IJCNN 2019. International Joint Conference on Neural Networks. Budapest, Hungary. 14-19 July 2019

Decoding of finger activation from ECoG data: a comparative study

Guillaume Jubien
LETI, CLINATEC
Univ. Grenoble Alpes, CEA,
Grenoble, France
guillaume.jubien@phelma.grenoble-inp.fr

Marie-Caroline Schaeffer
LETI, CLINATEC
Univ. Grenoble Alpes, CEA,
Grenoble, France
marie-caroline.schaeffer@centraliens-nantes.net

Stéphane Bonnet
LETI, DTBS, SEIVI, LS2P
Univ. Grenoble Alpes, CEA,
Grenoble, France
stephane.bonnet@cea.fr

Tetiana Aksenova
LETI, CLINATEC
Univ. Grenoble Alpes, CEA,
Grenoble, France
tetiana.aksenova@cea.fr
https://orcid.org/0000-0003-4007-234

***Abstract*— Motor Brain-Computer Interfaces (BCIs) are systems that allow severely motor-impaired patients to use their brain activity to interact with their environment. Electrocorticography (ECoG) arrays may be profitably used to develop safe and chronic motor BCI systems. BCI signal processing pipelines generally include neuronal signal pre-processing, feature extraction and classification/regression. The article presents a comparative study addressing the problem of neural feature classification in asynchronous multi-limb ECoG-driven BCIs. Several conventional classifiers often reported in the BCI literature were coupled with two preprocessing techniques and with a conventional feature extraction approach. They were compared to artificial neural network (ANN) end-to-end classifiers which mimic conventional BCI signal processing pipelines. Different initializations of ANNs were particularly studied. The comparison study was carried out using publicly available datasets (BCI competition IV).**

Keywords*— *Brain-Computer Interface, Electrocorticography, ANN, Deep Learning, Machine Learning

I. Introduction

Motor Brain-Computer Interfaces (BCIs) are systems that allow severely motor-impaired patients to use their brain activity to control external devices, and thereby to interact with their environment [1]. BCI systems estimate the user's intention from measurements of the brain activity and convert it into commands to drive the complex effector(s), e.g. upper- or lower-limb orthoses or prostheses.

To date, motor BCIs mostly rely on the measurement of electrophysiological signals of brain activity using intracortical Microelectrode Arrays (MEAs), intracranial Electrocorticography (ECoG) and noninvasive Electroencephalography (EEG) [2]. Depending on the invasiveness of the acquisition systems, sensors are located at a distance which ranges from a few μm (MEA), about 1.25mm for subdural and 1.4mm for epidural ECoG to several cm (EEG) from the cortical neurons generating the extracellular currents of interest [3], [4]. The resulting spatial and spectral characteristics of the acquired signals directly impact the feasibility of accurate multi-limb, multi- Degrees of Freedom complex effector control. In spite of promising results, the invasiveness of intracortical MEA arrays is for the moment a significant hindrance to their utilization for chronic clinical BCI because it induces immune reactions. The performances of EEG-based motor BCIs appear be limited by EEG low spatial resolution and frequency content. The ECoG invasiveness is limited when compared to the MEA's one because ECoG arrays are positioned on the cortex instead of being introduced into it. Chronic ECoG acquisition has been reported in a study completed with epileptic patients [5]. In addition, the wireless ECoG implant WIMAGINE® has recently been designed for long-term BCI applications [6]. These studies suggest that ECoG arrays may be profitably used to develop safe and chronic motor BCI systems.

Several processing steps are usually carried out by conventional BCI systems to translate the user's neural activity into commands for effector control, e.g. [2]. Features specifically related to the user's intentions are first extracted from measurements of the user's brain activity. A decoder is then applied to estimate the user's intention from these brain features. After being optionally enhanced by post-processing algorithms, this estimate is converted into commands used to drive the BCI effector(s), e.g. upper- or lower-limb orthoses or prostheses.

Classification of mental tasks is mostly used in noninvasive EEG-driven BCIs. In more invasive ECoG- and MEA-driven BCIs, both mental task classification and direct kinematic decoding were reported. Kinematic decoders extract continuously-valued kinematic parameters from the neural signals. Continuous and discrete decoders are often combined for asynchronous multi-limb effector control. Asynchronous BCIs are applied during both active and resting states, which are hereinafter referred to as Intentional Control (IC) and Non-Control periods (NC). Asynchronous multi-limb BCIs switch between NC and possibly multiple IC states, which are for example related to particular limbs activation. Switching/gating of state-specific kinematic decoders is generally based on classification algorithm(s). The respective interest of different classifiers for motor BCIs is still unclear and ultimately depends on the problem at hand. In particular, to the best of our knowledge, only limited comparative studies have been completed on ECoG data [7]. The article presents a comparative

978-1-7281-2009-6/$31.00 ©2019 IEEE paper N-20139.pdf

study addressing the problem of classification in asynchronous multi-limb ECoG-driven motor BCIs. Several conventional classifiers, which use is frequently reported in BCI research articles (namely, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), linear/nonlinear Support Vector Machine (SVM) and Logistic Regression (LR)), were coupled with two preprocessing techniques (Principal Component Analysis (PCA), and Partial Least Squares (PLS)) and with a conventional feature extraction approach based on Complex Continuous Wavelet Transform (CCWT). These standard machine learning (ML) algorithms were compared to end-to-end Artificial Neural Network (ANN) classifiers. The architecture of the proposed ANN mimics conventional BCI signal processing pipelines including spatial and temporal filtering, feature extraction and classification. Different initializations of ANNs were particularly studied.

This comparison study was carried out using publicly available datasets (BCI competition IV, dataset 4).

II. Materials and methods

A. Datasets and Preprocessing

The dataset contains ECoG recordings from three epileptic patients implanted for clinical monitoring with either 8x8 or 8x6 subdural electrode grids placed over parts of their sensorimotor cortex. During the experiment, subjects were repeatedly asked to execute a few movements with the finger specified via a visual cue. ECoG signals were filtered between 0.15 and 200 Hz and sampled at 1 kHz. In parallel, a data glove recorded each finger movement at 25 Hz. The motion signals were upsampled to 1 kHz. Each dataset contains about 600 s of ECoG recordings. A detailed description of the dataset can be found in [8]. Some channels were removed after visual inspection because they were corrupted by artifacts: #55 for subject 1, #21 and #38 for subject 2 and #50 for subject 3. A specific class was attributed to each finger activation: thumb (#1), index finger (#2), middle finger (#3), ring finger (#4) and little finger (#5). The 0-class was chosen to encode NC periods. The number of samples associated with each one of the 6 classes ranges from 10 to 30% of the total number of samples recorded for each subject. Input data were chronologically separated into a training, a validation, and a test dataset (30%, 17% and 33% of the full record, respectively).

B. Conventional Methods

Five classifiers, namely LDA, QDA, LR, linear and non-linear SVMs, were selected because of their frequent utilization in BCI studies and/or because the high decoding performance they reached in previous ECoG studies [9].

1) Generative classifiers: As generative classifiers, LDA and QDA model the way independent variables are generated within a class c, i.e. they model the joint probability $P(\mathbf{x}_n, y_n = c)$ for class c [10], where $\mathbf{x}_n$ and y_n refer to the n^{th} realization of the input (explanatory) variable and to the corresponding class label, respectively. Once the joint probability has been fitted for each class, the classification of a new observation sample $\mathbf{x}$ is performed by computing the posterior probability $P(y = c|\mathbf{x})$ with respect to each class. Using Bayes rule, this posterior probability is proportional to $P(\mathbf{x}|y = c) \cdot P(c)$ for class c. The most likely class label is assigned to the considered observation sample. The characteristics of the decision boundary are not explicitly chosen, but result from the distribution used to model data generation within each class.

Maximum Likelihood training of the parameters $\{\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c\}$ associated with the multivariate Gaussian distributions $P(\mathbf{x}|y = c) = N(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$, where $c = 0 \dots 5$, was performed on the training set. Contrary to QDA which relies on class-specific covariance matrices, the LDA covariance matrix $\boldsymbol{\Sigma}_c = \boldsymbol{\Sigma}$ is identical for all classes. Equiprobable class prior probabilities were utilized for both the LDA and QDA classifiers.

2) Discriminative classifiers: Logistic Regression and SVM: By contrast, LR and SVMs are discriminative classifiers. Discriminative classifiers directly model the posterior class probability $P(y_n = c|\mathbf{x}_n)$ [11], and do not model the intermediary probability $P(\mathbf{x}_n|y_n = c)$. This makes the use of discriminative classifiers advantageous when this distribution cannot be approximated with generic distributions. Discriminative classifiers are particularly relevant when $\mathbf{x}_n$ is high-dimensional or includes redundant (correlated) neural features [12], because non-discriminant features are not considered during model training.

a) Linear and nonlinear SVMs: Following [13], the nonlinear SVM was built using a Radial Basis Function (RBF) kernel. The *one-against-one* strategy was utilized to perform multi-class SVM classification because it has been reported as generally more efficient than the *one-against-all* strategy [14] and has been used in several BCI studies [13]. One SVM was thus trained for each possible class pair (i, j) $i \neq j$, and an Error-Correcting Output Code was used to combine the class estimate yielded by each binary classifier [15]. This approach consists in associating each class with a specific combination of the SVM outputs. During application, of the multi-class classifier, the output of each SVM is computed for the considered observation sample. The class which corresponds to the combination of SVM outputs which is the closest to the one obtained on the observation features (i.e., the one which minimizes the number of different outputs) is attributed to the input variable.

b) Logistic Regression: A multinomial LR was fitted on each session of the dataset. The Maximum Likelihood (ML) estimator of LR parameters is known to be unstable when it is applied to a training dataset with linearly separable classes [11]. Penalized LR training has been found to be efficient in several BCI studies, *e.*g. [13], [16]. Three training approaches were thus compared for the identification of LR and multinomial LR models: ML [16], LASSO (Least Absolute Shrinkage and Selection Operator) [18], and Elastic Net training [19]. LASSO training permits to stabilize parameter estimation in high dimensional settings. It relies on L1 penalization to introduce the prior knowledge that few parameters are non-zero. Elastic Net penalization, on the other hand, linearly combines L1 and L2 penalization. Integrating L2 penalization is particularly advantageous when the input variable contains highly correlated features. The amounts of L1 and L2 regularization were found by 6-fold cross-validation

3) Dimension reduction : BCI classifier performances depend on the characteristics of the input variable, in particular on its dimensionality [20]. The respective behaviors of the 5

classifiers were thus compared after different unsupervised and supervised dimensionality reduction procedures had been applied on the input variable.

a) PCA-based dimensionality reduction: PCA-based dimensionality reduction is frequently performed in BCI studies, e.g. [20]. PCA-reduced input variables of different sizes were here considered. The number of extracted principal components was chosen so that the resulting reduced variable explained a specific percentage of the original variance, here 60% and 80%, respectively. The corresponding reduced variables are henceforth referred to as "PCA 60" and "PCA 80".

b) PLS-based dimensionality reduction: The Partial Least Square algorithm has been designed for the identification of regression models. As such, it has been profitably applied for kinematic decoding in BCI studies [16], [21], [22]. PLS-based dimensionality reduction combined with classification has nevertheless been shown to be efficient for the decoding of discrete variables. Satisfying classification accuracies have been reported when such combinations have been utilized in BCI studies, e.g. [23], [24]. PLS-based projectors were therefore fitted to perform supervised dimensionality reduction before applying the considered generic classifiers. The first F PLS scores of $\mathbf{x}_n$, obtained by regressing the class dummy variable y_n against the neural features $\mathbf{x}_n$, were fed to the classifiers. F, i.e. the optimal number of PLS factors, was found using Wold's R criterion [25] with 6-fold cross-validation on the training dataset. The resulting reduced input variables are denoted by PLS_1 in the following sections. The PLS_1 reduced input variables and the corresponding classifiers were fitted on the same datasets. Another training strategy consists in using independent datasets to train the PLS projectors and the classifiers, thereby reducing the risk of overfitting. A reduced variable denoted by "PLS_2" was therefore built by training a PLS model on the first 40% of the training dataset. The second part of the training dataset (60%) was used to fit the classifiers.

4) Feature extraction : A conventional time-frequency feature extraction procedure (e.g. [22]) was used to compute the features fed to the conventional decoding methods. A neural feature vector (explanatory variable) was extracted from each 1s-long epoch of ECoG recording. For each time moment, the ECoG signal from each channel was mapped by the CCWT (Morlet wavelet) with 15 frequencies (10, 20, …, 150 Hz). Then the magnitude of the complex wavelet coefficients were decimated 100 times along the temporal modality. Thus 1000 points, representing 1 second, were decimated into 10 points to form the vector of observation $\mathbf{x}_n$, for the epoch n, of the explanatory variable $\mathbf{x} \in \mathbb{R}^{15 \cdot 10 \cdot N_c}$ where N_c is the number of ECoG grid electrodes.

C. Neural Network Architecture

The pipeline depicted in Fig. 1a is widely used in ECoG signal decoding [26], [27]. Spatial filters estimate source signals (spatial components). The time-domain filtering decomposes each spatial component into frequency subbands. The bandpower is finally computed and used as feature for classification/regression. A custom ANN architecture was chosen to reproduce the standard decoding pipeline (Fig. 1b)

N_c channels' ECoG signals were split into 1s-windows (with N_t points) with a sliding step of 40 ms. Windows of 1 second are widely used for ECoG decoding [28], [29] and the 40 ms time shift was chosen because of the initial data glove's sampling rate (25 Hz). N_e is the number of examples computed in the whole set. Each ECoG epoch is mean-centered. The class associated to an epoch is the one associated with its endpoint.

The first transposed convolution layer performs the spatial filtering step and estimates $N_s \leq N_c$ spatial components. The second convolution layer performs the temporal filtering with N_f time-domain filters. Finally, the average pooling layer computes the bandpower. These successive convolutional layers compute features in an end-to-end manner, and the last layers attribute the corresponding class.

1) Spatial convolution layer: This layer aims to find the optimal linear combinations of electrodes for the classification/regression task [30]. The input signal is multiplied by a $(N_s, 1, N_c)$ 3D matrix to compute spatial components. The transposed convolution layer reshapes the data matrix $(1, N_t, N_c) \rightarrow (N_s, N_t, 1)$, for accommodation reason in the next layer.

An unmixing matrix computed by FastICA [31] was used to initialize the spatial weights. FastICA permits to determine and isolate independent sources. FastICA was chosen according to a comparative study of performance of blind source separation algorithms for electrical brain neuronal activity recordings [32], and because of its frequent utilization in BCI studies. For feature reduction, it is necessary to select a subset of spatial components. The most discriminative spatial filters were selected for initialization using a LASSO regression [33].

2) Temporal convolutional layer: This layer uses temporal filters to realize a time-based convolution of the spatial components [34]. During training, the network learns to find the best appropriate weights (filter coefficients) to select the most informative frequency subbands. One filter is defined by $(1, F)$ matrix, where F is the number of filter coefficients. The layer uses each filter on data from the previous layer. Each computation was represented along the 3rd dimension $(N_s, N_t, 1) \rightarrow (N_s, N_t, N_f)$.

Biorthogonal wavelet 6.8 filters were chosen for filter initialization. Xie *et al.* [28] used three successive convolutional layers to mimic Discrete Wavelet Transform (DWT) at three scales. To simplify the architecture only one convolutional layer is used. Filters at each scale were convolved and 8 bandwidths were obtained. Four DWT filters were pruned by studying their cut of frequencies.

3) L2-pooling layers and output layers: These layers were used to calculate the bandpower of the previously extracted subbands. Under non-overlapping windows of 100 ms, L2 norm of signals is computed and the function $ln(1 + x)$ is applied.

A dropout layer $(p = 0.5)$ is used to reduce overfitting. [35]. Fully connected layer flatten the end-to-end features. Finally, softmax layer and output layer assign a class among the 6 possible choices.

.

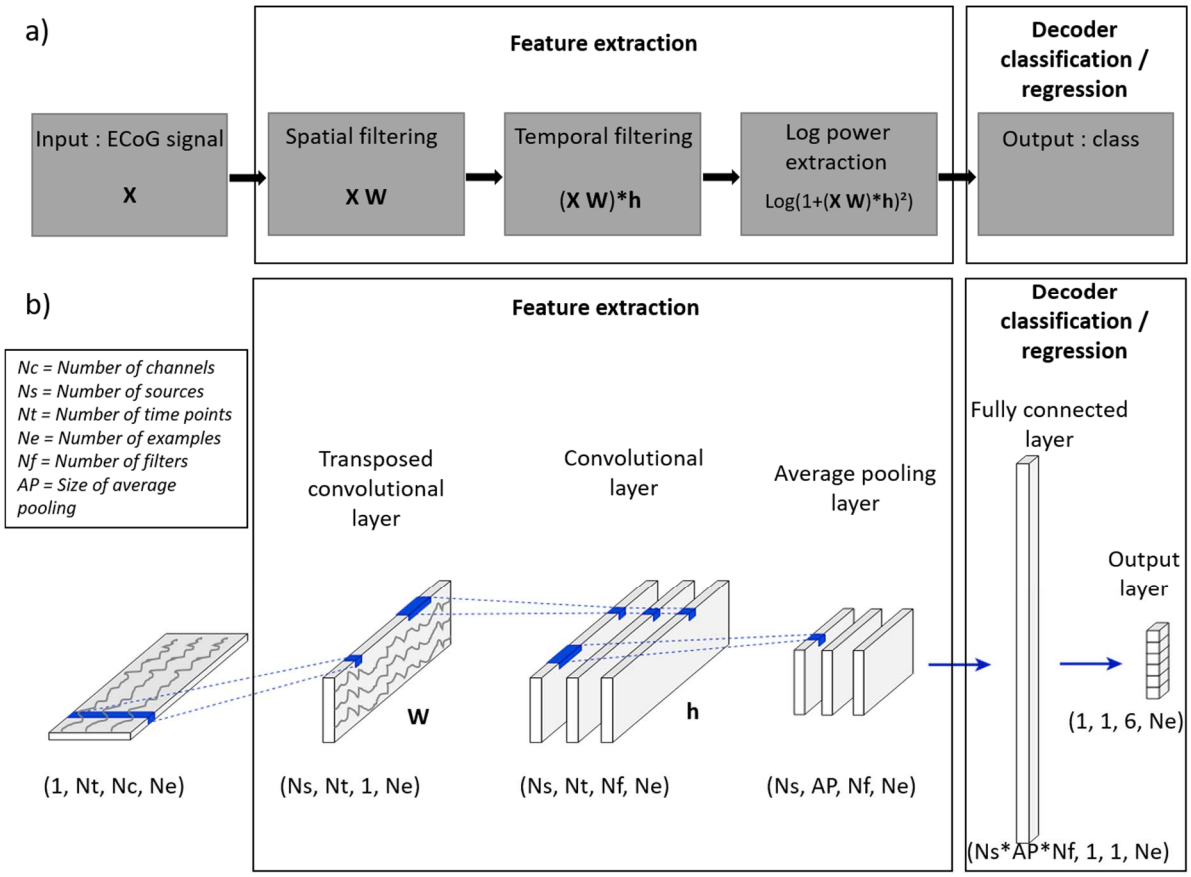

Fig. 1: a) Typical ECoG decoding pipeline. **X** represents the input ECoG epoch with size *Nt* x *Nc*, where *Nt* is the number of time points and *Nc* is the number of channels. The first block of feature extraction realizes a linear combination of channels. Then, the second block decomposes the signal into frequency subbands with time-domain filters **h**. The third block computes the log power of these signals. Finally, after the optional decimation, the extracted features are used by decoder to predict the output class. b) Proposed ANN architecture. The transposed convolutional layer performs spatial filtering and a data reshape, **W** is the spatial weight matrix. Then, a convolutional layer applies time-domain filtering with *Nf* filters, **h** represent filter weights. Average pooling layers gather "squareFct", "average pooling" and "L2Log" that compute the log bandpower features over a kernel of size poolSize. *AP* value is defined as 1000/*poolSize*. Finally, fully connected, softmax and output layers compute the output class.

D. Applications

Decoding performance was evaluated thanks to the accuracy criterion (ACC). True Positive (TP) value is the number of correct predictions. Accuracy is defined by the sum of TP values over the 6 classes, divided by the total number of examples ($Ttot$).

$$ACC = \frac{1}{Ttot} \sum_{i=0}^{5} TP_{class\ i} \quad (1)$$

Decoding performances of ANN were compared to accuracies obtained by conventional methods described in section II.B.

Influence of initialization was specially studied. Three decoding models were developed with different spatial and temporal initializations of ANN. The initialization using FastICA and bior6.8 DWT was compared to random initialization of special and temporal convolutional layers:

- Initialization 1: Spatial Filtering (SF) - Unmixing matrix of FastICA; Temporal filtering (TF) - bior6.8 DWT.
- Initialization 2: SF - random; TF – DWT.
- Initialization 3: SF - Unmixing matrix of FastICA; FT: random.

Spatial weights before and after training were analyzed to determine which channels were mostly selected by the neural network during the training. Finally, temporal weights and specifically the filters' bandwidth were studied after training. The aim was to determine if frequencies selected by the network correspond to frequencies referenced in the literature.

III. Results

A. Decoding performances

Table 1 presents decoding performances (ACC) for 6-class classification of finger activation from ECoG data of 3 subjects using conventional methods with different dimension reduction approaches and ANN with 3 different initialization. Grey cells indicate dimensionality reduction/classifier pairs which were discarded because the classifier was ill-suited for the corresponding feature dimensionality. In all cases, the best performance was achieved using full dataset. Comparing different dimension reduction methods, PLS1 mostly demonstrates best results. Dimension reduction may be

especially important for real time application and/or computing power limitations. Among conventional methods LR-LASSO or -Elastic net provided best results for all subjects. Among ANN, initialization 1 and 2 are suggested to be preferable. More experiments are needed to conclude the significance of difference.

B. Spatial weights in Neural Network

The two first columns in Fig. 2 illustrate normalized spatial weights after ANN training for all subjects. The same initialization (DWT) was used for temporal convolutional layer. Spatial convolutional layer was initialized by fastICA (Initialization 1) and random values (Initialization 2) respectively. Tables in Fig. 2 list the most frequently selected electrodes for a given subject. We remark that highlighted electrodes are quite similar for each subject, whatever the spatial initialization is.

C. Temporal weights in Neural Network

Frequencies selected by the network during training were studied for initialization 1 and 3. The same spatial initialization (fastICA) is coupled with bior6.8 DWT and random initialization for temporal convolutional layer. Fig. 3 represents the frequency responses of the four filters before and after training. For initialization 1, the network does a few modifications during training. Nevertheless, results suggest that it learned to put forward the following frequencies in filter 1: 21.5, 38.1 and 55.7 Hz. However, for initialization 3 (temporal weight initialized randomly) the network performs several modifications. It attenuates frequencies above 150 Hz and selects frequencies around 18, 34, 44 and 54 Hz.

TABLE I. DECODING PERFORMANCES (ACC) FOR CLASSIFICATION OF FINGER ACTIVATION FROM ECOG DATA USING CONVENTIONAL METHODS AND ANN WITH 3 DIFFERENT INITIALIZATIONS. GREY CELLS INDICATE DIMENSIONALITY REDUCTION/CLASSIFIER PAIRS WHICH WERE DISCARDED BECAUSE THE CLASSIFIER WAS ILL-SUITED FOR THE CORRESPONDING FEATURE DIMENSIONALITY

		PCA60	PCA80	PLS1	PLS2	Full
Subject 1	LDA			0,673	0,273	
	QDA			0,71	0,262	
	LR-ML			0,654	0,566	
	LR-LASSO	0,622	0,629	0,663	0,638	0,69
	LR-ElasticNet	0,628	0,631	0,66	0,633	**0,691**
	linear SVM	0,592	0,6	0,569	0,539	0,59
	rbf SVM			0,628	0,541	
	ANN init. 1	/	/	/	/	0,583
	ANN init. 2	/	/	/	/	0,574
	ANN init. 3	/	/	/	/	0,479
Subject 2	LDA			0,486	0,307	
	QDA			0,444	0,315	
	LR-ML			0,438	0,384	
	LR-LASSO	0,437	0,442	0,478	0,433	0,547
	LR-ElasticNet	0,452	0,442	0,464	0,435	0,55
	linear SVM	0,423	0,45		0,145	0,452
	rbf SVM			0,5	0,203	
	ANN init. 1	/	/	/	/	0,67
	ANN init. 2	/	/	/	/	**0,675**
	ANN init. 3	/	/	/	/	0,551
Subject 3	LDA			0,627	0,205	
	QDA			0,628	0,153	
	LR-ML			0,583	0,495	
	LR-LASSO	0,554	0,598	0,621	0,508	**0,685**
	LR-ElasticNet	0,551	0,612	0,625	0,51	**0,685**
	linear SVM	0,512	0,542	0,659	0,304	0,557
	rbf SVM			0,626	0,354	
	ANN init. 1	/	/	/	/	0,673
	ANN init. 2	/	/	/	/	0,67
	ANN init. 3	/	/	/	/	0,575

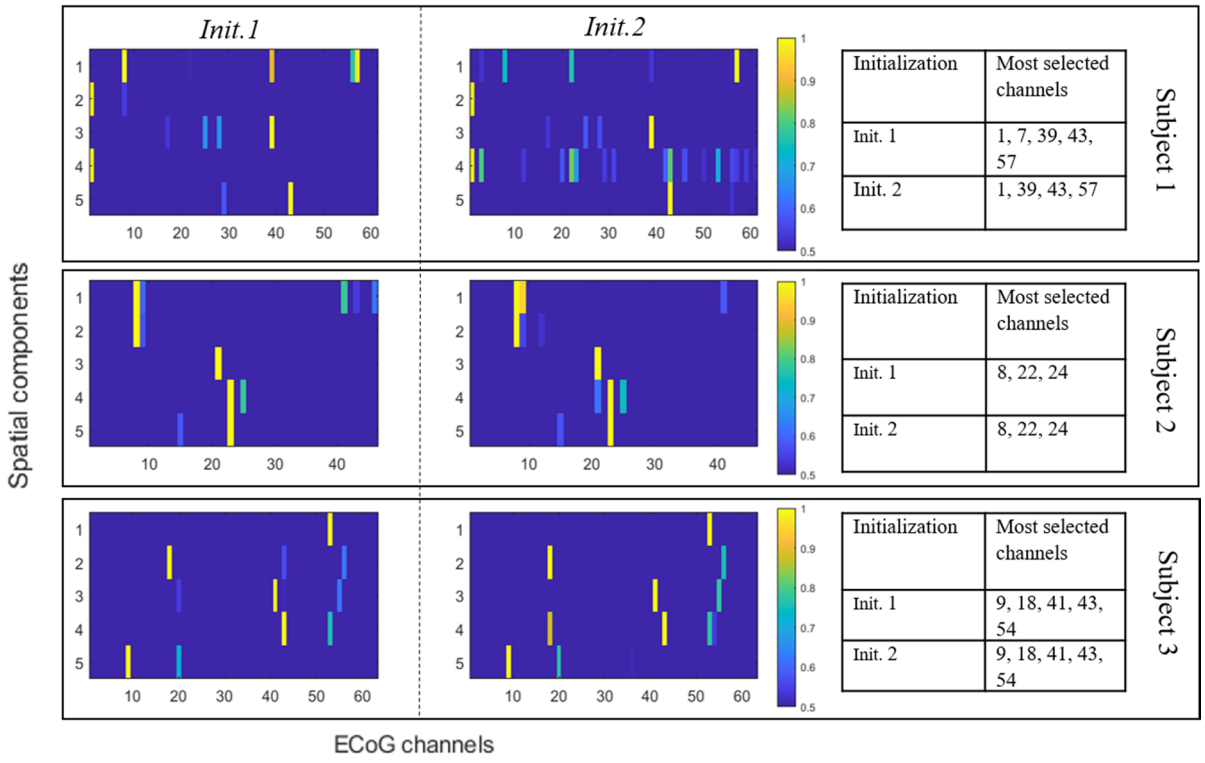

Fig. 2: Spatial weights after training for initialization 1 and 2, for all subjects. Initialization 1 and 2 have the same temporal initialization (WT) but are respectively initialized by fastICA and random values for spatial initialization. Weights were normalized along sources. Tables contain lists of electrodes with a normalized weight after training.

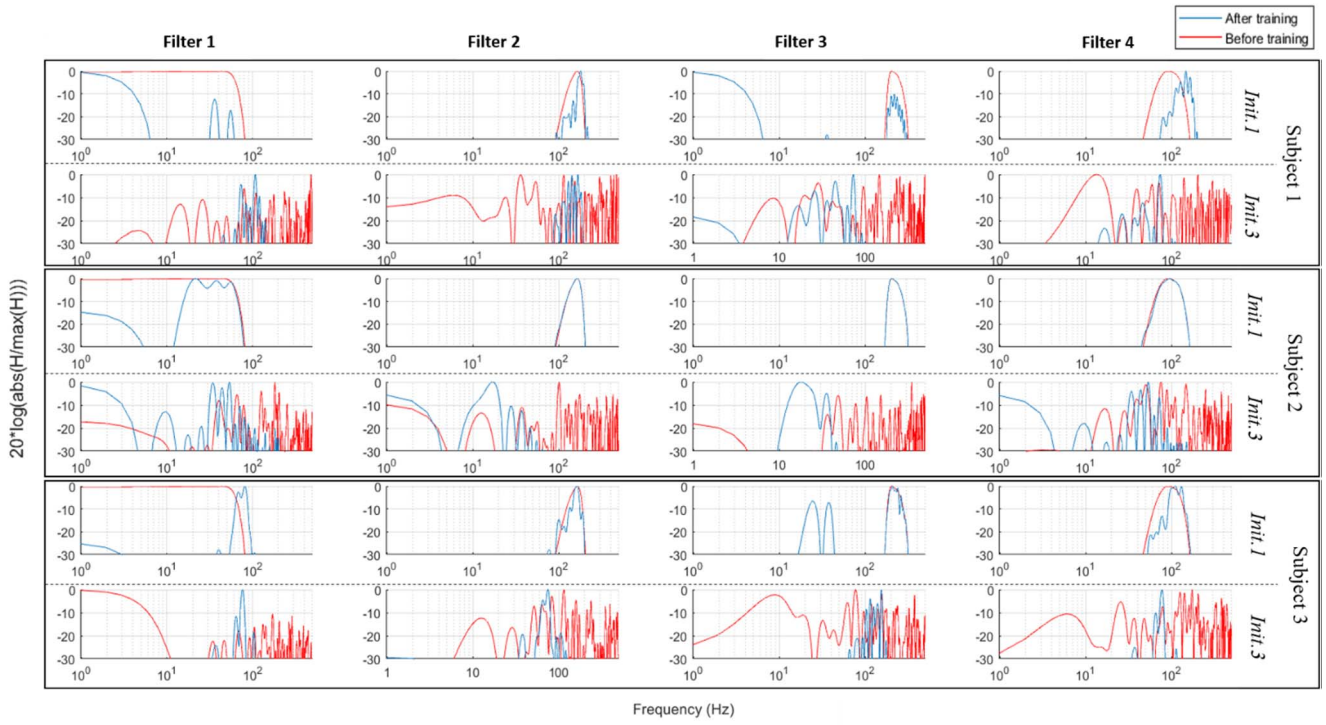

Fig. 3: Frequency reponses of four FIR filters before and after training (respectively the red and the blue curves) for all subjects. Initialization 1 and 3 have the same spatial initialization (fastICA) but are respectively initialized by bior6.8 wavelet and random time-filter coefficients.

IV. Discussion

This work addresses the problem of ECoG signal decoding by a deep learning (DL) approach in the context of BCI study. DL is widely used for image classification but still poorly explored for brain signal, and, particularly, ECoG signal decoding. The presented ANN mimics conventional BCI pipelines in an end-to-end decoding manner. Several simplifications and optimizations were conducted to obtain a simple ANN to be trained with highly restricted dataset (several minutes of recording). Spatial component reductions and filter suppressions were used to reduce features and prevent overfitting. In general, DL architecture is considered as "black box" and rarely interpreted in the literature. The major improvement of the current study is the interpretable weights evolution. Our work suggests that DL may be used to optimize neural feature extraction and ANN weights can help to understand which features characterize limbs movements for a subject.

On average over subjects, Linear SVM, LR-ElasticNet and LR-LASSO demonstrate 53.3% ± 7.2, 64.2% ± 7.8 and 64.1 % ± 8.1 of correct decoding, respectively. The ANN model obtains similar accuracy 64.2 % ± 5.1 with respect to these standard methods. In particular, our model seems particularly adapted for decoding subject 2 and less effective for subject 1. The short recording and, consequently, the simple ANN architecture may explain the moderate efficiency of DL compared to conventional ML approaches for the considered decoding task. Additional studies on larger datasets are required to further explore the decoders' performance and confirm the present findings.

It is known that deep learning is sensitive to initialization. On the contrary, in the current study, decoding performances vary only slightly according to the initialization. In particular, spatial weight initializations appear to have no influence on decoding performances. It is worth noting that spatial filtering is a key ingredient in EEG/ECoG decoding in BCI. The spatial filters are usually found in a data-driven way using PCA, ICA or CSP [36]. In CSP, the class label is taken into account to find the spatial filters that maximize the class-related signal variance. This variance becomes the feature of interest to be used in a second-stage classification. In unsupervised ICA and PCA, some assumptions are made on the source component statistics, without considering the class label. The proposed deep-learning-based approach suggests an alternative to conventional methods for supervised optimization of spatial filters. It allows computing spatial filters in an end-to-end decoding process, simultaneously learning temporal/spatial filters and the classifier. The spatial filters estimated using the proposed approach are quite robust to the weight initialization. Moreover, spatial weights identified by DL are in a good agreement with electrodes already identified in the literature and by the LASSO algorithm. This strategy seems thus adequate for identifying source components discriminant for motor tasks.

The frequencies picked up by the network during training seem to rely on the initialization. The first one using bior6.8 DWT led to a few modifications contrary to random initialization. However, some frequencies are common whatever the initialization is (e.g. filter 1 for subject 3 or filter 2 for subject 1). With the random initialization, we observed a consistent frequency selection, especially between 10 and 100 Hz, coherent with the BCI ECoG literature. Moreover, higher frequencies (<200Hz) were cut, which is coherent with collected data (data are filtered between 0.15 and 200 Hz).

This article presents both conventional classifiers frequently used in BCI studies and end-to-end ANN classifiers. The ability of these decoders to solve the classification of neural features in asynchronous multi-limb ECoG-driven motor BCI is compared. Further investigations are necessary to strengthen and, possibly, extend the study. Future work will consist in optimizing and testing conventional and DL decoders with clinical data collected in the frame of ECoG-based long-term BCI [6] clinical research protocol "BCI and Tetraplegia" (Principal Investigator Prof. A.-L. BENABID, https://clinicaltrials.gov/ct2/show/NCT02550522?term=clinatec).

Acknowledgment

This work received financial support through grants from the French National Research Agency (ANR-Carnot Institute), Fondation Motrice, Fondation Nanosciences, Fondation de l'Avenir, Fonds de dotation Clinatec, ASSYSTEM, KLESIA, and Fondation Philanthropique Edmond J. Safra.

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin Neurophysiol*, vol. 113, no. 6, pp. 767–791, Jun. 2002.

[2] M.-C. Schaeffer and T. Aksenova, "Data-Driven Transducer Design and Identification for Internally-Paced Motor Brain Computer Interfaces: A Review," *Frontiers in Neuroscience*, vol. 12, Aug. 2018.

[3] S. Waldert, T. Pistohl, C. Braun, T. Ball, A. Aertsen, and C. Mehring, "A review on directional information in neural signals for brain-machine interfaces," *Journal of Physiology-Paris*, vol. 103, no. 3–5, pp. 244–254, May 2009.

[4] G. Schalk and E. C. Leuthardt, "Brain-Computer Interfaces Using Electrocorticographic Signals," *IEEE Reviews in Biomedical Engineering*, vol. 4, pp. 140–154, 2011.

[5] E. S. Nurse *et al.*, "Consistency of Long-Term Subdural Electrocorticography in Humans," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 2, pp. 344–352, Feb. 2018.

[6] C. S. Mestais, G. Charvet, F. Sauter-Starace, M. Foerster, D. Ratel, and A. L. Benabid, "WIMAGINE: Wireless 64-Channel ECoG Recording Implant for Long Term Clinical Applications," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 1, pp. 10–21, Jan. 2015.

[7] P. Shenoy, K. J. Miller, J. G. Ojemann, and R. P. N. Rao, "Generalized Features for Electrocorticographic BCIs," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 273–280, Jan. 2008.

[8] G. Schalk *et al.*, "Decoding two-dimensional movement trajectories using electrocorticographic signals in humans," *Journal of Neural Engineering*, vol. 4, no. 3, pp. 264–275, Sep. 2007.

[9] F. Lotte *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, Jun. 2018.

[10] A. Y. Ng and M. I. Jordan, "On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive

Bayes," in *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. MIT Press, 2002, pp. 841–848.

[11] C. M. Bishop, *Pattern recognition and machine learning*, vol. 4. New York: Springer, 2006.

[12] C. Sutton, "An Introduction to Conditional Random Fields," *Foundations and Trends® in Machine Learning*, vol. 4, no. 4, pp. 267–373, 2012.

[13] H. Bashashati, R. K. Ward, G. E. Birch, and A. Bashashati, "Comparing Different Classifiers in Sensory Motor Brain Computer Interfaces," *PLoS ONE*, vol. 10, no. 6, p. e0129435, 2015.

[14] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[15] T. G. Dietterich and G. Bakiri, "Solving Multiclass Learning Problems via Error-correcting Output Codes," *Journal of Artificial Intelligence Research*, vol. 2, no. 1, pp. 263–286, Jan. 1995.

[16] D. T. Bundy, M. Pahwa, N. Szrama, and E. C. Leuthardt, "Decoding three-dimensional reaching movements using electrocorticographic signals in humans," *Journal of Neural Engineering*, vol. 13, no. 2, p. 026021, Apr. 2016.

[17] S. A Czepiel, "Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation," 2012.

[18] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *J Stat Softw*, vol. 33, no. 1, pp. 1–22, 2010.

[19] H. Zou, "Regularization and variable selection via the elastic net," 2004.

[20] S. Bhattacharyya, A. Khasnobish, A. Konar, D. N. Tibarewala, and A. K. Nagar, "Performance analysis of left/right hand movement classification from EEG signal by intelligent algorithms," in *2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, Paris, France, 2011, pp. 1–8.

[21] K. Shimoda, Y. Nagasaka, Z. C. Chao, and N. Fujii, "Decoding continuous three-dimensional hand trajectories from epidural electrocorticographic signals in Japanese macaques," *Journal of Neural Engineering*, vol. 9, no. 3, p. 036015, Jun. 2012.

[22] A. Eliseyev and T. Aksenova, "Stable and artifact-resistant decoding of 3D hand trajectories from ECoG signals using the generalized additive model," *Journal of Neural Engineering*, vol. 11, no. 6, p. 066005, Dec. 2014.

[23] A. Eliseyev *et al.*, "Iterative *N* -way partial least squares for a binary self-paced brain–computer interface in freely moving animals," *Journal of Neural Engineering*, vol. 8, no. 4, p. 046012, Aug. 2011.

[24] A. Eliseyev *et al.*, "L1-Penalized N-way PLS for subset of electrodes selection in BCI experiments," *Journal of Neural Engineering*, vol. 9, no. 4, p. 045010, Aug. 2012.

[25] B. Li, J. Morris, and E. B. Martin, "Model selection for partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 64, no. 1, pp. 79–89, Oct. 2002.

[26] Z. Y. Chin, K. K. Ang, C. Wang, C. Guan, and H. Zhang, "Multi-class filter bank common spatial pattern for four-class motor imagery BCI," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, MN, 2009, pp. 571–574.

[27] R. Flamary and A. Rakotomamonjy, "Decoding Finger Movements from ECoG Signals Using Switching Linear Models," *Frontiers in Neuroscience*, vol. 6, 2012.

[28] Z. Xie, O. Schwartz, and A. Prasad, "Decoding of finger trajectory from ECoG using deep learning," *J Neural Eng*, vol. 15, no. 3, p. 036009, Jun. 2018.

[29] N. X. R. Wang, A. Farhadi, R. Rao, and B. Brunton, "AJILE Movement Prediction: Multimodal Deep Learning for Natural Human Neural Recordings and Video," *arXiv:1709.05939 [cs, q-bio]*, Sep. 2017.

[30] Y. Liu, W. G. Coon, A. de Pesters, P. Brunner, and G. Schalk, "The effects of spatial filtering and artifacts on electrocorticographic signals," *Journal of Neural Engineering*, vol. 12, no. 5, p. 056008, Oct. 2015.

[31] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4–5, pp. 411–430, Jun. 2000.

[32] N. Oosugi, K. Kitajo, N. Hasegawa, Y. Nagasaka, K. Okanoya, and N. Fujii, "A new method for quantifying the performance of EEG blind source separation algorithms by referencing a simultaneously recorded ECoG signal," *Neural Networks*, vol. 93, pp. 1–6, Sep. 2017.

[33] R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective: Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 3, pp. 273–282, Jun. 2011.

[34] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," *arXiv:1603.07285 [cs, stat]*, Mar. 2016.

[35] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[36] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, Dec. 2000.