always positive, peaking in the low-alpha (approximately 8-9 Hz) and high-alpha (approximately 12-13 Hz) bands, and with near-zero relevance for low frequencies (delta and theta bands). In contrast, on average, non-improving patients showed positive relevance only in the delta and theta bands (peaking positively at about 3-4 Hz), followed by negative relevance in the high-alpha band (peaking negatively at about 12 Hz).

The frequency-level relevance was further aggregated and analyzed at the level of EEG bands (delta, theta, low-alpha, high-alpha). Fig. 7 presents the band-level relevance for non-improving patients (left panel) and improving patients (right panel), separately for each EEG band. In agreement with the previous results, in non-improving patients, the relevance values progressively shifted from positive to negative as the band moved from delta to high-alpha. Moreover, the delta, theta, and high-alpha bands had relevance values significantly different from zero

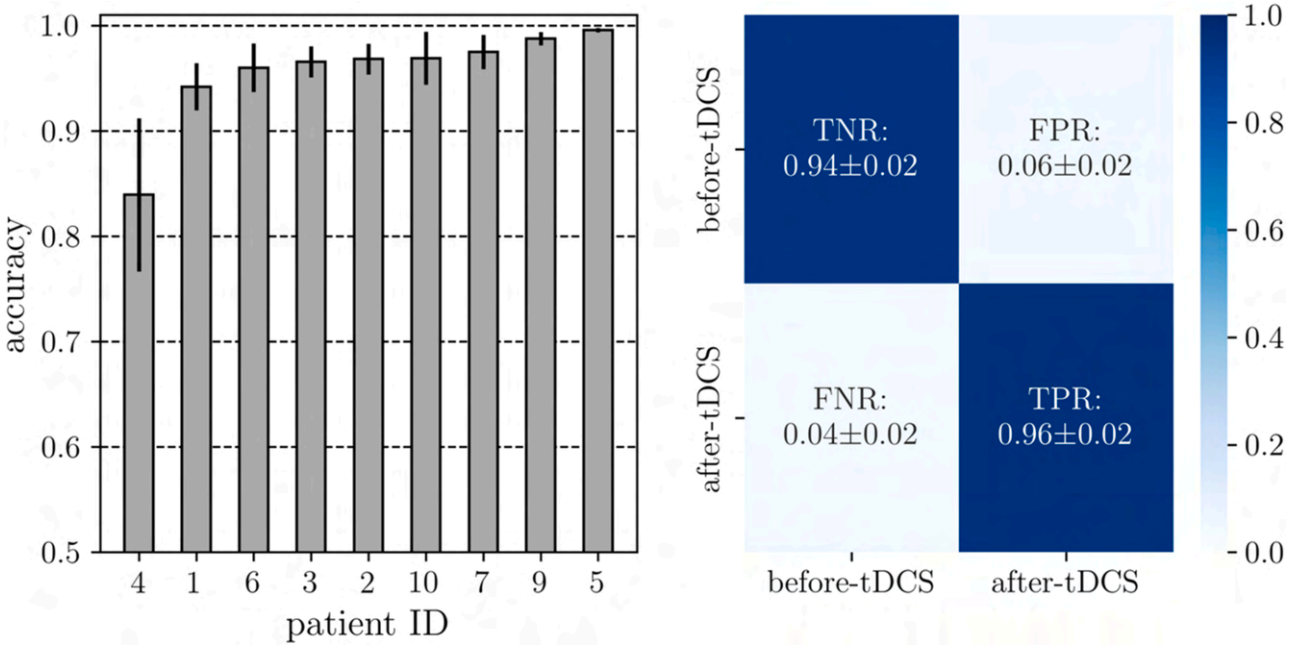

Fig. 3. Neural decoding performance: accuracy distribution and confusion matrix. The left panel shows the accuracy is displayed for each patient (height: mean across folds, error bar: std. across folds). The right panel displays the confusion matrix (normalized), averaged across folds (mean ± std. across the patients).

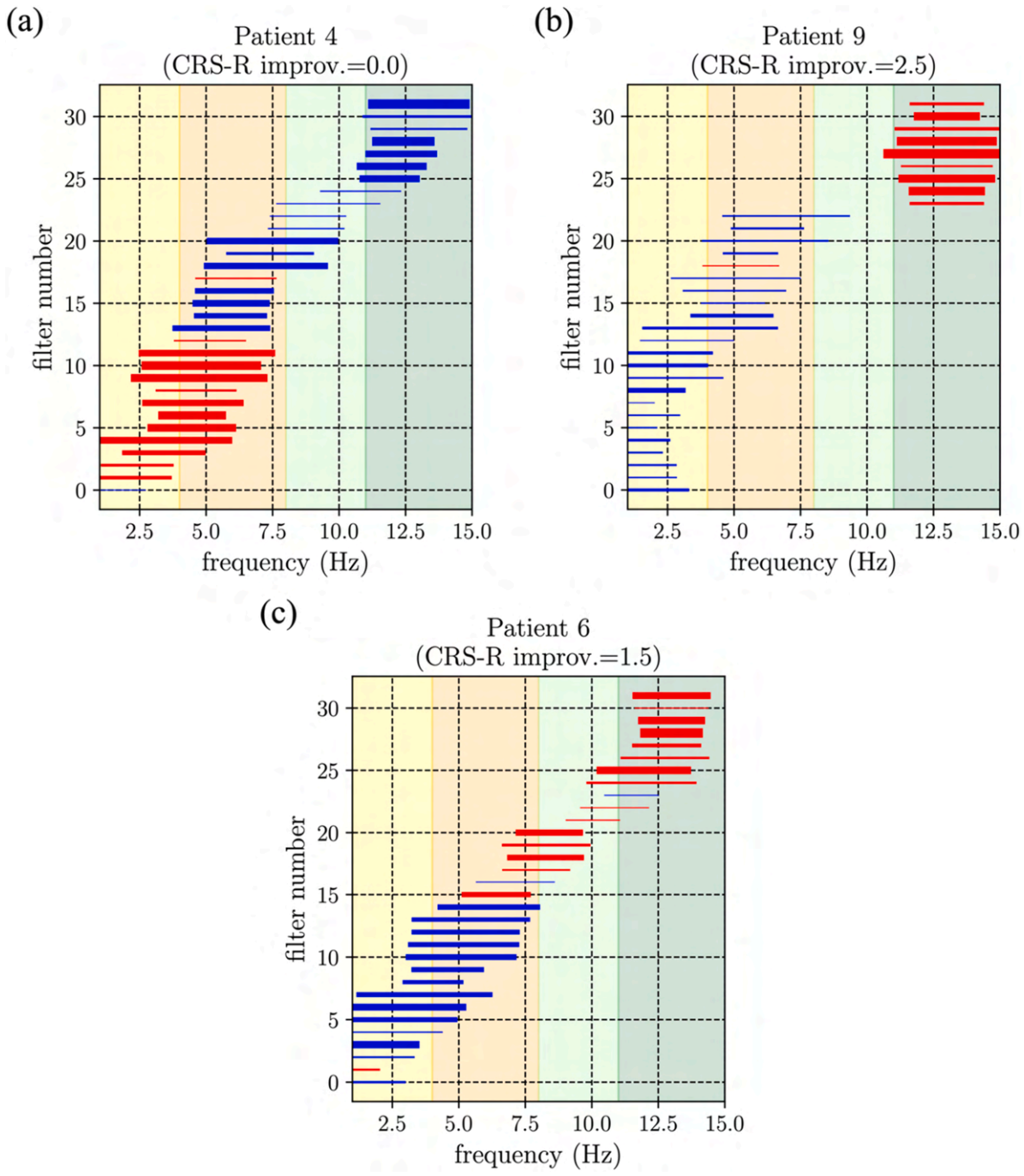

Fig. 4. Example of the learned interpretable spectral parameters. Panels a-c display the spectral parameters of one representative model trained on data from a non-improving patient (no. 4), highly improving patient (no. 9), and mildly improving patient (no. 6), respectively. Each bar represents a bandpass filter, with endpoints corresponding to the learned cutoff frequencies ($\theta^{(i)}, 0 \leq i \leq K-1, K = 32$). Bar thickness encodes the magnitude of the filter-level relevance score (s_i), quantifying the importance of each filter. Bar color indicates the sign of the relevance score ($s_i \geq 0$: blue; $s_i < 0$: red). EEG bands are color-coded as follows: yellow (delta, 1-4 Hz), orange (theta, 4-8 Hz), light green (low-alpha, 8-11 Hz), and dark green (high-alpha, 11-15 Hz).

($p < 0.05$), and all pairwise comparisons between EEG bands were significant ($p < 0.05$), with the delta band being the most positively relevant and the high-alpha band the most negatively relevant. In contrast, in the improving patients, only low- and high-alpha bands showed significantly positive relevance values compared to zero ($p < 0.01$), with the high-alpha band being significantly more relevant ($p < 0.05$) than delta and theta bands, thus emerging as the most relevant band.

Finally, in Fig. 8 the band-level relevance scores are related with the CRS-R improvement. Interestingly, the band-level relevance scores resulted significantly ($p < 0.001$) correlated with the CRS-R improvement only within the high-alpha band, showing a moderate [46] positive correlation ($r = 0.6$).

3.2. EEG spectral signatures of DOC intervention: traditional approach

The PSD across patients is reported in Fig. 9, separately for non-improving (left panels) and improving patients (right panels). Fig. 9a displays the percentage variation of the PSD (Δ_{PSD}), averaged across patients and EEG channels, separately for parietal, central, frontal, and all electrodes. In non-improving patients, the PSD after tDCS increased

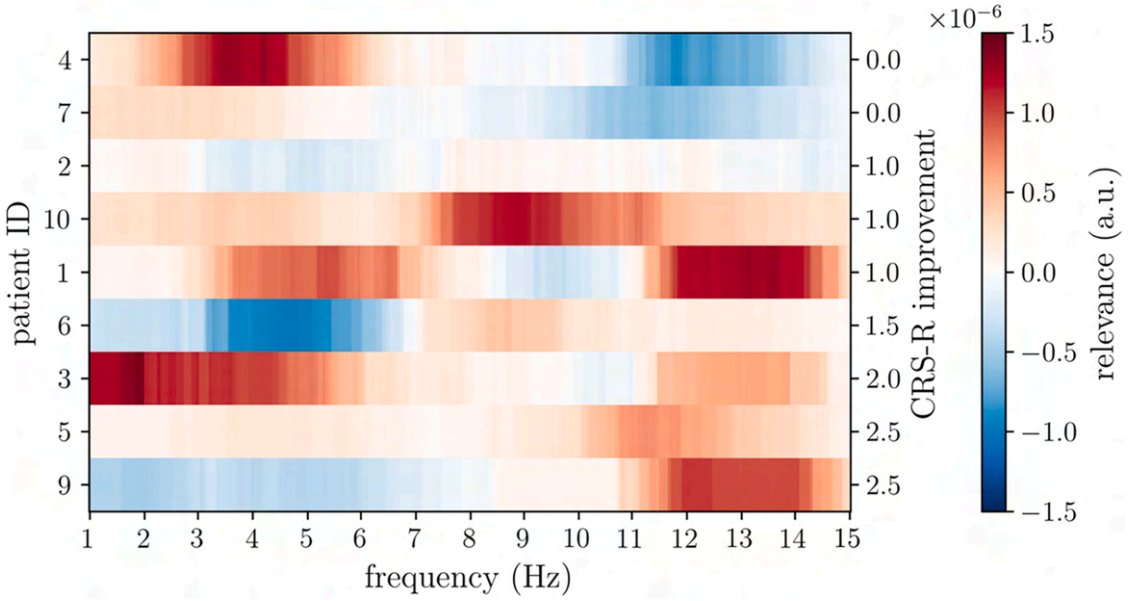

Fig. 5. Frequency-level relevance. The frequency-level relevance is shown as an heatmap, with each row corresponding to a different patient (mean value across cross-validation folds). Patients are sorted (see left y-axis) according to their CRS-R improvement (see right y-axis). Relevance values are expressed in arbitrary units (a.u.).

(up to approximately +150 %/Hz) at low frequencies, from about 1 to 6 Hz, and decreased (down to approximately -25 %/Hz) at higher frequencies, from about 6 to 15 Hz. This pattern was particularly pronounced in the frontal group of channels. Conversely, in improving patients, the PSD after tDCS showed an increase (up to approximately +50 %/Hz) in alpha band, from about 7 to 15 Hz, most evident over parietal channels. To provide a clearer visualization of this phenomenon, Fig. 9b displays the PSD separately for the before-tDCS and after-tDCS conditions, averaged across patients, for the electrode group exhibiting the strongest percentage variation in Fig. 9a, that is, frontal sites in non-improving patients, and parietal sites in improving patients.

Overall, these results are consistent with the findings from XAI (Fig. 6d). Specifically, the latter assigned high positive relevance to low-frequency components (delta- and theta-band frequencies) and high negative relevance to high-frequency components (alpha-band frequencies) for discriminating after-tDCS vs. before-tDCS in non-improving patients. In contrast, in improving patients, the XAI-based approach attributed high positive relevance to alpha-band frequency components. Taken together, the PSD results in Fig. 9 complement and strengthen the physiopathological interpretation of XAI results.

Despite the overall agreement between the traditional and the XAI-based approaches across patients, the advantages of the XAI-based EEG analysis become more evident at the individual patient level, as shown in Figs. 10 and 11 for three representative patients. These include one non-improving patient (no. 4, Figs. 10a and 11a), one highly improving patient (no. 9, Figs. 10b and 11b), and one mildly improving patient (no. 6, Figs. 10c and 11c). Figs. 10 and 11 display, respectively, the percentage variation of the PSD (Δ_{PSD}) and of the EEG band power (Δ_P), averaged across EEG channels for each electrode group (parietal, central, frontal, and all electrodes). In addition, the bottom plots in each panel of Fig. 11 also include the results of the permutation tests conducted on Δ_P at the individual patient level (see Section 2.4), by representing the null distribution and marking the statistically significant results ($p < 0.05$), separately for each EEG band and electrode group. As shown in Fig. 10, patients no. 4 and 9 exhibited Δ_{PSD} patterns resembling the mean pattern of the corresponding group (non-improving and improving patients) reported in Fig. 9. Conversely, patient no. 6 presented a less consistent pattern compared with the improving group, as Δ_{PSD} did not display the expected increase in alpha-band frequencies. The analysis on the variation of EEG band power (Fig. 11) confirmed a significant increase in delta and theta power and decrease in high-alpha power for the non-improving patient no. 4, and an opposite significant pattern for the highly improving patient no. 9, evident across most electrode groups. For the mildly improving patient no. 6, the analysis revealed an overall significant reduction of band power variation in delta and theta bands, and a general decreasing trend in both low- and high-alpha bands, across most electrode groups. This finding, obtained using the traditional approach, contrasts with the expected increase of alpha-band power typically associated with recovery in DOC patients.

4. Discussion

In this study, we present the design and application of an XAI-based approach aimed at uncovering the most salient frequency-domain EEG signatures that change during therapeutic intervention in DOC patients. To the best of the authors' knowledge, this is the first attempt of exploiting the feature learning performed by an XAI framework to analyze the spectral EEG correlates in DOC and their modulation by therapeutic intervention. The proposed approach is fully *automatic* (i.e., it does not discard potentially relevant information a priori), *patient-specific* (i.e., analyses are individualized for each DOC patient) and *intervention-specific* (i.e., tailored to DOC and DOC intervention). To showcase its potential, we applied our approach to real data from a pilot EEG-tDCS study involving DOC patients diagnosed with MCS, who underwent a tDCS treatment [9].

4.1. Spectral signatures of DOC intervention revealed via artificial intelligence

The deep learning-based measures derived from our XAI approach are based on the knowledge captured by a learning system (interpretable neural network), trained to discriminate resting-state EEG epochs recorded before vs. after the intervention, separately for each patient. The measures are obtained by coupling the learning system with an explanation technique. Here, we used Sinc-ShallowNet as learning system and DeepLIFT as explanation technique. Notably, the pipeline of our XAI approach is not specific of Sinc-ShallowNet, but rather holds for any interpretable neural network designed to enhance the interpretability in

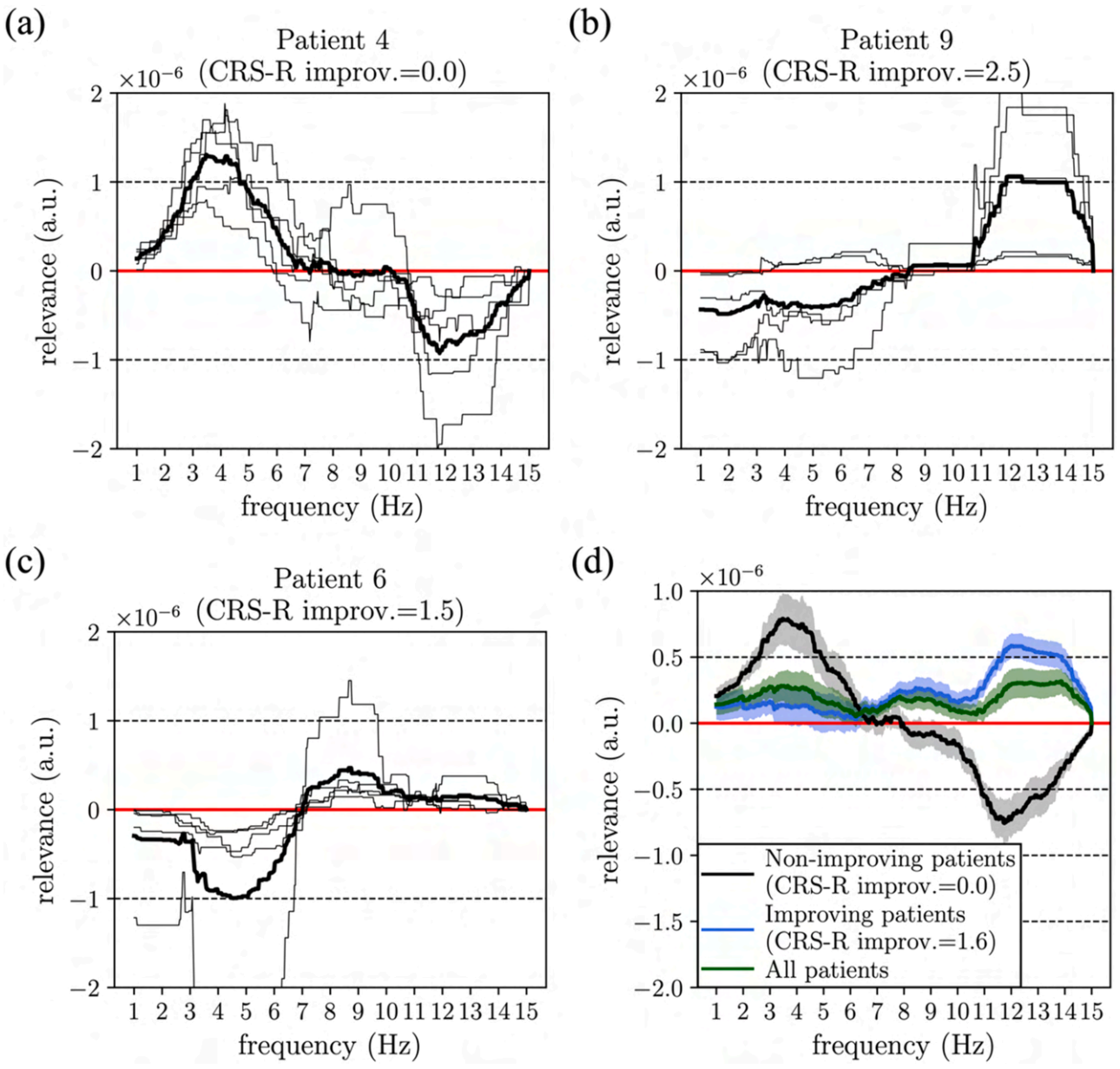

Fig. 6. Frequency-level relevance. Panels a-c – Representative examples of patient-specific relevance patterns. The frequency-level relevance is shown separately for one non-improving patient (no. 4, panel a), one highly improving patient (no. 9, panel b), and one mildly improving patient (no. 6, panel c). The relevance pattern is displayed for each cross-validation fold (thin black line) and as the average across folds (thick black line). Note that, to maintain consistent y-axis limits across panels, one curve in panel c was truncated (it peaks at approximately -3.5e-06). Panel d – Relevance patterns aggregated across patients, separately for non-improving patients (black, 2 patients, 10 relevance patterns), improving patients (blue, 7 patients, 35 relevance patterns), and all patients (green, 9 patients, 45 relevance patterns). Lines denote the mean value across patients and folds, while the shaded areas denote the standard error of the mean. In all panels, the red line represents the null relevance. Relevance values are expressed in arbitrary units (a.u.).

the frequency domain. For example, it could also be applied to the architectures proposed by Zhao et al. [22] (WaSFCNN) and by Ludwig et al. [21] (EEGminer), which employ a reparameterization of the first temporal convolutional layer similarly to Sinc-ShallowNet, enforcing the learning of band-pass filters with a different analytical formulation of the filter frequency response. Likewise, our XAI pipeline can incorporate any explanation technique capable of providing relevance attributions with respect to a hidden layer of the neural network (see Section 2.2.3), such as saliency maps or layer-wise relevance propagation [28].

The deep learning-based measures provided by our approach are the frequency-level relevance and band-level relevance. The reliability of these measures is reflected by the high decoding performance achieved by the neural network (Fig. 3 and Table 3). The network achieved an average accuracy of 95.6 % across patients, with accuracies of patient-specific decoders ranging from 83.9 % to 99.6 %, well-above the chance level for binary classification (50 %). The frequency-level relevance/band-level relevance quantifies how much each frequency component/EEG band contributes to the correct prediction of the resting-state EEG epochs acquired after tDCS intervention. Specifically, they capture how each specific frequency component/band affects the model's estimated probability of the after-tDCS condition, either positively (increasing $p(\text{after} - \text{tDCS}|X)$) or negatively (decreasing $p(\text{after} - \text{tDCS}|X)$). The frequency-level relevance patterns derived from the XAI approach (Fig. 6d) reveal distinctive EEG spectral differences depending

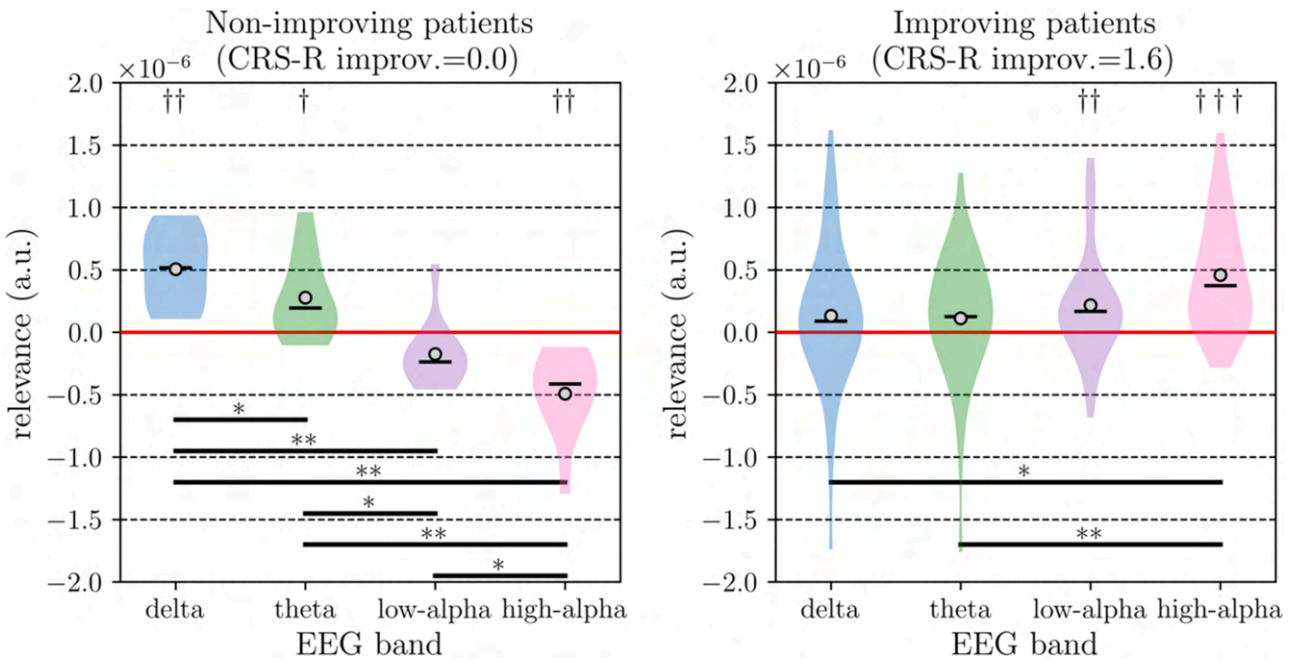

Fig. 7. EEG band-level relevance across patients. The band-level relevance is reported for non-improving patients (left panel, 2 patients, 10 relevance values) and improving patients (right panel, 7 patients, 35 relevance values), separately for each EEG band (delta, theta, low-alpha, high-alpha). Distributions are displayed as violin plots (horizontal black line: median; grey dot: mean). The red line represents the null relevance. Relevance values are expressed in arbitrary units (a.u.). The results from the performed statistical analysis are reported too, by marking the significant pairwise comparisons of band-level relevance between EEG bands (*$p < 0.05$, **$p < 0.01$) and against the null relevance (†$p < 0.05$, ††$p < 0.01$, †††$p < 0.001$).

on the neurobehavioral improvement, as measured by the CRS-R. Low frequencies (within delta and theta bands) are nearly irrelevant (near-zero relevance) in improving patients, but are positively relevant in non-improving patients. This result suggests that, in non-improving DOC patients, the after-tDCS condition relative to the before-tDCS condition was primarily associated with changes in low frequency EEG activity. By design of the XAI approach (see Section 2.2), this pattern likely reflects an increase in amplitude of low-frequency oscillations after the intervention, traduced into an increase of $p(\text{after} - \text{tDCS}|X)$ at low frequencies (positive relevance values). In addition to low frequency components, marked differences also emerged at higher frequencies. In improving patients, alpha-band components are positively relevant, whereas in non-improving patients they are negatively relevant. This suggests that the after-tDCS condition relative to the before-tDCS condition is associated with opposite alpha-band EEG changes in the two groups: an amplitude increase of alpha-band oscillations in improving patients, and a decrease in non-improving patients. These differences are even clearer in the EEG band-level relevance (Fig. 7). In non-improving patients, delta and theta bands show a significant positive relevance, and the high-alpha band shows a significant negative relevance. Conversely, in improving patients, only the low- and high-alpha bands have significant and positive relevance.

Overall, the XAI results indicate that the spectral characteristics of resting-state EEG change differently after the intervention relative to before, depending on the degree of neurobehavioral improvement. Specifically, non-improving patients showed a shift from alpha-band to low-frequency components, consistent with continued loss of consciousness and a progressive slowing of EEG oscillations over time [4]. In contrast, improving patients exhibited a shift toward alpha-band components, symptomatic of consciousness recovery (i.e., transition from delta/theta dominance to alpha dominance) [20]. Among all observed spectral changes, only the high-alpha band relevance showed a significant positive correlation with the CRS-R score, measuring improvement in awareness. In other words, the greater the relevance assigned by the XAI approach to high-alpha frequencies, shifting from negative to positive values, the larger the behavioral improvement observed in the patient. Interestingly, this finding aligns with our previous study on the same dataset [9], where we reported a significant positive correlation between the changes in the high-alpha relative power after tDCS and the CRS-R improvement. However, that prior analysis [9] was limited to only six EEG channels (F3, F4, C3, C4, P3, P4) grouped into parietal, frontal, and central sites (manually handcrafted measure), while the present XAI-based analysis considers the whole-scalp EEG data. Moreover, unlike our previous study [9], the present analysis automatically derives interpretable spectral measures that are tailored to DOC and to the therapeutic intervention by design. Therefore, the current results strengthen and generalize our earlier findings, providing a more robust and data-driven interpretation of the EEG modulations induced in DOC patients by therapeutic intervention.

Comparing the results of our XAI-based approach with those of a traditional approach based on the PSD and EEG band power, we observed a general match across patients (Figs. 6, 9, and 10). However, notable differences between the approaches emerge at the individual patient level. Considering patient no. 6 as a representative example, in the traditional approach, the PSD showed an overall significant reduction in delta and theta bands, also observed in low- and high-alpha bands across most of the electrode groups (Fig. 10c). In contrast, our XAI-based approach (Fig. 6c) revealed not only a negative relevance for low frequencies (from about 3 to 7 Hz) but also a consistently positive relevance for alpha-band frequencies (from about 7 to 15 Hz). These findings reflect the inherent ability of the XAI-based approach to automatically learn the most meaningful patient-specific features that change between before- and after-tDCS conditions, by optimally weighting EEG channels over the scalp. To further investigate this, we extracted the spatial filters (see point i. in Sect. 2.2.1) relative to alpha-band frequency components (alpha-band spatial filters). This procedure is detailed in Appendix A, and the results are presented in Appendix Figs. A1 and A2. Alpha-band spatial filters quantify the importance of each EEG electrode in the alpha

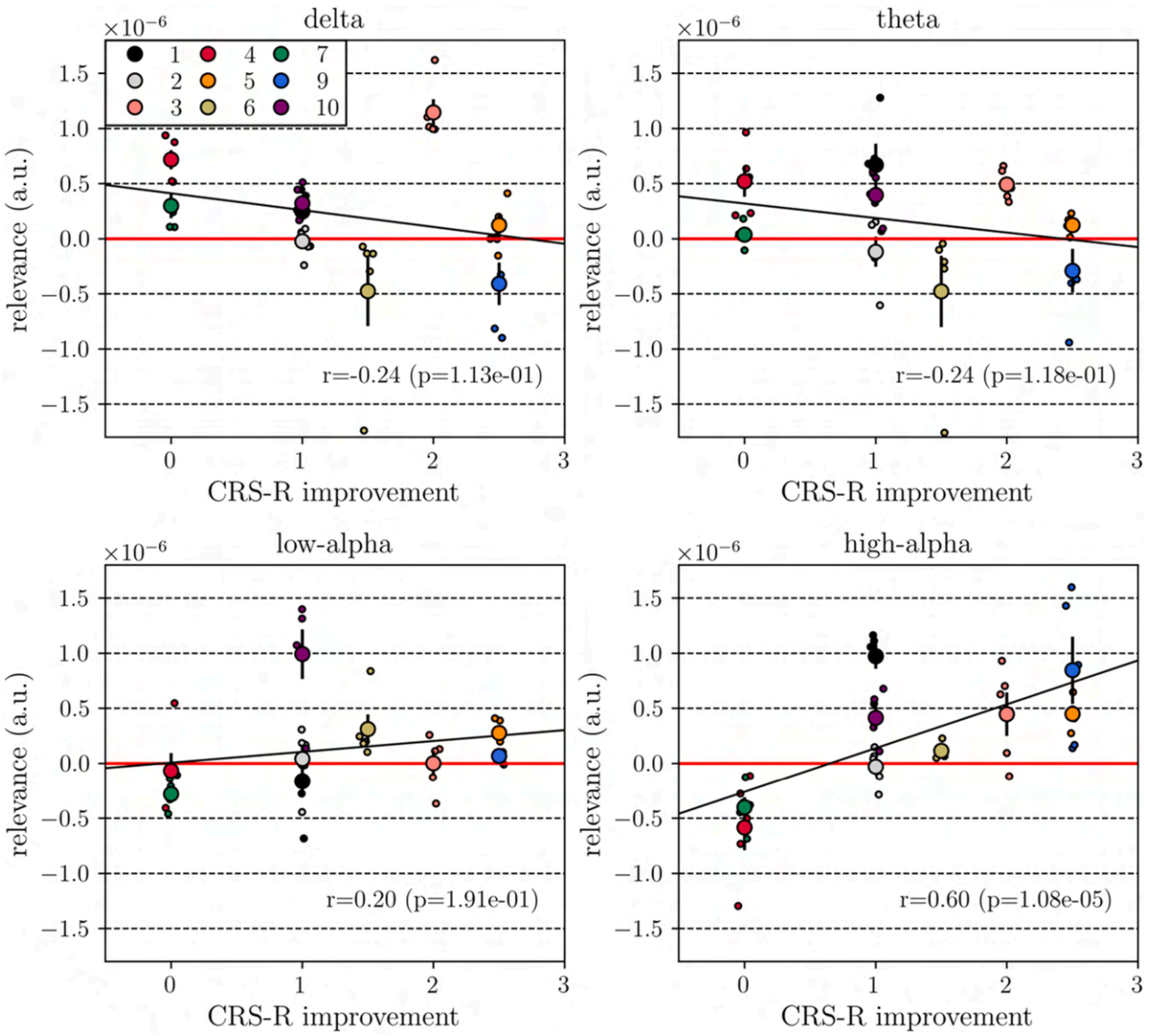

Fig. 8. Relation between EEG band-level relevance and improvements of consciousness (measured via CRS-R clinical scores) following intervention. Distributions of band-level relevance values are shown separately for each EEG band (delta, theta, low-alpha, high-alpha) and each level of CRS-R improvement. For each distribution, smaller dots represent single relevance observations obtained for a specific patient in each cross-validation fold, whereas the larger dot denotes the mean value across folds within each patient. Error bars represent the standard error of the mean across folds. Distributions are color-coded based on the patient ID. The red line represents the null relevance. Relevance values are expressed in arbitrary units (a.u.). Results from the correlation analysis are reported too, showing regression lines (black) along with the Pearson correlation coefficient (*r*) and the p-value for each EEG band.

band for discriminating before- vs. after-tDCS. Considering the alpha-band spatial filters of patient no. 6 (Fig. A2c), a small subset of channels was identified as most relevant, primarily in the centro-parietal (e.g., CP2) and parieto-occipital regions (e.g., O1 and O2). Interestingly, the application of the traditional analysis exclusively to these XAI-selected channels revealed a PSD increase in alpha band after the tDCS intervention (Fig. A3). Moreover, in the selected channels, the alpha-band power shows a significant increase in the low-alpha range after the tDCS intervention, rather than a decrease (as obtained in Fig. 10c across groups of channels). Therefore, by guiding the traditional analysis using insights from the XAI-based approach, the observed spectral modulation of patient no. 6 now aligns with the known increase of alpha band power in DOC recovery [9,47,48]. These results suggest that the neural network underlying our XAI framework can automatically identify electrode clusters that exhibit the most significant alpha-band changes. The main advantage of this approach is its ability to automatically extract spatio-spectral features that differentiate classes (before-tDCS vs. after-tDCS), without requiring manual intervention in defining the features and without introducing biases in the analysis due to expectation of results (e.g., averaged analyses on group of channels).

Overall, our findings underscore the central role of alpha-band oscillations in DOC recovery, corroborating evidence that successful interventions are associated with an increase of alpha-band spectral content [9,47,48]. Remarkably, this pattern was observed both for intervention-mediated recovery (as discussed in this section) and for spontaneous recovery (i.e., not mediated by intervention, see Appendix B). Our results also substantiate the significance of the alpha-band rhythm for awareness [49–51]. Indeed, the thalamus and thalamo-cortical interactions contribute to the origin of the alpha rhythm, and are involved in the generation and maintenance of awareness [52–54]. However, these interactions are impaired in DOC patients [55], for example due to the disruption of structural and

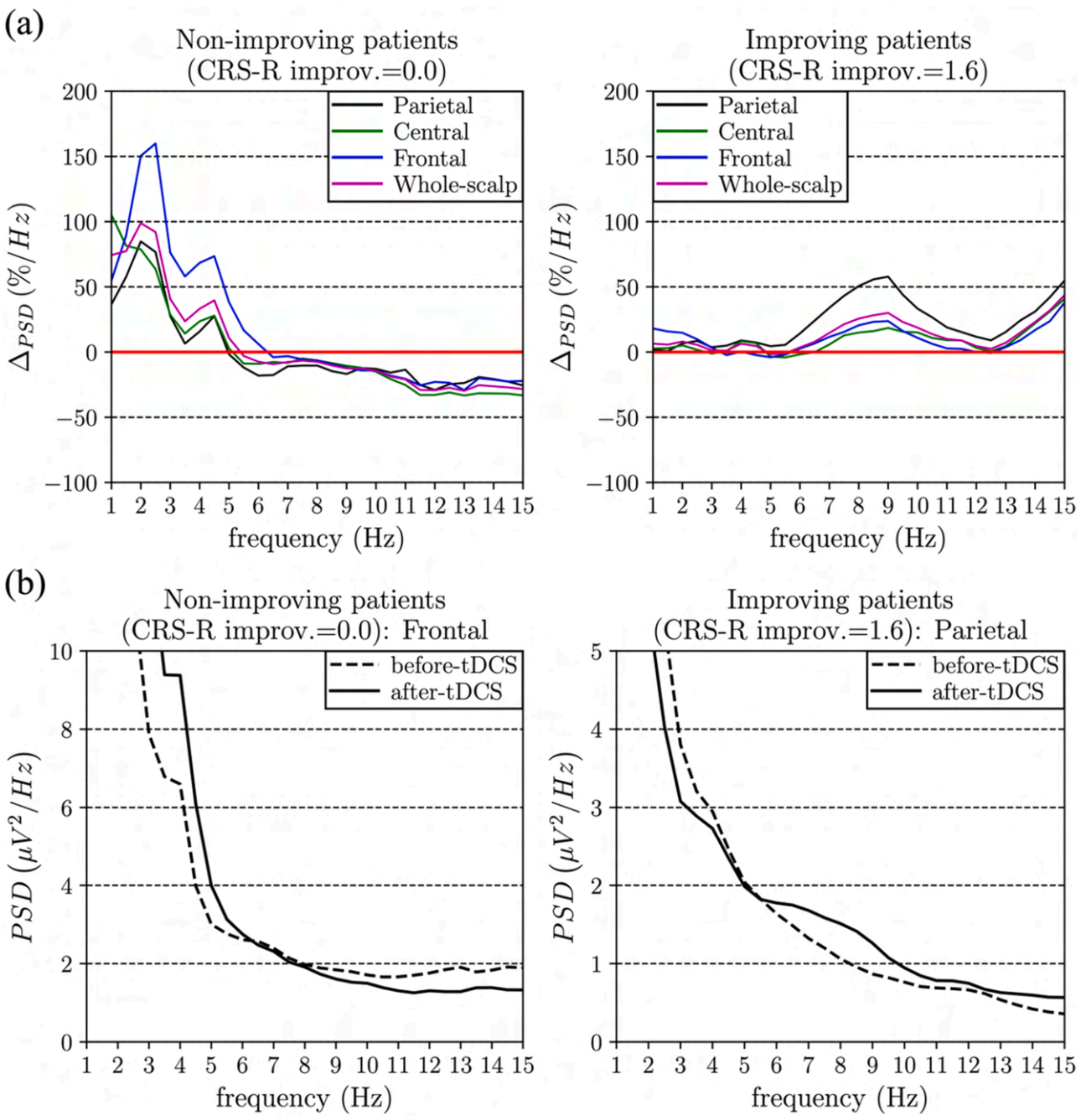

Fig. 9. Power spectral density (PSD) across patients. Panel a – Percentage variation of the PSD (Δ_{PSD}) between before- and after-tDCS conditions. The mean value of Δ_{PSD} across patients and electrodes (parietal, central, frontal, and all electrodes) is reported, separately for non-improving patients (left panel, 2 patients) and improving patients (right panel, 7 patients). The red line represents the null variation. Panel b – PSD. The PSD computed before tDCS (dashed black lines) and after tDCS (black lines) is displayed. The PSD is averaged across patients and across the frontal electrodes (non-improving patients) or parietal electrodes (improving patients).

functional connectivity between the thalamus and the cortex. The bilateral application of tDCS over the primary motor cortex, as applied in our pilot study [9], is known to modulate the functional connectivity of thalamo-cortical circuits [56]. Therefore, the brain stimulation could have transiently restored these impaired interactions in patients. In our XAI-based EEG analysis, this effect might have been reflected in the high relevance attributed to the high-alpha band, which was the only EEG band showing a correlation with behavioral signs of awareness.

4.2. Impact

The central novelty of this study lies in the development and application of an XAI algorithm for analyzing the EEG spectral changes in patients with DOC following an intervention. Traditional analyses of DOC neural correlates and intervention-induced neural changes typically rely on manually handcrafted measures, often discarding potentially relevant information a priori; indeed, handcrafted measures often select a limited subset of EEG channels or consider the average results at the level of group of channels, discarding potential differences across

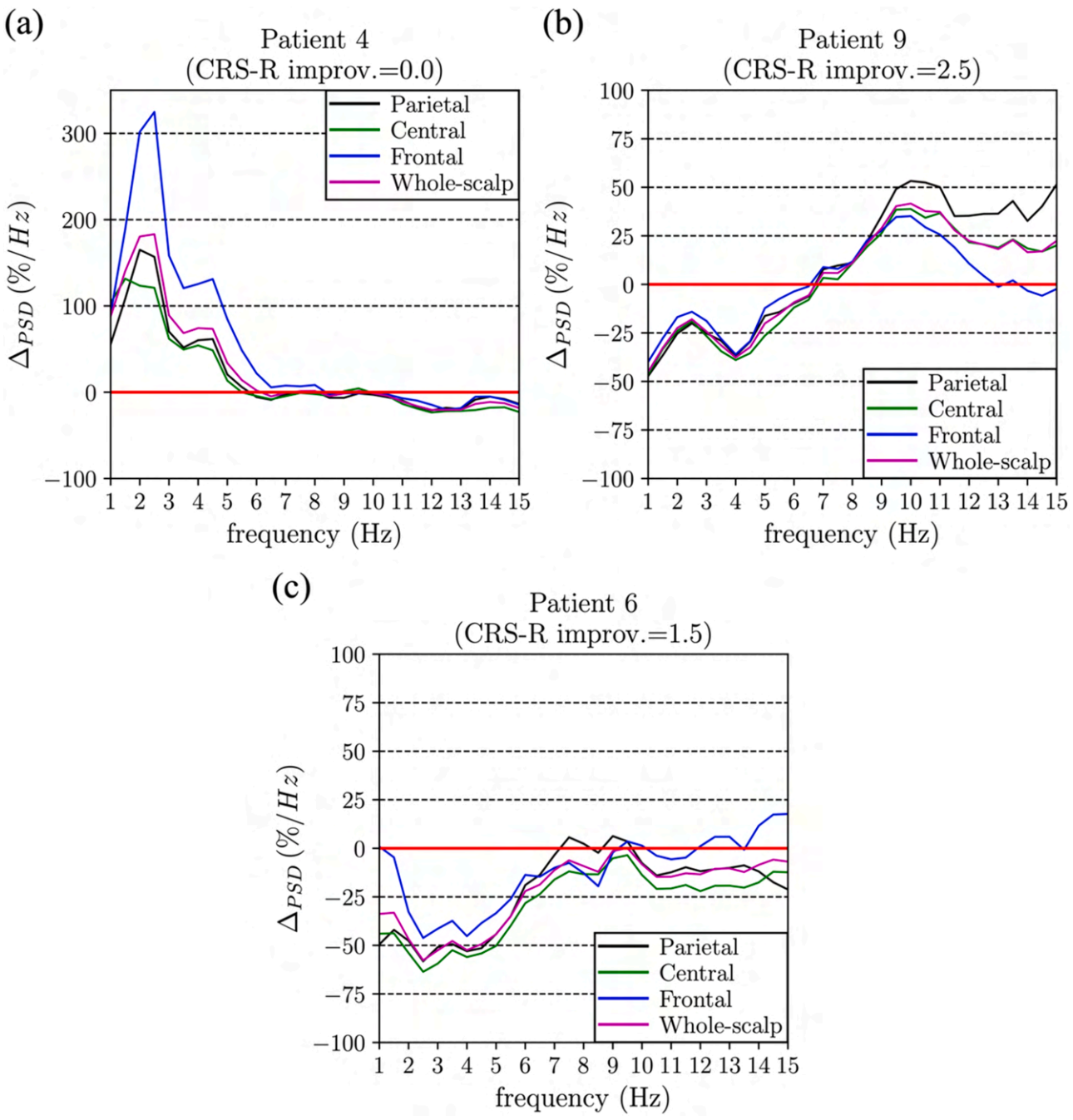

Fig. 10. Power spectral density (PSD) of representative patients. The percentage variation of the PSD (Δ_{PSD}) between before- and after-tDCS conditions is displayed. Panels a-c report the mean value of Δ_{PSD} across electrodes (parietal, central, frontal, and all electrodes) for one non-improving patient (no. 4), one highly improving patient (no. 9), and one mildly improving patient (no. 6), respectively.

single channels. Moreover, these measures are defined in a patient- and intervention-aspecific manner (e.g., by computing spectrum power measures in the same way across patients), without optimally tailoring them to individual neural signatures or to the intervention. In contrast, our XAI approach automatically derives frequency-domain EEG measures directly from the full input EEG, without prior assumptions about the spatial sites to analyze. This allows the framework to optimally identify EEG measures specific to each DOC patient and to track the possible effects of the intervention on an individual basis. Indeed, the measures produced by the XAI algorithm rely on the feature learning of an interpretable neural network, which automatically learns the optimal spectral features to discriminate post-intervention relative to pre-intervention EEG, based on each patient's resting-state EEG. These key properties make the XAI algorithm suitable to:

i. Track patient-sensitive changes in the frequency-domain, during DOC intervention. This could enable the development of novel and more precise frequency-domain EEG biomarkers to quantify and monitor the disorder of consciousness during therapeutic interventions, also facilitating the design of personalized intervention protocols.

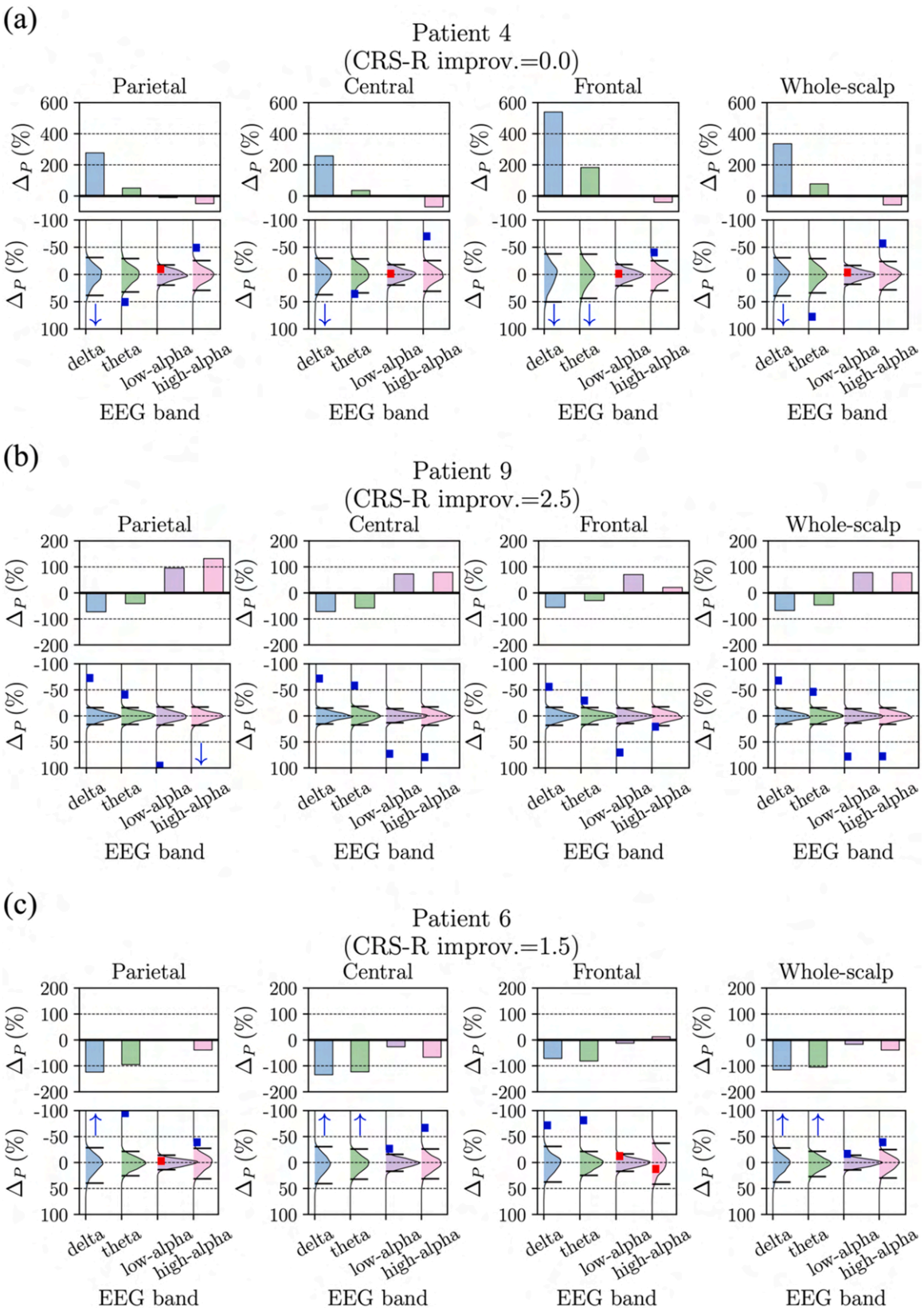

Fig. 11. EEG band power of representative patients. The variation of the EEG band power (Δ_P) between before- and after-tDCS conditions is reported separately for the considered EEG bands (delta, theta, low-alpha, and high-alpha) and groups of channels, by averaging the values across parietal, central, frontal, and all electrodes. Panels a-c display Δ_P for one non-improving patient (no. 4), one highly improving patient (no. 9), and one mildly improving patient (no. 6), respectively. Top panels display the variations using bar plots. Bottom panels display the results of the permutation tests performed on Δ_P, by displaying the null distribution, the observed Δ_P value (colored square, or upward/downward pointing arrow if outside the displayed limits), and the 2.5th and 97.5th percentiles of the null distribution. The color of the marker representing the observed value (square or arrow) indicates statistical significance: blue for significant results ($p < 0.05$), and red for non-significant results.

ii. Enable a more direct connection between multivariate neural recordings (e.g., EEG, as in this study) and key clinical variables related to DOC (e.g., scores of clinical scales assessed at different timelines during the intervention, as in this study). This way, our approach sharpens the EEG analysis on the neural features mostly linked to these clinical variables, by leveraging the capability of the XAI algorithm to automatically process the time series. This avoids relying on predefined processing choices that could bias or limit the results. These aspects may facilitate the data-driven identification of novel neural signatures associated with DOC, potentially revealing patterns that would remain undetected using conventional and hypothesis-driven approaches.

iii. Optimize intervention protocols based on non-invasive brain stimulation (e.g., tDCS in our study). The tDCS is among the most promising non-invasive brain stimulation techniques for DOC treatment [3]. However, its efficacy varies across patients, with improvements in consciousness often partial and/or transient [3]. This underscores the need to optimize stimulation parameters – such as target brain regions, stimulation duration, and intensity – to maximize the extent of functional recovery, the duration of recovery, and the proportion of patients showing neurobehavioral improvements. By integrating neuroimaging or neuroelectric techniques (e.g., EEG) with deep learning-based analytical approaches, the characterization of the neural effects of brain stimulation can be enhanced. Such integration may support and facilitate the refinement of stimulation protocols.

Finally, a secondary novelty of this study is the exploration of XAI-based EEG analysis in clinical population. While such analyses have been largely applied to healthy participants, the application of XAI methods to DOC patients offers a promising extension to clinical contexts.

4.3. Limitations and future directions

In this study, we presented an EEG analysis framework based on XAI, designed to automatically investigate the neural effects of DOC intervention. This approach exploits all the available information from neural recordings (e.g., all spatial sites), and operates in a patient- and intervention-specific manner. Given the novelty of this methodology, the primary goal of this study was to illustrate the analysis framework and provide an initial validation using real EEG data recorded from MCS patients during a pilot study on tDCS-based intervention. Although preliminary, the results are promising and motivate future studies to further enrich the validation on larger datasets. This work represents an initial test of the approach on a small-scale sample, specifically to assess its feasibility as a tool for exploring the patient-specific EEG correlates of intervention and recovery in MCS. While the sample size is limited, it is consistent with previous studies applying computational methods to resting-state EEG in MCS patients (e.g., 6, 7, 9, or 11 MCS patients) [57]. Additionally, the average recording duration per patient (46.6 minutes) also aligns with previous EEG deep learning studies (e.g., 7.1-33 minutes per subject in Lawhern et al. [39]). In the future, we will expand the validation of the XAI approach on other datasets, including patients with different etiologies and types of brain injury, and under different DOC intervention strategies. Moreover, since also spectral functional connectivity measures are often used together with spectrum power measures to monitor DOC changes following intervention [8,58], the framework will be extended to also identify the most salient changes in the spectral directional connectivity between EEG sensors. This enhancement will provide a more comprehensive tool for tracking the neural changes associated with DOC intervention.

5. Conclusions

In this study, we illustrated for the first time how an XAI framework can be applied to analyze the spectral EEG correlates of DOC patients and DOC intervention. We showcased the potential of this approach on MCS patients undergoing a tDCS-based intervention. Our results highlight the importance of alpha-band EEG oscillations in DOC intervention, revealing patient-specific changes in alpha activity associated with functional improvement. This indicates that deep learning-based measures can be used to track the functional DOC recovery on an individualized basis. The proposed analysis pipeline automatically identifies the frequency-domain EEG signatures of DOC that change most after the intervention, overcoming the limitations of traditional EEG analyses (measures manually handcrafted, patient-aspecific and intervention-aspecific). This framework could complement or replace conventional analyses, providing a more robust and transparent characterization of the EEG signatures in DOC patients, advancing our comprehension of neural modulations, and optimally tracking the intervention-induced neural modifications.

Ethics approval

The authors declare that the data used in this work was collected in a prior study, conducted according to the guidelines of the Declaration of Helsinki, approved on 27 September 2012 by the Institutional Review Board (or Ethics Committee) of Ferrara University Hospital (number 101-2012), and thus registered on the Clinicaltrial.gov database (ID protocol: NCT02288533).

Funding

This work is supported by #NEXTGENERATIONEU (NGEU) and funded by the Ministry of University and Research (MUR), National Recovery and Resilience Plan (NRRP), project MNESYS (PE0000006) — A Multiscale integrated approach to the study of the nervous system in health and disease (DN. 1553 11.10.2022).

This research was co-funded by the Italian Complementary National Plan PNC-I.1 “Research initiatives for innovative technologies and pathways in the health and welfare sector” D.D. 931 of 06/06/2022, “DARE - DigitAl lifelong pRevEntion” initiative, code PNC0000002, CUP: B53C22006450001.

Data availability

The data that has been used is confidential.

CRediT authorship contribution statement

Davide Borra: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Valentina Bonsangue:** Data curation. **Sofia Straudi:** Resources, Data curation. **Elisa Magosso:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Appendix A. Alpha-band spatial importance attributed by the artificial intelligence approach

The interpretable neural network adopted in this study (Sinc-ShallowNet) incorporates an interpretable spectral and spatial feature extractor (see block no. 1 description in Section 2.2.1). A set of two spatial filters is learned for each of the 32 band-pass filtered versions of the input. Each spatial filter optimally combines the EEG channels to provide the correct network output. After training, the network identifies the clusters of electrodes most useful for discriminating between before- and after-tDCS conditions. To quantify the contribution of each EEG electrode in processing alpha-band frequency components, we derived the alpha-band spatial filters as follows. For each trained model (45 in total, across 9 patients and 5 cross-validation folds), we extracted the 32 learned spatial filters and we computed their absolute value to assign a discriminatory score to each electrode, following prior studies [26,59]. The higher the score for one electrode the higher the importance of that electrode for solving the classification problem. Then, we selected the spatial filters linked to the band-pass filters containing alpha-band frequency components in their passband, and we averaged them together. The resulting representation was then averaged across cross-validation folds and normalized between 0 and 1, producing a patient-level representation of the EEG channels contributing most to alpha-band information processing (*alpha-band spatial filter*, one per patient). Spatial filters in alpha band were chosen because alpha-band EEG oscillations play a key role in DOC intervention [9,47,48], a finding also supported by our XAI results.

Fig. A1 shows the alpha-band spatial filters aggregated across patients, by averaging the filters separately for non-improving and improving patients. Fig. A2 presents the alpha-band spatial filters at the single patient level for one non-improving patient (no. 4), one highly improving patient (no. 9), and one mildly improving patient (no. 6). Finally, Fig. A3 reports the results from a traditional analysis (PSD and EEG band power) guided by the XAI results about the alpha-band spatial filters. In this case, we report the results for a representative patient (no. 6). From the alpha-band spatial filters reported in Fig. A2c, a small set of channels resulted the most relevant, mainly in the centro-parietal (e.g., CP2) and parieto-occipital regions (e.g., O1 and O2). Using the CP2 electrode, the traditional analysis was performed, computing the percentage variation of the PSD (Δ_{PSD}) and the variation of the EEG band power (Δ_P). The same permutation tests used in the main analyses on Δ_P were replicated here.

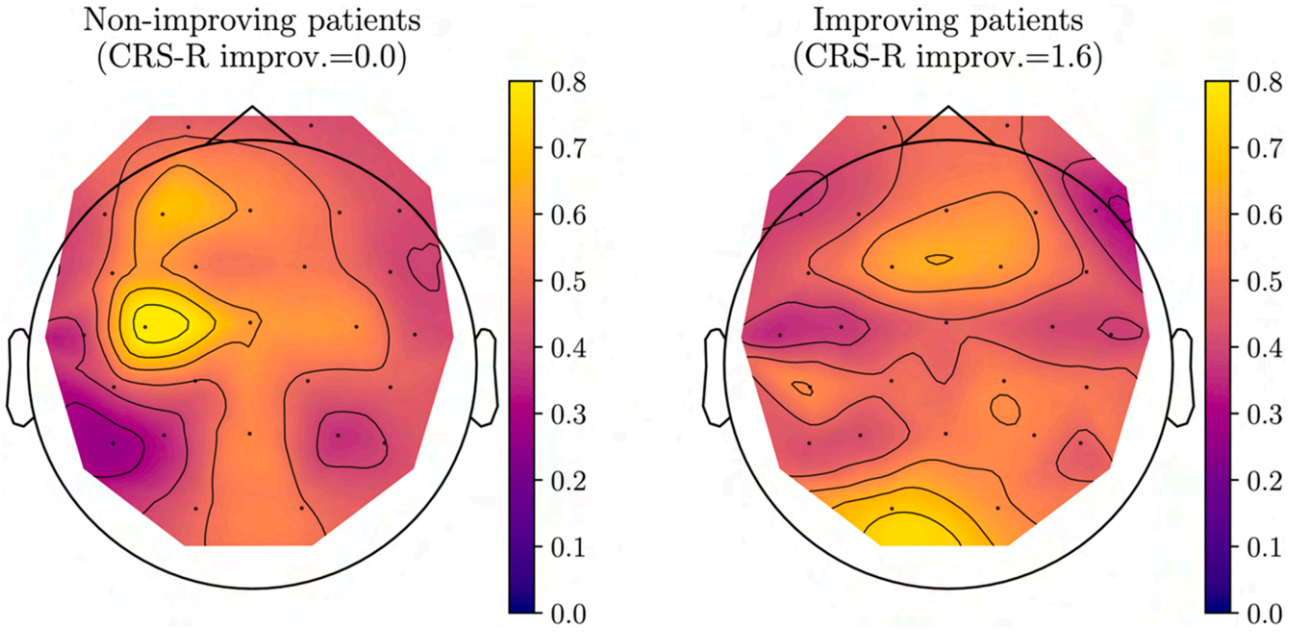

Fig. A1. Alpha-band spatial filters across patients. The spatial filter used for processing alpha-band frequencies (alpha-band spatial filter) is reported, averaged across patients. Results are shown separately for non-improving patients (left panel, 2 patients, spatial filters from 10 models) and improving patients (right panel, 7 patients, spatial filters from 35 models).

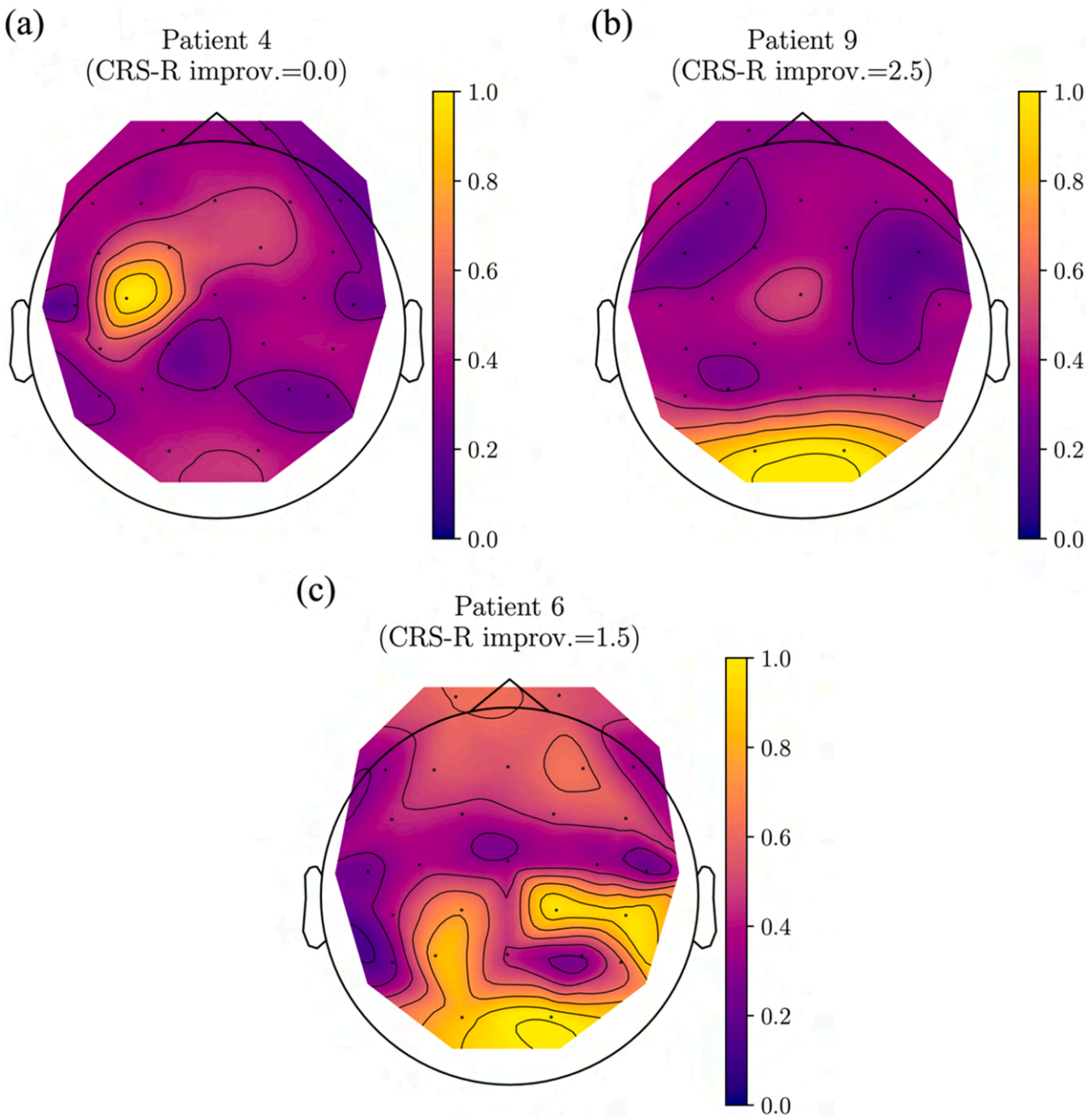

Fig. A2. Alpha-band spatial filters of representative patients. The spatial filter used for processing alpha-band frequencies (alpha-band spatial filter) is reported for one non-improving patient (no. 4), one highly improving patient (no. 9), and one mildly improving patient (no. 6), respectively in panels a-c.

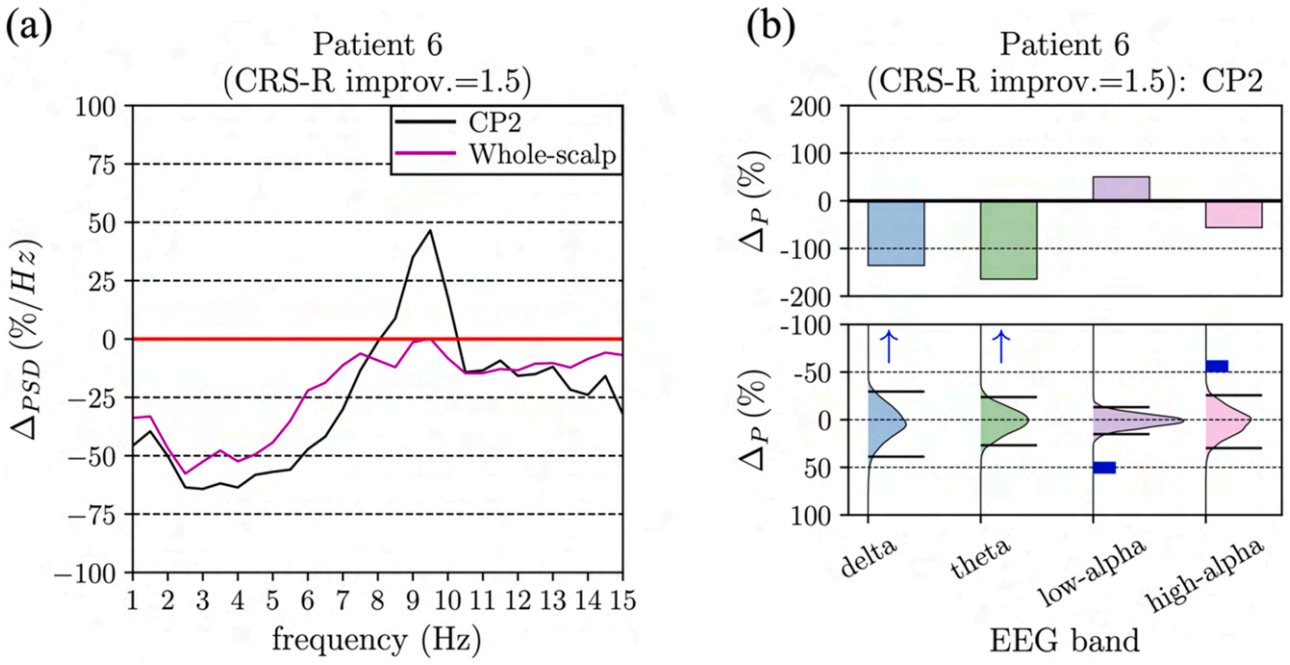

Fig. A3. Analysis of the power spectral density (PSD) and EEG band power guided by the XAI-based approach: patient no. 6. Panel a – PSD. The percentage variation of the PSD (Δ_{PSD}) between before- and after-tDCS conditions is displayed. The value of the Δ_{PSD} is taken from CP2. The average Δ_{PSD} pattern across all electrode sites is also shown for reference. Panel b – EEG band power. The variation of the EEG band power (Δ_P) between before- and after-tDCS conditions is reported separately for the considered EEG bands (delta, theta, low-alpha, and high-alpha) at CP2. The top panel displays the variation using a bar plot. The bottom panel displays the results of the permutation tests performed on Δ_P, by displaying the null distribution, the observed Δ_P value (colored square, or upward/downward pointing arrow if outside the displayed limits), and the 2.5th and 97.5th percentiles of the null distribution. The color of the marker representing the observed value (square or arrow) indicates the statistical significance: blue for significant results ($p < 0.05$), and red for non-significant results.

Appendix B. Analysis of the spectral EEG signatures of spontaneous recovery in DOC via artificial intelligence

In this section, the proposed XAI-based approach (see Section 2.2) was applied to EEG data collected before the tDCS intervention to investigate the spontaneous recovery in DOC patients. Specifically, neural decoders were trained and tested to discriminate between T-1 and T0 resting-state EEG epochs (see Section 2.1), which were recorded two weeks apart. During this interval, two patients showed spontaneous recovery, with CRS-R improvements of 1 and 2 points. The neural decoders achieved high decoding performance across patients, with an AUC of 97.4 ± 2.3 % (mean and standard deviation), an accuracy of 98.0 ± 2.0 %, a sensitivity of 97.6 ± 2.9 %, and a specificity of 97.1 ± 3.9 %. Fig. B1a displays the confusion matrix of neural decoders, while Fig. B1b reports the frequency-level relevance aggregated across patients, separately for spontaneously non-improving, spontaneously improving patients, and all patients. Interestingly, these spectral relevance patterns closely resembled those found when comparing before- vs. after-tDCS, except for a greater relevance of alpha-band components in spontaneously non-improving patients compared to non-improving patients who underwent the tDCS intervention. This overall consistency of results about DOC recovery across analyses is expected, as the XAI-based approach is designed to highlight spectral differences between the contrasted states under investigation, that is between before- and after-tDCS timelines or between two before-tDCS timelines (T-1 and T0). In other words, the framework is inherently capable of capturing recovery-related spectral neural signatures, irrespective of whether the recovery occurs spontaneously or is mediated by intervention. These findings further support the capability of the proposed XAI-based approach to reveal EEG spectral signatures associated with recovery in DOC patients.

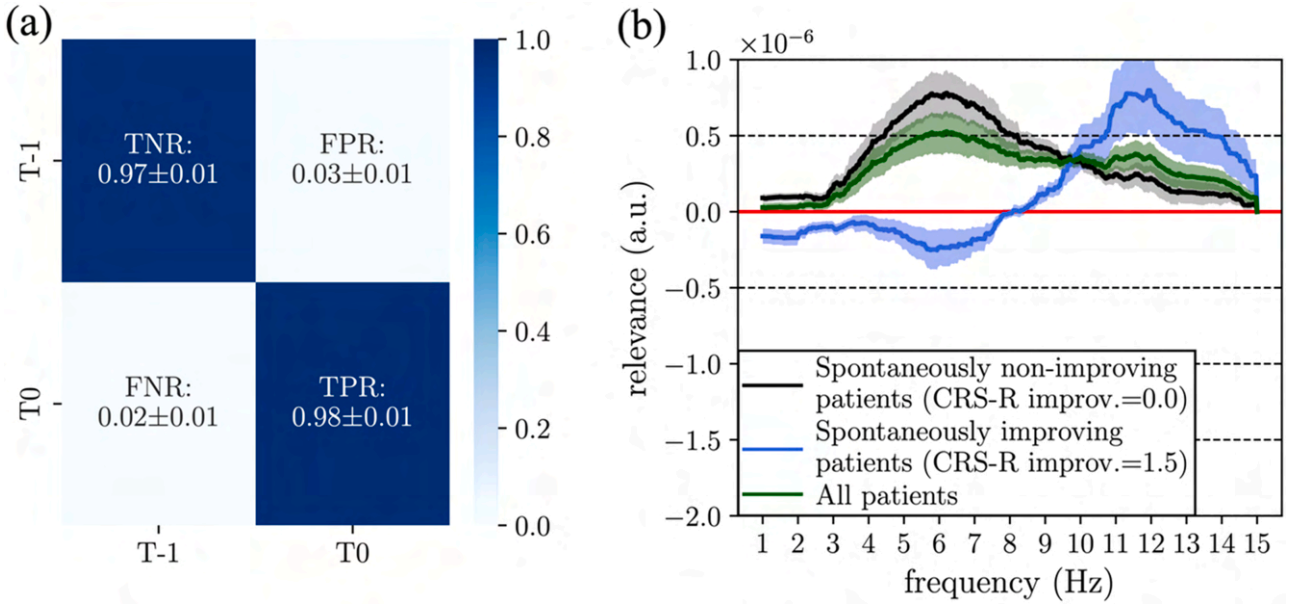

Fig. B1. Analysis of spontaneous recovery in DOC via artificial intelligence. Panel a – Neural decoding performance. The confusion matrix (normalized) is displayed, averaged across folds (mean ± std. across the patients). Panel b – Frequency-level relevance aggregated across patients. Relevance patterns are reported for spontaneously non-improving patients (black, 6 patients, 30 relevance patterns), spontaneously improving patients (blue, 2 patients, 10 relevance patterns), and all patients (green, 8 patients, 40 relevance patterns). Lines denote the mean value across patients and folds, while the shaded areas denote the standard error of the mean. The red line represents the null relevance. Relevance values are expressed in arbitrary units (a.u.).

References

[1] J.T. Giacino, J.J. Fins, S. Laureys, N.D. Schiff, Disorders of consciousness after acquired brain injury: the state of the science, Nat. Rev. Neurol. 10 (2014) 99–114, https://doi.org/10.1038/nrneurol.2013.279.

[2] O. Bodart, S. Laureys, O. Gosseries, Coma and disorders of consciousness: scientific advances and practical considerations for clinicians, Semin. Neurol. 33 (2013) 083–090, https://doi.org/10.1055/s-0033-1348965.

[3] A. Thibaut, N. Schiff, J. Giacino, S. Laureys, O. Gosseries, Therapeutic interventions in patients with prolonged disorders of consciousness, Lancet Neurol. 18 (2019) 600–614, https://doi.org/10.1016/S1474-4422(19)30031-6.

[4] Y. Bai, X. Xia, X. Li, A review of resting-state electroencephalography analysis in disorders of consciousness, Front. Neurol. 8 (2017) 471, https://doi.org/10.3389/fneur.2017.00471.

[5] M.C. Carboncini, A. Piarulli, A. Virgillito, P. Arrighi, P. Andre, F. Tomaiuolo, A. Frisoli, M. Bergamasco, B. Rossi, L. Bonfiglio, A case of post-traumatic minimally conscious state reversed by midazolam: clinical aspects and neurophysiological correlates, Restor. Neurol. Neurosci. 32 (2014) 767–787, https://doi.org/10.3233/RNN-140426.

[6] S.T. Williams, M.M. Conte, A.M. Goldfine, Q. Noirhomme, O. Gosseries, M. Thonnard, B. Beattie, J. Hersh, D.I. Katz, J.D. Victor, S. Laureys, N.D. Schiff, Common resting brain dynamics indicate a possible mechanism underlying zolpidem response in severe brain injury, Elife 2 (2013) e01157, https://doi.org/10.7554/eLife.01157.

[7] O. Gosseries, V. Charland-Verville, M. Thonnard, O. Bodart, S. Laureys, A. Demertzi, Amantadine, Apomorphine and Zolpidem in the Treatment of Disorders of Consciousness, CPD 999 (2013) 11–12, https://doi.org/10.2174/13816128113196660654.

[8] A. Naro, M. Russo, A. Leo, A. Cannavò, A. Manuli, A. Bramanti, P. Bramanti, R. S. Calabrò, Cortical connectivity modulation induced by cerebellar oscillatory transcranial direct current stimulation in patients with chronic disorders of consciousness: a marker of covert cognition? Clin. Neurophysiol. 127 (2016) 1845–1854, https://doi.org/10.1016/j.clinph.2015.12.010.

[9] S. Straudi, V. Bonsangue, S. Mele, L. Craighero, A. Montis, F. Fregni, S. Lavezzi, N. Basaglia, Bilateral M1 anodal transcranial direct current stimulation in post traumatic chronic minimally conscious state: a pilot EEG-tDCS study, Brain Inj. 33 (2019) 490–495, https://doi.org/10.1080/02699052.2019.1565894.

[10] E. Angelakis, E. Liouta, N. Andreadis, S. Korfias, P. Ktonas, G. Stranjalis, D.E. Sakas, Transcranial direct current stimulation effects in disorders of consciousness, Arch. Phys. Med. Rehabil. 95 (2014) 283–289, https://doi.org/10.1016/j.apmr.2013.09.002.

[11] A. Thibaut, M.-A. Bruno, D. Ledoux, A. Demertzi, S. Laureys, tDCS in patients with disorders of consciousness: sham-controlled randomized double-blind study, Neurology 82 (2014) 1112–1118, https://doi.org/10.1212/WNL.0000000000000260.

[12] F. Piccione, M. Cavinato, P. Manganotti, E. Formaggio, S.F. Storti, L. Battistin, A. Cagnin, P. Tonin, M. Dam, Behavioral and neurophysiological effects of repetitive transcranial magnetic stimulation on the minimally conscious state: a case study, Neurorehabil. Neural Repair. 25 (2011) 98–102, https://doi.org/10.1177/1545968310369802.

[13] P. Manganotti, E. Formaggio, S.F. Storti, A. Fiaschi, L. Battistin, P. Tonin, F. Piccione, M. Cavinato, Effect of high-frequency repetitive transcranial magnetic stimulation on brain excitability in severely brain-injured patients in minimally

conscious or vegetative state, Brain Stimul. 6 (2013) 913–921, https://doi.org/10.1016/j.brs.2013.06.006.
[14] X. Xia, Y. Bai, Y. Zhou, Y. Yang, R. Xu, X. Gao, X. Li, J. He, Effects of 10 Hz repetitive transcranial magnetic stimulation of the left dorsolateral prefrontal cortex in disorders of consciousness, Front. Neurol. 8 (2017) 182, https://doi.org/10.3389/fneur.2017.00182.
[15] L.R. Pisani, A. Naro, A. Leo, I. Aricò, F. Pisani, R. Silvestri, P. Bramanti, R. S. Calabrò, Repetitive transcranial magnetic stimulation induced slow wave activity modification: A possible role in disorder of consciousness differential diagnosis? Conscious. Cogn. 38 (2015) 1–8, https://doi.org/10.1016/j.concog.2015.09.012.
[16] O. Gosseries, F. Pistoia, V. Charland-Verville, A. Carolei, S. Sacco, S. Laureys, The role of neuroimaging techniques in establishing diagnosis, prognosis and therapy in disorders of consciousness, TONIJ 10 (2016) 52–68, https://doi.org/10.2174/1874440001610010052.
[17] G. Buzsáki, A. Draguhn, Neuronal Oscillations in Cortical Networks, Science 304 (2004) 1926–1929, https://doi.org/10.1126/science.1099745.
[18] M. Boly, M.-E. Faymonville, P. Peigneux, B. Lambermont, P. Damas, G. Del Fiore, C. Degueldre, G. Franck, A. Luxen, M. Lamy, G. Moonen, P. Maquet, S. Laureys, Auditory processing in severely brain injured patients: differences between the minimally conscious state and the persistent vegetative state, Arch. Neurol. 61 (2004) 233, https://doi.org/10.1001/archneur.61.2.233.
[19] E.J. Kobylarz, N.D. Schiff, Neurophysiological correlates of persistent vegetative and minimally conscious states, Neuropsychol. Rehabil. 15 (2005) 323–332, https://doi.org/10.1080/09602010443000605.
[20] S. Bagnato, C. Boccagni, C. Prestandrea, A.A. Fingelkurts, A.A. Fingelkurts, G. Galardi, Changes in standard electroencephalograms parallel consciousness improvements in patients with unresponsive wakefulness syndrome, Arch. Phys. Med. Rehabil. 98 (2017) 665–672, https://doi.org/10.1016/j.apmr.2016.09.132.
[21] S. Ludwig, S. Bakas, D.A. Adamos, N. Laskaris, Y. Panagakis, S. Zafeiriou, EEGminer: discovering interpretable features of brain activity with learnable filters, J. Neural Eng. 21 (2024) 036010, https://doi.org/10.1088/1741-2552/ad44d7.
[22] D. Zhao, F. Tang, B. Si, X. Feng, Learning joint space–time–frequency features for EEG decoding on small labeled data, Neural Netw. 114 (2019) 67–77, https://doi.org/10.1016/j.neunet.2019.02.009.
[23] D. Borra, S. Fantozzi, E. Magosso, Interpretable and lightweight convolutional neural network for EEG decoding: Application to movement execution and imagination, Neural Netw. 129 (2020) 55–74, https://doi.org/10.1016/j.neunet.2020.05.032.
[24] D. Borra, E. Magosso, Deep learning-based EEG analysis: investigating P3 ERP components, J. Integr. Neurosci. 20 (2021) 791–811, https://doi.org/10.31083/j.jin2004083.
[25] D. Borra, E. Magosso, M. Castelo-Branco, M. Simoes, A Bayesian-optimized design for an interpretable convolutional neural network to decode and analyze the P300 response in autism, J. Neural Eng. 19 (2022), https://doi.org/10.1088/1741-2552/ac7908.
[26] A. Farahat, C. Reichert, C. Sweeney-Reed, H. Hinrichs, Convolutional neural networks for decoding of covert attention focus and saliency maps for EEG feature visualization, J. Neural Eng. (2019), https://doi.org/10.1088/1741-2552/ab3bb4.
[27] A. Vahid, M. Mückschel, S. Stober, A.-K. Stock, C. Beste, Applying deep learning to single-trial EEG data provides evidence for complementary theories on action control, Commun. Biol. 3 (2020) 112, https://doi.org/10.1038/s42003-020-0846-z.
[28] G. Montavon, W. Samek, K.-R. Müller, Methods for interpreting and understanding deep neural networks, Digit. Signal Process. 73 (2018) 1–15, https://doi.org/10.1016/j.dsp.2017.10.011.
[29] A. Shrikumar, P. Greenside, A. Kundaje, Learning important features through propagating activation differences, ArXiv Preprint (2017), https://doi.org/10.48550/ARXIV.1704.02685.
[30] F. Lombardi, G. Gatta, S. Sacco, A. Muratori, A. Carolei, The italian version of the coma recovery scale-revised (CRS-R), Funct. Neurol. 22 (2007) 47–61.
[31] Y.G. Bodien, C.A. Carlowicz, C. Chatelle, J.T. Giacino, Sensitivity and Specificity of the coma recovery scale–revised total score in detection of conscious awareness, Arch. Phys. Med. Rehabil. 97 (2016) 490–492.e1, https://doi.org/10.1016/j.apmr.2015.08.422.
[32] E. Magosso, D. Borra, The strength of anticipated distractors shapes EEG alpha and theta oscillations in a Working Memory task, Neuroimage 300 (2024) 120835, https://doi.org/10.1016/j.neuroimage.2024.120835.
[33] D. Borra, S. Fantozzi, M.C. Bisi, E. Magosso, Modulations of cortical power and connectivity in alpha and beta bands during the preparation of reaching movements, Sensors 23 (2023) 3530, https://doi.org/10.3390/s23073530.
[34] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (1981) 381–395, https://doi.org/10.1145/358669.358692.
[35] D.-A. Clevert, T. Unterthiner, S. Hochreiter, Fast and accurate deep network learning by exponential linear units (elus), in: ArXiv Preprint, 2015, https://doi.org/10.48550/ARXIV.1511.07289.
[36] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: ArXiv Preprint, 2015, https://doi.org/10.48550/ARXIV.1502.03167.
[37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014) 1929–1958.
[38] S. Saha, M. Baumert, Intra- and inter-subject variability in EEG-based sensorimotor brain computer interface: a review, Front. Comput. Neurosci 13 (2020) 87, https://doi.org/10.3389/fncom.2019.00087.
[39] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, B.J. Lance, EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces, J. Neural Eng. 15 (2018) 056013, https://doi.org/10.1088/1741-2552/aace8c.
[40] R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for EEG decoding and visualization, Hum. Brain Mapp. 38 (2017) 5391–5420, https://doi.org/10.1002/hbm.23730.
[41] A. Sujatha Ravindran, J. Contreras-Vidal, An empirical comparison of deep learning explainability approaches for EEG using simulated ground truth, Sci. Rep. 13 (2023) 17709, https://doi.org/10.1038/s41598-023-43871-8.
[42] F. Wilcoxon, Individual comparisons by ranking methods, Biom. Bull. 1 (1945) 80, https://doi.org/10.2307/3001968.
[43] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. B (Methodol.) 57 (1995) 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
[44] M. Friedman, The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance, J. Am. Stat. Assoc. 32 (1937) 675–701, https://doi.org/10.1080/01621459.1937.10503522.
[45] T.E. Nichols, A.P. Holmes, Nonparametric permutation tests for functional neuroimaging: a primer with examples, Hum, Brain Mapp 15 (2002) 1–25, https://doi.org/10.1002/hbm.1058.
[46] D.E. Hinkle, W. Wiersma, S.G. Jurs, Applied statistics for the behavioral sciences, 5th ed, Houghton Mifflin, Boston, 2003.
[47] S. Bagnato, C. Boccagni, A. Sant'Angelo, C. Prestandrea, R. Mazzilli, G. Galardi, EEG predictors of outcome in patients with disorders of consciousness admitted for intensive rehabilitation, Clin. Neurophysiol. 126 (2015) 959–966, https://doi.org/10.1016/j.clinph.2014.08.005.
[48] R. Lehembre, B. Marie-Aurélie, A. Vanhaudenhuyse, C. Chatelle, V. Cologan, Y. Leclercq, A. Soddu, B. Macq, S. Laureys, Q. Noirhomme, Resting-state EEG study of comatose patients: a connectivity and frequency analysis to find differences between vegetative and minimally conscious states, Funct. Neurol. 27 (2012) 41–47.
[49] A. Naro, P. Bramanti, A. Leo, A. Cacciola, A. Bramanti, A. Manuli, R.S. Calabrò, Towards a method to differentiate chronic disorder of consciousness patients' awareness: The low-resolution brain electromagnetic tomography analysis, J. Neurol. Sci. 368 (2016) 178–183, https://doi.org/10.1016/j.jns.2016.07.016.
[50] C. Babiloni, M. Sarà, F. Vecchio, F. Pistoia, F. Sebastiano, P. Onorati, G. Albertini, P. Pasqualetti, G. Cibelli, P. Buffo, P.M. Rossini, Cortical sources of resting-state alpha rhythms are abnormal in persistent vegetative state patients, Clin. Neurophysiol. 120 (2009) 719–729, https://doi.org/10.1016/j.clinph.2009.02.157.
[51] J. Leon-Carrion, J.F. Martin-Rodriguez, J. Damas-Lopez, J.M. Barroso Y Martin, M. R. Dominguez-Morales, Brain function in the minimally conscious state: A quantitative neurophysiological study, Clin. Neurophysiol. 119 (2008) 1506–1514, https://doi.org/10.1016/j.clinph.2008.03.030.
[52] W. Klimesch, EEG-alpha rhythms and memory processes, Int. J. Psychophysiol. 26 (1997) 319–340, https://doi.org/10.1016/S0167-8760(97)00773-3.
[53] J.M. Shine, L.D. Lewis, D.D. Garrett, K. Hwang, The impact of the human thalamus on brain-wide information processing, Nat. Rev. Neurosci. 24 (2023) 416–430, https://doi.org/10.1038/s41583-023-00701-0.
[54] Y.D. Van Der Werf, M.P. Witter, H.J. Groenewegen, The intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness, Brain Res. Rev. 39 (2002) 107–140, https://doi.org/10.1016/S0165-0173(02)00181-9.
[55] S. Laureys, The neural correlate of (un)awareness: lessons from the vegetative state, Trends Cogn. Sci. 9 (2005) 556–559, https://doi.org/10.1016/j.tics.2005.10.010.
[56] R. Polanía, W. Paulus, A. Antal, M.A. Nitsche, Introducing graph theory to track for neuroplastic alterations in the resting human brain: a transcranial direct current stimulation study, Neuroimage 54 (2011) 2287–2296, https://doi.org/10.1016/j.neuroimage.2010.09.085.
[57] S. Corchs, G. Chioma, R. Dondi, F. Gasparini, S. Manzoni, U. Markowska-Kaczmar, G. Mauri, I. Zoppis, A. Morreale, Computational methods for resting-state EEG of patients with disorders of consciousness, Front. Neurosci 13 (2019) 807, https://doi.org/10.3389/fnins.2019.00807.
[58] A. Naro, P. Bramanti, A. Leo, M. Russo, R.S. Calabrò, Transcranial alternating current stimulation in patients with chronic disorder of consciousness: a possible way to cut the diagnostic gordian knot? Brain Topogr. 29 (2016) 623–644, https://doi.org/10.1007/s10548-016-0489-z.
[59] H. Cecotti, A. Graser, Convolutional neural networks for p300 detection with application to brain-computer interfaces, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 433–445, https://doi.org/10.1109/TPAMI.2010.125.