International Journal of Neural Systems, Vol. 0, No. 0 (2005) 1–??

A HYBRID-DOMAIN DEEP LEARNING-BASED BCI FOR DISCRIMINATING HAND MOTION PLANNING FROM EEG SOURCES

COSIMO IERACITANO
DICEAM, University Mediterranea of Reggio Calabria,
Via Graziella Feo di Vito,
Reggio Calabria, 89124, Italy
E-mail: cosimo.ieracitano@unirc.it

FRANCESCO CARLO MORABITO
DICEAM, University Mediterranea of Reggio Calabria,
Via Graziella Feo di Vito,
Reggio Calabria, 89124, Italy
E-mail: morabito@unirc.it

AMIR HUSSAIN
School of Computing, Edinburgh Napier University
Edinburgh EH10 5DT, Scotland, U.K.,
E-mail: a.hussain@napier.ac.uk

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

In this paper, a hybrid-domain deep learning (DL) based neural system is proposed to decode hand movement preparation phases from electroencephalographic (EEG) recordings. The system exploits information extracted from the temporal-domain and time-frequency-domain, as part of a hybrid strategy, to discriminate the temporal windows (i.e., EEG epochs) preceding hand sub-movements (open/close) and the resting state. To this end, for each EEG epoch, the associated cortical source signals in the motor cortex and the corresponding time-frequency (TF) maps are estimated via beamforming and Continuous Wavelet Transform (CWT), respectively. Two Convolutional Neural Networks (CNN) are designed: specifically, the first CNN is trained over a dataset of temporal (T) data (i.e., EEG sources), and is referred to as T-CNN; the second CNN is trained over a dataset of TF data (i.e., TF-maps of EEG sources), and is referred to as TF-CNN. Two sets of features denoted as T-features and TF-features, extracted from T-CNN and TF-CNN respectively, are concatenated in a single features vector (denoted as TTF-features vector) which is used as input to a standard multi-layer perceptron for classification purposes. Experimental results show a significant performance improvement of our proposed hybrid-domain DL approach as compared to temporal-only and time-frequency-only based benchmark approaches, achieving an average accuracy of 76.21 ± 3.77%.

Keywords: Deep Learning; Brain Computer Interface; Electroencephalography; Beamforming; Wavelet Transform; Feature Fusion.

1. Introduction

The present work deals with decoding hand motion planning from electroencephalographic (EEG) signals. The aim is to assess if EEG holds enough information to reveal, ever since the motion planning phase preceding motion onset, which hand movement the subject is going to perform. In this context, works in the literature are usually focused on Motor Imagery (MI)[1] (which consists in investigating the brain potentials generated by the imagination of movements) or Motor Execution (ME)[2] (which consists in finding cortical representation of different movements). Motion preparation is far less investigated than motor execution or imagery, although it may be the keystone to achieving a natural Brain Computer Interface (BCI) control paradigm. In the state-of-the-art literature related to EEG-based motor imagery decoding systems, most papers are focused on right vs. left hand MI.[3] Right and left hand motor imagery EEG signals can indeed be somehow reliably discriminated, thanks to the asimmetry in the sensorimotor rhythms generated by motion planning, which is due to the contralaterality in motion control by the two brain hemispheres.[4] The current rehabilitation systems are based on right hand vs. left hand vs. feet sustained motor imagery and rely on the stimulus to neural plasticity that derives from watching the desired movement implemented virtually, by means of an avatar, or really, by means of an external device like a prosthetic arm.[5] Unfortunately, sustained movement imagination is an unnatural mechanism that requires prolonged and intensive user's training. If the desired movement could be decoded ever since its motion preparation phase, a motor impaired subject could only need to attempt to move his/her hand in order to activate the device that is in charge of performing the desired action, with no need of any sustained motor imagery. Motion attempt is a topic related to motor execution, because it deals with the motion planning phase of motor execution. In the state-of-the-art literature related to EEG-based motor execution decoding, most works dealing with the analysis of motion from EEG signals for BCI applications has focused on drawing information about the ongoing movement, which seems to be to some extent possible by analyzing lower frequencies (<3 Hz).[6–8] In particular, reaching directions/targets[9,10] appear decodable from Motor Related Cortical Potentials (MRCP), brain potentials associated with voluntary movements.[11,12] Researchers have hypothesized that different sub-movements of the same hand, such as open, hand close, palmar and lateral grasp,[13,14] are embedded in MRCP, which could be detected by an EEG-based BCI and transformed into the control input to a neuroprosthesis or a computer application. Sensorimotor rhythms (SMR), brain waves detectable in the central area of the scalp in the range 13-15 Hz, and beta waves (nearly 13-40 Hz) are also considered relevant to movement analysis.[12] In contrast, little is known regarding the possibility to decode motor preparation of sub-movements of the same hand,[5] which is a topic of paramount importance as a reliable decoding of motor planning would allow for a more prompt and friendly way to interface the brain with the machine. Furthermore, being able to decode different sub-movements of the same hand would empower the current EEG-based BCI systems used for rehabilitation (mainly based on right vs. left hand movement intention discrimination) with possible benefit to neural plasticity. This would also lead to a more natural control paradigm. In conclusion, this work aims to respond to the following questions: Can motor preparation be reliably drawn from EEG signals before motion onset? Can the intention to perform different sub-movements of the hand be drawn from EEG signals? Papers in the literature have provided some answers to the first question,[2] whereas the second one is still mostly unanswered except in Ref.[13] The proposed work aims to contribute to the development of a system that is able to detect the preparation a movement and to decode which specific hand sub-movement the brain is planning. To this end, a dataset of EEG recordings preceding hand motion onset was analyzed. Herein, by "hand motion planning" the paper will refer to the preparation of hand opening/closing movements. Many more complex movements like, for example, finger flexion/extension or palmar/lateral grasp, are based on the simple, basic, task of hand opening/closing which involves all the palmar muscles of the hand. Among the possible hand sub-movements, open and close are indeed of basic relevance.[5] Since EEG signals related to hand motion analysis appear to hold relevant properties both in the time (T) and in the time-frequency (TF) domains,[12] a *hybrid-domain* strategy can be developed that in principle could outperform the two single domain approaches. Specifi-

cally, as Deep Learning (DL[15]) achieved relevant results in several research fields,[16–19] we propose the use of a *hybrid-domain* DL neural architecture capable of combining features extracted from T and TF domains, for improving hand's movement planning classification. Two customized Convolutional Neural Networks (CNN) are developed to extract T and TF features from EEG sources' time series (estimated by solving the inverse problem through beamforming) and from the corresponding TF maps (estimated via Continuous Wavelet Tranform, CWT). The extracted T and TF features are then fused together and used as input to a standard multi-layer neural network to perform the ultimate 3-way classification task: pre-hand close (HC) vs. pre-hand open (HO) vs. rest (RE). Experimental results show that the proposed hybrid DL-based approach (hereby denoted as TTF-NET) outperforms the two single-domain systems (T-NET and TF-NET). As a consequence, the fusion of features extracted from different domains appears promising for the investigation of motor planning based on EEG.

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

(i) Development of an innovative hybrid-domain DL-based approach capable of discriminating EEG segments preceding the execution of hand's submovements (open/close) and EEG resting segments
(ii) Development of a CNN-based system able to automatically extract the most significant features from time (i.e., EEG sources' time series) and time-frequency (i.e., TF maps) domain
(iii) Development of a hybrid-domain DL-based system with potential for deployment in BCI applications

The manuscript is organized as follows: in Section 2, first, the EEG dataset used in the present research is described. Then, the proposed methodology (including premotor EEG epoch selection, extraction of EEG sources in the motor cortex via beamforming, the CWT analysis and the proposed DL-based open/close hand movement planning detection features fusion framework) is introduced. Section 3 reports the achieved results, then discussed in Section 4. Section 5 concludes the paper and addresses some future research perspectives.

2. Materials and Methodology

2.1. *Dataset description*

In this work, the publicly available database provided by Ofner et al.[13] was adopted. This database consists of EEGs recordings co-registered with signals collected from motion sensors (a glove and an exoskeleton). To the best of our knowledge, this is the very first public collection of high density EEGs co-registered with motion data during the execution of BCI experiments of motor execution and imagery. The dataset can be downloaded from the website BNCI Horizon 2020, all the details about how data can be downloaded and how they were acquired, as well as information regarding the protocol approved by the Ethics Committee, are provided in Ref.[13] The dataset includes 15 healthy subjects (aged 27±5 years, nine females). All of them, except S01, are right-handed. EEG was recorded by means of 61 active electrodes and four 16-channel amplifiers (g.tec medical engineering GmbH, Austria), electrodes are embedded in a cap that ensures correct positioning. Reference channel is placed on the right mastoid and the ground is placed on AFz. Further details about channels' montage can be found in Ref.[13] Since the present study is focused on motor preparation, the motor execution part of the database was selected. The protocol[13] consists in executing cue-based movements of the right upper limb starting from a common neutral position (lower arm extended to 120 degree and in a neutral rotation, hand half open). Throughout the experiment, subjects had been sitting on a comfortable chair with their right arm supported by an anti-gravity exoskeleton (Hocoma, Switzerland) to prevent muscle fatigue. Subjects had a computer screen in front of them, displaying the cues of the paradigm (movements to be performed). Every session consisted of 10 runs. Every run included 6 trials, every trial included one hand open, one hand close and one rest cues. The paradigm followed by Ofner et al.[13] is structured as follows: at second 0, a beep sounded and a fixation cross appeared on the computer screen (subjects were instructed to focus their gaze on it to avoid eye movements). At second 2, a cue appeared on the computer screen, showing the task the subject was required to perform in the next seconds. At the end of task execution, the subject moved her/his hand back to the starting neutral position.

2.2. *Methodology*

Fig. 1 illustrates the procedure of the proposed methodology framework. Notably, it includes the following stages:

(i) *Premotor EEG epochs selection.* Temporal windows (i.e., epochs) of 1s preceding hand motion onset and resting epochs of the same duration are extracted from the available 61-channels EEG recordings.

(ii) *Beamforming and extraction of EEG sources in the motor cortex.* Given the e^{th} EEG epoch, inverse problem is solved through beamforming to reconstruct cortical electrical sources (there are 2000 possible source locations in the model adopted in the present work). Next, source locations belonging to the motor cortex, Brodmann areas (BA) 4 and 6, are selected (210 source locations out of 2000).

(iii) *Time-Frequency Analysis.* For every epoch, the 210 EEG sources are analyzed through CWT and hence, for each source signal, the corresponding time-frequency map is estimated. The resulting 210 TF maps are then stratified in a single 3D matrix (time x frequency x source location).

(iv) *Temporal-only (T) domain based classification.* The 210-EEG source signals (corresponding to the e^{th} EEG epoch) is the input to the first CNN (here denoted as T-CNN), developed to extract the T-features. A standard NN (referred as T-NN) follows the T-CNN and performs the classification based on T features only. It is to be noted that the sub-system (T-CNN + T-NN) performs the time-only domain based classification and is referred as T-NET.

(v) *Time-frequency-only (TF) domain based classification.* The 3D-TF maps (estimated from the 210-EEG source signals and corresponding to the e^{th} EEG epoch) is the input to the second CNN (here denoted as TF-CNN) developed to extract TF-features. Similarly, a standard NN (referred as TF-NN) follows TF-CNN and performs the classification based on TF features only. The sub-system (TF-CNN + TF-NN) performs the time-frequency-only domain based classification and is referred as TF-NET.

(vi) *Hybrid TTF domain based classification.* T-features and TF-features, extracted from T-CNN and TF-CNN respectively, are passed to a features fusion layer. The resulting vector, here denoted as *TTF-features fusion vector*, is the input to a multi-layer TTF-NN that performs the ultimate 3-way classification task. The architecture composed of the two CNNs, together with the features fusion layer and the TTF-NN classifier, performs our proposed hybrid TTF-domain DL based classification approach, referred as TTF-NET.

2.3. *Premotor EEG epochs selection*

Signals' frames prior to hand's open/close motion onset were extracted and included in the analysis together with resting frames of equal duration. In particular, 900 (= 10 runs x 6 trials x 15 subjects) EEG epochs per movement class (hand open/close) were selected. In order to come up with a balanced dataset, a comparable number of resting state EEG epochs was included in the analysis. Overall, N_e=2700 (= 10 runs x 6 trials x 3 classes x 15 subjects) EEG epochs were analyzed: 1800 of them were pre-motion epochs, 900 were resting epochs preceding no motion. The exact timing of motion onset was derived from the signals collected from motion sensors, following the procedure described in Ref.[13] In order to assess that motion onset was correctly detected, the marked timing was visually checked for all of the 1800 pre-motion epochs under examination. The choice of the length of frames preceding motion onset was made by taking into account the theory of MRCPs, brain potentials that come with planning and initiation of movements.[12] A MRCP consists of three sub-potentials: the readiness potential (RP) or Bereitschaft potential (BP, related to movement preparation), the motor potential (MP, related to movement execution) and the movement-monitoring potential (MMP, related to performance control).[12] Since BP arises around 1.5 to 1s prior to movement onset, the length of the selected EEG frames (hereby named "epochs") preceding motion onset was set at 1s. The analyses described in Section 2.4 and Section 2.5, were carried out epoch by epoch.

2.4. *Time-domain analysis: extraction of EEG sources in the motor cortex*

A single dipole, located in the cortex, models the activity of a large group of simultaneously active neu-

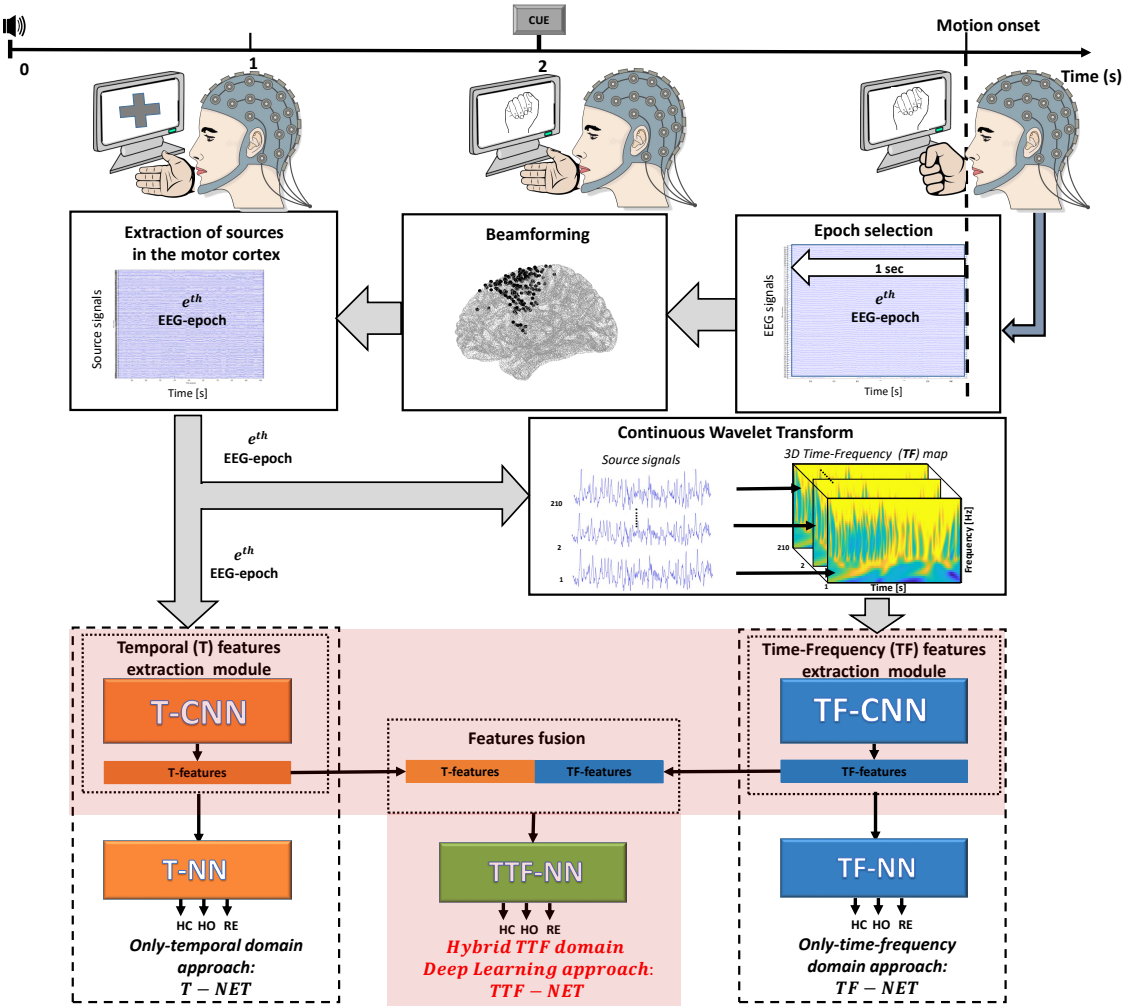


Figure 1: Hybrid-domain deep learning approach. First, the paradigm followed by Ofner et al.[13] is depicted. Next, the EEG epoch of 1s preceding motion onset is selected. Given the e^{th} EEG-epoch, inverse problem is solved (through beamforming) and EEG source signals are reconstructed. Specifically, 210 source signals related to motor cortex (Brodmann's Areas 4 and 6) are selected. Finally, the extracted EEG source signals are projected to the time-frequency (TF) domain and embedded into a volume (time x frequency x source) of TF maps. EEG sources' time series and TF maps are passed as input to two deep CNNs-based extraction modules, denoted as T-CNN and TF-CNN, that extract the most relevant features from both T and TF domain. T-features and TF-features are are fused together and fed into a standard multi-layer NN (denoted as TTF-NN). The architecture highlighted in red (T-CNN, TF-CNN, features fusion, TTF-NN) represents the proposed hybrid-domain deep learning model (TTF-NET).

rons.[20,21] Unfortunately, because of volume conduction effects, a single EEG channel will not be able to selectively collect the activity of a limited cortical region underneath the sensor, but the activity of relatively far away regions will be detected as well. This results in a poor spatial resolution of EEG signals. In order to deal with volume conduction, the proposed methodology relies on the projection of EEG signals into the cortex,[22] which is of great importance as our goal is to selectively analyze the electrical activity of a specific part of the brain: the motor cortex. The cortical electrical activity can be estimated from the analysis of the scalp one, provided that a model of fields' propagation through cerebral tissues is de-

fined to reconstruct the cortical electrical sources. Reconstructing sources consists in solving the "inverse problem" starting from a "forward model" definition. Solving the *forward problem* consists in calculating the scalp potentials resulting from the propagation of the fields generated by the current dipoles located in the cortex. Sources are projected into the scalp through the forward model, which takes into account the structural and conductive properties of the cerebral tissues, that get involved in fields' propagation. Considered that the capacitive component of tissue impedance can be discarded in the frequency range of bioelectric signals generated by the body, the quasi-static approximation of Maxwell's equations can be adopted thus the forward model can be assumed linear[23] and can be formulated in the following way:

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

where $\mathbf{q}_r$(t) is a 3 dimensional directed current dipole associated to location "r" (with r=1,...,Ns, where Ns is the number of possible source locations in the cortex); $\mathbf{L}$ is knows as "lead field" matrix, which represents the head model and projects the current dipole $\mathbf{q}_r$(t) to the scalp potential $\mathbf{x}$(t).[23] Since the number of cortical locations Ns (2000 in this work) is much larger than the number of available scalp channels Nc (61 in this work), the estimation of $\mathbf{q}_r$(t) from $\mathbf{x}$(t) is an ill posed problem and infinite solutions exist. Solving the inverse problem allows to estimate cortical sources $\mathbf{q}_r$(t) given the collected signals $\mathbf{x}$(t) and a lead field matrix $\mathbf{L}$. In this work, the finite element model (FEM) of the cortical surface with 2000 nodes (i.e., source locations) and the lead field $\mathbf{L}$ generated by Haufe et al.[23,24] was adopted. Furthermore, the New York Head (NYH) forward model, introduced in Ref,[24] was used. Once the forward model is defined, the inverse problem is solved starting from EEG signals collected at the scalp by beamforming technique. The advantages of using beamforming in EEG-based BCI applications were demonstrated by Grosse-Wentrup et al.[25] In the present paper, linearly constrained minimum variance (LCMV) beamformers were employed to reconstruct source activity at each of the 2000 cortical locations.[23] Beamformers assume the EEG collected at a given sensor is a linear combination of components originating from different source locations. Each component is estimated by maximizing the contribution of one of the sources while suppressing contributions originating from the other source locations. Here, the BA of every source location (node) of the model was determined starting from its Montreal Neurological Institute (MNI) stereotaxic coordinates. MNI coordinates of the 2000 nodes were first converted into Talairach coordinates through the algorithm developed by Lancaster et al.[26] After coordinates' conversion, *Talairach Daemon* software[27,28] was used to match the Tailarach coordinates with Talairach Atlas labels, in order to estimate the BA of every source location. Finally, the 210 locations belonging to BAs 4 and 6 were selected.

2.5. *Time-Frequency Analysis of EEG sources*

The electrical activity of the 210 source locations of interest were then analyzed through Continuous Wavelet Transform (CWT),[29] in order to investigate the behaviour of the motor cortex in the phase of motor planning in the time-frequency domain. Given an epoch preceding hand open/close motion or a resting epoch of the dataset under analysis, source signals in the motor cortex were estimated according to Section 2.4. Every source signal was then passed through CWT in order to project it into the time-frequency domain and construct a TF map. Given a source signals x, CWT is defined as:

$$CWT(s,\tau) = \frac{1}{\sqrt{s}} \int x(t)\psi^*(\frac{t-\tau}{s}) \ . \tag{2}$$

where $CWT(s,\tau)$ is the wavelet coefficient associated to scale s and shift τ, the symbol * represents the complex conjugate operator,[29] ψ is the wavelet mother function (a template basis function with zero mean, finite duration and variable frequency content). The two variables s and τ determine how much the *wavelet mother* is dilated (scale s) and shifted (translation τ). In this work, *db4* was selected as mother wavelet as it allowed to cover the range of interest (0.6-45 Hz) with a uniformly distributed vector of frequencies. The range of interest was set between 0.6Hz and 45Hz because it covers the motor related cortical potentials (<5 Hz),[30] the sensory motor rhythms (nearly 13-15 Hz) and the beta band (nearly 13-30 Hz), which are of paramount importance in the analysis of neural correlates of

movements,[12] while mitigating the effects of muscular artifacts (20-300 Hz).[31] Given a mother wavelet function, the scale s is inversely proportional to frequency. Given a set of scale values, the related set of frequencies depends on the mother wavelet. Once the mother wavelet is selected (*db4* was selected in this work), it is therefore necessary to identify the most suitable set of scales to represent the frequency range of interest. As addressed above, the aim of the present work was to process EEG segments preceding motion onset in order to extract Time-Frequency maps representative of the frequencies in the range 0.6-45 Hz, as uniformly as possible and to give equal emphasis to the various sub-rhythms involved in movement preparation (MRCP, mu, beta, etc). Given the selected mother wavelet *db*4, the frequency band under consideration (0.6-45Hz), and the sampling rate of 160Hz, we analysed the pseudo-frequencies corresponding to the scales (the pseudo-frequency being equal to the central frequency of the mother wavelet divided by the scale s) by means of the *Matlab* function $scal2freq$. This led to a vector of scales being defined that was suitable to uniformly represent the frequencies in the range under examination, resulting in 43 constituent elements. In the end, for every epoch and for every source location, a **CWT** matrix was calculated. By calculating $|CWT(s,\tau)|^2$ the scalogram is determined and denoted herein as "Time-frequency" (TF) map.

2.5.1. *Comparison of pre-hand motion TF images*

For every subject and every pre-motion epoch, the TF maps of the 210 source signals were extracted as described in Sections 2.4. In order to evaluate the potential of pre-motion TF maps in providing information relevant to the classification of the movement under preparation, a similarity comparison was carried out between pre-hand opening and pre-hand closing TF maps, for every subject and for every source location. Comparing the similarity between the pre-motion TF maps could indeed help to understand whether any difference is detectable in the time-frequency behaviour of the electrical sources in the motor cortex during the pre-hand opening and pre-hand closing phases. The Structural Similarity Index (SSIM) was selected[32] to estimate the similarity between HC-TF and HO-TF images. SSIM was introduced by Wang et al.[33] and quantifies the similarity between an image and a reference one. SSIM=1 means perfect match whereas SSIM=0 means no match. Let x and y be two images to be compared, SSIM index is defined as follows:

$$SSIM(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2\mu_y^2 + C_1} \cdot \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2\sigma_y^2 + C_2} \cdot \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \,. \tag{3}$$

where μ_x and σ_x represent, respectively, the mean and the standard deviation of image x (similarly, μ_y and σ_y for image y), σ_{xy} represents the cross-covariance between x and y. C_1, C_2, C_3 are small positive constants values, introduced to guarantee numerical stability of μ_x,μ_y, σ_x, σ_y, σ_{xy}. C_1, C_2, C_3.[32] For comparing images, it is often useful to apply SSIM index locally rather than globally. Localized quality measurement can indeed provide a similarity map of the two images and hence provide information about where in the images they differ most from each other. According to Ref.,[32] local statistic are computed within a circular-symmetric Gaussian weighting function, SSIM index is then calculated within the local window in this way producing a *SSIM map*.

2.6. *Hybrid-domain deep neural architecture for open-close hand movement planning detection*

Fig. 2 shows a schematic block diagram of the proposed hybrid-domain DL based neural architecture for open-close hand movements planning detection. Specifically, the proposed system includes mainly: two DL modules, that receive input data in the time domain (i.e., EEG-source time-series) and in time-frequency domain (i.e., 3D-TF maps), to extract the T and TF features; a contextual features fusion component followed by a multi-layer TTF-NN that exploits the fused T and TF features for enhanced classification. Further details of each processing module of the developed hybrid-domain DL architecture is comprehensively described in the following Sections.

2.7. *Deep Learning based features extraction modules*

The core building blocks of the two processing features extraction modules are based on a DL architecture. Specifically, two CNNs are employed:

one used to extract temporal features directly from EEG sources' time series; one used to extract time-frequency features directly from 3D-TF maps of EEG sources. It is worth mentioning that a typical CNN includes a features extractor stage, consisting of several layers of *convolutional*, *activation* and *pooling*; and a classification stage consisting of a standard artificial neural network (NN).[34] The convolution layer performs the dot product between a set of F learnable filters and the input representation $\mathbf{R}$ sized r_1 x r_2. Notably, each filter (sized f_1 x f_2) is spatially small and scans the input data with a step (or stride) s, generating a corresponding set of F activation maps (or features maps) sized a_1 x a_2 where $a_1 = \frac{r_1 - f_1 + 2p}{s} + 1$ and $a_2 = \frac{r_2 - f_2 + 2p}{s} + 1$ (with p zero padding parameter). The activation layer applies a nonlinear transfer function to every estimated feature map. In particular, here, the *Rectified Linear Unit* (ReLu) is used, as offers good generalization performance and is not computationally expensive.[35] The pooling reduces the spatial dimension of the input representation. In this work, the max pooling operation is employed as it provides robustness to small fluctuation and consequently, better translation-invariant features.[36] It consists of a filter sized $\bar{f}_1$ x $\bar{f}_2$ that sweeps over the feature map (resulting from the previous layer) with step size $\bar{s}$, producing a downsampled representation sized of$\bar{r}_1$ x $\bar{r}_2$, with: $\bar{r}_1 = \frac{r_1 - \bar{f}_1}{\bar{s}} + 1$ and $\bar{r}_2 = \frac{r_2 - \bar{f}_2}{\bar{s}} + 1$.

2.7.1. *Temporal features extraction: T-CNN*

The CNN developed for the extraction of temporal features, denoted as *T-CNN*, consists of three convolutional layers (each followed by a ReLU), three max pooling layer, and a final 2-layer NN (denoted as T-NN) employed to perform the 3-way classification task: HC vs. HO vs. RE. The network is designed to receive as input epochs of EEG sources $\mathbf{I}$ sized i_1 x i_2, with i_1=210 (number of sources estimated) and i_2=512 (samples in 1s EEG window, preceding the movement onset). The $\mathbf{I}^{th}$ EEG sources epoch feeds into the first convolutional layer consisting of 8 filters (sized 3 x 3) and p=1. Each filter slides over the input representation $\mathbf{I}$ with stride s=1, coming up with a corresponding feature map of the same input dimension. Next, the ReLU activation function is applied to each feature map and the max pooling layer, consisting of filters sized 3 x 2 and stride s=3, reduces the input space from 210 x 512 to 70 x 171. The second convolutional layer has 16 filters sized 3 x 3, padding parameter and stride of 1. The features maps generated, after applying ReLU function, are downsampled through max pooling operation (filter size 4 x 3, stride s=3), delivering 16 features maps sized 23 x 57. The last convolutional layer includes 32 filters sized 3 x 3, padding parameter and stride of 1. At this stage, the max pooling layer has filter size of 5 x 6, stride of 3 and produces features maps of 7 x 18. The feature maps are flattened into a vector sized 32 x 7 x 18 = 4032 (referred as *T-features vector*) processed by a standard feedforward NN (T-NN with 2-hidden layers of 1000 and 500 hidden neurons, respectively, with softmax output layer, used to discriminate among HC/HO/RE). In this study, the sub-system composed of T-CNN and T-NN is denoted as *T-NET* (Fig. 2 (a)).

2.7.2. *Time-frequency features extraction: TF-CNN*

The CNN developed for the extraction of time-frequency features, denoted as *TF-CNN*, has a similar structure of T-CNN. In particular, it consists of three convolutional layers (with 8, 16, 32 filters, respectively), three max pooling layers and a final 2-hidden layers NN (denoted as TF-NN) with softmax output to perform HC vs. HO vs. RE discrimination task. In contrast to T-CNN, TF-CNN receives as input a volume of time-frequency maps $\mathbf{J}$ sized j_1 x j_2 x j_3, with j_1=43 (number of frequencies ranging 0.6-45 Hz), j_2=i_1=210 and j_3=i_2=512. Each convolutional layer uses filters size of 3 x 3, boundary padding of 1, stride of 1 and is followed by a ReLU activation function; whereas, the first max pooling layer has filter size of 5 x 4, stride 2 and produces 8 features maps sized 20 x 255; the second max pooling layer has filter size of 6 x 5, stride 2 and delivers 16 features maps sized 8 x 126; the last max pooling layer has filter size of 6 x 6, stride 2 and generates 32 features maps sized 2 x 61. Finally, the 32 feature maps are reshaped into a vector sized 32 x 2 x 61 = 3904 (referred as *TF-features vector*) and used as input to a 2-hidden layer feedforward NN (TF-NN) topologically similar to T-NN (Fig. 2 (a)). Similarly, the sub-system composed of proposed TF-CNN and TF-NN is referred as *TF-NET* (Fig. 2 (b)). Both the T-CNN and the TF-CNN were implemented with MATLAB R2018a

and trained by using Adaptive Moment (ADAM) optimization algorithm (with minibatch size of 21), on two NVIDIA GeForce RTX 2080 Ti GPU installed on a high performance workstation Intel(R) Core(TM) i7-8700K CPU 3.70 GHz with 64 GB RAM. After running several training/test sessions, we empirically observed that a mini batch size of 21 delivered fast computation and good generalization ability and was thus adopted in our experiments. Learning parameters of T-CNN and TF-CNN were set-up according to a *trial and error* strategy. Specifically, it was observed that the best classification performances were achieved with the default ADAM hyper-parameters (first moment exponential decayβ_1=0.9, second moment exponential decay β_2=0.999), learning rate α= 10^{-3} for T-CNN and α=10^{-4} for TF-CNN. Each CNN was trained by taking into account one subject at time for about 120 iterations, until the loss function converged. Notably, the training time for T-CNN and TF-CNN was of about 30 minutes and 1,5 hours, respectively, per subject and using the 10-fold cross validation technique.

2.7.3. *Features fusion for enhanced classification*

The T-features (sized 1 x 4032) and TF-features (sized 1 x 3904) vectors extracted from T-CNN and TF-CNN, respectively, are fused into a single vector, denoted as *TTF-features fusion vector*, which is further classified. Specifically, the TTF-features fusion vector sized 1 x 7936 (4032 + 3904) is the input to a multi-layers NN (denoted as TTF-NN), trained and tested to obtain enhanced classification performances. NN_3 is composed of 2-hidden layers with 4000 and 1000 units, respectively, trained with supervised learning modality via conventional back-propagation procedure.[37] The saturating ReLU activation function is used in each layer and the proposed network is trained for 50 iterations until the cross-entropy loss function converges. It is to be noted that, the features fusion layer and the proposed TTF-NN is referred as *TTF-NET* (Fig. 2 (c)).

2.8. *Performance parameters*

The effectiveness of our proposed systems (T-NET, TF-NET, TTF-NET) was assessed by applying the k-fold cross validation. It is generally preferable to split the data sample into k groups with the same number of samples, so that the sample of model skill scores are all equivalent. In our work, the dataset consisted of three subsets: Resting (RE), preparation of hand opening (HO), preparation of hand closing (HC), each comprising 60 samples. The overall dataset of a given subject was therefore made up of 180 examples. Choosing k=10 allowed us to split up each dataset into 10 folds of 18 samples (balanced with respect to the classes RE, HC, HO) taking into account the limited size of the original dataset (i.e., 180 samples per subject). Other trials have been carried out with different k values that yielded similar performance. F-score, precision, recall and accuracy parameters, defined as follows:

$$F - score = 2 * \frac{Precision * Recall}{Precision + Recall} . \quad (4)$$

$$Precision = \frac{tp}{tp + fp} . \quad (5)$$

$$Recall = \frac{tp}{tp + fn} . \quad (6)$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} . \quad (7)$$

where *tp*, *fp*, *tn*, *fn* are the true positive, false positive, true negative, false negative, respectively.

2.9. *Permutation Analysis*

In order to demonstrate the classification results are not achieved by chance (due to the randomness of the 10 folds taken into account) and in order to better analyze the behavior of the proposed classifier, we carried out the standard *permutation-based p-value* statistical test.[38] Such test assesses the probability that the estimated performance parameter would be obtained by chance, evaluating the p-value under a certain textitnull hypothesis. In particular, the null hypothesis assumes that features and class labels are independent. This results in permuting the labels of the dataset under analysis several times and evaluating for each iteration the statistical metric of interest (S).[38] The p-value represents the fraction of permuted instances where the classifier performed at least as extreme as or better in the random configuration (X^*) than in the original dataset (X) and is

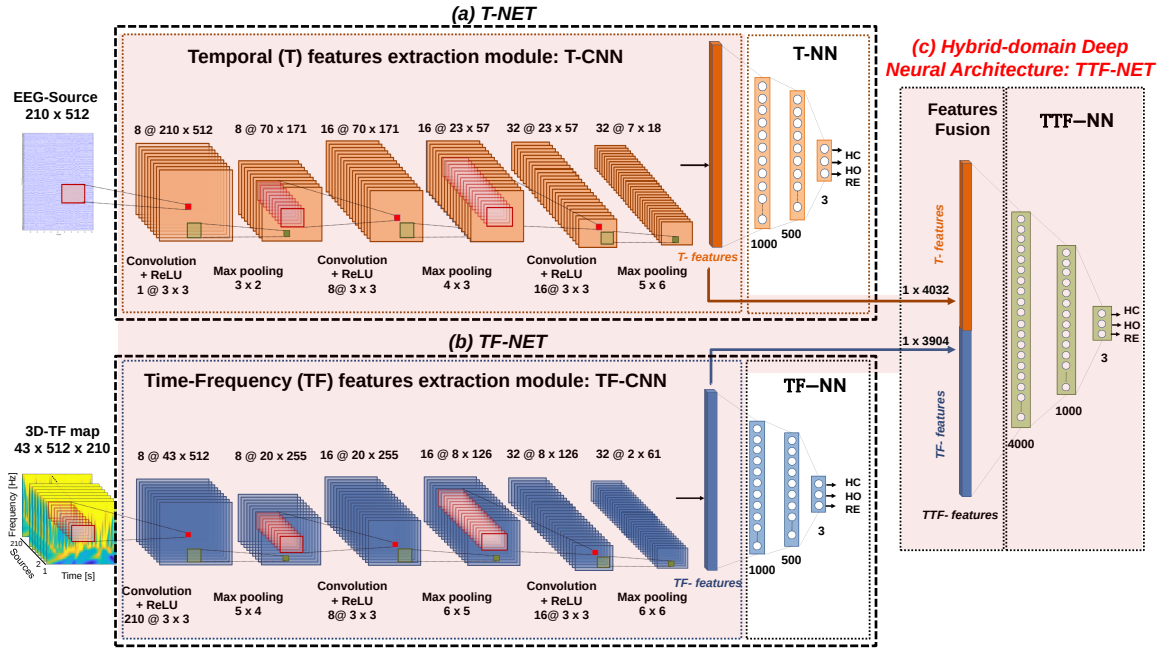


Figure 2: (a) Temporal-only domain based classification approach: T-NET. (b) Time-frequency-only domain based classification approach: TF-NET. (c) Highlighted in red, the hybrid temporal and time-frequency (TTF) domain based classification approach: TTF-NET.

empirically defined as:

$$p - value = \frac{1}{N} \sum_{i=1}^{N} I(S^* \geq S) \ . \qquad (8)$$

where N is the number of permutations, I is the indicator function, S is the statistic evaluated for the original data X; whereas, S^* is the statistic evaluated for permuted dataset X^*. Low p-value, typically smaller than a specific threshold α, means that the null hypothesis (independence of features and labels) is rejected and consequently the classifier is statistically significant.

3. Results

3.1. *Analysis of pre-hand motion TF images*

For every subject and every source location, the available 60 pre-opening (HO) TF images as well as the 60 pre-closing (HC) TF images were entered in joint a pre-movement collection of 120 matrices (Premov) meant for comparison with the available 60 resting (RE) TF images. Then, the SSIM between every pair of Premov and RE TF images was computed. In this way, every source location was assigned a set of 7200 (=120x60) SSIM values. Such values were then averaged for every source location. For every subject, the 10% of source locations exhibiting the lowest average SSIM values were selected in order to pinpoint which cortical locations minimized the similarity between motion planning and resting conditions. As detailed in Section 2.4, cortical source locations belonging to the premotor cortex (BA6) or to the primary motor cortex (BA4) were selected and included in the analysis. In Fig. 3, such cortical locations are denoted by dots: BA6 is depicted in yellow dots and BA4 in green dots, the remaining locations of the cortex model are denoted by black dots. For every subject, the source locations of BA4 and BA6 that exhibited the lowest SSIM between Premov and RE conditions are depicted with red dots. Such an analysis brought to light an interesting recurrent trend across the subjects: most of source locations minimizing SSIM between TF images in the Promov and RE conditions were located astride BA6 and BA4, in line with the cortical projection of hand motion in the *motor homunculus* representation.[39] Such sources are mainly located in the left hemisphere, in line with the contralaterality of the mechanisms of movement planning (the subject was

indeed planning to move his/her right hand).

In order to provide to further endorse the choice to work at source rather at scalp level, two source locations (named S_{left} and S_{right} hereinafter), located symmetrically in the two hemispheres, were selected to show the role of source estimation in emphasizing the contralaterality of motion planning. Top subplots in Fig. 4 show an example (subject S03) of the temporal trend of the EEG scalp channels C3 and C4 (typically analyzed in BCI EEG-based applications related to the execution or imagination of movements[1]) in a segment astride the onset of the right hand opening movement (1s before, 1s after). Similarly, the bottom subplots in Fig. 4 show the temporal trend of the EEG sources' time series related to S_{left} and S_{right} (the two central source locations shared by all subjects in Fig. 3). In particular, the thick black line indicates the mean, whereas the vertical blue bars indicate the variance over the epochs. It is worth to highlight the presence of slow oscillations that precede movement onset (related to MRCPs), which endorses the decision to develop a classifier based on features extracted in the time domain. When the brain plans or imagines the execution of the movement of a limb, the contralateral hemisphere is activated.[39] Such contralaterality is indeed more relevant in the EEG at source level rather than at scalp level, mainly because projecting EEGs into the cortex allows to improve EEG's spatial resolution, as discussed in Section 2.4.

The idea of combining the information deriving from the features extracted by a T-based CNN classifier with those extracted by a TF-based CNN classifier originated from the consideration that MRCP and SMR represent complementary aspects of motor preparation, which is also confirmed in the literature.[40] One of the two aforementioned source locations, S_{left} and S_{right}, was selected to show an example of how the SSIM map between the pre-opening and pre-closing average TF images looks (Fig. 5). It is worth to note that the dissimilarity between the two pre-motion conditions is more marked (blue regions of the SSIM map) in two regions of the map: 1) below 1.5 Hz in throughout the time interval of 1s preceding motion onset; 2) between 13 and 30 Hz in the interval 1 to 0.5 s prior to motion onset. The aforementioned observations suggest that time-frequency maps of motor cortex sources may help to discriminate hand sub-movement motion planning and endorsed the decision to develop DL classifier based on such maps.

3.2. *Sub-movement planning classification*

A hybrid-domain deep learning based neural architecture is developed to extract T and TF features, automatically, from EEG sources signals (time-domain) and 3D-TF maps (time-frequency domain), in order to perform a 3-way classification: pre-hand open (HC) vs. pre-hand close (HC) vs. resting (RE). Specifically, two DL-based networks, denoted as T-NET and TF-NET (that included two custom CNNs: T-CNN and TF-CNN, respectively) were developed. T-CNN received as input the EEG sources' time series, whereas TF-CNN received as input the volume of TF maps. The following ablation analysis is carried out: first, the ability of T-NET and TF-NET to perform the 3-way (HC vs. HO vs. RE) classification task in the time-only and time-frequency-only domain, respectively, was evaluated. Next, the hybrid TTF-domain DL-based approach that combines T and TF features, extracted from T-CNN and TF-CNN, respectively, was evaluated to assess its ability to enhance HC vs. HO vs. RE classification (TTF-NET). It is worth noting that, 15 3-way classifiers (one per subject) were trained and tested independently. Each classifier included 180 EEG epochs (60 of HC, 60 of HO and 60 of RE) and for each class. The k-fold cross validation technique (with k=10) was employed, hence, performance parameters are expressed as mean value ± standard deviation. Since F-score includes information of precision and recall, results are discussed in terms of F-score and accuracy.

- *Temporal-only (T) domain based classification: T-NET.* Table 1 (a) reports F-score performance for each class (i.e. HC/HO/RE) of the developed T-NET (composed of T-CNN and T-NN). Specifically, HC and HO classes were discriminated very well in S04 with values of 86.29 ± 3.78% and 91.32 ± 3.51% respectively, while RE category showed rate up of to 99.43 ± 1.28% in S10. It is worth noting that the worst classification performance of RE class was only 79.03 ± 14.79% (S06); whereas, lower values were observed for HC and HO: 41.09 ± 20.45% (S11) and 35.07 ± 17.87% (S10), respectively. The average performance per discrimi-

nation task was also estimated, achieving F-scores of 57.05 ± 10.59%, 52.90 ± 12.38%, 89.47 ± 5.96% for HC, HO and RE, respectively. Overall, the average accuracy score was of 67.90 ± 5.72%. In particular, the best performance was observed in S04 with 88.52 ± 4.22% (Table 2).

- *Time-frequency-only (TF) domain based classification: TF-NET.* Table 1 (b) reports the F-score performance for each class (i.e. HC/HO/RE) of the developed TF-NET (consisted of TF-CNN and TF-NN). The highest F-score values of 90.16 ± 7.10% (HC), 92.62 ± 5.91% (HO) were observed in S04 and 97.81 ± 2.32% (RE) in S10. In contrast, the worst performance were of 32.24 ± 12.15% (HC), 29.46 ± 14.30% (HO) and 72.17 ± 7.54% (RE) for S13, S07 and S06, respectively. Here, the average values of HC, HO and RE were of 53.22 ± 11.80%, 54.28 ± 9.10%, 84.72 ± 6.95% respectively. Finally, the achieved mean accuracy was of 65.68 ± 5.38%. In particular, the best performance was observed in S04 with an accuracy of 92.22 ± 7.34% (Table 2).
- *Hybrid Temporal and time-frequency (TTF) domain based classification: TTF-NET.* Table 1 (c) reports the F-score performance for each class (i.e. HC/HO/RE) of the developed TTF-NET (consisted of the two developed CNNs, the features fusion component and the TTF-NN classifier). The maximum HC, HO and RE discrimination value was of 97.14 ± 2.08% (S04), 97.11 ± 2.94% (S04) and 98.89 ± 2.48% (S10), respectively; whereas, the minimum HC, HO and RE classification performance was of 52.62 ± 11.65% (S02), 51.79 ± 13.10% (S02) and 89.81 ± 2.94% (S12), respectively. Overall, higher average values than T and TF based classifications, in terms of both F-score and accuracy, were achieved. Indeed, average F-scores were of 66.00 ± 6.17% for HC, 67.21 ± 6.87% for HO and 94.24 ± 3.57% for RE tasks. Finally, the average accuracy was of 76.21 ± 3.77% (Table 2).

Fig. 6 shows the histogram of the accuracies achieved by T-NET, TF-NET, TTF-NET for each subject. As can be seen, TTF-NET outperformed T-NET, TF-NET in all subjects, except for S12, where T-NET reported similar accuracy, but with higher standard deviation. It is also to be noted that, T-NET, TF-NET, TTF-NET delivered accuracies significantly higher than chance level (33%). It is to be noted that the 10-folds cross validation technique reported very good results in terms of average accuracy and showed that the fusion of T and TF features improved the classification performance (TTF-NET, Table 2).

Furthermore, in order to demostrate that such results are statistically significant, the permutation analysis (described in Section 2.9) was performed. Specifically, in this study, the TTF-features dataset (extracted from T-CNN and TF-CNN) and used as input to the TTF-NET was taken into account. First, for each subject, the k^{th} fold, comprising of the TTF-features corresponding to the accuracy closest to the achieved average value, was used as original dataset. Next, N permutations of the class label were produced. Ideally, the entire set of permutations should be considered to estimate the exact p-value. However, this was not computationally feasible. Hence, here, N=1000 permutations were performed due to the limited computational power available. For each iteration, the statistic metric (i.e., accuracy) was calculated. Finally, the p-value was estimated according to 8 and compared with a small threshold (α = 0.05[41]). It is to be noted that S in the definition 8 here corresponds to the average accuracy value of the subject under analysis. For example, for subject 02, the average accuracy was of 66.30% (Table 2). Hence, the p-value was estimated by counting how many estimated accuracy values were equal or greater than 66.30%, over 1000 permutations. In this case p-value=0.00/100=0.00 $<$ 0.05. Overall, experimental results reported that for all subjects p-value $<$ 0.05. Thus, the null hypothesis is rejected and the proposed classifier is statistically significant.

3.3. *Comparison with the state-of-the-art*

Other groups have used the dataset shared by Ofner et al.[13] In particular, Cho et al.,[42] who proposed a subject-specific time interval selection and applied common spatial patterns (CSP) and regularized linear discriminant analysis (RLDA) to classify motor execution EEGs, reporting an average classification accuracy of 56.83%. Namazi et al.[43] explored how the complexity of EEG signal changes during the execution or the imagination of different upper limb's movements. They found out that EEG exhibits high level of complexity in elbow flexion

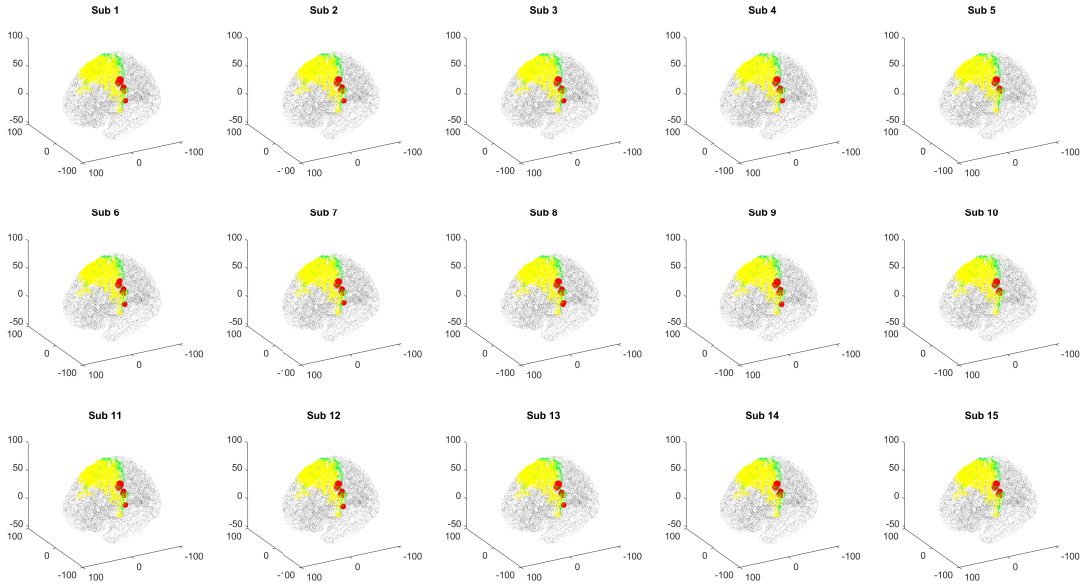


Figure 3: Source locations that exhibited the lowest SSIM values (highest dissimilarity) between the average time-frequency behaviour in pre-movement (hand's opening/closing planning) and resting conditions. The nodes of the head model of the cortex are represented as black dots. Yellow dots represent BA6 whereas green dots represent BA4 (premotor and primary motor cortex). The red thick dots show the 10% of source locations that exhibited the lowest average SSIM. Every subplot is associated to a subject.

and hand-close movements in ME and the lowest level of complexity in hand-open and rest condition in ME. Jeong et al.[44] proposed a method to enhance the MRCP decoding performance by introducing a subject-dependent and a section-wise spectral filtering (SSSF) method that takes into account the subjects' peculiar MRCP characteristics. They performed a binary motor execution classification, achieving an average accuracy of 0.72±0.09 in HC vs. RE and of 0.76±0.06 in HO vs. RE. However, the aim of the aforementioned works was to discriminate the movement, performed or only imagined, by analyzing the EEG signals astride the movement onset (pre-movement+movement execution), but not to predict whether the subject was going to perform a movement by processing his/her pre-movement EEG signals only, which is the aim of the present work. Analyzing pre-movement EEG signals only, makes the classification inherently more difficult as neural correlates of motor initiation are missing, nevertheless, for sake of comparison with the literature, the performance of the proposed method (pre-movement analysis) will be compared to the performance achieved in the literature in decoding movements by analysis both pre-movement+movement execution EEG signals. In particular, Ofner et al.[13] estimated the performance of their method for varying length of the processed EEG segments, for sake of comparison, the performance of the present method will be compared to the best performance achieved by them: the largest accuracy they achieved in movement execution vs. resting was 80% whereas the largest accuracy in movement execution vs. movement execution was 40%. As regards further differences with the literature, our approach aims at performing a 3-way classification (preparation of hand opening, preparation of hand closing, resting with no movement preparation: HO vs. HC vs. RE) in this way, a single classifier would be able to perform motion planning detection while decoding hand opening vs. hand closing planning. In our work, an average accuracy of 76.21±3.77% was achieved (chance level = 33.33%) in HO vs. HC vs. RE discrimination. Also Jeong et al.[45] performed a 3-way classification, but over forearm supination vs. forearm pronation vs. rest class, they introduced a Hierarchical Flow CNN, achieving

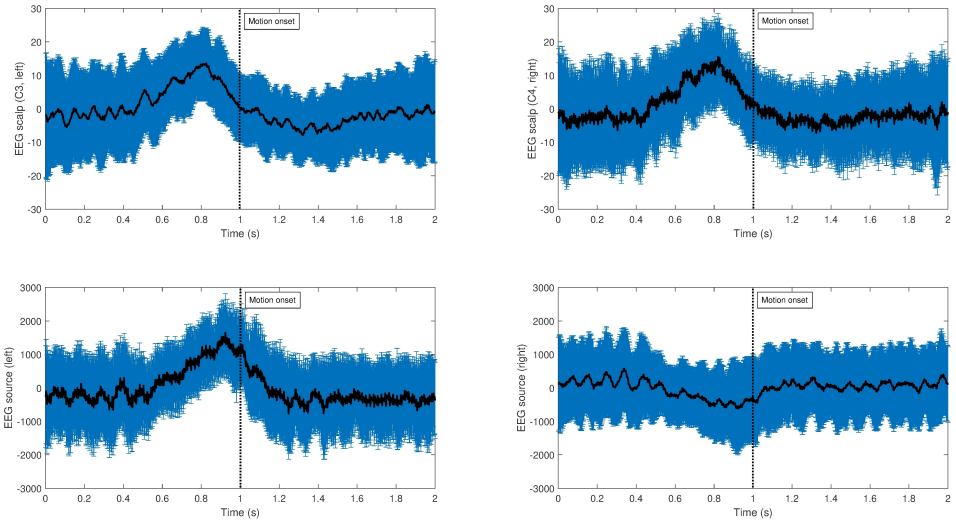


Figure 4: Example of the temporal behaviour of EEG signals astride the onset of hand opening (1s before, 1s after) observed in Subject 3. The two top subplots show the EEG signals collected at scalp sensors C3 (top, left) and C4 (top, right). The two bottom plots show the EEG sources' time series estimated for the two locations mentioned in Section 3.1, namely the source in the left hemisphere (bottom, left) and in the right one (bottom, right). Every subplot shows the mean trend (black thick line), averaged over the 60 pre-opening epochs of the subject, together with the variance (blue vertical bars).

Table 1: F-score values estimated on test sets, when: (a) temporal-only data (i.e., EEG sources) are used as input to T-NET; (b) time-frequency-only data (i.e., 3D-TF maps) are used as input to TF-NET; (c) the temporal and time-frequency features are used as input to TTF-NET. All the outcomes are reported as average value ± standard deviation %. The highest values are typed in bold face.

	(a) T-NET (T-CNN + NN_1)			**(b) TF-NET (TF-CNN + NN_2)**			**(c) TTF-NET (features fusion layer + NN_1)**		
Subject	**HC**	**HO**	**RE**	**HC**	**HO**	**RE**	**HC**	**HO**	**RE**
S01	55.35 ± 18.41	43.57 ± 13.12	89.73 ± 5.17	52.76 ± 12.16	41.07 ± 6.98	85.91 ± 8.15	62.94 ± 10.74	64.14 ± 9.04	95.01 ± 4.14
S02	44.62 ± 10.56	39.25 ± 10.50	91.50 ± 3.79	44.32 ± 17.26	43.88 ± 11.17	95.00 ± 4.36	52.62 ± 11.65	51.79 ± 13.10	94.24 ± 4.54
S03	74.56 ± 6.42	82.21 ± 2.24	90.05 ± 3.47	85.13 ± 2.46	85.37 ± 6.51	88.97 ± 5.77	88.29 ± 5.80	92.79 ± 3.20	96.30 ± 2.23
S04	**86.29 ± 3.78**	**91.32 ± 3.51**	87.83 ± 7.70	**90.16 ± 7.10**	**92.62 ± 5.91**	93.63 ± 10.48	**97.14 ± 2.08**	**97.11 ± 2.94**	96.84 ± 2.79
S05	53.42 ± 12.65	45.70 ± 19.13	89.39 ± 7.81	45.11 ± 6.54	51.81 ± 4.09	82.04 ± 6.58	59.82 ± 3.37	66.26 ± 10.67	93.99 ± 6.06
S06	49.66 ± 8.68	42.82 ± 14.41	79.03 ± 14.79	39.80 ± 17.85	50.89 ± 10.84	72.17 ± 7.54	56.62 ± 3.98	61.93 ± 11.88	92.74 ± 2.74
S07	53.60 ± 6.68	41.18 ± 9.01	80.69 ± 9.73	34.37 ± 20.51	29.46 ± 14.30	58.85 ± 19.33	55.63 ± 11.13	54.28 ± 4.67	89.07 ± 2.91
S08	58.35 ± 4.48	57.97 ± 11.99	96.81 ± 2.19	54.61 ± 12.97	54.87 ± 10.91	90.70 ± 5.76	71.37 ± 7.02	73.31 ± 5.61	96.36 ± 3.47
S09	62.25 ± 8.90	55.66 ± 13.70	96.70 ± 1.26	55.07 ± 16.47	47.11 ± 15.34	95.04 ± 2.35	67.26 ± 6.43	63.71 ± 3.95	96.70 ± 1.26
S10	49.67 ± 13.76	35.07 ± 17.87	**99.43 ± 1.28**	53.10 ± 4.62	49.05 ± 10.31	**97.81 ± 2.32**	56.57 ± 3.99	55.07 ± 8.31	**98.89 ± 2.48**
S11	41.09 ± 20.45	46.77 ± 13.12	82.84 ± 7.41	43.26 ± 19.77	51.26 ± 0.94	77.52 ± 12.67	60.40 ± 4.87	64.09 ± 8.52	95.54 ± 6.12
S12	62.89 ± 11.16	55.91 ± 16.84	93.61 ± 5.44	52.31 ± 8.63	36.63 ± 16.47	75.11 ± 4.63	62.35 ± 4.08	60.61 ± 5.64	89.81 ± 2.94
S13	46.47 ± 15.10	50.28 ± 18.40	88.94 ± 7.80	32.24 ± 12.15	58.51 ± 11.70	86.40 ± 4.02	64.80 ± 7.35	61.69 ± 6.92	91.18 ± 7.06
S14	62.44 ± 9.22	54.89 ± 11.43	87.99 ± 7.54	64.87 ± 8.70	69.34 ± 3.90	88.95 ± 5.36	70.13 ± 5.94	76.18 ± 3.73	94.65 ± 3.20
S15	55.04 ± 8.58	50.89 ± 10.37	87.46 ± 3.95	51.22 ± 9.79	52.27 ± 7.09	82.71 ± 4.95	63.99 ± 4.17	65.20 ± 4.86	92.23 ± 1.56
Average	57.05 ± 10.59	52.90 ± 12.38	89.47 ± 5.96	53.22 ± 11.80	54.28 ± 9.10	84.72 ± 6.95	66.00 ± 6.17	67.21 ± 6.87	94.24 ± 3.57

an average accuracy of 0.52±0.03.

4. Discussion

The present research aims at exploring the potential of DL in detecting whether the brain is in a resting condition or is planning to perform a specific movement of the hand, from the analysis of its EEG recordings. This is a challenging issue in the field of BCI. To this end, the dataset made freely available by Ofner et al.,[13] consisting of 61-channels EEG recordings acquired during the execution of hand open/close sub-movements (of the same limb) and also during resting state, was taken into account. Next, source signals in the motor cortex were extracted from 1s EEG epochs preceding the onset of motion and from 1s resting state epochs. The EEG source signals of every epoch under analysis were projected into the time-frequency (TF) domain and the estimated TF maps were arranged in a 3D configuration so that the three directions of the matrix represented time, frequency and source location. To the best of our knowledge, this is the first work that attempts to develop a hybrid-domain deep learning system, using T and TF features extracted from EEG sources and TF maps by means of two custom CNNs, with the aim to discriminate hand's opening/closing motor preparation (i.e. HC/HO) and rest (RE). It is worth mentioning that the dataset used to perfom the 3-way classification is in principle too small. However, each sample consisted of a high-dimesionsional matrix in the time-frequency as well as in the temporal domain. Specifically, a 3D-TF-map sized 43 x 512 x 210 and a 2D-T map (i.e., EEG-sources) sized 210 x 512, were processed. Hence, here, we exploited DL techniques (i.e. CNN) for automatic extraction of the most significant features from such high-dimensional inputs (i.e., 3D-TF-maps sized 43 x 512 x 210 and EEG-sources sized 210 x 512, as shown in Fig. 2), which are otherwise difficult to process with standard ML approaches. Furthermore, a total of 7936 features were extracted through the developed CNNs, including 3904 features from 43 x 512 x 210=4.623.360 TF-input data, and 4032 features from 210 x 510= 107.100 T-input data. Hence, since the results achieved with the concatenation of both T and TF features (extracted from T-CNN and TF-CNN respectively) were promising, we decided to explore the potential of DL in decoding motion planning of hand's sub-movements from EEG signals by processing the available dataset shared by Ofner et al.[13] Experimental results show that the proposed hybrid-domain deep learning approach outperformed the temporal-only domain and time-frequency-only domain classification approaches, reporting higher performances in discriminating pre-motion planning (pre-hand open vs. pre-hand close) and rest condition (no movement). Notably, T-NET and TF-NET achieved averaged accuracies of 67.90 ± 5.72% and 65.68 ± 5.38%, respectively. In contrast, TTF-NET reported averaged rate up to 76.21 ± 3.77% (Table 2). As regards the complexity in terms of number of learnable parameters, the two branches, T-NET and TF-NET, involve a feature extraction part (named T-CNN and TF-CNN, respectively) and a classifying part (named T-NN and TF-NN, respectively). Both T-NN and TF-NN involve approximately 4.5M parameters. Following initial independent training of the T-NET branch and of the TF-NET branch, the T-NN and TF-NN are discarded and the trained T-CNN and TF-CNN are included in the double branch approach (TTF-NET). The TTF-NET merges the features extracted by T-CNN and TF-CNN and feeds them to another classifier (named, TTF-NN) that involves 35M parameters. Overall, the double branch approach implies 35M-4.5M-4.5M=26M more parameters, compared to the two single branches approach. This increased complexity can be acceptable considering the average accuracy improvement provided by TTF-NET is 5%, compared to T-NET, and 10%, compared to TF-NET.For further information about the number of network's parameters, please refer to Figure 2.

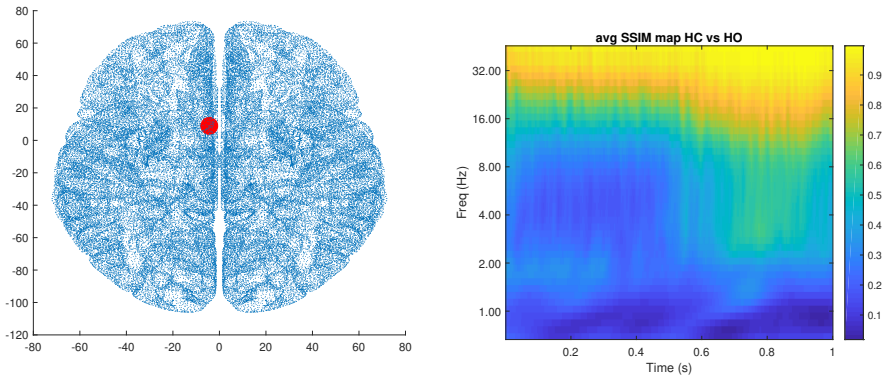
Table 2: Accuracy values estimated on test sets, when: temporal-only data (i.e., EEG sources) are used as input to T-NET; time-frequency- only data (i.e., 3D-TF maps) are used as input to the TF-NET; the temporal and time-frequency features are used as input to TTF-NET. All the outcomes are reported as average value ± standard deviation The highest values are typed in bold face.

Subject	(a) T-NET (T-CNN + NN_1)	(b) TF-NET (TF-CNN + NN_2)	(c) TTF-NET (features fusion layer + NN_1)
S01	65.56 ± 6.88	61.48 ± 4.61	74.07 ± 6.55
S02	60.00 ± 4.06	62.96 ± 5.56	66.30 ± 6.98
S03	82.59 ± 2.81	86.67 ± 2.75	92.59 ± 3.46
S04	**88.52 ± 4.22**	**92.22 ± 7.34**	**97.04 ± 2.11**
S05	64.44 ± 7.57	59.26 ± 4.14	74.07 ± 3.46
S06	58.52 ± 9.41	56.67 ± 2.81	71.48 ± 5.34
S07	59.26 ± 6.80	45.19 ± 6.49	66.67 ± 4.54
S08	71.85 ± 4.22	68.15 ± 5.77	80.74 ± 4.26
S09	72.59 ± 4.42	67.41 ± 6.75	76.30 ± 3.31
S10	64.07 ± 3.84	67.41 ± 2.81	70.37 ± 3.46
S11	60.37 ± 6.36	58.15 ± 8.55	73.70 ± 3.56
S12	71.85 ± 8.53	57.04 ± 7.90	71.85 ± 2.41
S13	63.33 ± 7.45	62.22 ± 6.63	72.96 ± 2.81
S14	69.63 ± 5.49	63.51 ± 3.37	80.74 ± 2.48
S15	65.89 ± 3.71	65.52 ± 5.37	74.20 ± 1.76
Average	67.90 ± 5.72	65.68 ± 5.38	76.21 ± 3.77

Most of the previous studies on EEG-based BCI were focused on the discrimination of imagined movements of the right hand, the left hand, of feet, tongue and most of them were based on the analysis of public datasets made available by the international BCI competitions. DL could lead to a significant step forward in EEG-based BCI, provided that large databases are shared by the international scientific community, as happened for image analysis and Natural Language Processing, which experienced an extraordinary development right thanks to the availability of very large shared dataset. Another key step forward would be to make a distinction between pure "motor imagery" and "motor intention" by giving the latter a more restrictive meaning as compared to what has been done far in the literature. "Motor intention" should indeed mean intention to actually perform a movement and not merely to imagine performing a movement. To the best of our knowledge, to this day, the concepts of motor imagery and motor intention are quite overlapped in the literature. In Ref.,[46] the group from Graz emphasized the importance of complex movement intention decoding which has been far less investigated than classic motor imagery. The ability to decode the intention of complex movements would lead to a big step forward in the development of user-friendly BCI and encourage the development of a DL-based movement-assist devices (i.e., BCI-rehabilitation systems) for those who are capable of planning movements but that are (totaly or partially) incapable to perform motor control. The main limitation of the present work derived from the poor spatial resolution of EEG signals. We have been pursuing here the challenging goal of discriminating the preparation phase of two different sub-movements of the same limb (hand open/hand close), thus inherently mapping onto the same sub-region of the primary motor cortex, at least in the phase of movement implementation. In order to improve the EEG spatial resolution and reduce volume conduction effects, EEG signals were mapped onto the cortex by solving the inverse problem through beamforming. In particular, a very detailed head model was adopted including 6 types of different tissues (scalp, skull, cerebrospinal fluid (CSF), gray matter, white matter, air cavities) and 2000 cortical locations. Nevertheless, in order to fully exploit the inverse problem solution, a very high-density EEG (128 channels or more) will be used in the future. Another limitation is the size of the dataset. A relatively small number of examples is available, specifically, 60 trials per class per subject. The number of participants (15) is small as well. In our future work, a larger cohort of subjects will be involved and more trials will be recorded per subject. The results achieved within the present study, albeit limited to the analysed dataset, look promising and endorse that EEG source signals recorded during the preparation of different sub-movements of the same limb have different projections in the time-frequency domain, which could encourage to extend the application of the proposed system to electrocorticography, thus to implantable BCI.

5. Conclusions

The main objective of the present study was to realise early detection and decoding of motor preparation of sub-movements (open/close) of the same hand. To this end, EEG segments (i.e., epochs) relative to 1s before movement onset and to resting state, were extracted and preprocessed through beamforming to estimate the EEG sources in the motor cortex. Next, 3D-TF maps (time x frequency x source) were also generated. A *hybrid-domain* deep neural architecture was proposed by combining the features obtained from a time-based CNN and a time-frequency-based CNN, delivering the highest accuracy rate of up to 76.21 ± 3.77% when compared to temporal-only and time-frequency only based frameworks. However, we expect a higher number of EEG channels could allow for a more accurate inverse problem solution and consequently a better reconstruction of source signals. This, would entail an im-



(a) (b)

Figure 5: Example of a source location in the motor cortex (a) and its SSIM map between the pre-opening and pre-closing mean TF images (b).

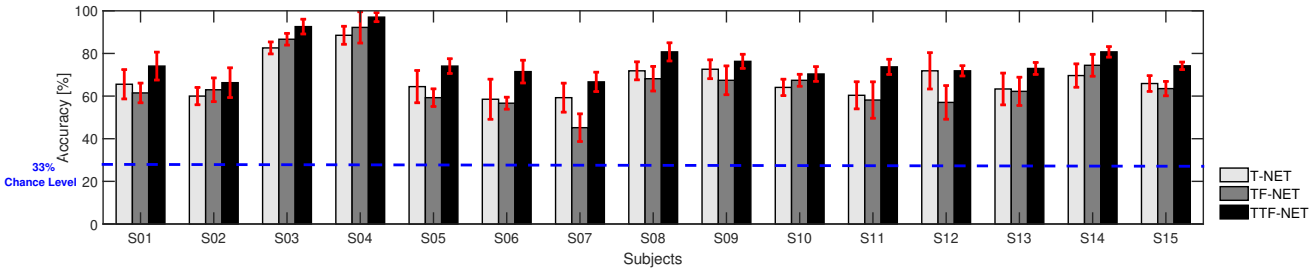


Figure 6: Histogram of the accuracy achieved with the temporal-only approach (T-NET, light grey bars), time-frequency-only approach (TF-NET, dark gray bars) and our features fusion approach (TTF-NET, back bars), for each subject. The vertical red line of each box represents the standard deviation, while the horizontal dash blue line the chance level (that is of 33%).

proved classification. For these reasons we plan BCI experiments of motor execution by collecting high-density EEGs (with 256 channels) during the execution of different sub-movement of the single hand as well as of both hands. A large of number of participants will be enrolled in our future studies and many trials will be recorded per subject so that we will be able to develop and test both intra-subject and cross-subjects classifiers, with the very ambitious goal to contribute to development of BCI systems that do not require calibration. BCI systems are indeed usually calibrated to fit individual users which implies that the classifier learns specific neural characteristics of the subject under analysis, and hardly adapts to other individuals (whether healthy or even with some neurological injury). Furthermore, newer and more powerful supervised machine learning/classification algorithms will be considered in the future extension of our research such as Enhanced Probabilistic Neural Network, Neural Dynamic Classification algorithms, Dynamic Ensemble Learning Algorithm, and Finite Element Machines for fast learning.[47–53]

6. Acknowledgements

The authors also wish to thank the anonymous reviewers for their insightful comments and suggestions. This paper was supported by POR Calabria FESR FSE 2014-2020 - Grant Ref. C39B18000080002. Hussain would like to acknowledge the support of the UK Engineering and Physical Sciences Research Council (EPSRC) - Grants Ref. EP/M026981/1, EP/T021063/1, EP/T024917/1.

Bibliography

1. 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) (2018) p. 031005.
2. 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, *Progress in brain research*, **228** (Elsevier, 2016), pp. 39–70.
3. F. Li, W. Peng, Y. Jiang, L. Song, Y. Liao, C. Yi, L. Zhang, Y. Si, T. Zhang, F. Wang *et al.*, The dynamic brain networks of motor imagery: time-varying causality analysis of scalp EEG, *International Journal of Neural Systems* **29**(01) (2019) p. 1850016.
4. K. Brodmann, *Brodmann's: Localisation in the cerebral cortex* (Springer Science & Business Media, (2007)).
5. P. Ofner, A. Schwarz, J. Pereira, D. Wyss, R. Wildburger and G. 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) (2019) p. 7134.
6. P. Ofner and G. R. Müller-Putz, Using a noninvasive decoding method to classify rhythmic movement imaginations of the arm in two planes, *IEEE Transactions on Biomedical Engineering* **62**(3) (2014) 972–981.
7. J.-H. Kim, F. Bießmann and S.-W. Lee, Decoding three-dimensional trajectory of executed and imagined arm movements from electroencephalogram signals, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **23**(5) (2014) 867–876.
8. N. Robinson and A. Vinod, Noninvasive brain-computer interface: decoding arm movement kinematics and motor control, *IEEE Systems, Man, and Cybernetics Magazine* **2**(4) (2016) 4–16.
9. E. Y. Lew, R. Chavarriaga, S. Silvoni and J. d. R. Millán, Single trial prediction of self-paced reaching directions from EEG signals, *Frontiers in neuroscience* **8** (2014) p. 222.
10. A. Úbeda, J. M. Azorín, R. Chavarriaga and J. d. R. Millán, Classification of upper limb center-out reaching tasks by means of EEG-based continuous decoding techniques, *Journal of neuroengineering and rehabilitation* **14**(1) (2017) p. 9.
11. H. Shibasaki and M. Hallett, What is the bereitschaftspotential?, *Clinical neurophysiology* **117**(11) (2006) 2341–2356.
12. A. Shakeel, M. S. Navid, M. N. Anwar, S. Mazhar, M. Jochumsen and I. K. Niazi, A review of techniques for detection of movement intention using movement-related cortical potentials, *Computational and mathematical methods in medicine* (2015) p. 346217.
13. 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) (2017) p. e0182578.
14. A. Schwarz, P. Ofner, J. Pereira, A. I. Sburlea and G. R. Mueller-Putz, Decoding natural reach-and-grasp actions from human EEG, *Journal of neural engineering* **15**(1) (2017) p. 016005.
15. Y. LeCun, Y. Bengio and G. Hinton, Deep learning, *nature* **521**(7553) (2015) p. 436.
16. C. Hua, H. Wang, H. Wang, S. Lu, C. Liu and S. M. Khalid, A novel method of building functional brain network using deep learning algorithm with application in proficiency detection, *International Journal of Neural Systems* **29**(01) (2019) p. 1850015.
17. A. H. Ansari, P. J. Cherian, A. Caicedo, G. Naulaers, M. De Vos and S. Van Huffel, Neonatal seizure detection using deep convolutional neural networks, *International Journal of Neural Systems* **29**(04) (2019) p. 1850011.
18. C. Ieracitano, A. Paviglianiti, M. Campolo, A. Hussain, E. Pasero and F. C. Morabito, A novel automatic classification system based on hybrid unsupervised and supervised machine learning for electrospun nanofibers, *IEEE/CAA Journal of Automatica Sinica* **8**(1) (2020) 64–76.
19. G. B. Martins, J. P. Papa and H. Adeli, Deep learning techniques for recommender systems based on collaborative filtering, *Expert Systems* **37**(6) (2020) p. e12647.
20. R. Grech, T. Cassar, J. Muscat, K. P. Camilleri, S. G. Fabri, M. Zervakis, P. Xanthopoulos, V. Sakkalis and B. Vanrumste, Review on solving the inverse problem in EEG source analysis, *Journal of neuroengineering and rehabilitation* **5**(1) (2008) p. 25.
21. H. Hallez, B. Vanrumste, R. Grech, J. Muscat, W. De Clercq, A. Vergult, Y. D'Asseler, K. P. Camilleri, S. G. Fabri, S. Van Huffel *et al.*, Review on solving the forward problem in EEG source analysis, *Journal of neuroengineering and rehabilitation* **4**(1) (2007) p. 46.
22. 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) (2015) 4–14.
23. S. Haufe and A. Ewald, A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies, *Brain topography* (2016) 1–18.
24. S. Haufe, Y. Huang and L. C. Parra, A highly detailed FEM volume conductor model based on the ICBM152 average head template for EEG source imaging and TCS targeting, *Conf Proc IEEE Eng Med Biol Soc*, 2015.
25. M. Grosse-Wentrup, C. Liefhold, K. Gramann and M. Buss, Beamforming in noninvasive brain-computer interfaces, *IEEE Transactions on Biomedical Engineering* **56**(4) (2009) 1209–1219.

26. J. L. Lancaster, D. Tordesillas-Gutiérrez, M. Martinez, F. Salinas, A. Evans, K. Zilles, J. C. Mazziotta and P. T. Fox, Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template, *Human brain mapping* **28**(11) (2007) 1194–1205.
27. J. L. Lancaster, L. Rainey, J. Summerlin, C. Freitas, P. T. Fox, A. Evans, A. Toga and J. Mazziotta, Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method, *Human brain mapping* **5**(4) (1997) 238–242.
28. J. L. Lancaster, M. G. Woldorff, L. M. Parsons, M. Liotti, C. S. Freitas, L. Rainey, P. V. Kochunov, D. Nickerson, S. A. Mikiten and P. T. Fox, Automated Talairach atlas labels for functional brain mapping, *Human brain mapping* **10**(3) (2000) 120–131.
29. I. Daubechies, *Ten lectures on wavelets* (SIAM, 1992).
30. J. N. Spring, N. Place, F. Borrani, B. Kayser and J. Barral, Movement-related cortical potential amplitude reduction after cycling exercise relates to the extent of neuromuscular fatigue, *Frontiers in human neuroscience* **10** (2016) p. 257.
31. S. Muthukumaraswamy, High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations, *Frontiers in human neuroscience* **7** (2013) p. 138.
32. Z. Wang and A. C. Bovik, Mean squared error: Love it or leave it? a new look at signal fidelity measures, *IEEE signal processing magazine* **26**(1) (2009) 98–117.
33. Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE transactions on image processing* **13**(4) (2004) 600–612.
34. A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, 2012, pp. 1097–1105.
35. V. Nair and G. E. Hinton, Rectified linear units improve restricted boltzmann machines, *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
36. D. Scherer, A. Müller and S. Behnke, Evaluation of pooling operations in convolutional architectures for object recognition, *International conference on artificial neural networks*, Springer2010, pp. 92–101.
37. R. Hecht-Nielsen, Theory of the backpropagation neural network, *Neural networks for perception*, (Elsevier, 1992), pp. 65–93.
38. M. Ojala and G. C. Garriga, Permutation tests for studying classifier performance, *Journal of Machine Learning Research* **11**(Jun) (2010) 1833–1863.
39. M. Catani, A little man of some importance, *Brain* **140**(11) (2017) 3055–3061.
40. D. Liu, W. Chen, Z. Pei and J. Wang, Detection of lower-limb movement intention from eeg signals, *2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, IEEE2017, pp. 84–89.
41. J. Ludbrook and H. Dudley, Why permutation tests are superior to t and f tests in biomedical research, *The American Statistician* **52**(2) (1998) 127–132.
42. J.-H. Cho, J.-H. Jeong, K.-H. Shim, D.-J. Kim and S.-W. Lee, Classification of hand motions within EEG signals for non-invasive bci-based robot hand control, *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE2018, pp. 515–518.
43. H. Namazi, T. S. Ala and V. Kulish, Decoding of upper limb movement by fractal analysis of electroencephalogram (EEG) signal, *Fractals* **26**(05) (2018) p. 1850081.
44. J.-H. Jeong, N.-S. Kwak, C. Guan and S.-W. Lee, Decoding movement-related cortical potentials based on subject-dependent and section-wise spectral filtering, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28**(3) (2020) 687–698.
45. J.-H. Jeong, B.-H. Lee, D.-H. Lee, Y.-D. Yun and S.-W. Lee, Eeg classification of forearm movement imagery using a hierarchical flow convolutional neural network, *IEEE Access* **8** (2020) 66941–66950.
46. 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, *Progress in brain research* **228** (2016) 39–70.
47. A. Ortiz-Rosario and H. Adeli, Brain-computer interface technologies: from signal to action, *Reviews in the Neurosciences* **24**(5) (2013) 537–552.
48. A. Burns, H. Adeli and J. A. Buford, Brain–computer interface after nervous system injury, *The Neuroscientist* **20**(6) (2014) 639–651.
49. M. Ahmadlou and H. Adeli, Enhanced probabilistic neural network with local decision circles: A robust classifier, *Integrated Computer-Aided Engineering* **17**(3) (2010) 197–210.
50. M. H. Rafiei and H. Adeli, A new neural dynamic classification algorithm, *IEEE transactions on neural networks and learning systems* **28**(12) (2017) 3074–3083.
51. D. R. Pereira, M. A. Piteri, A. N. Souza, J. P. Papa and H. Adeli, Fema: a finite element machine for fast learning, *Neural Computing and Applications* **32**(10) (2020) 6393–6404.
52. K. M. R. Alam, N. Siddique and H. Adeli, A dynamic ensemble learning algorithm for neural networks, *Neural Computing and Applications* **32**(12) (2020) 8675–8690.
53. F. Hu, H. Wang, Q. Wang, N. Feng, J. Chen and T. Zhang, Acrophobia quantified by eeg based on cnn incorporating granger causality., *International Journal of Neural Systems* (2020) 2050069–2050069.