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

Pierpaolo Croce[1,2], Filippo Zappasodi[1,2], Francesco di Pompeo[1,2] Arcangelo Merla[1,2] & Antonio Maria Chiarelli[1,2]

E-mail: pierpaolo.croce@unich.it

[1] Department of Neuroscience, Imaging and Clinical Sciences, "G.dAnnunzio" University, Chieti, Italy
[2] Institute of Advanced Biomedical Technologies, "G.dAnnunzio" University, Chieti, Italy

Abstract.
Objective. to be filled.
Approach. to be filled.
Main Results. to be filled.
Significance. to be filled.

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). EEG provides information regarding 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 (Gert Pfurtscheller et al. 2010). Among these, functional Near Infrared Spectroscopy (fNIRS) provided encouraging results (Fazli et al. 2012b).

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, depending on the slow dynamics of the hemodynamic response, yields low temporal resolution but, 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 physiological information provided by and characteristic of EEG and fNIRS (Croce et al. 2017), higher BCI performances of combined measurements with respect to standalone EEG were reported extensively (**chiarelliREVdaaggiungere**; Fazli et al. 2012b; Hong et al. 2015; M Jawad 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 ($8-15Hz$), 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 beta activity is suppressed in related brain areas(Event Related Desynchronization, ERD) (Pfurtscheller 2001; Gert Pfurtscheller et al. 2006). fNIRS features are generally computed from HbO and Hb variations in the brain which are dependent, among others, on the Blood Oxygen Level Dependant (BOLD) effect (Naseer et al. 2015; Steinbrink et al. 2006).

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

Different experiment and algorithms have been applied to combined EEG-fNIRS BCI. Fazli et al. (2012a) proposed Linear Discriminant Analysis to classify ERD EEG 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 amplitude of the cerebral blood oxygen signal. Lee et al. (2014) employed LDA on combined EEG and fNIRS features to classify three conditions: right and left motor imagery and 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 two methods: 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 M Jawad Khan et al. (2014) and Muhammad Jawad Khan et al. (2017) 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 and 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 (DNN) 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 algorithms related and they can be divided in three 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, DNN 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 functions 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 were 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 were 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 Deep Learning Network 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 task through DNN with an accuracy of 94%. Finally, H. T. 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 (Gert 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 in the multimodal recording modality was estimated and compared to other classification algorithms and to standalone EEG and fNIRS.

2. Methods

2.1. Experimental Paradigm

Six healthy subjects (males, average age of 34 years $\pm$ 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 set 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. 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 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 1a,c.). Skin/electrode impedance was measured before recordings and kept below $50k\Omega$. EEG data were sampled at 250 Hz and processed in a real time fashion. Raw data were stored over 1 second window and filtered between 13 Hz and 30 Hz (2nd order Digital Butterworth filter). The beta-band filtered signal was squared and averaged over the second (β_{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:

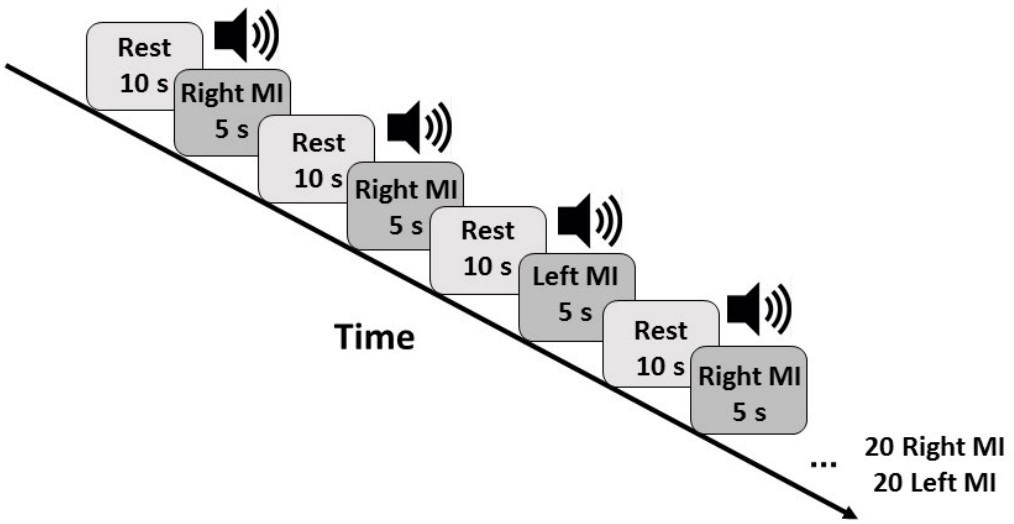

Figure 1. Motor imagery task sequence. The squeezing imagery consisted of 5 seconds of task and 10 seconds of rest. Right or Left imagery instruction was presented in a pseudo-random order. During the 5 seconds of task, the subjects were instructed to perform the squeezing imagery with a repetition frequency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

$$ERD/ERS = \frac{\beta_{PowMI} - \beta_{PowBas}}{\beta_{PowBas}} \tag{1}$$

where β_{PowMI} is the average power over 1 second window during the task and β_{PowBas} 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

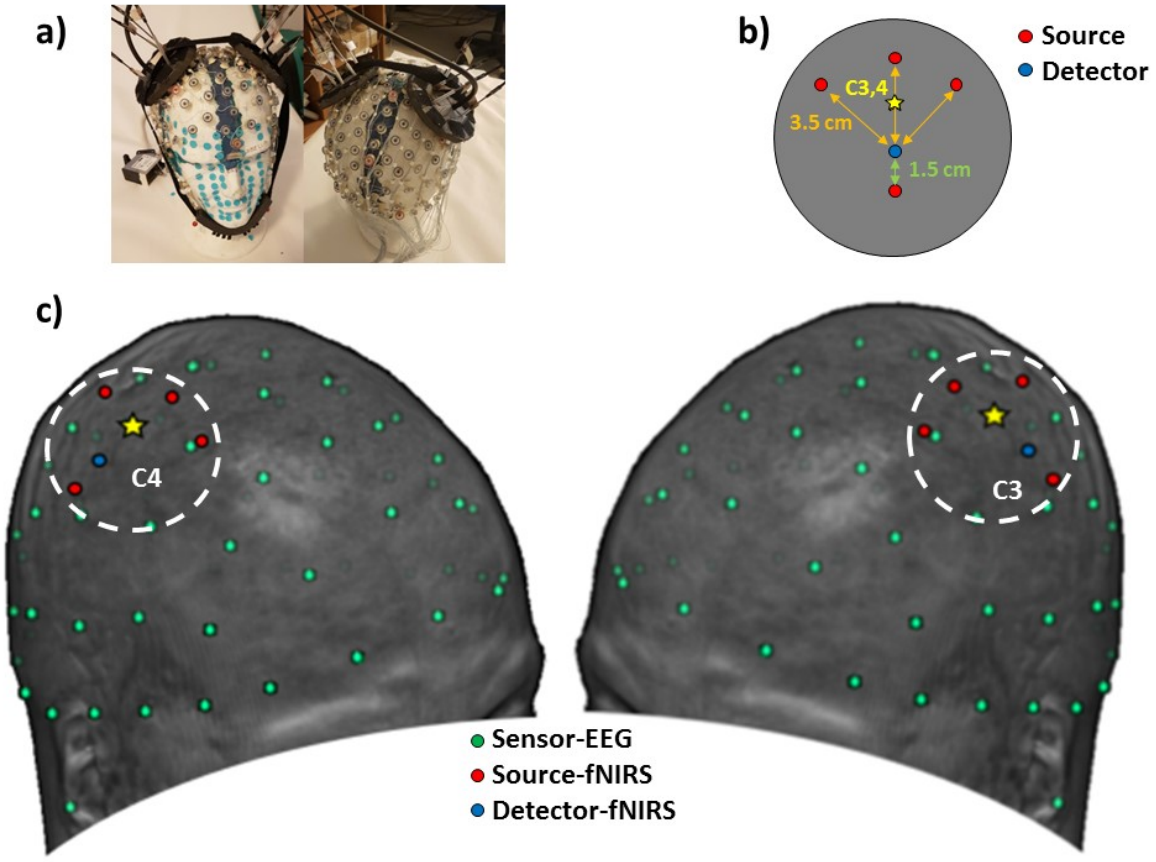

Figure 2. a) Pictures of the full head EEG cap and optical probes located on a dummy head. The EEG was a Electrical Geodesic Net 300 system. The fNIRS spectrometer was an Imagent from ISS with 32 laser diodes and 4 PMTs. Light was sent to the scalp using multimodal optical fibers (0.4 mm core) and from 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 does not reach the brain cortex), provided information regarding scalp-related hemodynamic oscillations. c) EEG electrodes and fNIRS optodes employed in the study overlayed onto a rendered structural Magnetic Resonance Image of a representative subject.

tubes (PMTs). 8 light sources (4 injection point, 2 wavelengths) and 1 detectors 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 multimodal 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 1a, with an optical layout reported in figure 1b, 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 does

not reach the brain cortex), provided information regarding scalp-related hemodynamic oscillations (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:

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

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):

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

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 ΔO_2Hb and deoxy-hemoglobin ΔHHb changes 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 Networks and Classification

Deep Neural Networks (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. Deep Neural Networks 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

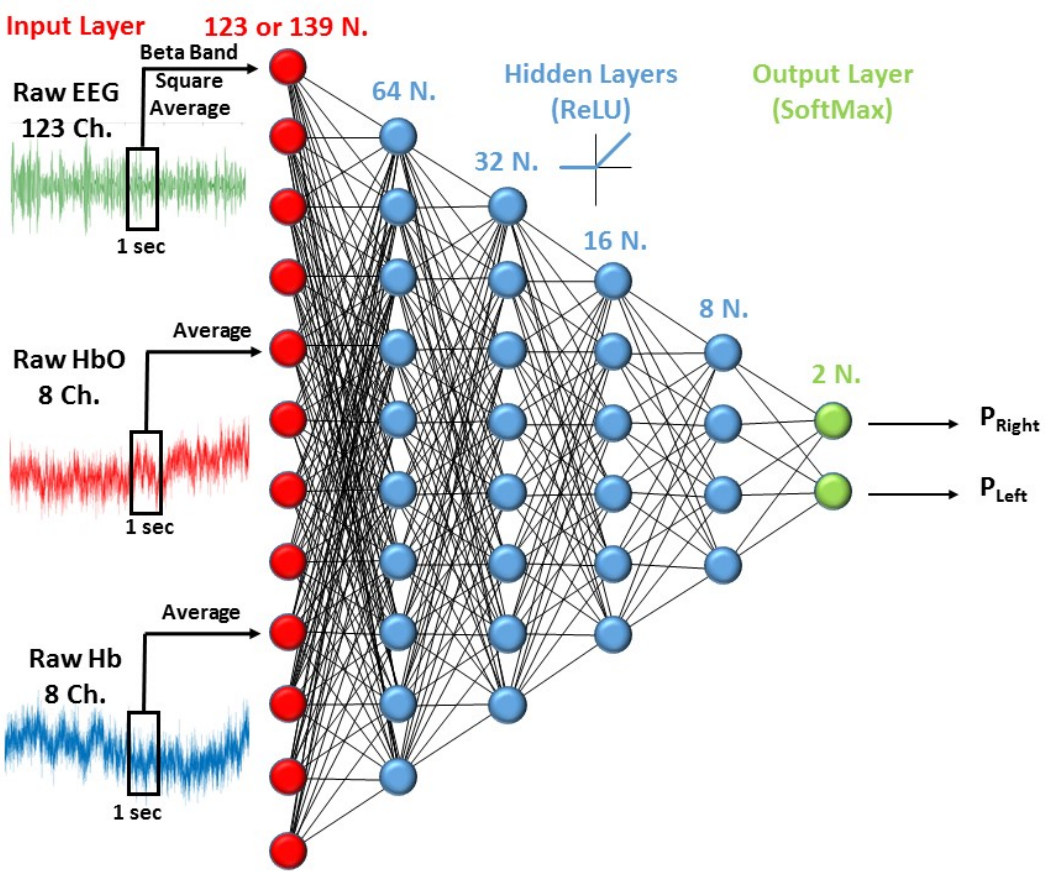

Figure 3. DNN Network employed for motor imagery classification. The network was a full connected feed-forward DNN . Input neurons were 123 (when standalone EEG classification was performed), or 139 (when both EEG and fNIRS were employed). The feature of the input neurons were 1 second average ERD/ERS, 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. 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.

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 we performed a first investigatory comparison between DNN performance and other classifier performances, we decided to fix the DNN structure (also called the DNN hyperparameters), without investigating multiple DNN architectures. The Network employed was a full connected feed-forward DNN and its structure is reported in figure 3 Our data set consisted of 123 (when standalone EEG classification was performed), or 139 (when both EEG and fNIRS were employed) input neurons. The feature of the input neurons were 1 second average ERD/ERS, ΔO_2Hb and ΔHHb, expressed in 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. This is an important aspect to account when training DNNs. Each neuron of the hidden layers performs 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 gradient problem providing better performance than other non-linear functions (such as the hyperbolic tangent or the sigmoid functions) (Dahl et al. 2013). Hidden neurons response when ReLU 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:

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

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 $W_2(2)$ 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. Random initialization of weights was proven to be another factor that avoids local minima during training (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:

$$CE = -\sum_{i} y'_i \ln y_i \tag{7}$$

where y is the output vector of the DNN ($[P_{Right}\ P_{Left}]$ in the study) 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 optimize network's weights and biases by exploring the space of such parameters relying on local the slope (gradient) of the objective function. As optimization algorithm we employed the state-of the art Adam Optimizer (Kingma et al. 2014). Adam Optimizer is an algorithm that is different from classical stochastic gradient descent since the methods computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients. 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} (ibid.). The optimization procedure was iterated for 1000 epochs. 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 seconds of Right or Left motor imagery for each subject. The training and testing set for each validation were selected from different imagery trials. The cross-validation procedure was performed 1000 times. 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 classificator 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 performances were estimated employing a 10-fold cross validation procedure. In accordance with the DNN analysis, the cross-validation procedure was performed 1000 times. LDA ans SVM analysis were performed in Matlab.

2.5. Statistical Analysis

The statistical analysis was threefold:

- It compared multimodal EEG-fNIRS with standalone EEG BCI performances (it estimated the Recording Effect);
- It compared DNN with other classification algorithms (LDA and SVM, it estimated the Classificator Effect).
- It investigated possible interactions or synergies between the Recording and the Classificator Effects. Results from the 6 subjects (classification accuracy) were computed for all possible combination of recordings and classificators. Since standalone EEG recordings coupled with LDA classificator can be considered the simplest BCI configuration, the analysis that investigated main effects, synergies and interactions of recording modality and classificator were performed

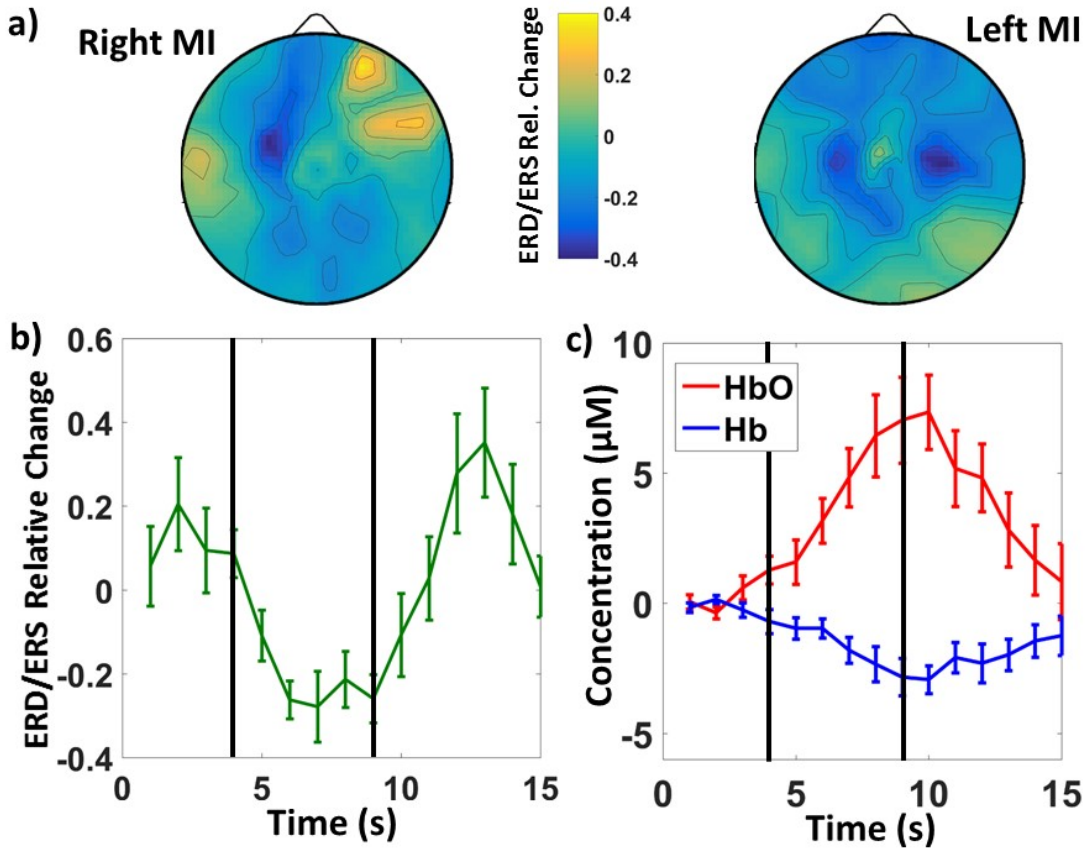

Figure 4. a) Examples of average EEG Right and Left Motor Imagery responses for a representative subject across all the 20 trials. The motor cortex controlateral ERD activation (lower β power with respect to baseline) is appreciable. b) Average timecourse and related standard errors of the ERD/ERS relative 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. 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.

on differential accuracies with respect to EEG-LDA. This differential analysis was conducted to dampen the problem of the intrinsic variability of the classification outcome among subjects, highlighting changes in accuracy caused by Recording and the Classificator modalities. A two-way anova was performed to investigate possible main effects and interactions of Recording (factor one, EEG vs. EEG-fNIRS with respect to EEG-LDA), and Classificator (factor two, SVM vs. DNN with respect to EEG-LDA) at a subjects' group level. Moreover, post-hoc analysis were conducted to provide average differences and their statistical significance among different conditions. Statistical analysis was performed in Matlab.

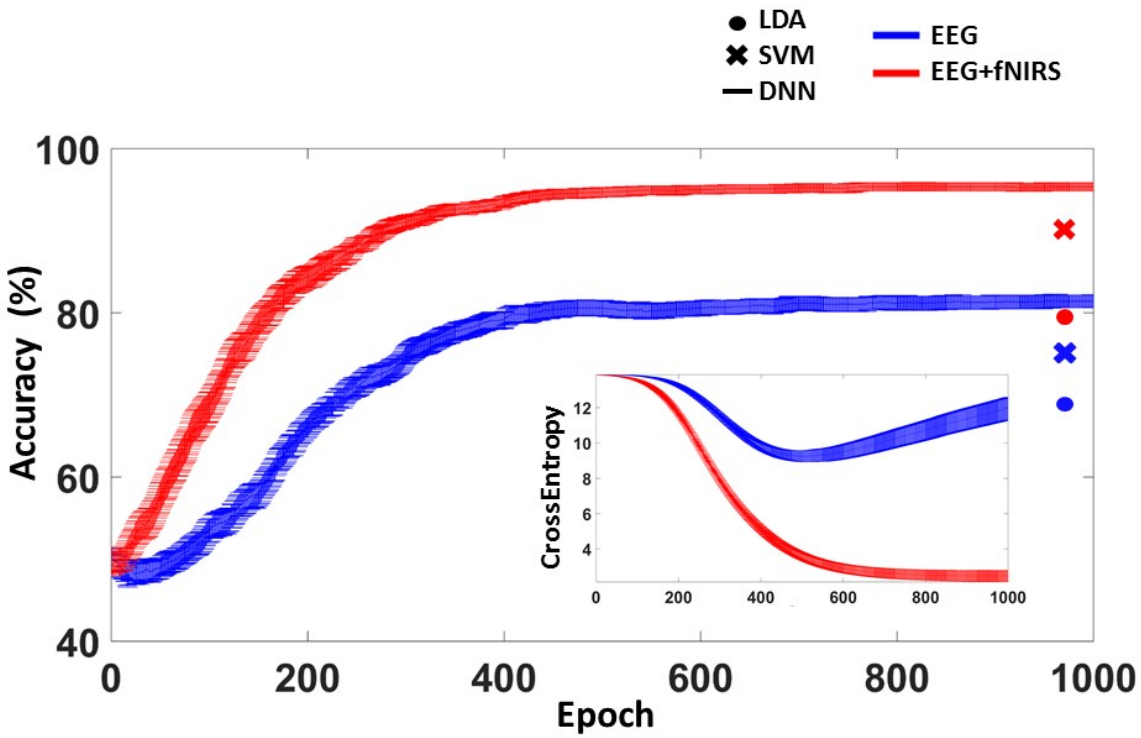

Figure 5. Average accuracy and related standard errors of the DNN employed as a function of training epoch when standalone EEG or combined EEG-fNIRS recordings were analyzed. The average accuracy and variability were computed among cross-validations. The figure also reports related cross-entropies estimates, along with cross-validated LDA and SVM average accuracies.

3. Results

Figure 4a shows examples of average Right and Left Motor Imagery responses for a representative subject across all the 20 trials. A typical controlateral ERD activation (lower β power with respect to baseline) on the motor cortices is appreciable. 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 a highly activated electrode located over the left sensorimotor cortex .Figure 4c shows the average timecourses and related standard errors of oxy- and deoxy-hemoglobin changes considering the same time frame in an activated left-located channel during Right hand MI. A typical Blood-Oxygen-Level Dependent (BOLD) effect can be identified in response to the motor imagery. Figure 5 show average accuracy performances and related standard errors of the DNN employed as a function of training epoch when standalone EEG or combined EEG-fNIRS recordings were analyzed. The average accuracy and variability were computed among cross-

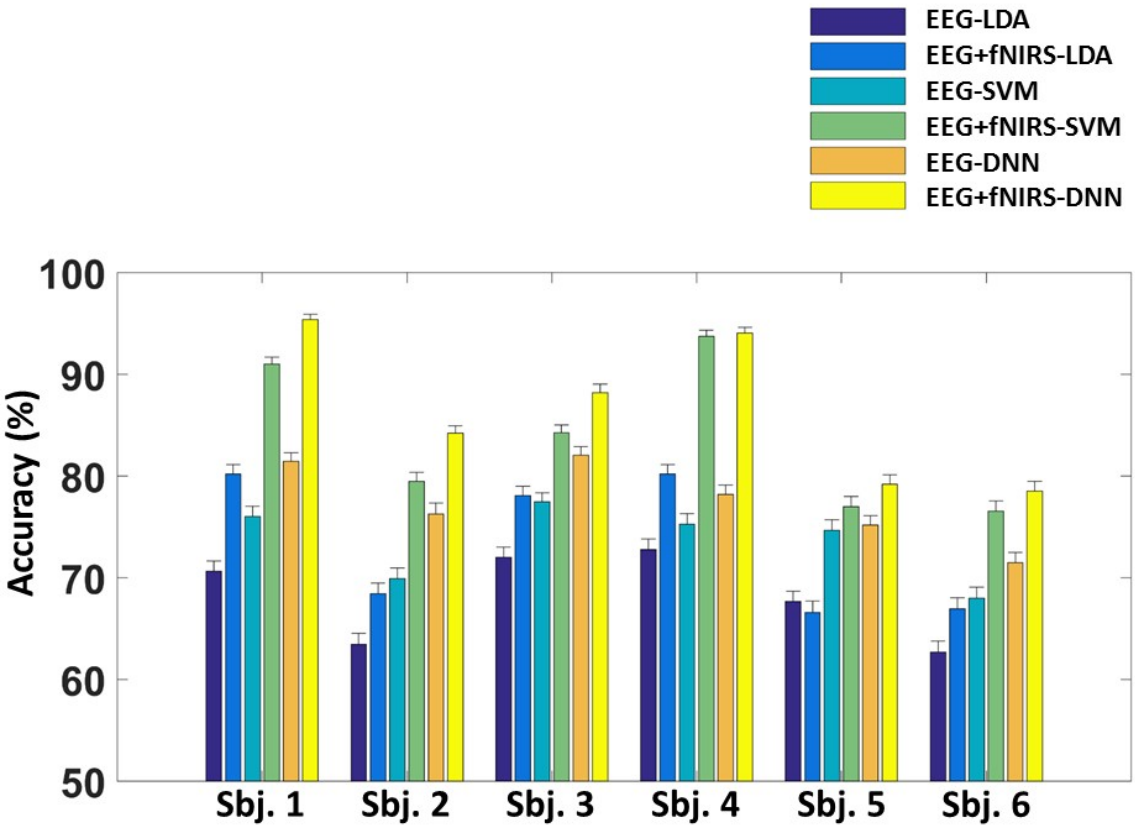

Figure 6. Average and related standard errors of cross-validated accuracies of all combination of recording procedure and classificator technology for each of the six subjects.

validations. Figure 5 also reports related cross-entropies estimates. Moreover, LDA and SVM average cross-validated performances are reported on top-right corner of the figure. For the reported subject, which was the one providing the overall highest classification performance, a clear increase in accuracy can be appreciated when fNIRS is employed with EEG (95.40% with respect to standalone EEG, 81.50%), independently of the classificator, and a clear different performance of classification accuracy as a function of classificator (LDA,SVM, DNN in increasing classification accuracy order), independently of the recording procedure with apparent cumulative effects. Figure 6 reports the average and related standard errors, cross-validated accuracies of all combination of recording procedure and classificator technology for each of the six subjects. A similar pattern as the one reported for the representative subject in figure 5 is present for all of the subjects examined. However, a clear variance in the average classification performance among subjects is appreciable with a best intra-subject performance ranging from 78.50% to 95.40%. Thus, in order to highlight specific effects of recordings and classificator the two way anova described in the Statistical Analysis section was performed on accuracies increases with respect to EEG-LDA.

Before the anova, a Shapiro-Wilk normality test proved the gaussian-like distributions of the accuracy changes within each conditions (all p's>0.05). The two-way anova highlighted a strong effect of Recording ($F(1,20)=42.22$, $p<0.05$) and a significant effect of Classificator ($F(1,20)=5.27$, $p<0.05$). No interaction between Recording and Classificator was found ($F(1,20)=0.11$, $p>0.05$). The anova results suggest a main effects of both Classificator procedure and recording on the accuracy of the BCI with a possible cumulative effect without interaction. Post-hoc analysis tested pairwise differences of conditions and it provided statistical significance for practically all possible combinations (p's<0.05). We obtained an average increase of EEG-SVM with respect to EEG-LDA of 5.35% (SD=1.56%), an increase of EEG+fNIRS-SVM of 15.47% (SD=4.58%), an increase of EEG-DNN of 9.24% (SD=2.59%) and an increase of EEG+fNIRS-DNN of 18.39% (SD=4.75%). Notably the increase in accuracy from EEG to EEG+fNIRS, without considering classificator, was of 9.63%, whereas the increase in accuracy from SVM to DNN, without considering recordings, was of 3.41%. This results, combined with an overall increase of EEG+fNIRS-DNN vs EEG-SVM of 13.04% highlighted the synergic effect of multimodal recording and DNN classification procedure.

4. Discussion

Short re-caption of main DNN concepts and EEG-fNIRS for MI. Discussion of obtained results: increased performance with fNIRS and with DNN classification without interaction (cumulative effect). Non linear space separation complex, In addition to performing linear classification, SVMs can efficiently perform non-linear classification by mapping their inputs into high-dimensional feature spaces. We tried radial basis Limitations: fixed hyperparameters (overall length of experiment, frequency band, response time (1sec), DNN structure). Report spatial CNN results were poorer (low spatial information content of EEG), may work better with full-head fNIRS. Time CNN for feature extration worked not well Future direction RNN for self paced

5. Conclusion

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.

Chiarelli, Antonio M, 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, Antonio M, 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.

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, 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, M Jawad, Melissa Jiyoun Hong, 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.

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

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, Hai Thanh, Cuong Quoc Ngo, KHOA Truong Quang Dang, and Van Toi Vo (2013). "Temporal hemodynamic classification of two hands tapping using functional nearffffdffffdfffdinfrared 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.

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

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

— (2006). "Future prospects of ERD/ERS in the context of brain–computer interface (BCI) developments". In: *Progress in brain research* 159, pp. 433–437.

Pfurtscheller, Gert et al. (2010). "The hybrid BCI". In: *Frontiers in neuroscience* 4.

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.

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.

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.

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.