A novel explainable machine learning approach for EEG-based brain-computer interface systems

Noname manuscript No.
(will be inserted by the editor)

Explainable Machine Learning Inverse Problem Solution for EEG-based BCI

Cosimo Ieracitano · Nadia Mammone · Amir Hussain · Francesco Carlo Morabito

Received: date / Accepted: date

Abstract Electroencephalographic (EEG) recordings can be of great help in decoding the open/close hand's motion preparation. To this end, cortical EEG source signals in the motor cortex (evaluated in the 1s window preceding movement onset) are extracted by solving inverse problem through beamforming. EEG sources epochs are used as source-time maps input to a custom deep Convolutional Neural Network (CNN) that is trained to perform 2-ways classification tasks: pre-hand close (HC) vs. resting state (RE) and pre-hand open (HO) vs. RE. The developed deep CNN works well (accuracy rates up to 89.65±5.29% for HC vs. RE and 90.50±5.35% for HO vs. RE), but the core of the present study was to explore the interpretability of the deep CNN to provide further insights into the activation mechanism of cortical sources during the preparation of hands' sub-movements. Specifically, *occlusion sensitivity analysis* was carried out to investigate which cortical areas are more relevant in the classification procedure. Experimental results show a recurrent trend of spatial cortical activation across subjects. In particular, the central region (close to the longitudinal fissure) and the right temporal zone of the premotor together with the primary motor cortex appear to be primarily involved. Such

C. Ieracitano (Corresponding author)
DICEAM University Mediterranea of Reggio Calabria, Italy, Via Graziella Feo di Vito, 89124
E-mail: cosimo.ieracitano@unirc.it

N. Mammone
DICEAM University Mediterranea of Reggio Calabria, Italy, Via Graziella Feo di Vito, 89124
E-mail: nadia.mammone@unirc.it

A. Hussain
School of Computing, Edinburgh Napier University, Edinburgh EH10 5DT, Scotland, UK
E-mail: a.hussain@napier.ac.uk

F.C. Morabito
DICEAM University Mediterranea of Reggio Calabria, Italy, Via Graziella Feo di Vito, 89124
E-mail: morabito@unirc.it

findings encourage an in-depth study of cortical areas that seem to play a key role in hand's open/close preparation.

Keywords Brain Computer Interface · Beamforming · Deep Learning · Explainable Machine Learning

1 Introduction

Scalp electroencephalography (EEG) is a non-invasive technique that collects the electrical fields produced by the brain and indirectly reveals its underlying activity [1]. EEG is the gold standard diagnostic technique for several neurological diseases, as well as in neuroscience and cognitive research [2,3,4,5]. EEG is commonly exploited in brain-computer interface systems (BCI), whose ultimate goal is to allow the brain to directly communicate with an external device by decoding subject's intentions from EEG signals and converting them into a set of suitable commands [6]. EEG is relatively affordable, widely spread, easy to use, and generally well tolerated by subjects. Unfortunately, EEG entails some non trivial limitations: 1) a poor signal-to-noise ratio (SNR) which reflects on the the brain's waves of interest to be corrupted by multiple sources of noise called *artifacts* [7]; 2) EEG recordings are non-stationary signals, thus their statistical characteristics vary across time [8]; 3) poor spatial resolution caused by volume conduction effects [9]; 4) high inter-subject variability which limits the ability of a classifier, trained over a cohort of subjects, to generalize well across subjects [10]. One of the greatest potentials of Deep Learning (DL) is the ability to generalize well even in presence of complex inputs [11]. In the context of the EEG analysis, this results in the possibility of identifying patterns relevant to classification even in presence of additional irrelevant waves in the EEGs. However, the main limitation in the application of DL to EEG processing lays in the relatively small number of samples generally available in EEG databases, as compared to the number of samples that are typically available in databases meant for computer vision or natural language processing (NLP) applications, which made DL so powerful in such fields. The DL has been applied so far to EEG mainly in the following fields: motor imagery (MI) (22%), mental workload (16%), emotion recognition (16%), seizure detection (14%), event related potential detection (10%), sleep stage scoring (9%) and other studies (13%) [12]. MI is the topic more related to the focus of the present study: motor preparation.

BCI systems based on MI generally require the user to imagine performing a given movement, in a sustained way, in order to allow the system to learn how to classify the imagined movement with good accuracy [13]. However, sustained motor imagery is not natural neither comfortable for the user; moreover, it requires an intensive training and causes a delay between the onset of imagination and the time the desired control is issued [14]. Conversely, in motor preparation investigation, the subject performs or attempts to perform the movement and the behavior of EEG signals collected before motion onset/attempt is investigated to predict the intended movement [15]. Decoding

the preparation of the movement, whether it is actually implemented or just attempted (for example, in case the subject has a motor disability hindering motor implementation), would be far more natural and immediately decodable [16]. Furthermore, the mechanisms of motor preparation are still not clear to scientists. Previous studies showed that premotor cortex is activated contralaterally during motor preparation, which was observed by fMRI/NIRS as well as in EEG signals [17], however it is not clear whether and how the different subareas of premotor cortex work together to develop motion planning. For the all the aforementioned motivations, it is reasonable to consider motor preparation a key field of research and, in line with such consideration, motor preparation of different sub-movements of upper limbs was investigated in [18]. Frames of EEG signals preceding motion's onset were compared with frames of EEGs collected in absence of any motion planning (resting) [18]. Mammone et al. [18] reached an accuracy of 90.30±5.6% in pre-movement vs. resting discrimination and of 62.47±6.7% in the discrimination of the preparation of different sub-movements. A deep Convolutional Neural Network (CNN) was designed and trained through stratified time-frequency maps of 210 EEG source locations in the premotor and primary motor cortex. However, no interpretation of the achieved results was provided. In the present work, we aim at achieving good accuracy in motor planning classification of hands' movements, by designing a deep CNN to be trained over single channel images (time vs. sources), and also to provide an intrpretation of the achieved results in terms of involvement of the different cortical areas. As artificial neural networks act as a black-box, an explainable machine learning (EML) approach is here proposed to interpret the achieved results by exploring the behavior of the trained network [19]. In summary, the aim of the present work is twofold: 1) To design a novel deep CNN that, by processing EEG source signals in the motor cortex, is able to discriminate the phases of preparation of hands' movements (open/close) from resting (no movement planning); 2) To explain the achieved results by means of EML, in order to assess which EEG sources (i.e., which cortical locations) play a decisive role in the classification of hand's motor preparation phases. The final aim is to find out possible areas in the motor cortex that are mainly involved in planning hands' movements. In fact, while it is well known which areas are most involved in the implementation of movements of the different parts of the body [20], it is indeed known that the activation of the movement is triggered by relatively well localized areas in the primary motor cortex, it is not well known whether and how motor planning is spatially organized. EML could yield a significant contribution in this field [21]. To this end, a deep CNN was designed and trained to discriminate hand's opening (HO) and hand's closing (HC) motion preparation phases from resting (RE) phases. The present analysis is focused on the classification of HC vs. RE and HO vs. RE. The training database was constructed by processing EEG signals collected from 15 subjects recruited within a BCI study conducted by Ofner et al. [22]. The paradigm introduced in [22] provided that the subject performed cue-based movements starting from a neutral rest position. The developed CNN receives as input epochs of EEG source signals in the time interval of 1 s preceding

motion onset (named *time-source maps* herein). Such source signals were estimated by solving the inverse problem starting from EEG scalp signals. Source locations belonging to the motor cortex are then included in the analysis. The developed CNN was able to discriminate premov (HC or HO) vs. RE with an average accuracy of 90%. An occlusion sensitivity analysis was subsequently carried out by passing the time-source maps as input to the network to evaluate which sources (i.e., cortical locations) are estimated to be more relevant in the classification of HC/HO from RE. We could observe a recurrent spatial pattern across subjects that show greater activation of the left part of the motor cortex in the central area, close to the longitudinal fissures between the two hemispheres, together with the extreme right part of the motor cortex belonging to the temporal lobe.

The paper is organized as follows: in Section 2 the proposed method is presented. The preprocessing steps including beamforming technique and the cortical sources extraction are also described. Section 3 shows the proposed deep CNN for the pre-movements tasks classification, whereas, Section 4 introduces the salient cortical source recovery procedure by means of EML (i.e. occlusion sensitivity analysis). In Section 5 experimental results are reported. Section 6 discusses the achieved findings and Section 7 concludes the paper.

2 Methodology

The proposed methodology is shown in Figure 1. It includes the following processing modules:

1. *Extraction of premotor EEG epochs*: EEG segments preceding the (open/close) movements of the right hand as well as the rest condition are extracted from the available EEG dataset.
2. *Inverse problem solution and extraction of the cortical EEG sources*: beamforming technique is applied to reconstruct EEG sources corresponding to a given EEG epoch. It is worth noting that among the 2000 cortical locations of the adopted head model, those related to the areas 4 and 6 of Brodmann (i.e., the primary motor and the premotor cortex) are considered. Specifically, 210 EEG sources are extracted.
3. *DL-based system for pre-movements tasks classification*: the extracted 210-EEG sources are fed into a DL-based classifier. Specifically, a deep CNN is developed to perform the following binaries classifications: HC vs. RE and HO vs. RE. The deep CNN receives time-source maps as input (i.e., matrices sized 210 x 512, where 210 is the number of the source locations under consideration and 512 is the sampling rate).
4. *Explainable DL system, salient cortical source recovery*: occlusion sensitivity analysis is applied to investigate which cortical area is more relevant to the open/close hand motion planning. Specifically, portions of the input maps (i.e., time-source maps) are occluded with the aim of detecting the cortical area that mostly affected the classification. K-means cluster-

ing technique is then used to segment the saliency maps and automatically detect the cortical source locations most relevant to the task prediction.

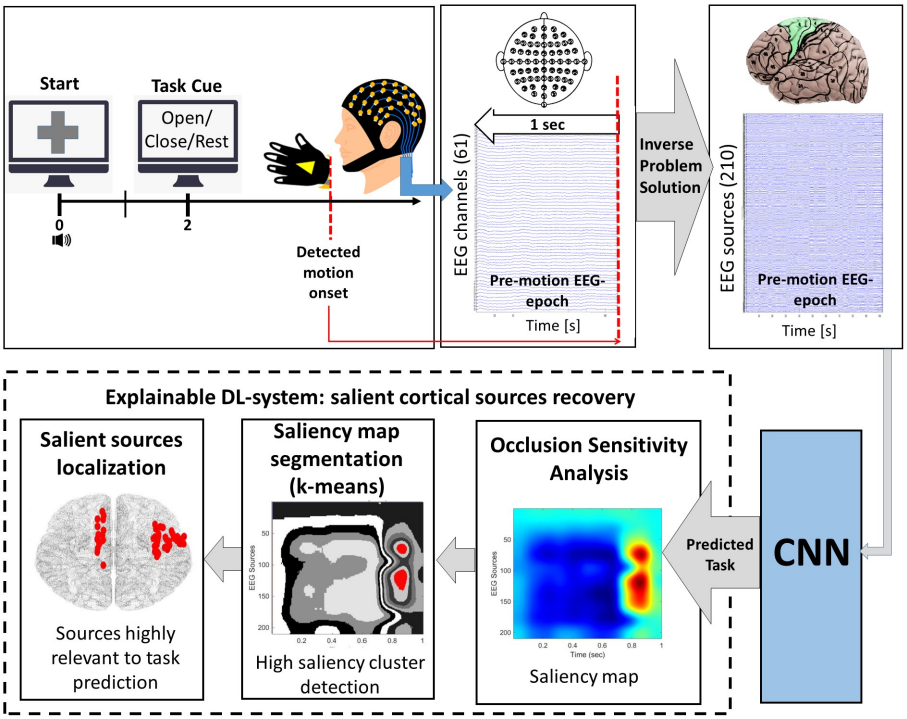

Fig. 1 Flowchart of the proposed framework. The subject is initially in a rest condition watching a monitor in front of him/her. After a beep (at 0 s), a fixation cross is produced on the monitor in order to make him/her attention on it. Next (at 2 s), the cue of the desired task (i.e., HC/HO) is shown. The subject performs the cued movement and then returns to the original neutral position. The temporal window of 1 s preceding motion execution (i.e., pre-motion EEG epoch) is extracted and the inverse problem is solved by means of beamforming technique in order to reconstruct the EEG sources. Notably, 210-EEG cortical sources (i.e., pre-motion EEG sources epoch) belonging to the motor cortex are taken into account and used as input to the proposed deep CNN for the following binary classification tasks: HC vs. RE and HO vs. RE. Finally, an occlusion sensitivity analysis is carried out to detect the most significant regions in the input map (i.e., epoch of the EEG signals) involved in the classification process. The achieved saliency map is then segmented by k-means clustering and the cluster associated to highest saliency is identified. Sources belonging to that cluster interpreted as providing the largest contribution in decoding the specific movement.

2.1 Extraction of premotor EEG epochs

In the present research, a publicly available database of EEG recordings co-registered with signals collected from motion sensors [22] was used to construct

the train and test dataset. The database can be found at $http://bnci-horizon-2020.eu/database/data-sets$ together with detailed information about channel layout, recording settings and paradigm description. The study involved 15 healthy subjects (aged 27±5 years, nine of the them are females). EEG signals were acquired by Ofner et al. [22] by means of 61 active EEG electrodes and four 16-channel amplifiers (g.tec medical engineering GmbH, Austria). Right mastoid channel was used as reference one and AFz was set as ground channel. EEG signals were band-pass filtered between 0.01 Hz and 200 Hz (8-th Chebyshev filter), notch filtered at 50 Hz and sampled at 512 Hz. The database consists of a motor execution and a motor imagery part. Since the goal of the present study was to investigate motor preparation, the first part was included in the analysis. During the experiment, subjects remained seated on a comfortable chair and an anti-gravity exoskeleton (Hocoma, Switzerland) supported their right arm. The paradigm consisted in executing cue-based movements of the right upper limb starting from a neutral position (lower arm extended to 120 degree and in a neutral rotation, hand half open) [22]. The experiment consisted of 10 runs, every run included 6 trials and each trial included one hand open (HO), one hand close (HC) and one rest (RE) cues. The timeline of the paradigm can be summarized as follows: at second 0, a fixation cross appeared on a computer screen, positioned in front of the subject, to attract her/his gaze on it and limit eye movements. At second 2, the cue of the task to be performed (HC/HO/RE) appeared on the computer screen. After task execution, the subject moved her/his hand back to the starting neutral position. In order to train a neural network to decode motor preparation phases from EEG signals, a dataset of EEG epochs preceding motion onset was necessary. To this purpose, the onset of movement was estimated by processing motion data collected by glove sensors. Specifically, the onset of movement was detected by processing the signals recorded from motion sensors embedded in the glove, following the procedure described in [22]. The marked onset timing was manually checked for all of the 1800 pre-motion epochs under examination and the frames (epochs) of EEG signals preceding the marked onset were extracted accordingly. Such epochs were included in the analysis together with a balanced set of resting EEG epochs. Specifically, 900 EEG epochs (derived from 10 runs x 6 trials x 15 subjects) per movement class (hand open/hand close) were taken into account. In order to generate a balanced dataset, a comparable number of resting state EEG epochs was extracted. In the end, 2700 EEG epochs (derived from 10 runs x 6 trials x 15 subjects x 3 classes) were extracted from the EEG recordings and included in the dataset. As regards the choice of the length of the frame preceding motion onset, it was set at 1 s after taking into account the typical timeline of motor related cortical potentials (MRCP) which are brain waves that arise together with movements' preparation and initiation [23, 18].

2.2 Inverse problem solution and extraction of the cortical EEG sources

It is known that the EEG has a very good temporal resolution but a poor spatial resolution, due to volume conduction effects [1,24,25]. Inverse problem solution is a possible way to deal with such effects. In the proposed methodology, EEG signals are used to reconstruct a set of source signals where every source signal represents the contribution of a source location (current dipole) located in the cortex [24,25]. Solving the "inverse problem" means reconstructing source locations' contribution to the overall EEG signals collected at the scalp. EEGs can be hypothesized to be the projection of sources' contributions from cortical locations to scalp sensors through a "forward model" [26]. Such forward model takes into account the structural and conductive properties of brain tissues. In the frequency range of EEG signals, the quasi-static approximation of Maxwell's equations can be assumed hence the forward model becomes linear [27] and be formulated as follows:

$$\mathbf{x}(t) = \mathbf{L}\mathbf{q}_r(t) \tag{1}$$

$\mathbf{q}_r$(t) is the 3 dimensional directed current dipole associated to cortical location "r" (where r=1,...,Ns and Ns is the number of possible source locations in the cortex); $\mathbf{L}$ is knows as "lead field" matrix, which represents the head model that projects the current dipole $\mathbf{q}_r$(t) into the scalp potential $\mathbf{x}$(t) [27]. The number of sources Ns is typically larger than the number of channels Nc thus estimating $\mathbf{q}_r$(t) from x(t) is inherently an ill posed problem. The adopted head model consists of 2000 cortical locations (Ns=2000) whereas the number of scalp channels of the EEG recordings analysed in the present work is Nc=61. In this work, the *New York Head* (NYH) forward model, developed by Haufe et al. [28], was adopted. Such head model is based on the popular ICBM152 anatomy, a nonlinear average of T1-weighted structural MR images collected from 152 adults. By solving the inverse problem, cortical current dipoles $\mathbf{q}_r$(t)) are estimated starting from the recorded EEG signals $\mathbf{x}$(t) and from the lead field matrix $\mathbf{L}$. Several inverse problem solution approaches can be found in the literature on EEG source imaging: minimum-norm solutions, beamformers, and dipole modeling [24,29]. Beamforming solves the inverse problem by maximizing the contribution of a given source location while suppressing contributions from the other ones and was proved very effective in BCI applications by Grosse-Wentrup et al. [30]. The premotor and primary motor cortex are considered crucial in movement planning and execution [31, 32]. Such regions fall in the Brodmann's Areas 4 and 6 of the brain [33]. Each one of the 2000 available source locations was associated to the corresponding Brodmann Area through its Montreal Neurological Institute (MNI) stereotaxic coordinates. MNI coordinates of every source locations were known. First of all, they were indeed converted into Talairach coordinates [34] and then matched with Talairach Atlas labels [35], in order to come up with the corresponding Brodmann area of every source location. In the end, 210 locations belonging to Brodmann areas 4 and 6 were selected out of the 2000 available ones.

3 Deep Learning-based system for pre-movements tasks classification

A deep learning classifier is proposed to discriminate EEG source epochs belonging to the pre-movement (close/open) of the right hand (HC/HO) or to the resting state (RE) class. Notably, a deep Convolutional Neural Network (CNN) is developed to perform the 2-ways classification tasks: the HC vs. RE and HO vs. RE.

3.1 Convolutional Neural Network

CNN is a well-known deep learning model widely used especially in computer vision [36,37] and image recognition [38,39]. It is composed of subsequence layers of *convolution*, *activation*, *pooling* followed by a multi-layer fully connected neural network for classification purpose [40]. The *convolutional* layer includes a bank of J filters used to estimate the dot product (i.e., covolution operation) with the input map $\mathbf{T}$ sized t_1 x t_2. More specifically, each filter (sized j_1 x j_2) performs the convolution with the selected local area and sweeps over the input representation with a specific stride using the same values of weights. This operation results in the so called *features maps* $\mathbf{Z}$ of size z_1 x z_2:

$$z_1 = \frac{t_1 - j_1 + 2 \times p}{s} + 1 \tag{2}$$

and

$$z_2 = \frac{t_2 - j_2 + 2 \times p}{s} + 1 \tag{3}$$

where p is the zero padding parameter. The *activation* layer introduces nonlinearity in the model. Specifically, here, the *Rectified Linear Unit* (ReLU) is employed for its ability to achieve good generalization and training time [41]. The *pooling* performs a downsampling operation of the feature maps resulting from the previous layer. The max pooling operation is used for its good translation-invariant properties [42]. It has a filter sized $\bar{j}_1$ x $\bar{j}_2$ that scans the input feature map with stride $\bar{s}$. This operation outputs a reduced map sized $\bar{t}_1$ x $\bar{t}_2$, with:

$$\bar{r}_1 = \frac{t_1 - \bar{j}_1}{\bar{s}} + 1 \tag{4}$$

and

$$\bar{r}_2 = \frac{t_2 - \bar{j}_2}{\bar{s}} + 1 \tag{5}$$

The CNN ends with a standard feed-forward MLP composed of a softmax output function used for performing the discrimination tasks.

3.2 Design of the deep CNN

The proposed CNN is developed to accept as input time-source maps (i.e., EEG sources epochs) sized t_1 x t_2, where t_1=210 represents the number of sources taken into account in this study, and t_2=512 represents the number of samples included in 1 s temporal epoch before the movement onset. The deep learning model consists of three stacked modules of convolutional ($conv_i$, with i=1,2,3), ReLU and max pooling ($mpool_i$) layers followed by a common MLP for performing the 2-ways classification: HC vs. RE and HO vs. RE. Figure 2 shows the architecture of the proposed deep CNN. Note that the topology of the developed model was set-up after several tests and using a *trial-and-error* strategy. Specifically, the temporal data (210 x 512) are firstly convolved with 4 learnable filters sized 5 x 5 ($conv_1$). Every filter scans the input representation with stride 1 resulting in 4 feature maps sized 206 x 508 (according eq. 2, with p=0). After applying the ReLU activation function, max pooling is used to squeeze the input space from 206 x 508 to 68 x 169 ($mpool_1$). This layer has the filter size 5 x 4 and stride equal to 3. Next, the four features maps extracted from $mpool_1$ are fed into a new layer of convolution ($conv_2$) composed of 8 filters sized 5 x 5 and unit stride, producing 8 maps of dimension 64 x 165 to which the ReLU activation function is applied. The second max pooling layer ($mpool_2$) has filter size 4 x 6, stride=3 and outputs 8 downsampled representation of size 21 x 54. Similarly, the last convolutional layer ($conv_3$) has 16 filters sized 5 x 5 and unit stride. The resulting features maps have dimension of 17 x 50. Finally, the ReLU activation function is applied and the third max pooling layer ($mpool_3$) generates 16 maps sized 5 x 16 (filter size 5 x 5, stride 3). The flatten vector of dimension5 x 16 x 16 = 1280 inputs a standard 2-hidden layers MLP consisted of 500 and 50 hidden units respectively, employed to perform the binaries classifications: HC vs. RE and HO vs. RE.

3.3 Learning parameters set-up

The Adaptive Moment (Adam) optimization procedure [43] was used to train the proposed deep CNN (Figure 2), using mini-batches size of 28. Training options were set-up by using the practical recommendations reported in [44,43], specifically: learning rate α=10^{-2}, first moment decay rate β_1=0.9, and second moment decay rate β_2=0.999. The network was implemented with MATLAB R2019a and trained over about 50 iterations on two Nvidia GeForce RTX 2080 Ti GPU, each with 11 GB memory, installed on a processor Intel(R) Core(TM) i7-8000K CPU @ 3.70GHz and RAM of 64 GB. Furthermore, the proposed deep CNN was trained and tested iteratively by using epochs of the each subject under analysis. Thus, 15 CNN classifiers were trained and tested. The estimated training time was of about 10 min per subject (applying the k-folds cross validation procedure, with k=10).

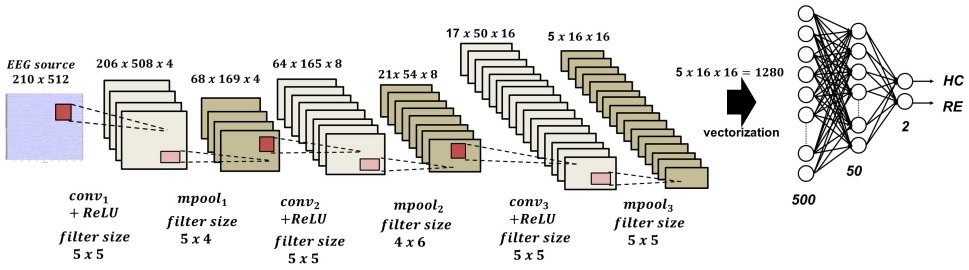

Fig. 2 Architecture of the proposed deep CNN. It consists of three convolutional layers (followed by ReLU non-linearity) and three max pooling layers. The network ends with a 2-hidden layers MLP employed to perform the 2-ways classifications: HC vs. RE and HO vs. RE.

4 Explainable Deep Learning system: salient cortical source recovery

4.1 Occlusion sensitivity analysis

Occlusion analysis has been widely used in image classification to show the sensitivity of a pre-trained CNN to different areas of an input image [45]. It consists in systematically occluding different patches of the input data with a grey mask and estimating the related effect on the network output. For each mask location the discrimination is performed using a pre-trained CNN and estimating the change in classification score for a specific class than the initial prediction (input without occlusion). Such changes in classification result in the so called *heatmap* or *saliency map* **H** with a coloration ranging from blue to red and with the same input dimension. This representation reveals which area of the image is the most essential for the classification. Specifically, red color corresponds to higher values and consequently represents the most significant area that contributed to identify the specified class. When this region is occluded the classification performance decreases. Blue color corresponds to lower values and represents the areas not relevant during the discrimination task. In this study, the occlusion technique is applied to recover the cortical sources that are activated during the (open/close) hand's movement preparation. Given a subject under analysis, the e^{th} EEG source epoch (sized 210 x 512) is repeatedly occluded with a 42 x 256 pixel grey mask that moves across the input data with a vertical and horizontal stride of 21 and 51, respectively. It is worth noting that the dimension and stride of the mask has been set-up empirically, after several experiments. For each position of the mask, the 2-way discrimination task (i.e., HC vs. RE or HO vs. RE) is performed by using the proposed pre-trained CNN (Sect. 3.2). The output is a heatmap sized 210 x 512. As an example, Figure 3a shows the input map (i.e., EEG sources epoch) when a portion is occluded by a grey mask; whereas, Figure

3b reports the achieved saliency representation map. In this case, as can be seen, the red area roughly corresponding to the sources ranged between (130-180) in the temporal window (0.7-0.9), denotes the most relevant zone in the classification process. Further considerations and analyses are reported in the Experimental Results Sect. 5.2.

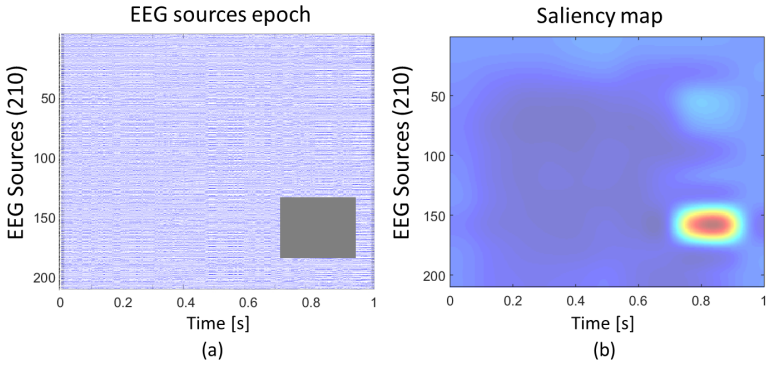

Fig. 3 (a) EEG sources epoch when a sub-region is occluded by a grey mask, that slides along the whole input size. (b) Saliency map with a coloration ranging from blue to red. Blue color represents low values that correspond to regions not significant for the classification task, vice-versa, red color represents high values that correspond to regions significant for the specified discrimination class.

4.2 Saliency maps segmentation through k-means

In order to provide a deeper understanding of which sub-areas in the motor cortex gave the largest contribution to decode movements' planning, a segmentation of the saliency maps was necessary in order to extract the high saliency zones automatically. To this end, the k-means clustering algorithm was applied to partition each saliency map to k=10 clusters. K-means is a widely applied clustering algorithm [46]. Its aim is to gather data points into a given number of clusters by following an iterative four-steps procedure:

1. the initial cluster centers are set randomly;
2. data points are assigned to the nearest cluster by estimating the euclidean distance between the data point and the cluster centers, in this way clusters are redefined;
3. update the clusters' centers;
4. go back to step 2. and repeat the procedure from 2. to 4. until the cluster centers do not change or the specified iteration number is reached.

After applying k-means to the saliency maps, the cluster associated to the highest saliency values was detected, the corresponding highly salient sources were extracted and mapped onto the cortex by red dots.

5 Experimental Results

5.1 Classification performance of the pre-movements tasks

The classification performance of the proposed deep CNN were evaluated using standard metrics (accuracy, recall, precision, F1-score):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

where TP and TN are true positive and negative, respectively; whereas, FP and FN are false positive and negative, respectively [47]. Notably, TP is the number of EEG sources epochs belonging to HC/HO category and correctly classified as HC/HO; TN is the number of EEG sources epochs belonging to RE category and correctly classified as RE; FP is the number of EEG sources epochs belonging to RE misclassified as HC/HO and vice-versa FN are EEG epochs of HC/HO class erroneously identified as RE. In this study, 15 CNNs were trained and tested, 1 per subject, and the following 2-ways discrimination tasks were performed: HC vs. RE and HO vs. RE. The dataset of each binary classifier was composed of 120 EEG epochs (i.e., 60 belonging to HC/HO and 60 to RE). The k-fold cross validation technique was applied (with k=10), in particular: the train set included 70% of data (i.e. EEG sources epochs) and the test set the remaining 30%.

Table 1 reports results of the HC vs. RE classification. Remarkable discrimination values were observed in all subjects, reporting average recall, precision, F1-score and accuracy of 89.14±7.24% and 91.19±7.88%, 89.69±4.98%, 89.65±5.29%, respectively. It is worth noting that the highest individual classification performance was achieved by Sb_{08} with accuracy of 98.02±2.10%, F1-score of 97.94±2.21%, recall of 96.03±4.20% and precision of 100%; while the lowest individual classification performance was achieved by Sb_{07}. However, also in this case high discrimination scores were observed, but with higher standard variation: accuracy of 79.76±10.11%, F-score of 79.86±9.60%, recall of 80.16±12.77% and precision of 80.83±12.36%.

Table 2 reports results of the HO vs. RE classification. Also in this scenario very good performance were observed (average recall of 89.31±8.02%, average precision of 93.04±7.66%, average F1-score of 90.41±5.32% and average accuracy of 90.50±5.35%). Notably, Sb_{08} and Sb_{02} achieved the best performances in terms of accuracy and F1-score; while, Sb_{07} reported the worst individual classification performance with an accuracy value of 75.79±9.72% and F1-score of 76.95±5.87%.
Hence, the proposed deep CNN reported very good performance for both HC vs. RE and HO vs. RE discrimination task. This result was also confirmed by the analysis of the Precision-Recall Curve (PRC, shown in Figure 5) and in particular by measuring the corresponding Area Under the PRC. Specifically, the average AUPRC over all subjects was estimated, reporting AUPRC rates up to 89.36±9.02% and 90.1±8.93%, for the HC vs. RE and HO vs. RE classification tasks, respectively.

5.2 Salient cortical source locations recovery

Occlusion sensitivity analysis was performed by using the proposed pre-trained deep CNN to estimate the saliency of every EEG source. Specifically, for each subject, an averaged saliency map was estimated by averaging the saliency maps corresponding to the HC/HO EEG epochs correctly classified during the testing procedure and herein denoted as $\tilde{\mathbf{H}}^{tk}_{Sb_i}$, where tk represents the pre-movement task (i.e., HC/HO) and Sb_i is the subject under analysis (with i=1,2,..15). As an example, Figure 4a illustrates the average saliency representation of Subject 08 while preparing to perform hand closing ($\tilde{\mathbf{H}}^{HC}_{Sb_08}$). Saliency is encoded with a coloration going from blue (low saliency) to red (high saliency). Highly salient EEG source locations can be recovered by detecting the EEG sources associated to red areas. Notably, red areas denote that the classification score decreases when the corresponding local regions of the input were hidden by the mask, which means that the occluded area is relevant to classification. In the example map shown in Figure 4a, the area located around 0.85s and approximately associated to the EEG sources ranging from 70 to 170, looks colored in red. Hence, it resulted relevant to the decision making. Average saliency maps were then segmented as described in Sect. 4.2. Figure 4b depicts the clustered saliency map of Subject 08. Notably, the red area represents the cluster with the highest saliency and refers to the most relevant EEG sources, which were then mapped onto the cortex (red dots in Figure 4c). In the example shown in Figure 4c), EEG sources located in the left central zone (close to the longitudinal fissure) and in the right-temporal zone contributed the most to decoding hand closing motor planning. Following the aforementioned procedure, salient source locations in HC and HO motor preparation were estimated for every subject and are shown in Figure 6. It is worth to note the a recurrent spatial pattern of cortical activation (left central zone close to the longitudinal fissure and right-temporal zone) which occurred similarly during HC or HO motor preparation. Such pattern occurred in 10 out

of 15 subjects during hand closing preparation and in 11 out of 15 subject in hand opening preparation. Nine out of 15 subjects exhibited the same spatial pattern in HC as well as in HO motor planning.

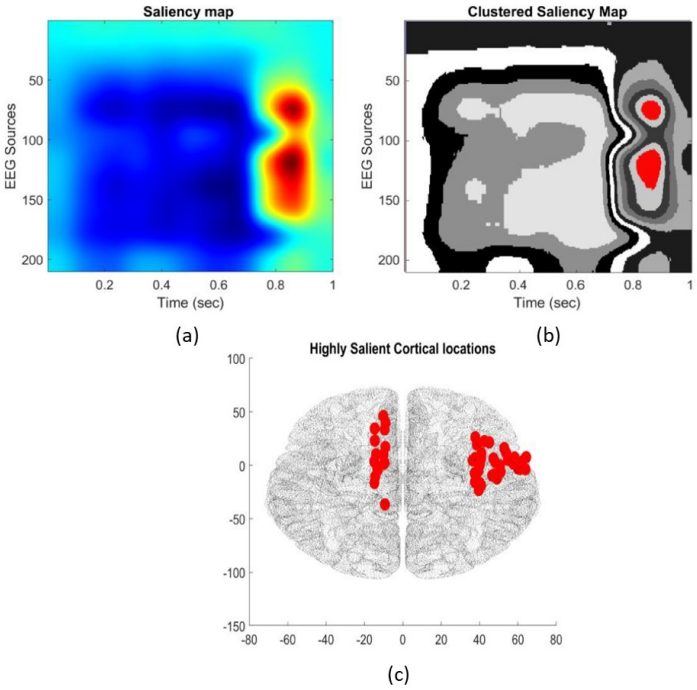

Fig. 4 (a) Saliency map of Subject 08 while planning to perform HC task, achieved by occlusion sensitivity analysis. (b) Clustered Saliency map achieved by k-means technique with k=10. Note that red areas represent the cluster with the highest saliency. (c) Cortical surface representation with highly salient cortical sources depicted with red dots.

6 Discussion

The present research aims at exploring the interpretability and explainability of the proposed DL-based system in order to provide further insight into the hidden mechanism of cortical sources activation when the brain is preparing hand's open/close movement. To this end, the dataset [22] composed of EEG signals recorded from 15 subjects who performed several repetitions of hand's open and close always starting from a common neutral resting position. EEG signals recorded during resting condition (i.e., no motion planning) were also analyzed. First, a dataset of EEG epochs of 1s preceding motion onset was constructed. The onset of motion was determined through the signals collected by motion sensors embedded in a glove that the participant had worn

throughout the experiment. Beamforming was then applied to EEG epochs to solve the inverse problem and reconstruct the electrical sources in the cortical locations belonging to the primary motor and premotor cortex (i.e., 210 cortical locations, as described in Sect.2.2). Next, such premotor EEG source epochs (1s width) were used as input to a customized deep CNN to perform the following binary classifications: HC vs. RE and HO vs. RE, reporting very good discrimination performance: average accuracy rate up to 89.65±5.29% and 90.50±5.35%, respectively. Hence, the temporal trend of electrical sources in the motor cortex allows in principle for motion planning discrimination from resting phases. However, since the ultimate goal of the present study was to provide an in-depth understanding of which cortical locations contributed the most to the discrimination of HC/HO motion planning from resting phase in the frame of 1s preceding the execution of the movement, an occlusion sensitivity analysis was proposed. Specifically, after training and testing the proposed CNN, EEG time-source epochs were systematically occluded with a grey mask and used as input to the pre-trained deep CNN, producing the so called *heatmaps* or *saliency maps*. This technique allowed to highlight which areas of the input map (i.e., EEG time-sources epochs) relevant to the decision making process. In order to detect the high saliency regions in the map, k-means clustering technique was applied and the high-saliency cluster was identified as described in Sect. 5.2. By detecting the high saliency areas in the time-sources maps, the corresponding highly relevant source locations could be pinpointed. The more relevant sources were then mapped onto the cortical surface and represented with red dots. As can been seen in Figure 6, a recurring pattern can be detected in each subject. Specifically, the cortical sources located in the central area of the motor cortex (close to the longitudinal fissure) and the temporal zone of the right motor cortex resulted highly relevant during HC/HO movement's planning preparation. It is to be noted that, to date, it is still not well known whether and how motion planning is spatially organized over the motor cortex. A contralateral involvement of the premotor cortex in motion planning was reported in the literature [17] but further details about sub-areas involvement are still to be investigated. Hence, our findings may shed a new light on motor preparation and suggest that the aforementioned motor cortical regions (i.e., central and temporal right) are the mostly involved in the HC/HO sub-movement preparation. It is also worth noting that intra-subject differences can be observed. For example, in Subjects 01 and 07 (Figure 6), only the right-central sub-region resulted highly relevant to HC detection; whereas, the left-central and also the right-temporal sub-regions look involved in HO detection. To the best of our knowledge, this is the first attempt to study motor preparation through explainable machine learning. Furthermore, this is the first work that attempts to detect the subareas of motor cortex that are more salient to the preparation open/close hand's movement. Recurrent spatial patterns of cortical activation could be detected across subjects, namely, the central area close to the longitudinal fissure and the right temporal area of the premotor and primary motor cortex. However, the proposed methodology has some limitations. First, the number of EEG

channels used in this study was 61, we think that using a higher number of electrodes would have a positive impact on inverse problem solution, leading to a more accurate cortical source reconstruction. Second, the number of EEG epochs used to train the proposed CNN was limited. Overall, each class (HC, HO, RE) included only 60 EEG epochs. Third, movement's onset was marked by processing the data collected by the motion sensors embedded in the glove that the participant used to wear during the experiment. Motion data collected through the glove are smooth and do not allow to detect onset instantaneously, which means the epochs used for training may have captured the early ms of motion implementation, causing, in principle, the similar activation patterns visible in HC and HO (Figure 6). For the aforementioned reasons, in the future, we intend not only to enroll a larger cohort of subject and record high-density EEGs (128-256 channels) but, for a more precise motion onset detection, EEG will be co-registered with electromiography (EMG). In addition, the analysis of EEG data is always a spatio-temporal process that is first related to the spiking activities of cortical circuits, i.e. by individual neurons cooperating to a task. This process has also a spectral component that is superimposed in the global approach proposed in [48], the NeuCube computational architecture based on Brain-Inspired approaches. The concept of the Spiking Neural Networks (SNN) is at the basis of the complex model that allows for on-going learning and classification over time [49,50]. NeuCube allows for generating a deep unsupervised learning spatio-temporal spike sequences in a scalable 3D SNN reservoir. This can be relevant for adaptation to new data, possibly in real-time modality, which is one of the future objectives also for the architecture here proposed.

7 Conclusion

In this paper we proposed a novel deep CNN capable of classifying time-source maps (i.e., EEG sources epochs) related to hands' sub-movements (open/close) phase from resting state, achieving remarkable results, namely, average accuracy of 89.65±5.29% and 90.50±5.35% in HC vs. RE and HO vs. RE discrimination task, respectively. Furthermore, in order to investigate which cortical source has mostly contributed in the classification of hand's motor preparation phase, EML was applied. Occlusion sensitivity analysis allowed to produce suitable *saliency maps*, from which to identify the most relevant areas of the input. The highest saliency region was detected though k-means clustering technique and the enclosed cortical sources were mapped onto the cortical surface. Experimental results mainly showed that the central and the right-temporal cortical sub-regions are activating while the subject was planning hand's movements (i.e., HC/HO). It is to be noted that the cortical activation rules that govern the motion planning are still not well known. Hence, on the basis of the achievements here reported, we believe that the proposed approach may be considered to be an interesting breakthrough in BCI applications.

Table 1 Performance of the proposed deep CNN in terms of recall, precision, F1-score and accuracy for the HC vs. RE classification task.

Subject	Recall [%]	Precision [%]	F1-score [%]	Accuracy [%]
S01	86.51±7.76	89.45±11.45	87.36±6.16	87.30±6.58
S02	93.65±5.00	95.92±6.97	94.46±2.00	94.44±2.27
S03	87.30±9.47	91.96±13.10	88.67±7.36	88.49±8.70
S04	92.06±6.30	93.67±7.89	92.48±3.33	92.46±3.48
S05	89.68±8.13	81.42±6.57	84.96±4.00	84.13±4.45
S06	88.10±8.13	90.69±6.58	89.09±5.51	89.29±5.18
S07	80.16±12.77	80.83±12.36	79.86±9.60	79.76±10.11
S08	96.03±4.20	100.00±0.00	97.94±2.21	98.02±2.10
S09	96.83±4.37	98.41±2.73	97.57±3.05	97.62±2.97
S10	96.83±4.37	97.86±5.67	97.25±4.09	97.22±4.24
S11	81.75±14.24	86.29±15.31	82.29±8.76	82.14±10.25
S12	90.48±4.20	91.76±8.15	90.95±5.04	90.87±5.25
S13	87.30±8.91	90.99±9.62	88.72±7.06	88.89±7.17
S14	83.33±5.56	89.37±5.94	86.05±3.50	86.51±3.37
S15	87.14±5.24	89.18±5.87	87.68±2.97	87.65±3.29
Average	89.14±7.24	91.19±7.88	89.69±4.98	89.65±5.29

Table 2 Performance of the proposed deep CNN in terms of recall, precision, F1-score and accuracy for the HO vs. RE classification task.

Subject	Recall [%]	Precision [%]	F1-score [%]	Accuracy [%]
S01	84.13±14.50	96.10±8.13	88.62±8.14	89.68±6.15
S02	92.86±6.18	99.25±1.99	95.82±2.97	96.03±2.71
S03	85.71±16.31	94.24±8.11	88.57±10.49	89.68±7.81
S04	89.68±5.00	100.00±0.00	94.50±2.80	94.84±2.50
S05	92.86±7.67	91.52±9.13	91.83±6.10	91.67±6.42
S06	96.03±4.20	87.63±11.93	91.05±5.50	90.08±6.96
S07	77.78±11.11	80.13±18.01	76.95±5.87	75.79±9.72
S08	94.44±4.54	97.86±5.67	96.00±3.75	96.03±3.88
S09	90.48±8.91	97.74±5.97	93.70±6.11	94.05±5.65
S10	92.06±5.42	97.02±5.73	94.32±3.91	94.44±3.93
S11	76.98±12.18	94.62±9.21	84.04±7.29	85.71±6.09
S12	95.24±5.94	83.83±11.60	88.53±4.83	87.30±5.97
S13	88.89±8.49	94.39±6.71	91.35±5.97	91.67±5.78
S14	95.24±3.83	90.23±7.10	92.42±2.79	92.06±3.37
S15	87.31±6.01	91.03±5.66	88.40±3.32	88.50±3.35
Average	89.31±8.02	93.04±7.66	90.41±5.32	90.50±5.35

8 Declarations

8.1 Funding

This work was co-funded by the European Commission, the European Social Fund and the Calabria Region (code: C39B18000080002). The authors are the only responsible for this publication and the European Commission and the Calabria Region decline any responsibility for the use that may be made of the information in it held. This work was also supported by the UK Engi-

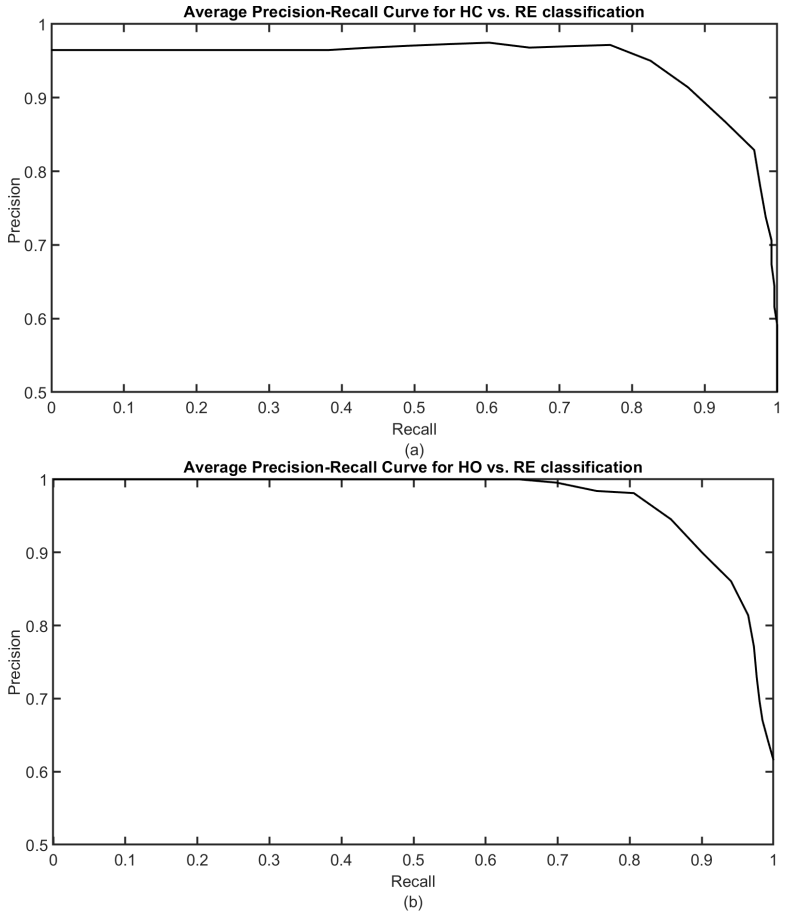

Fig. 5 Average precision-recall curve for (a) HC vs. RE and (b) HO vs. RE classification task.

neering and Physical Sciences Research Council (EPSRC) (EP/M026981/1, EP/T021063/1, EP/T024917/1).

8.2 Conflict of interest

The authors declare that they have no conflict of interest.

8.3 Availability of data and material

Data are publicly available from the BNCI Horizon 2020 database at http://bnci-horizon-2020.eu/database/data-sets (accession number 001-2017).

8.4 Code availability

Custom code will be available on request to the corresponding author.

References

1. P. L. Nunez and R. Srinivasan. *Electric fields of the brain, The neurophysics of* EEG. New York, Oxford University Press, 2006.
2. Zohreh Gholami Doborjeh, Maryam G Doborjeh, and Nikola Kasabov. Attentional bias pattern recognition in spiking neural networks from spatio-temporal EEG data. *Cognitive Computation*, 10(1):35–48, 2018.
3. Ateke Goshvarpour and Atefeh Goshvarpour. A novel approach for EEG electrode selection in automated emotion recognition based on lagged poincare's indices and sloreta. *Cognitive Computation*, pages 1–17, 2019.
4. Cosimo Ieracitano, Nadia Mammone, Amir Hussain, and Francesco C Morabito. A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. *Neural Networks*, 123:176–190, 2020.
5. Cosimo Ieracitano, Nadia Mammone, Alessia Bramanti, Amir Hussain, and Francesco C Morabito. A convolutional neural network approach for classification of dementia stages based on 2d-spectral representation of EEG recordings. *Neurocomputing*, 323:96–107, 2019.
6. F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3):031005, 2018.
7. N. Mammone and F. C. Morabito. Enhanced Automatic Wavelet Independent Component Analysis for Electroencephalographic Artifact Removal. *Entropy*, 16(12):6553–6572, 2014.
8. H. Zeng and A. Song. Optimizing single-trial EEG classification by stationary matrix logistic regression in brain–computer interface. *IEEE transactions on neural networks and learning systems*, 27(11):2301–2313, 2016.
9. J. Malmivuo and R. Plonsey. *Bioelectromagnetism, Principles and Applications of Bioelectric and Biomagnetic Fields.* New York, Oxford, Oxford University Press, 1995.
10. Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, 16(5):051001, 2019.
11. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.
12. Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering*, 16(3):031001, 2019.
13. Fabien Lotte and Camille Jeunet. Online classification accuracy is a poor metric to study mental imagery-based BCI user learning: an experimental demonstration and new metrics. In *7th International BCI Conference*, pages hal–01519478, 2017.
14. G. R. Müller-Putz, A. Schwarz, J. Pereira, and P. Ofner. From classic motor imagery to complex movement intention decoding: The noninvasive Graz-BCI approach. In *Progress in brain research*, volume 228, pages 39–70. Elsevier, 2016.
15. Giovanni Vecchiato, Maria Del Vecchio, Luca Ascari, Sergey Antopolskiy, Fabio Deon, Luca Kubin, Jonas Ambeck-Madsen, Giacomo Rizzolatti, and Pietro Avanzini. Electroencephalographic time-frequency patterns of braking and acceleration movement preparation in car driving simulation. *Brain research*, 1716:16–26, 2019.
16. Patrick Ofner, Andreas Schwarz, Joana Pereira, Daniela Wyss, Renate Wildburger, and Gernot R Müller-Putz. Attempted arm and hand movements can be decoded from low-frequency EEG from persons with spinal cord injury. *Scientific reports*, 9(1):7134, 2019.
17. Takuro Zama and Sotaro Shimada. Simultaneous measurement of electroencephalography and near-infrared spectroscopy during voluntary motor preparation. *Scientific reports*, 5(1):1–9, 2015.
18. Nadia Mammone, Cosimo Ieracitano, and Francesco C Morabito. A deep cnn approach to decode motor preparation of upper limbs from time–frequency maps of EEG signals at source level. *Neural Networks*, 124:357–372, 2020.
19. Grégoire Montavon, Wojciech Samek, and Klaus-Robert Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15, 2018.

20. Marco Catani. A little man of some importance. *Brain*, 140(11):3055–3061, 2017.
21. Jean-Marc Fellous, Guillermo Sapiro, Andrew Rossi, Helen S Mayberg, and Michele Ferrante. Explainable artificial intelligence for neuroscience: Behavioral neurostimulation. *Frontiers in Neuroscience*, 13:1346, 2019.
22. P. Ofner, A. Schwarz, J. Pereira, and G. R. Müller-Putz. Upper limb movements can be decoded from the time-domain of low-frequency EEG. *PloS one*, 12(8):e0182578, 2017.
23. Aqsa Shakeel, Muhammad Samran Navid, Muhammad Nabeel Anwar, Suleman Mazhar, Mads Jochumsen, and Imran Khan Niazi. A review of techniques for detection of movement intention using movement-related cortical potentials. *Computational and mathematical methods in medicine*, page 346217, 2015.
24. Roberta Grech, Tracey Cassar, Joseph Muscat, Kenneth P Camilleri, Simon G Fabri, Michalis Zervakis, Petros Xanthopoulos, Vangelis Sakkalis, and Bart Vanrumste. Review on solving the inverse problem in EEG source analysis. *Journal of neuroengineering and rehabilitation*, 5(1):25, 2008.
25. Hans Hallez, Bart Vanrumste, Roberta Grech, Joseph Muscat, Wim De Clercq, Anneleen Vergult, Yves D'Asseler, Kenneth P Camilleri, Simon G Fabri, Sabine Van Huffel, et al. Review on solving the forward problem in EEG source analysis. *Journal of neuroengineering and rehabilitation*, 4(1):46, 2007.
26. B. J. Edelman, B. Baxter, and B. He. EEG source imaging enhances the decoding of complex right-hand motor imagery tasks. *IEEE Transactions on Biomedical Engineering*, 63(1):4–14, 2015.
27. Stefan Haufe and Arne Ewald. A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies. *Brain topography*, pages 1–18, 2016.
28. Stefan Haufe, Yu Huang, and Lucas C Parra. A highly detailed FEM volume conductor model based on the ICBM152 average head template for EEG source imaging and TCS targeting. In *Conf Proc IEEE Eng Med Biol Soc*, 2015.
29. François Tadel, Sylvain Baillet, John C Mosher, Dimitrios Pantazis, and Richard M Leahy. Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational intelligence and neuroscience*, page 8, 2011.
30. Moritz Grosse-Wentrup, Christian Liefhold, Klaus Gramann, and Martin Buss. Beamforming in noninvasive brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 56(4):1209–1219, 2009.
31. Dora Hermes, Mariska J Vansteensel, Anke Marit Albers, Martin G Bleichner, Marije R Benedictus, C Mendez Orellana, EJ Aarnoutse, and NF Ramsey. Functional MRI-based identification of brain areas involved in motor imagery for implantable brain–computer interfaces. *Journal of Neural Engineering*, 8(2):025007, 2011.
32. Shailesh S Kantak, James W Stinear, Ethan R Buch, and Leonardo G Cohen. Rewiring the brain: potential role of the premotor cortex in motor control, learning, and recovery of function following brain injury. *Neurorehabilitation and neural repair*, 26(3):282–292, 2012.
33. Korbinian Brodmann. *Brodmann's: Localisation in the cerebral cortex*. Springer Science & Business Media, (2007).
34. Jack L Lancaster, Diana Tordesillas-Gutiérrez, Michael Martinez, Felipe Salinas, Alan Evans, Karl Zilles, John C Mazziotta, and Peter T Fox. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human brain mapping*, 28(11):1194–1205, 2007.
35. Neucube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural Networks*, 52:62 – 76, 2014.
36. Liangliang Liu, Shaowu Chen, Fuhao Zhang, Fang-Xiang Wu, Yi Pan, and Jianxin Wang. Deep convolutional neural network for automatically segmenting acute ischemic stroke lesion in multi-modality mri. *Neural Computing and Applications*, pages 1–14, 2019.
37. Wei Tang, Dongsheng Zou, Su Yang, Jing Shi, Jingpei Dan, and Guowu Song. A two-stage approach for automatic liver segmentation with faster r-cnn and deeplab. *Neural Computing and Applications*, pages 1–10, 2020.
38. Fei Gao, Teng Huang, Jinping Sun, Jun Wang, Amir Hussain, and Erfu Yang. A new algorithm for sar image target recognition based on an improved deep convolutional neural network. *Cognitive Computation*, 11(6):809–824, 2019.

39. Zhenwen Liang, Jie Shao, Dongyang Zhang, and Lianli Gao. Traffic sign detection and recognition based on pyramidal convolutional networks. *Neural Computing and Applications*, pages 1–11, 2019.
40. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
41. Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.
42. Dominik Scherer, Andreas Müller, and Sven Behnke. Evaluation of pooling operations in convolutional architectures for object recognition. In *International conference on artificial neural networks*, pages 92–101. Springer, 2010.
43. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
44. Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. *in: Neural networks: Tricks of the trade*, pages 437–478, 2012.
45. Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.
46. Kiri Wagstaff, Claire Cardie, Seth Rogers, Stefan Schrödl, et al. Constrained k-means clustering with background knowledge. In *Icml*, volume 1, pages 577–584, 2001.
47. David Martin Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.
48. Nikola K Kasabov. *Time-space, spiking neural networks and brain-inspired artificial intelligence*. Springer, 2019.
49. Alessandro Vato, Laura Bonzano, Michela Chiappalone, S Cicero, F Morabito, Antonio Novellino, and G Stillo. Spike manager: a new tool for spontaneous and evoked neuronal networks activity characterization. *Neurocomputing*, 58:1153–1161, 2004.
50. Nikola K Kasabov. Neucube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural Networks*, 52:62–76, 2014.

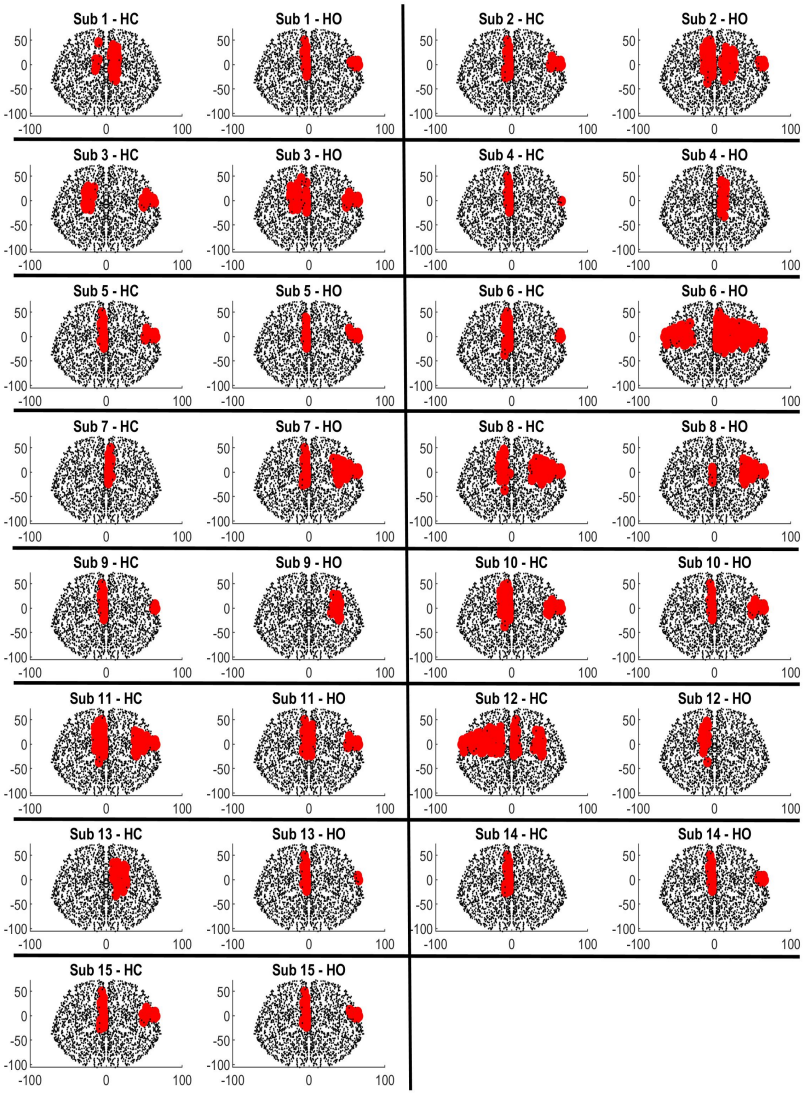

Fig. 6 Cortical surface representations of each subject. Red dots indicate the highly salient cortical sources locations that have mostly involved in the prediction task (i.e. HC/HO).