Deep Learning for hybrid EEG-fNIRS Brain-Computer Interface: application to Motor Imagery Classification.

To cite this article before publication: Antonio Maria Chiarelli *et al* 2018 *J. Neural Eng.* in press https://doi.org/10.1088/1741-2552/aaaf82

Deep Learning for hybrid EEG-fNIRS Brain-Computer Interface: application to Motor Imagery Classification

Antonio Maria Chiarelli[1,2*], Pierpaolo Croce[1,2*], Arcangelo Merla[1,2], Filippo Zappasodi[1,2]

E-mail: antonio.chiarelli@unich.it

[1] Department of Neuroscience, Imaging and Clinical Sciences, "G. d'Annunzio University, Chieti, Italy
[2] Institute of Advanced Biomedical Technologies, "G. d'Annunzio" University, Chieti, Italy
[*] The authors contributed equally and are displayed in alphabetical order

Abstract. *Objective.* Brain-Computer Interface (BCI) refers to procedures that link the central nervous system to a device. BCI was historically performed using Electroencephalography (EEG). In the last years, encouraging results were obtained by combining EEG with other neuroimaging technologies, such as functional Near Infrared Spectroscopy (fNIRS). A crucial step of BCI is brain state classification from recorded signal features. Deep Artificial Neural Networks (DNNs) recently reached unprecedented complex classification outcomes. These performances were achieved through increased computational power, efficient learning algorithms, valuable activation functions, and restricted or back-fed neurons connections. By expecting significant overall BCI performances, we investigated the capabilities of combining EEG and fNIRS recordings with state-of-the-art Deep Learning procedures. *Approach.* We performed a guided Left and Right Hand Motor Imagery task on 15 subjects with a fixed classification response time of 1 second and overall experiment length of 10 minutes. Left vs. Right classification accuracy of a DNN in the multi-modal recording modality was estimated and it was compared to standalone EEG and fNIRS and other classifiers. *Main Results.* At a group level we obtained significant increase in performance when considering multi-modal recordings and DNN classifier with synergistic effect. *Significance.* BCI performances can be significantly improved by employing multi-modal recordings that provide electrical and hemodynamic brain activity information, in combination with advanced non-linear Deep Learning classification procedures.

1. Introduction

Brain Computer Interface (BCI) refers to a group of procedures that directly link central nervous system to a computer or a device (Wolpaw et al. 2000). BCI can

focus on mapping, assisting, augmenting, or repairing human cognitive and sensory-motor functions. Historically, BCI was performed using Electroencephalography (EEG) (Lotte et al. 2007) which provides information about brain electrical activity with very high temporal resolution (ms scale) (Hallez et al. 2007).

In the last years, BCI studies investigated the possibility of combining EEG with other neuroimaging technologies (Pfurtscheller et al. 2010). Among these, functional Near Infrared Spectroscopy (fNIRS) provided encouraging results (Fazli et al. 2012b; Shin et al. 2016, 2017).

EEG and fNIRS are both flexible, scalp located procedures that can be employed for monitoring multiple populations in ecological conditions (Costantini et al. 2013; Farroni et al. 2013; Watanabe et al. 1999; Zappasodi et al. 2017). Whereas EEG captures the macroscopic temporal dynamics of brain electrical activity through passive voltages evaluation, fNIRS estimates brain hemodynamic oscillations relying on spectroscopic measurements of oxy- and deoxy-hemoglobin (HbO and Hb, respectively) fluctuations in the cortex (Ferrari et al. 2012; Villringer et al. 1997). Orthogonally with respect to EEG, fNIRS depends on the slow dynamics of the hemodynamic response and it yields low temporal resolution. However, because of the fast exponential decay of light sensitivity, it provides good spatial resolution (around 1 cm) (Chiarelli et al. 2015, 2016).

Because of different characteristics and physiological information provided by EEG and fNIRS (Croce et al. 2017), higher BCI performances of combined measurements with respect to standalone EEG were reported extensively (Chiarelli et al. 2017; Fazli et al. 2012b; Hong et al. 2015; Khan et al. 2014).

Two main processing steps are involved in BCI: feature extraction and classification.

EEG features are usually extracted based on the power of the signal frequency bands. Indeed, well-distinct behaviors of EEG signal have been identified based on signal frequencies [delta ($< 4Hz$), theta ($4 - 7Hz$), alpha or mu ($8 - 15Hz$), beta ($16 - 31Hz$), and gamma ($> 31Hz$) (Nuwer 1988)]. For example, during the execution of a motor task (or during the imagination or observation of the movement), the mu-beta activity is suppressed in related brain areas (Event Related Desynchronization, ERD) (Pfurtscheller 2001). fNIRS features are generally computed from HbO and Hb variations in the brain which are dependent, among others, on the Blood Oxygen Level Dependent (BOLD) effect (Naseer et al. 2015; Steinbrink et al. 2006).

The classification procedure aims to accurately classify the brain state based on the extracted signal features and it is a fundamental step of BCI processing.

Different experimental settings and algorithms have been applied to combined EEG-fNIRS BCI. Fazli et al. (2012a) proposed Linear Discriminant Analysis (LDA) to classify ERDs and time average fNIRS concentration changes during executed movements as well as motor imagery. In Ma et al. (2012) a Gaussian Radial-Basis kernel Support Vector Machine (SVM) was used to classify a motor imagery BCI based on EEG power spectral densities and fNIRS signal amplitudes. Lee et al. (2014) employed LDA on combined EEG and fNIRS features to classify three conditions: right and left motor imagery plus the idle status. They reached a classification accuracy of about 65%. In Buccino et

al. (2016) the features to be submitted to a LDA were extracted combining different metrics: Regularized Common Spatial Patterns (RCSP) for EEG and combination of average and slope indicators for fNIRS signals. In this case an accuracy between $72-79\%$ was reached in a movement recognition task. In Khan et al. (2017, 2014) LDA was used to classify control commands based on EEG peak amplitudes of selected motor area channels and mean values of HbO and Hb for fNIRS with accuracy ranging between $80-95\%$. For all of the above mentioned studies, the authors recognized an increase BCI performance of combined measurements with respect to standalone fNIRS and EEG.

Recently, Deep Learning Classifiers are increasing their popularity. In the simplest fashion, Deep Learning refers to artificial Neural Networks (NN) (LeCun et al. 2015; Schmidhuber 2015) that are composed of many layers. Deep NNs (DNNs) use a cascade of layers of nonlinear processing units (neurons). Each successive layer uses the output from the previous layer as input and all, or part, of the neurons from consecutive layers are connected. DNN can perform very complex, non-linear, transformations-classifications, greatly increasing shallow NN (Bianchini et al. 2014) and other classifiers performances (LDA, SVM, etc.). In fact, they can reach unprecedented classification outcomes when applied to signals (e.g. speech and language processing) and or images (Collobert et al. 2008; Hinton et al. 2012; Krizhevsky et al. 2012; Simonyan et al. 2014). Because of their performances, these algorithms are also receiving attention within the biomedical field (Ciresan et al. 2012; Hudson et al. 2000; Ronneberger et al. 2015). Multiple technological development allowed for Deep Learning evolution. Among them, the increased computation power clearly played an important role. However, the major improvements are related to algorithms improvement and they can be divided in four categories:

- Implementation of efficient learning algorithms that avoid local minima in the objective function and poor generalization (over-fitting) (Kingma et al. 2014); because of the presence of many free parameters (sometimes millions or more), and the possibility to represent very complex functions, DNNs were usually affected by local minima in the objective function and over-fitting during training.
- Development of new Neuron's activation functions (such as Rectified Linear Unit Function, ReLU function (Dahl et al. 2013; Maas et al. 2013)) that dampen the vanishing gradient problem (Pascanu et al. 2013); in fact traditional activation functions such as the hyperbolic tangent or the sigmoid function had wide ranges of the independent variables with small gradients; this aspect, combined with the back propagation algorithm (Hecht-Nielsen et al. 1988), exponentially dampened weight update rate going from the last to the first layers, heavily slowing the overall learning rate of the network.
- Implementation of Neural Networks where neurons are connected to portions of signals and or images that are close in time and/or space (Convolutional Neural Networks, CNNs (Kalchbrenner et al. 2014; Krizhevsky et al. 2012)), encoding

temporal and/or spatial information; standard, full connected DNN did not encode any spatio-temporal information.

- Development of Neural Networks where outputs are fed back into the network in a sequential manner that allow information storage (Recurrent Neural Networks, RNNs (Hochreiter et al. 1997; Mikolov et al. 2010)). Standard, full connected DNN did not provide memory capabilities and sequential information control.

DNN have been successfully applied to both EEG and fNIRS BCI classification problem. In Jirayucharoensak et al. (2014) a DNN was used to classify three levels of valence and arousal based on EEG power spectral densities features. They reached an accuracy of about 50%. Hajinoroozi et al. (2015) employed Deep Belief Network to EEG signals for the classification of driver's cognitive states. In An et al. (2014) Left vs. Right motor imagery classification was performed by employing few EEG recording channels via DNN with an average accuracy of about 80%. Bashivan et al. (2015) trained a CNN using EEG power in three different frequency bands of interest. They reported a best-performance accuracy of about 92%.

Regarding fNIRS, only few studies were performed employing Deep Learning. Hennrich et al. (2015) investigated DNN classification performances of three mental task reporting accuracy values similar to other classification algorithms (such as LDA and SVM). Abibullaev et al. (2011) classified four mental tasks through DNN with an accuracy of 94%. Finally, Nguyen et al. (2013) classified Left vs. Right motor imagery fNIRS activity with average accuracy of 85%. To the best of our knowledge, no studies implementing deep learning algorithm for BCI classifications in a combined EEG-fNIRS framework were performed.

In this paper, by expecting significant overall BCI performances, we investigated the capabilities of combining multi-modal EEG-fNIRS brain recordings with state-of the art Deep Learning classification procedures. As a first investigation step, we performed a guided Left and Right Hand Motor Imagery task (Pfurtscheller et al. 1997) and, by employing a common temporal frame of 1 second between technologies (Govindan et al. 2016), Left vs Right classification accuracy of a DNN in the multi-modal recording modality was estimated and compared to standalone EEG, fNIRS and other classification algorithms.

2. Methods

2.1. Experimental Paradigm

Fifteen healthy subjects (males, average age of 32 ± 5 years) were recruited for the study. All subjects were right handed, reported no history of neurological or psychiatric disease and did not receive psychoactive medications. Subjects sat on a chair with the arms comfortably resting on a desk and were asked to perform right or left hand squeezing imagery guided by an acoustic stimulus. The motor imagery task sequence is depicted in figure 1. The squeezing imagery consisted of 5 seconds of task and 10 seconds of rest.

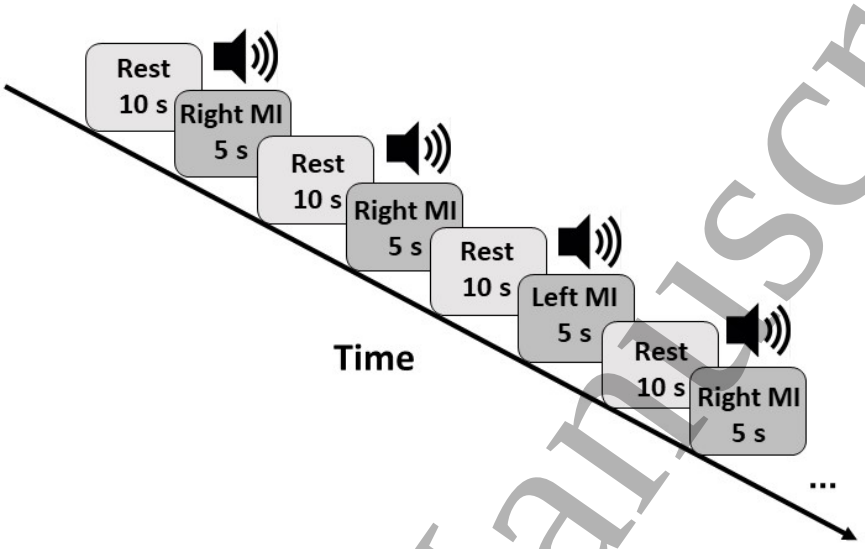

Figure 1. Motor imagery task sequence. The squeezing imagery consisted of 5 seconds of task and 10 seconds of rest. Right or Left imagery instruction was presented in a pseudo-random order. During the 5 seconds of task, the subjects were instructed to perform a hand squeezing imagery with a repetition frequency of $\sim 1Hz$. The task provided a total number of 20 Left-hand and 20 Right-Hand 5 seconds trials, for a total of 40 trials that, by considering a response time of 1 second, provided a total number of 200 training set per subject in an experiment time of 10 minutes.

Right or Left imagery instruction was presented in a pseudo-random order. During the 5 seconds of task, the subjects were instructed to perform the squeezing imagery with a repetition frequency of $\sim 1Hz$. The task provided a total number of 20 Left-hand and 20 Right-Hand 5 seconds trials for a total of 40 trials. We estimated the motor-imagery state every second during each trial that, considering a trial length of 5 seconds, provided a total number of 200 training set per subject in an experiment time of 10 minutes.

2.2. ElectroEncephalography Recordings

Brain electric activity was recorded with a full-head, 128 channels EEG system (Electrical Geodesic Inc, EEG System Net 300, figure 2a,c.). Skin/electrode impedance was measured before recordings and kept below $50k\Omega$. Notice that, for the EEG system employed, $50k\Omega$ is an impedance within the optimal range recommended by

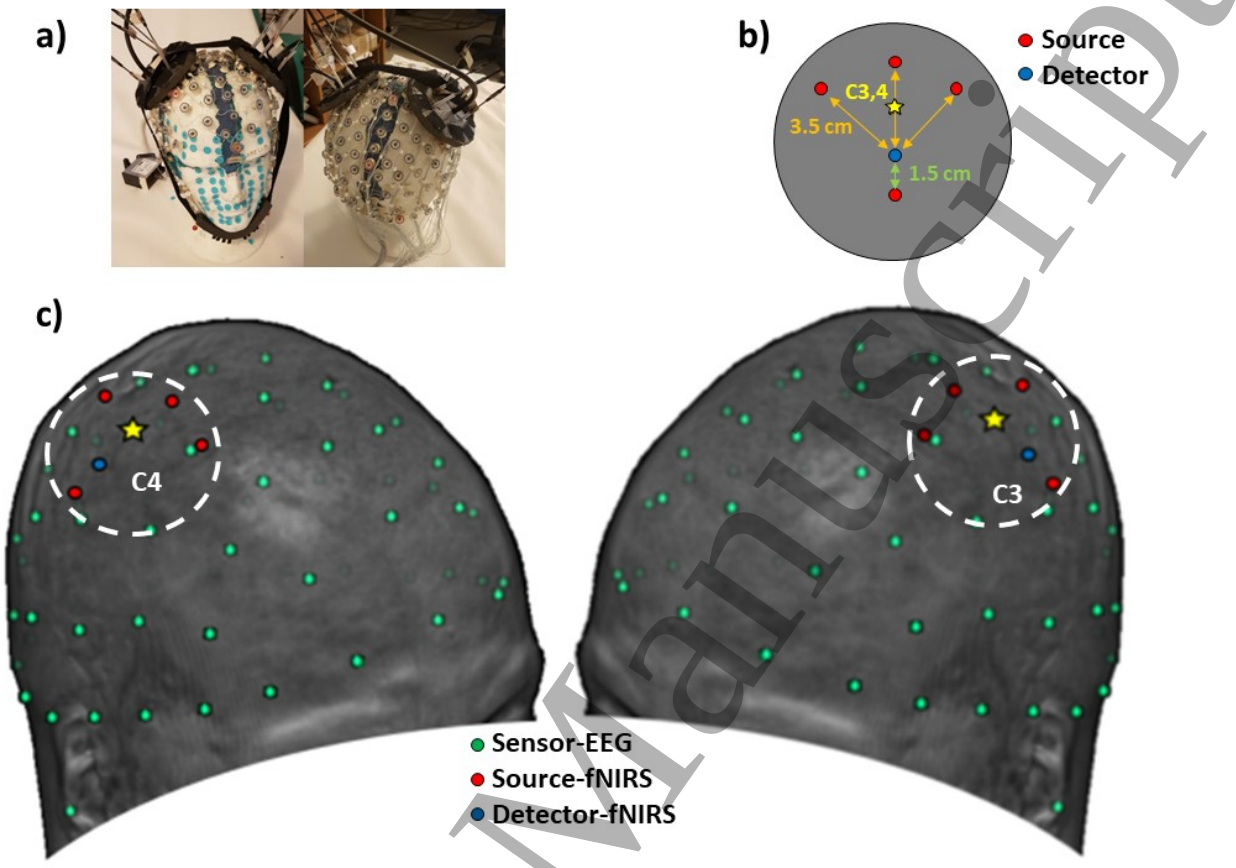

Figure 2. a) Pictures of the full head EEG cap and optical probes located on a dummy head. The EEG was a Electrical Geodesic Net 300 system. The fNIRS spectrometer was an Imagent from ISS. Light was sent to the scalp using multi-modal optical fibers (0.4 mm core) and from the scalp back to the PMTs using fiber bundles (3 mm diameter). The fibers were held in place using soft, but rigid, custom-built optical patches located on top of the EEG. b) Optical layout employed for each hemisphere. The optical layout consisted of 3 fNIRS channels over C3 and C4 with an interoptode distance (3.5 cm) that allowed sensitivity to brain activity. One short distance channel (1.5 cm), not sensitive to brain activity (its sensitivity pattern did not reach the brain cortex), provided information regarding scalp-related hemodynamic oscillations to the classifiers. c) EEG electrodes and fNIRS optodes employed in the study overlayed onto a rendered structural Magnetic Resonance Image of a representative subject.

the manufacturer. In fact, although a sensor-to-scalp impedance below $5k\Omega$ is generally required for EEG recordings, the HydroCel Geodesics Sensor Net provides excellent signals with impedances up to $50-100k\Omega$ thanks to the high input impedance amplifiers (Tucker 1993).

EEG data were sampled at 250 Hz and processed in a real time fashion. Raw data were stored over a 1 second window and filtered between 8 and 30 Hz (within the 1 second window, zero-phase 2nd order Digital Butterworth filter). The mu+beta-band filtered signal was squared and averaged over the second to obtain a power estimate (Pow). The Event-Related Synchronizations (ERSs) or Event-Related Desynchronizations (ERDs)

(Pfurtscheller and Lopes da Silva, 1999; Neuper and Pfurtscheller, 2001) were obtained as relative changes in power during the motor imagery execution with respect to rest:

$$ERD/ERS = \frac{Pow_{MI} - Pow_{Bas}}{Pow_{Bas}} \quad (1)$$

where Pow_{MI} is the average power over 1 second window during the task and Pow_{Bas} is the average power in a 1 second window prior to the task onset. ERD/ERS values for each second during the task were fed to the learning algorithms for classification. Only 123 of the 128 EEG electrodes were employed in the learning process (5 auxiliary signal electrodes were removed).

2.3. functional Near Infared Spectroscopy Recordings

Brain hemodynamic activity was recorded over the sensorimotor regions (C3 and C4, 10-20 System) employing a commercial NIR spectrometer from ISS (Imagent, Champaign, Illinois). The apparatus is a Frequency Domain system equipped with 32 laser diodes ($\sim$1 mW power, 16 emitting light at 690 nm and 16 at 830 nm) and 4 photo-multiplier tubes (PMTs). 8 light sources (4 injection point, 2 wavelengths) and 1 detector were employed for each hemisphere. Time-multiplexing was employed for source coding with a total system sampling frequency of 10 Hz. Light was sent to the scalp using multi-modal optical fibers (0.4 mm core) and from the scalp back to the PMTs using fiber bundles (3 mm diameter). The fibers were held in place using soft, but rigid, custom-built optical patches located on top of the EEG, figure 2a, with an optical layout reported in figure 2b, c. The optical layout consisted of 3 fNIRS channels over C3 and C4 with an interoptode distance (3.5 cm) that allowed sensitivity to brain activity. One short distance channel (1.5 cm), not sensitive to brain activity (its sensitivity pattern did not reach the brain cortex), provided information regarding scalp-related hemodynamic oscillations to the classifiers (Gagnon et al. 2014). The raw Continuous Wave intensity (I) was averaged at a 1 sec pace . The Optical Densities (ODs)over time were computed as:

$$OD = -\ln \frac{I(t)}{I(t_0)} \quad (2)$$

where $I(t)$ is the signal intensity at second t and $I(t_0)$ is the average signal intensity in the first second of recording. Variations in the concentration of oxy-hemoglobin and deoxy-hemoglobin were derived for each channel based on the Modified Lambert Beer Law (Sassaroli et al. 2004):

$$\begin{bmatrix} O_2Hb \\ HHb \end{bmatrix} = \frac{1}{\rho} \begin{bmatrix} \epsilon_{O_2Hb}(\lambda_1) \cdot DPF(\lambda_1) & \epsilon_{HHb}(\lambda_1) \cdot DPF(\lambda_1) \\ \epsilon_{O_2Hb}(\lambda_2) \cdot DPF(\lambda_2) & \epsilon_{HHb}(\lambda_2) \cdot DPF(\lambda_2) \end{bmatrix}^{-1} \times \begin{bmatrix} OD(\lambda_1) \\ OD(\lambda_2) \end{bmatrix} \quad (3)$$

where O_2Hb and HHb represent the changes in oxy-hemoglobin and deoxy-hemoglobin concentrations, ρ is the interoptode distance, ϵ and DPF are, respectively, the extinction coefficients for the two chromophores and the Differential Pathlength

Factors at the wavelengths of interest (λ_1 and λ_2). The extinction coefficients of the two forms of hemoglobin at the different wavelengths were extracted from (Zijlstra et al. 1991) ($\epsilon_{O_2Hb}(690nm) = 0.0096\ mm^{-1}$, $\epsilon_{O_2Hb}(830nm) = 0.021\ mm^{-1}$, $\epsilon_{HHb}(690nm) = 0.05\ mm^{-1}$, $\epsilon_{HHb}(830nm) = 0.017\ mm^{-1}$). The DPFs were derived fron Scholkmann and Wolf ($DPF(690nm) = 6.5$, $DPF(830nm) = 5.5$) (Scholkmann et al. 2013). Oxy-hemoglobin and deoxy-hemoglobin changes (ΔO_2Hb, ΔHHb, respectively) during the motor imagery were obtained with respect to rest (8 channels and 2 Hemoglobin forms for a total of 16 features per second):

$$\begin{bmatrix} \Delta O_2Hb \\ \Delta HHb \end{bmatrix} = \begin{bmatrix} O_2Hb_{MI} - O_2Hb_{Bas} \\ HHb_{MI} - HHb_{Bas} \end{bmatrix} \tag{4}$$

where O_2Hb_{MI} and HHb_{MI} are the average hemoglobin concentrations over 1 second window during the task and O_2Hb_{Bas} and HHb_{Bas} are the average hemoglobin concentrations in a 1 second window prior to the task onset. Hemoglobin change in both short and long distance channels for each second during the task were fed to the learning algorithms for classification.

2.4. Deep Neural Network and Classification

DNNs allow computational models that are composed of multiple processing layers of non-linear units, called neurons, able to learn representations of data with multiple levels of abstraction. DNNs find complex structure in data-sets by using the backpropagation algorithm (Hecht-Nielsen et al. 1988) that guides changes in Networks' parameters that are sequentially updated in each layer from the representation in the previous layer. Whereas Network parameters are learned from the data, the DNN structure have to be heuristically selected a priori or determined through computationally demanding hyper-parameters optimization algorithms (Bengio 2000; MacKay 1996; Snoek et al. 2012)

Since a first investigatory comparison between DNN performance and other classifier performances was carried out, we decided to fix the DNN structure without investigating multiple DNN architectures. The network employed was a full connected feed-forward DNN and its structure is reported in figure 3. The data set consisted of 123 (when standalone EEG classification was performed), 16 (when standalone fNIRS classification was performed), or 139 (when both EEG and fNIRS were employed) input neurons. The features of the input neurons were 1 second average ERD/ERS (expressed as relative change), ΔO_2Hb and ΔHHb [expressed as microMolar (μM) change]. Notice that the average value of the features were of the order of 1 and all features' values were within one order of magnitude. In fact, the input values order of magnitude is an important aspect to take into account when training DNNs. Each neuron of the hidden layers performed a non-linear transformation of a linear combination of all the output from the previous layer. As a non-linear processing function we decided to employ the Rectified Linear Unit (ReLU) function, which was proven to dampen the vanishing

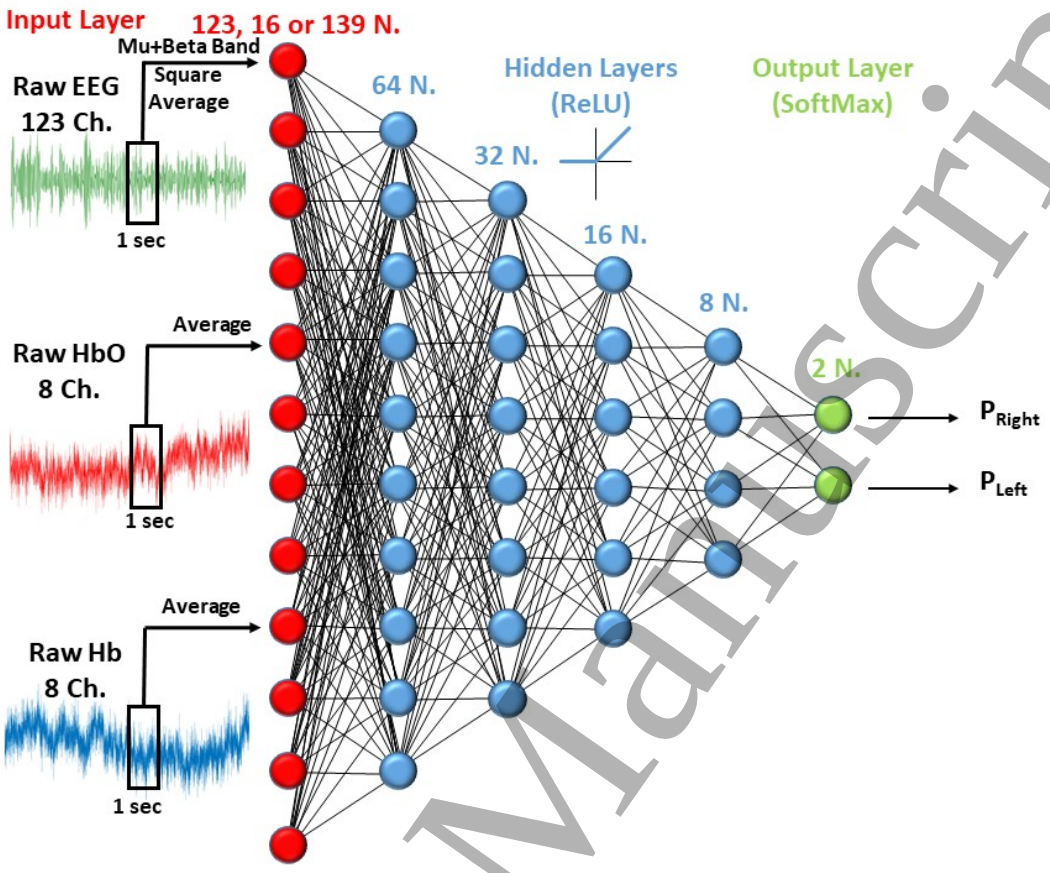

Figure 3. DNN Network employed for motor imagery classification. The network was a full connected feed-forward DNN . Input neurons were 123 (when standalone EEG classification was performed), 16 (when standalone fNIRS classification was performed), or 139 (when both EEG and fNIRS were employed). The feature of the input neurons were 1 second average ERD/ERS, ΔO_2Hb and ΔHHb, expressed in microMolar (μM) change. As a non-linear processing function of the neurons' hidden layers the Rectified Linear Unit (ReLU) function was employed. The number of hidden layers (4) and neurons were selected to approximately decrease the number of processing unit (thus compressing information) by a factor of 2 between successive layers when multi-modal recordings were employed. The output layer was composed of two neurons performing a softmax transformation. The softmax function outputs for the two neurons the predicted probability of being in the right (P_{Right}) or left (P_{Left}) imagery state.

gradient problem providing better performance than other non-linear functions (such as the hyperbolic tangent or the sigmoid function) (Dahl et al. 2013). Hidden neurons output, when ReLU function is employed, can be written as:

$$y = \begin{cases} 0, & \text{if } wx + b \leq 0 \\ wx + b, & \text{if } wx + b > 0 \end{cases} \tag{5}$$

where x is the input vector, w and b are the weight vector and bias, respectively, and y is the output vector. Since the classifier had to discriminate between two states,

namely Left or Right Motor Imagery state, the output layer was composed of two neurons performing a softmax transformation:

$$\begin{bmatrix} P_{Right} \\ P_{Left} \end{bmatrix} = \begin{bmatrix} \frac{e^{w_1 x}}{\sum_{k=1}^{2} e^{w_k x}} \\ \frac{e^{w_2 x}}{\sum_{k=1}^{2} e^{w_k x}} \end{bmatrix}. \quad (6)$$

The softmax function outputs for the two neurons the predicted probability of being in the right (P_{Right}) or left (P_{Left}) imagery state. x is the input vector of the softmax layer and W_1 and W_2 are the weight vectors of the neurons. The number of hidden layers (4) and neurons (refer to figure 3) were selected to approximately decrease the number of processing unit (thus compressing information) by a factor of 2 between successive layers. The weights were initialized in a pseudo-random approach employing a truncated normal distribution (0 mean, 0.1 SD, 2 SD truncation), whereas the biases were initialized to 0 (Sutskever et al. 2013).

The DNN was trained in a supervised learning approach (Hastie et al. 2009). In the supervised learning, DNN parameters, i.e. weights ws and biases bs, are adjusted relying on an objective function minimization procedure. The objective function measures the error (or distance) between the output scores and the desired scores . We employed the cross-entropy error as objective function. Cross-entropy (CE) is defined as:

$$CE = -\sum_{i} y'_i \ln y_i \quad (7)$$

where y is the output vector of the DNN ($[P_{Right}\ P_{Left}]$ in the study) and y' is the known state ([1 0] for Right Hand Imagery or [0 1] for Left Hand Imagery). Cross-entropy metric takes into account the closeness of a prediction and is a more granular way to compute error than Classification Error or Mean Squared Error (Murphy 2012).

The choice of optimization algorithm for deep learning model is extremely important. In general, learning algorithms are procedures that optimize network's weights and biases by exploring the parameters' space relying on the local slope (gradient) of the objective function. As optimization algorithm we employed the Adam Optimizer (Kingma et al. 2014). Adam Optimizer is a state-of the art learning algorithm that is different from classical stochastic gradient descent since it computes individual adaptive learning rates from estimates of the first and second moments of the gradients, dampening slow learning rate and/or local minima issues. Adam Optimizer parameters were set to: learning rate=10^{-4}, first moment exponential decay rate=$9 \cdot 10^{-1}$, second moment exponential decay rate=$9.99 \cdot 10^{-1}$, constant=10^{-8} (Kingma et al. 2014). As a regularization procedure for avoiding overfitting we employed a dropout approach on all the hidden layers with a dropout of 0.75. The optimization procedure was iterated for 2000 iteration. In order to address the DNN performance, we decided to perform a 10-fold cross validation procedure (Kohavi et al. 1995) employing all the 200 trials of Right or Left motor imagery for each subject which meant a training set of 180 (batch

size of 90) and a testing set of 20 for each iteration. The training and testing set for each validation were selected from different imagery trials. The cross-validation procedure was performed 1000 times and the average outcome for each subject was considered. This analysis provided an estimate of the performance achieved by the DNN after a training of ∼9 minutes employing the chosen task design. The accuracy of the DNN was evaluated by counting the number of correct DNN predictions after an argmax evaluation of probabilities of the DNN output vector. The described DNN architecture, training and validation were implemented in Python through the open-source software library Tensorflow (Abadi et al. 2016).

DNN performances were compared to LDA and linear SVM. LDA is a classical closed-solution linear classifier that rely on classes means and covarancies (Balakrishnama et al. 1998). SVM is a supervised learning algorithm that can be employed for binary linear classification. SVM model is a representation of points in the feature space, mapped so that the points of different classes are divided by a gap that is as wide as possible (Cortes et al. 1995). LDA and SVM analysis were computed in Matlab. For the SVM the following settings were employed: Autoscale, 1; Boxconstraint,1; Kernel function, Linear; Method, Sequential Minimum Optimization. LDA and SVM performances were estimated employing a 10-fold cross validation procedure. In accordance with the DNN analysis, the cross-validation procedure was performed 1000 times.

2.5. Statistical Analysis

The statistical analysis was threefold:

- It compared multi-modal EEG-fNIRS with standalone EEG or fNIRS BCI performances.
- It compared DNN with other classification algorithms such as LDA and SVM.
- It investigated possible interaction effects between the recording modality and the classifiers. Results from the 15 subjects (classification accuracy) were computed for all possible combination of recordings and classifiers.
  In order to highlight differential effects a two-way repeated measurement ANOVA (rANOVA) was performed. Two within subjects factors were considered: recording modality (three levels: standalone EEG, standalone fNIRS, combined EEG-fNIRS) and classification algorithm (three levels: LDA, SVM, DNN). Moreover, post-hoc analysis were conducted to provide average differences and their statistical significance between different interesting conditions. Statistical analysis was performed in Matlab and SPSS.

3. Results

Figure 4a shows examples of average Right and Left Motor Imagery responses for a representative subject across all the tasks. A typical controlateral ERD activation (lower

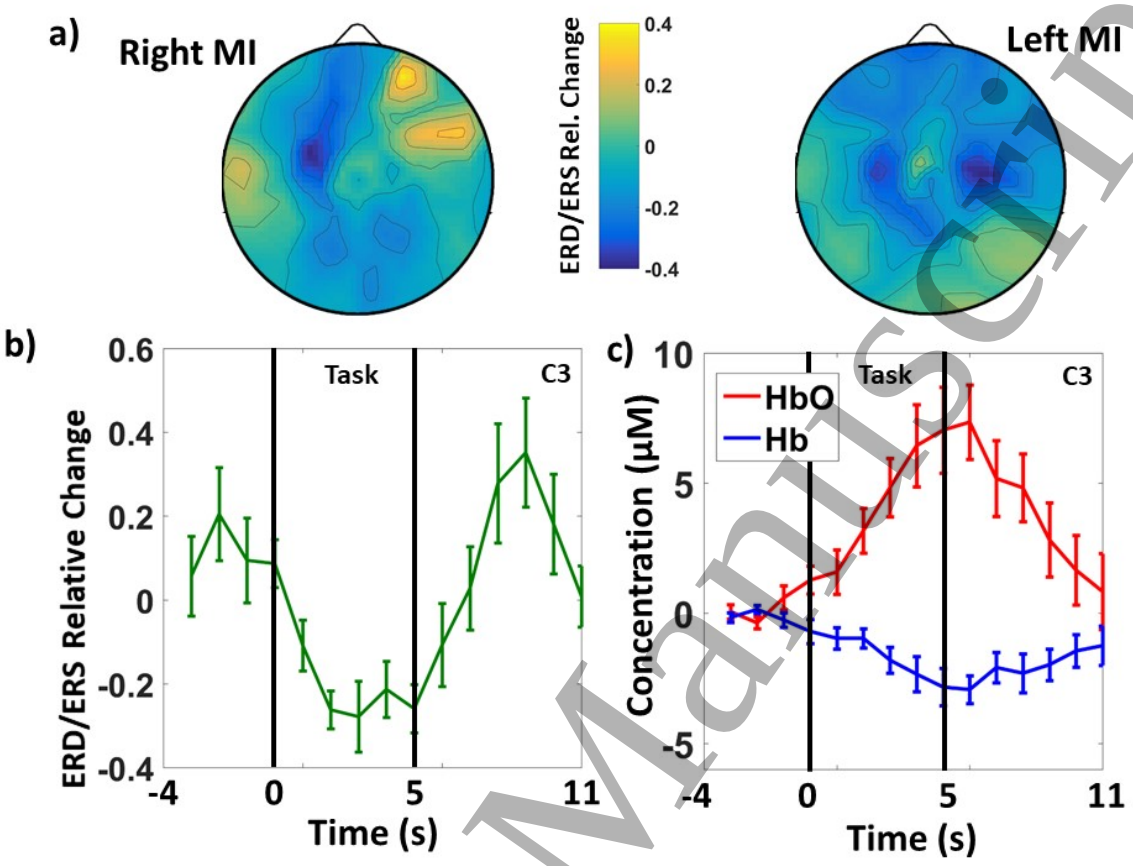

Figure 4. a) Examples of average EEG Right and Left Motor Imagery responses for a representative subject across all the data set. The motor cortex controlateral (stronger than ipsilateral) ERD activation (lower power with respect to baseline) is appreciable. b) Average timecourse and related standard errors of the ERD/ERS relative change from 5 seconds prior to 5 seconds post right motor imagery task for the same subject in a highly activated electrode located over the left sensorimotor cortex (C3). c) Average timecourses and related standard errors of O_2Hb and HHb change from 5 seconds prior to 5 seconds post right motor imagery task for the same subject in an activated left located channel for the same subject (over C3).

power with respect to baseline) on the motor cortices is appreciable. In fact, although the left hand Motor Imagery map shows also an ipsilater motor cortex activation, which was present in many subjects, the image still shows an higher activation in the contralateral cortex. Figure 4b shows the average timecourse and related standard errors of the ERD/ERS relative change from 5 seconds prior to 5 seconds post to the right motor imagery task for the same subject in the C3 electrode. Figure 4c shows the average timecourses and related standard errors of oxy- and deoxy-hemoglobin changes considering the same time frame in the optical channel over C3 during Right hand MI. A typical Blood-Oxygen-Level Dependent (BOLD) effect can be identified in response to the motor imagery. Figure 5a shows average accuracy performances and related standard errors of the DNN employed on the testing set as a function of training iteration when

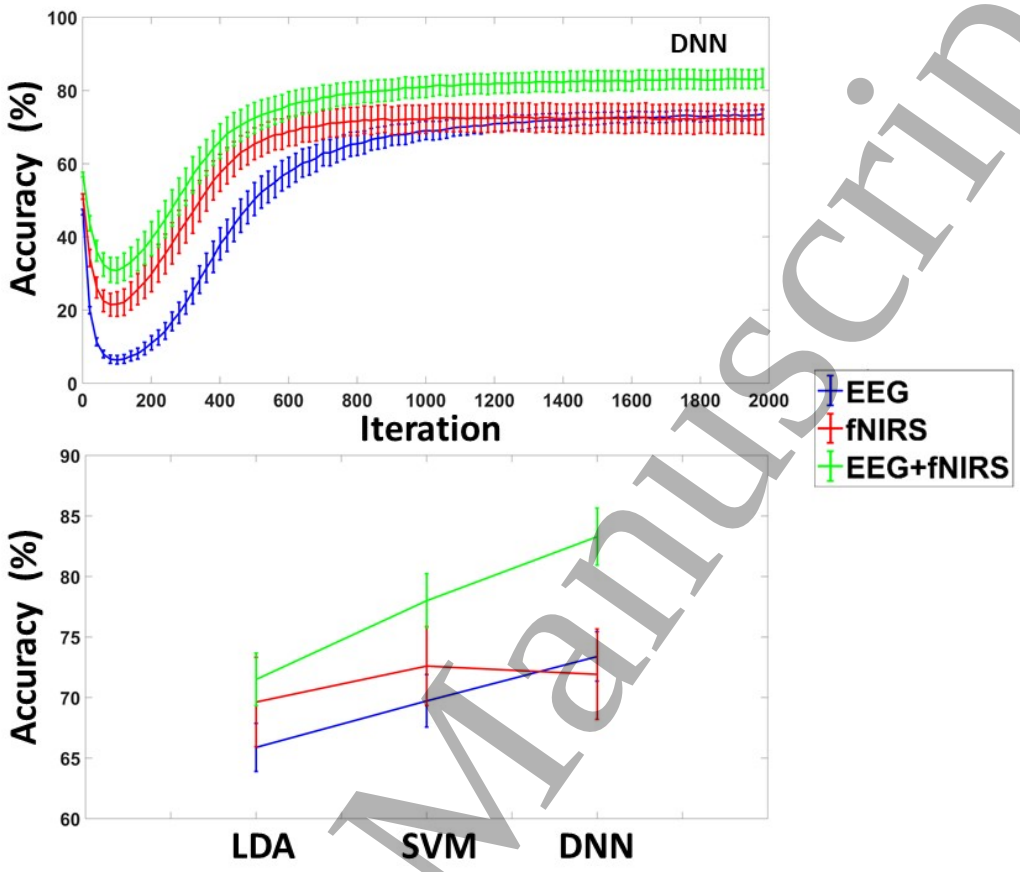

Figure 5. a) Average accuracies and related standard errors of the DNN employed as a function of training iteration when standalone EEG, standalone fNIRS or combined EEG-fNIRS recordings were analyzed. b) Average accuracy and related standard errors of standalone EEG, standalone fNIRS, and combined EEG-fNIRS as a function of classifier technology (LDA, SVM, DNN). The average accuracies and variabilities were computed among the 15 subjects.

standalone EEG, standalone fNIRS or combined EEG-fNIRS recordings were analyzed. Single subject accuracies for all the examined combinations of recordings (EEG,fNIRS, EEG+fNIRS) and classifiers (LDA,SVM, DNN) are reported in Table1. The average accuracies and variabilities were computed among the 15 subjects.

A clear increase in accuracy can be appreciated when fNIRS is employed with EEG when compared to standalone modalities (average EEG-DNN accuracy 73.38%, average fNIRS+DNN accuracy 71.92%, average fNIRS+EEG-DNN accuracy 83.28% at iteration 2000). Figure 5b reports the average accuracies and related standard errors of standalone EEG, standalone fNIRS, and combined EEG-fNIRS as a function of classifier technology (LDA, SVM, DNN). The rANOVA described in the Statistical Analysis section was performed on accuracies. Before the rANOVA, although the bounded nature of the accuracy metric, a Shapiro-Wilk normality test proved the approximatively gaussian-like distributions of the accuracy within each condition (all p's>0.05). The rANOVA highlighted a strong effect of Classifier ($F(2,28)=22.18$,

Subj. N.	EEG-LDA (%)	fNIRS-LDA (%)	EEG+fNIRS-LDA (%)	EEG-SVM (%)	fNIRS-SVM (%)	EEG+fNIRS-SVM (%)	EEG-DNN (%)	fNIRS-DNN (%)	EEG+fNIRS-DNN (%)
1	62,4	63,8	64,1	69,2	71,6	76,2	67,1	65,0	69,0
2	70,9	64,3	66,6	74,6	67,5	69,7	74,7	78,0	88,0
3	74,2	59,2	76,9	67,4	60,3	72,5	77,2	54,9	74,2
4	67,3	88,3	73,0	75,8	89,7	91,7	76,8	95,0	95,5
5	67,7	74,3	70,5	74,7	72,2	73,5	73,5	76,2	81,8
6	75,6	75,1	76,3	79,0	77,7	84,0	82,7	69,6	86,5
7	75,8	70,2	80,3	78,9	70,9	82,8	80,6	63,9	83,7
8	61,8	55,0	62,1	65,9	59,6	71,4	74,1	61,6	74,3
9	69,5	52,4	63,2	71,0	63,6	66,5	68,1	53,6	73,4
10	45,4	60,7	89,9	45,2	71,7	89,8	49,0	64,5	93,9
11	61,4	44,6	59,3	72,7	46,7	65,6	78,8	51,8	70,9
12	64,7	74,2	67,5	68,9	72,9	73,0	75,2	76,3	81,7
13	69,2	80,1	72,5	69,8	81,4	78,7	76,3	82,8	88,6
14	61,1	94,4	82,0	60,1	93,6	86,5	72,2	90,2	93,1
15	61,2	87,6	68,1	72,5	89,5	88,0	74,7	95,4	94,6

Table 1. Single subject accuracies for all the examined combinations of recordings (EEG, fNIRS, EEG+fNIRS) and classifiers (LDA, SVM, DNN)

$p<0.05$) and an almost significant effect of Recording ($F(2,28)=3.30$, $p=0.05$). An interaction between Recording and Classifier was found ($F(4,56)=8.37$, $p<0.05$). The rANOVA results suggest a main effects of both Classifier procedure and Recording on the accuracy of the BCI with a possible synergistic effect. Average performances of the different recordings, partialling out the effect of classifier were: EEG=$69.66\% \pm 1.94\%$, fNIRS=$71.37\% \pm 3.47\%$, fNIRS+EEG=$77.59\% \pm 2.02\%$. Average performances of the different classifiers, partialling out the effect of recording were: LDA=$68.99\% \pm 1.77\%$, SVM=$73.43\% \pm 1.80\%$, DNN=$76.19\% \pm 2.07\%$. Best average performance was obtained when combining fNIRS+EEG with DNN classifier: fNIRS+EEG-DNN=$83.28\% \pm 2.36\%$. A single subject best performance was obtained for the fNIRS+EEG-DNN recording-classifier with an accuracy of 95.50%.

Post-hoc analysis were further performed through pairwise contrasts between conditions of interest. With respect to multi-modal performances, fNIRS+EEG-DNN vs. EEG-DNN provided an accuracy increase of $9.89\% \pm 3.31\%$ (paired t-test, $t = 2.99$, $df = 14$, $p < 0.05$, Bonferroni corrected) whereas fNIRS+EEG-DNN vs. fNIRS+DNN provided an accuracy increase of $11.36\% \pm 2.31\%$ (paired t-test, $t = 4.91$, $df = 14$, $p < 0.05$, Bonferroni corrected). These results highlighted the advantages of multi-modal recording further supported by the small difference found between EEG-DNN and fNIRS+DNN $1.46\% \pm 4.02\%$ (paired t-test, $t = 0.36$, $df = 14$, $p > 0.05$). With respect to classifier performances, fNIRS+EEG-DNN vs. fNIRS+EEG-SVM provided an accuracy increase of $5.28\% \pm 1.42\%$ (paired t-test, $t = 3.71$, $df = 14$, $p < 0.05$, Bonferroni corrected) whereas fNIRS+EEG-DNN vs. fNIRS+EEG-LDA provided an accuracy increase of $11.79\% \pm 2.00\%$ (paired t-test, $t = 5.89$, $df = 14$, $p < 0.05$, Bonferroni corrected). This results highlighted the advantages of DNN classifier over LDA or SVM.

4. Discussion

Brain activity recordings, feature extraction and state classification are required steps that allow to interface the human brain with a device (Wolpaw et al. 2000).

Within the neural activity recording modalities, multi-modal monitoring is receiving

increased interest. In fact, because of the multiple physiological information that brain activity can provide, several techniques have been developed over the years to study brain signals coming from different neurophysiological mechanisms (Croce et al. 2016; Fazli et al. 2012b; Hong et al. 2015; Khan et al. 2014; Pfurtscheller et al. 2010). However, due to the absence of a specific technology that can record the whole spectrum of the information generated by these signals, simultaneous multi-modal monitoring of brain state has become particularly useful. Among multi-modal monitoring, the integration of EEG and fNIRS showed good potentialities. EEG and fNIRS are particularly suited for BCI since they are both flexible, scalp located recording procedures that can records electrical (EEG) and hemodynamic (fNIRS) brain activity (Chiarelli et al. 2017) .

With respect to classification procedure, general current approaches to pattern recognition-classification make essential use of advanced machine learning methods such as the one employed in the DNNs (LeCun et al. 2015). In this paper we performed a first exploratory study that investigated the BCI capabilities of combining multi-modal EEG-fNIRS data with advanced Deep Learning Classifiers (DNNs).

A supervised Left and Right Hand Motor Imagery task (figure 1) was performed and Left vs Right classification accuracy improvements of EEG-fNIRS recordings and DNN Classifier were estimated and compared to standalone EEG, fNIRS and other classifiers (LDA, SVM). We found that baseline classification accuracy was dependent on the subject examined. However, as expected, when estimating repeated measurement effects, we obtained an improvement in classification performance when fNIRS recordings were added to the EEG and vice-versa (at the chosen threshold of statistical significance, $p = 0.05$), and a strong increase in performance when DNN Classifier was employed ($p < 0.05$). Moreover, an interaction between recording modalities and classifiers was found suggesting a synergistic effect ($p < 0.05$). Importantly, the size of our results were comparable with the one reported extensively in literature (Buccino et al. 2016; Khan et al. 2017, 2014; Lee et al. 2014; Ma et al. 2012).

Some further consideration should be reported. With respect to classification accuracy of multi-modal recordings, we should outline that we considered a possible effect of the different numerosity and head coverage of electrodes and optodes in our experiment. For that reason we applied the same algorithms only considering few electrodes (16) around the motor cortices (in similarity with the fNIRS measurement). However, the results obtained in this modality (not reported in the manuscript) were practically identical to the results obtained when considering the whole EEG headset (within 1% variability, same statistical results). This suggested the ability of the different classifiers to select the electrodes/optodes of interest.

With respect to classification accuracy as a function of classifier employed, we expect that the improvement obtained through the DNN classifier should be unbiased or, at worst, biased towards underestimation of DNN performance. In fact, because of the investigatory nature of the study, hyperparameters of the different classifiers were not optimized and they were heuristically chosen. For this reason some over-fitting,

which in theory could be reduced, might have occurred. However, over-fitting should intrinsically affect more the non-linear (DNN) classifier .

Furthermore, some more consideration should be reported with respect to the DNN structure. The DNN employed was a full connected feed-forward DNN employing state of the art neurons' transfer functions and trained with up to date learning algorithms and objective functions. The DNN employed was not a CNN (encoding spatial or temporal information, (Kalchbrenner et al. 2014; Krizhevsky et al. 2012)) nor an RNN (encoding the sequential order of features (Hochreiter et al. 1997; Mikolov et al. 2010)). We actually implemented different CNN architectures. These architecture performed both spatial (by applying 2D filters on images derived from the electrode and/or optode topographic layout) or temporal (by applying 1D filters on signal timecourses) filtering and pattern recognition. However, CNN architectures did not reach the performances of the full connected DNN (between 1% and 5% lower accuracy among different subjects). We think that the poor results obtained with the spatial CNN could be caused by the intrinsic low resolution of the EEG technology (Pfurtscheller et al. 1997) (by knowing the spatial distribution of electrodes did not seems to provide increased information content to the NN), and by the limited number of optodes employed for the fNIRS. We think that, by relying on the higher spatial information content of the fNIRS, it could be interesting to investigate spatial CNN performances in a multi-modal acquisition setting where a denser and bigger field of view optical array is employed. Regarding CNN employing temporal filters, we tried to feed to the CNN 1 second window raw data. However, the CNN performance were poorer when compared to the DNN fed with filtered (in the mu-beta-band) signals. These results could be dependent on the limited number of training data for BCI, which is an intrinsic limiting factor when acquiring data of brain activity. In fact, this aspect can possibly limit the automatic selection of signal feature capabilities of the CNN, forcing some feature extraction prior to the classification machinery.

Moreover, it should be highlighted that RNN architectures were not tested on the data set. In fact we think that RNNs may be highly suited in a self paced motor imagery BCI, where the sequential temporal information may be fundamental for the imagery classification. Being the experiment guided, thus differential measurements with respect to baseline were provided to classifiers, RNN were not strictly suited for the experiment layout. Further, we should outline that, in order to compare DNN with other highly performing classifiers, we implemented SVM in a non-linear classification fashion (employing a radial basis function kernel (Park et al. 1991)). However, non-linear SVM provided poorer results with respect to linear SVM suggesting worst generalization of non-linear SVMs classifiers with respect to the DNN. Finally we should also report that we tried to apply classifiers in a subject to subject transfer fashion without re-training the classifier on each subject. However, we obtained poor results with this particular approach (slightly above chance classification), independently of the procedure employed, suggesting a high degree of variability in the spatio-temporal feature of the signal recorded among the subjects examined and further need

of investigation about this interesting topic (Cecotti et al. 2017; Sturm et al. 2016).

5. Conclusion

Results reported in this paper suggest potentially high BCI performances of combined multi-modal EEG-fNIRS recording and deep learning classifiers with synergistic effects. The higher performances of the multi-modal acquisition with respect to standalone EEG and fNIRS highlight the higher information content of combining both hemodynamic and electrical brain activity recordings. The higher performances of DNN with respect to linear SVM, suggests the non linearity involved in BCI classification and, when compared to the poor results obtained implementing non-linear SVM, the capabilities of state of the art DNN learning procedures to avoid poor generalization. The limiting factor of low spatial resolution of the EEG technology and the low number of data points obtainable in-vivo did not allow an efficient feature extraction through CNN forcing some pre-processing prior to the DNN. Multi-modal electrical and hemodynamic brain imaging, coupled with some feature extraction procedures and DNN may represent a great improvement in BCI research.

6. Acknowledgements

This study was partially funded by grant: H2020, ECSEL-04-2015-Smart Health, Advancing Smart Optical Imaging and Sensing for Health (ASTONISH).

References

Abadi, Martín et al. (2016). "Tensorflow: Large-scale machine learning on heterogeneous distributed systems". In: *arXiv preprint arXiv:1603.04467*.

Abibullaev, Berdakh, Jinung An, and Jeon-Il Moon (2011). "Neural network classification of brain hemodynamic responses from four mental tasks". In: *International Journal of Optomechatronics* 5.4, pp. 340–359.

An, Xiu, Deping Kuang, Xiaojiao Guo, Yilu Zhao, and Lianghua He (2014). "A deep learning method for classification of EEG data based on motor imagery". In: *International Conference on Intelligent Computing*. Springer, pp. 203–210.

Balakrishnama, Suresh and Aravind Ganapathiraju (1998). "Linear discriminant analysis-a brief tutorial". In: *Institute for Signal and information Processing* 18.

Bashivan, Pouya, Irina Rish, Mohammed Yeasin, and Noel Codella (2015). "Learning representations from EEG with deep recurrent-convolutional neural networks". In: *arXiv preprint arXiv:1511.06448*.

Bengio, Yoshua (2000). "Gradient-based optimization of hyperparameters". In: *Neural computation* 12.8, pp. 1889–1900.

Bianchini, Monica and Franco Scarselli (2014). "On the complexity of neural network classifiers: A comparison between shallow and deep architectures". In: *IEEE transactions on neural networks and learning systems* 25.8, pp. 1553–1565.

Buccino, Alessio Paolo, Hasan Onur Keles, and Ahmet Omurtag (2016). "Hybrid EEG-fNIRS Asynchronous Brain-Computer Interface for Multiple Motor Tasks". In: *PloS one* 11.1, e0146610.

Cecotti, Hubert and Anthony J Ries (2017). "Best practice for single-trial detection of event-related potentials: Application to brain-computer interfaces". In: *International Journal of Psychophysiology* 111, pp. 156–169.

Chiarelli, Edward L Maclin, Kathy A Low, Monica Fabiani, and Gabriele Gratton (2015). "Comparison of procedures for co-registering scalp-recording locations to anatomical magnetic resonance images". In: *Journal of biomedical optics* 20.1, pp. 016009–016009.

Chiarelli, Edward L Maclin, Kathy A Low, Kyle E Mathewson, Monica Fabiani, and Gabriele Gratton (2016). "Combining energy and Laplacian regularization to accurately retrieve the depth of brain activity of diffuse optical tomographic data". In: *Journal of biomedical optics* 21.3, pp. 036008–036008.

Chiarelli, F Zappasodi, F Di Pompeo, and A Merla (2017). "Simultaneous functional near-infrared spectroscopy and electroencephalography for monitoring of human brain activity and oxygenation: a review." In: *Neurophotonics* 4.4, p. 041411.

Ciresan, Dan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber (2012). "Deep neural networks segment neuronal membranes in electron microscopy images". In: *Advances in neural information processing systems*, pp. 2843–2851.

Collobert, Ronan and Jason Weston (2008). "A unified architecture for natural language processing: Deep neural networks with multitask learning". In: *Proceedings of the 25th international conference on Machine learning*. ACM, pp. 160–167.

Cortes, Corinna and Vladimir Vapnik (1995). "Support vector machine". In: *Machine learning* 20.3, pp. 273–297.

Costantini, Marcello, Assunta Di Vacri, Antonio Maria Chiarelli, Francesca Ferri, Gian Luca Romani, and Arcangelo Merla (2013). "Studying social cognition using near-infrared spectroscopy: the case of social Simon effect". In: *Journal of biomedical optics* 18.2, pp. 025005–025005.

Croce, Pierpaolo, Alessio Basti, Laura Marzetti, Filippo Zappasodi, and Cosimo Del Gratta (2016). "EEG? fMRI Bayesian framework for neural activity estimation: a simulation study". In: *Journal of Neural Engineering* 13.6, p. 066017.

Croce, Pierpaolo, Filippo Zappasodi, Arcangelo Merla, and Antonio Chiarelli (2017). "Exploiting neurovascular coupling: A Bayesian Sequential Monte Carlo approach applied to simulated EEG fNIRS data." In: *Journal of Neural Engineering*.

Dahl, George E, Tara N Sainath, and Geoffrey E Hinton (2013). "Improving deep neural networks for LVCSR using rectified linear units and dropout". In: *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, pp. 8609–8613.

Farroni, Teresa, Antonio M Chiarelli, Sarah Lloyd-Fox, Stefano Massaccesi, Arcangelo Merla, Valentina Di Gangi, Tania Mattarello, Dino Faraguna, and Mark H Johnson (2013). "Infant cortex responds to other humans from shortly after birth". In: *Scientific reports* 3.

Fazli, Siamac, Jan Mehnert, Jens Steinbrink, Gabriel Curio, Arno Villringer, Klaus-Robert Mfffdfffdller, and Benjamin Blankertz (2012a). "Enhanced performance by a hybrid NIRS–EEG brain computer interface". In: *NeuroImage* 59.1, pp. 519–529. DOI: 10.1016/j.neuroimage.2011.07.084. URL: https://doi.org/10.1016%2Fj.neuroimage.2011.07.084.

Fazli, Siamac, Jan Mehnert, Jens Steinbrink, Gabriel Curio, Arno Villringer, Klaus-Robert Müller, and Benjamin Blankertz (2012b). "Enhanced performance by a hybrid NIRS–EEG brain computer interface". In: *Neuroimage* 59.1, pp. 519–529.

Ferrari, Marco and Valentina Quaresima (2012). "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application". In: *Neuroimage* 63.2, pp. 921–935.

Gagnon, Louis, Meryem A Yücel, David A Boas, and Robert J Cooper (2014). "Further improvement in reducing superficial contamination in NIRS using double short separation measurements". In: *Neuroimage* 85, pp. 127–135.

Govindan, Rathinaswamy B, An Massaro, Taeun Chang, Gilbert Vezina, and Adré du Plessis (2016). "A novel technique for quantitative bedside monitoring of neurovascular coupling". In: *Journal of neuroscience methods* 259, pp. 135–142.

Hajinoroozi, Mehdi, Tzyy-Ping Jung, Chin-Teng Lin, and Yufei Huang (2015). "Feature extraction with deep belief networks for driver's cognitive states prediction from

EEG data". In: *Signal and Information Processing (ChinaSIP), 2015 IEEE China Summit and International Conference on*. IEEE, pp. 812–815.

Hallez, Hans et al. (2007). "Review on solving the forward problem in EEG source analysis". In: *Journal of neuroengineering and rehabilitation* 4.1, p. 46.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). "Overview of supervised learning". In: *The elements of statistical learning*. Springer, pp. 9–41.

Hecht-Nielsen, Robert et al. (1988). "Theory of the backpropagation neural network." In: *Neural Networks* 1.Supplement-1, pp. 445–448.

Hennrich, Johannes, Christian Herff, Dominic Heger, and Tanja Schultz (2015). "Investigating deep learning for fNIRS based BCI". In: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. IEEE, pp. 2844–2847.

Hinton, Geoffrey et al. (2012). "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups". In: *IEEE Signal Processing Magazine* 29.6, pp. 82–97.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long short-term memory". In: *Neural computation* 9.8, pp. 1735–1780.

Hong, Keum-Shik, Noman Naseer, and Yun-Hee Kim (2015). "Classification of prefrontal and motor cortex signals for three-class fNIRS–BCI". In: *Neuroscience letters* 587, pp. 87–92.

Hudson, Donna L and Maurice E Cohen (2000). *Neural networks and artificial intelligence for biomedical engineering*. Wiley Online Library.

Jirayucharoensak, Suwicha, Setha Pan-Ngum, and Pasin Israsena (2014). "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation". In: *The Scientific World Journal* 2014.

Kalchbrenner, Nal, Edward Grefenstette, and Phil Blunsom (2014). "A convolutional neural network for modelling sentences". In: *arXiv preprint arXiv:1404.2188*.

Khan and Keum-Shik Hong (2017). "hybrid eeg–fnirs-Based eight-command Decoding for Bci: application to Quadcopter control". In: *Frontiers in neurorobotics* 11.

Khan, Melissa Jiyoun, and Keum-Shik Hong (2014). "Decoding of four movement directions using hybrid NIRS-EEG brain-computer interface". In: *Frontiers in human neuroscience* 8, p. 244.

Kingma, Diederik and Jimmy Ba (2014). "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980*.

Kohavi, Ron et al. (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection". In: *Ijcai*. Vol. 14. 2. Stanford, CA, pp. 1137–1145.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*, pp. 1097–1105.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). "Deep learning". In: *Nature* 521.7553, pp. 436–444.

Lee, Min-Ho, Siamac Fazli, Jan Mehnert, and Seong-Whan Lee (2014). "Hybrid brain-computer interface based on EEG and NIRS modalities". In: *Brain-Computer Interface (BCI), 2014 International Winter Workshop on*. IEEE, pp. 1–2.

Lotte, Fabien, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi (2007). "A review of classification algorithms for EEG-based brain–computer interfaces". In: *Journal of neural engineering* 4.2, R1.

Ma, Lan, Lixin Zhang, Lu Wang, Minpeng Xu, Hongzhi Qi, Baikun Wan, Dong Ming, and Yong Hu (2012). "A hybrid brain-computer interface combining the EEG and NIRS". In: *Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS), 2012 IEEE International Conference on*. IEEE, pp. 159–162.

Maas, Andrew L, Awni Y Hannun, and Andrew Y Ng (2013). "Rectifier nonlinearities improve neural network acoustic models". In: *Proc. ICML*. Vol. 30. 1.

MacKay, David JC (1996). "Hyperparameters: optimize, or integrate out?" In: *Fundamental theories of physics* 62, pp. 43–60.

Mikolov, Tomas, Martin Karafiát, Lukas Burget, Jan Cernockỳ, and Sanjeev Khudanpur (2010). "Recurrent neural network based language model." In: *Interspeech*. Vol. 2, p. 3.

Murphy, Kevin P (2012). *Machine learning: a probabilistic perspective*. MIT press.

Naseer, Noman and Keum-Shik Hong (2015). "fNIRS-based brain-computer interfaces: a review". In: *Frontiers in human neuroscience* 9, p. 3.

Nguyen, Cuong Quoc Ngo, KHOA Truong Quang Dang, and Van Toi Vo (2013). "Temporal hemodynamic classification of two hands tapping using functional nearfffdfffdfffdinfrared spectroscopy". In: *Frontiers in human neuroscience* 7, p. 516.

Nuwer, Marc R (1988). "Quantitative EEG: I. Techniques and problems of frequency analysis and topographic mapping." In: *Journal of Clinical Neurophysiology* 5.1, pp. 1–44.

Park, Jooyoung and Irwin W Sandberg (1991). "Universal approximation using radial-basis-function networks". In: *Neural computation* 3.2, pp. 246–257.

Pascanu, Razvan, Tomas Mikolov, and Yoshua Bengio (2013). "On the difficulty of training recurrent neural networks". In: *International Conference on Machine Learning*, pp. 1310–1318.

Pfurtscheller (2001). "Functional brain imaging based on ERD/ERS". In: *Vision research* 41.10, pp. 1257–1260.

Pfurtscheller, Brendan Z Allison, Clemens Brunner, Gunther Bauernfeind, Teodoro Solis-Escalante, Reinhold Scherer, and Thorsten O Zander (2010). "The hybrid BCI". In: *Frontiers in neuroscience* 4.

Pfurtscheller and Christa Neuper (1997). "Motor imagery activates primary sensorimotor area in humans". In: *Neuroscience letters* 239.2, pp. 65–68.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 234–241.

Sassaroli, Angelo and Sergio Fantini (2004). "Comment on the modified Beer–Lambert law for scattering media". In: *Physics in Medicine and Biology* 49.14, N255.

Schmidhuber, Jürgen (2015). "Deep learning in neural networks: An overview". In: *Neural networks* 61, pp. 85–117.

Scholkmann, Felix and Martin Wolf (2013). "General equation for the differential pathlength factor of the frontal human head depending on wavelength and age". In: *Journal of biomedical optics* 18.10, pp. 105004–105004.

Shin, Jaeyoung, Alexander von Luhmann, Benjamin Blankertz, Do-Won Kim, Jichai Jeong, Han-Jeong Hwang, and Klaus-Robert Muller (2016). "Open Access Dataset for EEG+ NIRS Single-Trial Classification". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering*.

Shin, Jaeyoung, Klaus-Robert Müller, Christoph H Schmitz, Do-Won Kim, and Han-Jeong Hwang (2017). "Evaluation of a Compact Hybrid Brain-Computer Interface System". In: *BioMed research international* 2017.

Simonyan, Karen and Andrew Zisserman (2014). "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556*.

Snoek, Jasper, Hugo Larochelle, and Ryan P Adams (2012). "Practical bayesian optimization of machine learning algorithms". In: *Advances in neural information processing systems*, pp. 2951–2959.

Steinbrink, Jens, Arno Villringer, Florian Kempf, Daniel Haux, Stefanie Boden, and Hellmuth Obrig (2006). "Illuminating the BOLD signal: combined fMRI–fNIRS studies". In: *Magnetic resonance imaging* 24.4, pp. 495–505.

Sturm, Irene, Sebastian Lapuschkin, Wojciech Samek, and Klaus-Robert Müller (2016). "Interpretable deep neural networks for single-trial eeg classification". In: *Journal of neuroscience methods* 274, pp. 141–145.

Sutskever, Ilya, James Martens, George Dahl, and Geoffrey Hinton (2013). "On the importance of initialization and momentum in deep learning". In: *International conference on machine learning*, pp. 1139–1147.

Tucker, Don M (1993). "Spatial sampling of head electrical fields: the geodesic sensor net". In: *Electroencephalography and clinical neurophysiology* 87.3, pp. 154–163.

Villringer, Arno and Britton Chance (1997). "Non-invasive optical spectroscopy and imaging of human brain function". In: *Trends in neurosciences* 20.10, pp. 435–442.

Watanabe, Kazuyoshi, Fumio Hayakawa, and Akihisa Okumura (1999). "Neonatal EEG: a powerful tool in the assessment of brain damage in preterm infants". In: *Brain and Development* 21.6, pp. 361–372.

Wolpaw, Jonathan R et al. (2000). "Brain-computer interface technology: a review of the first international meeting". In: *IEEE transactions on rehabilitation engineering* 8.2, pp. 164–173.

Zappasodi, Filippo, Pierpaolo Croce, Alessandro Giordani, Giovanni Assenza, Nadia M Giannantoni, Paolo Profice, Giuseppe Granata, Paolo M Rossini, and Franca Tecchio (2017). "Prognostic Value of EEG Microstates in Acute Stroke". In: *Brain topography*, pp. 1–13.

Zijlstra, WG, A Buursma, and WP Meeuwsen-Van der Roest (1991). "Absorption spectra of human fetal and adult oxyhemoglobin, de-oxyhemoglobin, carboxyhemoglobin, and methemoglobin." In: *Clinical chemistry* 37.9, pp. 1633–1638.