NIRS signals: Motor Imagery (MI)

Rigazio Sofia (282247), Romano Anastasio (289707),
Roccaro Lucia (289423), Ruzzante Elena (292194)

1 Introduction

Near Infrared Spectroscopy (NIRS) is an optical neuroimaging technique that uses near-infrared light to measure changes in oxygenated (O_2Hb) and deoxygenated (HHb) hemoglobin levels based on differences in the light absorption spectra.
The key advantages of this technique is to be non-invasive, to have a relatively low cost, to be easily integrated with other techniques (such as EEG and fMRI), and to be both portable and safe in its use [1]. In addition, NIRS' low sensitivity to body movements and the systems' portability make it suitable for monitoring cortical hemodynamics during everyday life activities and to assess the effects of neurorehabilitation, for both very active and bedridden people [2].

Because of all these advantages of NIRS in respect to other techniques, such as EEG and MRI, we want to investigate whether NIRS signal acquired during a Motor Imagery task can be used in clinical applications on its own. This would lead to significant benefits in term of time, cost and comfort.
Specifically, our goal is to understand whether NIRS signals on its own can be used to create a classifier that discriminates between a left and right Motor Imagery task and if this classfier can control a prosthesis that performs the desired task. To achieve this result, we took our cue from some of the best-performing methods available in literature [3], [4], [5] and we attempted to compare the two different paradigms of features used (temporal and statistical) to create a classifier.

2 Materials and Methods

2.1 Dataset

The project is based on an open-access dataset provided by Shin et al. [6]. In particular, from this repository only the NIRS signals present in dataset A has been used, consisting of three sessions of left and right hand motor imagery.
NIRScout (NIRx GmbH, Berlin, Germany) was used to collect the signals. A total of 14 sources and 16 detectors were placed over the head in frontal, motor and visual areas resulting in 36 NIRS channels (9 frontal channels, 24 motor and somatosensorial channels and 3 occipital channels).
NIRS set of data were recorded using two different wavelengths (760 nm and 850 nm) at 12.5 Hz. Then, signals were down-sampled to 10 Hz.

Data were collected from a total of 29 healthy individuals. Each participant performed three sessions of left and right hand motor imagery. Hand motor imagery tasks were repeated 10 times for each side in each session (with a total of 30 left tasks and 30 right tasks for all three sessions). Each task period lasted 10 s and was followed by 15-17 s rest.
The observation window, on which we have based all the further evaluations, consists of the rest interval that begins after the task is concluded. In this time frame, we may expect an increase in O_2Hb and a parallel decrease in HHb starting from a few seconds after task end until 20-30 s after task end.

2.2 Preprocessing of data

The whole data processing was performed using MATLAB R2021b [7] and its toolboxes (in particular, we used Martinez-Cagigal's function [8] for the Topographic map).

Calculation of HHb and O2Hb concentration changes Relative changes in the absorption of near-infrared light were converted to changes in the concentration of O_2Hb and HHb using the modified Beer-Lambert law, that means solving the linear system in eq. (1).

$$\begin{cases} \Delta A_{\lambda_1} = \alpha_{\lambda_1}^{O_2Hb} \cdot \Delta c^{O_2Hb} \cdot d \cdot B + \alpha_{\lambda_1}^{HHb} \cdot \Delta c^{HHb} \cdot d \cdot B \\ \Delta A_{\lambda_2} = \alpha_{\lambda_2}^{O_2Hb} \cdot \Delta c^{O_2Hb} \cdot d \cdot B + \alpha_{\lambda_2}^{HHb} \cdot \Delta c^{HHb} \cdot d \cdot B \end{cases} . \tag{1}$$

The inter-optode distance d was 30 mm. Molar Extinction Coefficient for Hemoglobin were taken from [9], and are respectively equal to the ones visible hereafter in eq. (2):

$$\begin{aligned} \alpha_{\lambda_1}^{HHb} &= 1548.52\,\mu\text{M}^{-1}\,\text{cm}^{-1} \qquad & \alpha_{\lambda_2}^{HHb} &= 691.32\,\mu\text{M}^{-1}\,\text{cm}^{-1} \\ \alpha_{\lambda_1}^{O_2Hb} &= 586\,\mu\text{M}^{-1}\,\text{cm}^{-1} \qquad & \alpha_{\lambda_2}^{O_2Hb} &= 1058\,\mu\text{M}^{-1}\,\text{cm}^{-1} \end{aligned} \tag{2}$$

Differential path lengths factor were estimated separately for the two wavelengths using eq. (3). The formula takes age in years (the mean age was calculated) and λ in nm [10].

$$B = 223.3 + 0.05624 * age^{0.8493} - 5.723 * 10^{-7} * \lambda^3 + 0.001245 * \lambda^2 - 0.9025 * \lambda; \tag{3}$$

Signal filtering Thereafter, Parametric Power Spectral Density was evaluated using Burg method with the purpose of estimating noise frequencies. Consequently, O_2Hb and HHb signals were band-pass filtered between 0.01 Hz and 0.1 Hz using a 6^{th} order Butterworth filter in order to remove high frequency physiological noises (which, for adults, have values of $\sim$1 to 1.5 Hz for heart beat and $\sim$0.4 Hz for respiration) [3]. Anticausal filtering was applied to avoid the introduction of latencies and phase distortion.

Task segmentation For each task of each session we extracted 35 s of the filtered signal consisting of: 5 s before the task start, 10 s of task duration and 20 s after the task end. Thus, we extracted for each channel of each subject a total of 30 epochs for the left task and 30 epochs for the right task.

Then, we performed a baseline correction on each epoch for both HHb and O_2Hb signals with the purpose of removing the offset present at the beginning of each task. To achieve this, we subtracted from each sample of the signal the mean calculated between -5 s and -2 s before the start of the task period.

Creation of averaged signals For each subject and each channel, we calculated the mean of the epochs of each side task. A result of this procedure is shown in Figure 1.

We also tried evaluating the mean of the channels between different subjects: however, this procedure introduces a huge flattening of our signals, resulting in an excessive loss of information. For this reason, this computation was discarded and not used in the following parts of our work.

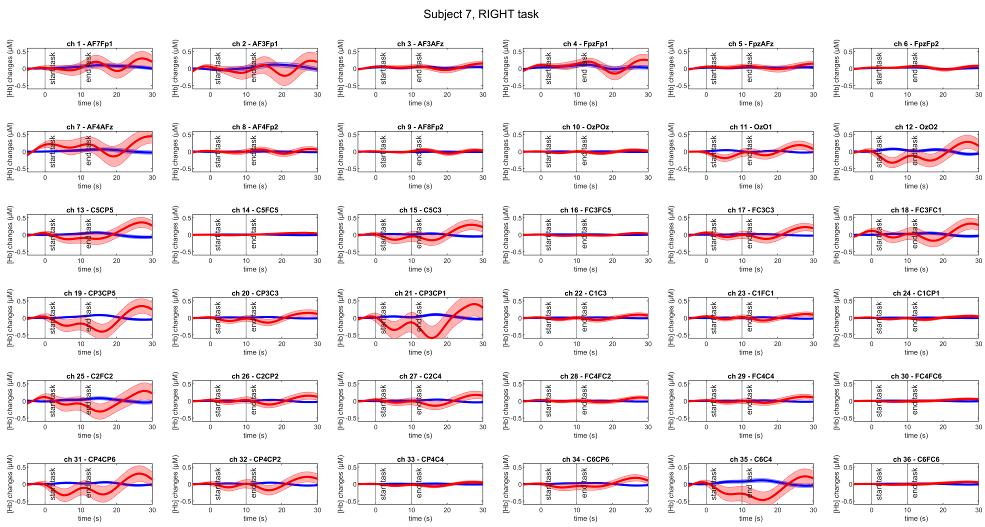


Figure 1: Averaged signals of all Right task for each channel of subject 7. Y-axis limits have been set to equal values for all channels, to better highlight those showing higher activations.

2.3 Classification

Features extraction: a dual approach Since we didn't notice any significant activation in O_2Hb or deactivation in HHb at a first sight, we tried to quantify the changes by extracting some features from the data. The window of observation on which we evaluated the features starts at the end of the task and continues until the end of the segmented epoch.

In particular, we tried two different approaches. On one side, we calculated 5 parameters: mean, variance, kurtosis, skewness and slope, following other studies [3], [4]. Concurrently, we tested a second approach that maintains temporal information of every epoch, instead of considering some global parameters [5]. To achieve this, the window of observation was partitioned in a series of 2 s consecutive segments and then the mean of each segment was calculated.

The parameters described above were calculated out of O_2Hb and HHb channels during the whole observation window, and first visualized in a topography on the scalp (colormap) [8]. For each single channel, the resulting arrays of features consist respectively of 5 statistical and 10 temporal features, respectively for the two approaches.

The procedure was separately implemented for both averaged and non-averaged epoch signals.

Classifier construction - Support Vector Machine The extracted parameters were then used to train some classifiers. After consulting the literature, we decided to use a SVM classifier that has shown considerable performances for motor imagery tasks [3]–[5], [11].

Throughout this project, by *Mediated classifiers* we will refer to classifiers that receive as input features extracted from averaged tasks, whereas we will call *Non Mediated classifiers* those in which features are extracted by all tasks considered individually. Classification was performed separately on both non-averaged (NM) and averaged (M) signals, then further subdividing the analysis between statistical (S) and temporal (T) features.
Each kind of classifier (Mediated and Non Mediated) was trained and tested separately on different groups of channels: frontal, occipital, motor and all channels. The whole procedure was repeated in the same way for both HHb and O_2Hb signals. After the training phase, we tested our classifiers on all tasks at the same time, then only on right tasks an only on left ones, to see if our classifiers showed better accuracies on one side. For better clarity on all the evaluated cases, see Figure 3.

The feature matrix was slightly different for the Mediated and Non Mediated case. In the Mediated classifiers, each subject is represented by 2 rows (one for the left and one for the right task), while in the Non Mediated classifiers every subject is represented by 60 rows (10 left and 10 right tasks for each of the three sessions). The columns of the feature matrix are the same for both Mediated and Non Mediated cases, so respectively:

- max. 180 features (= 5 features $\times$ number of O_2Hb (or HHb) channels) for the statistical approach;
- max. 360 features (= 10 features $\times$ number of O_2Hb (or HHb) channels) for the temporal approach.

The dataset was divided into Training Set and Test Set using two different approaches: 70-30 % random division for non-averaged signals, while for averaged signals we preferred a leave-one-out strategy, due to the reduced number of data. The leave-one-out approach was applied 29 times, each time choosing a different subject for testing. The overall accuracy of the Mediated classifier was evaluated by averaging the resulting 29 classifiers accuracies.

MANOVA p-value and classification consistency problem We then performed a Multivariate analysis of variance (MANOVA), in order to correctly interpret our results and to understand whether the selected features were statistically significant and thus discriminant for our classifiers.

MANOVA is a statistical test that allows taking into account multiple continuous dependent variables. We chose MANOVA test, indeed, because we needed to consider and compare more parameters at same time, as our classifiers are based on the overall variation of all features.
Due to the high dimensionality of our feature matrix, MANOVA was performed for each channel separately in all our classifiers. The results are shown through some topographical visualizations in Figure 2.

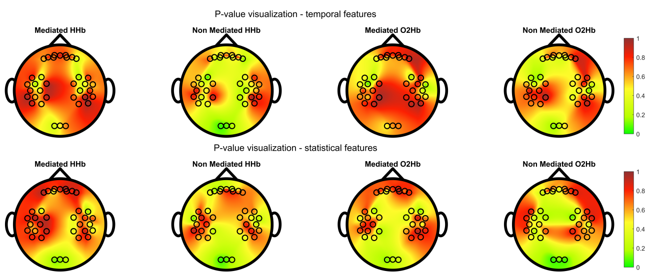


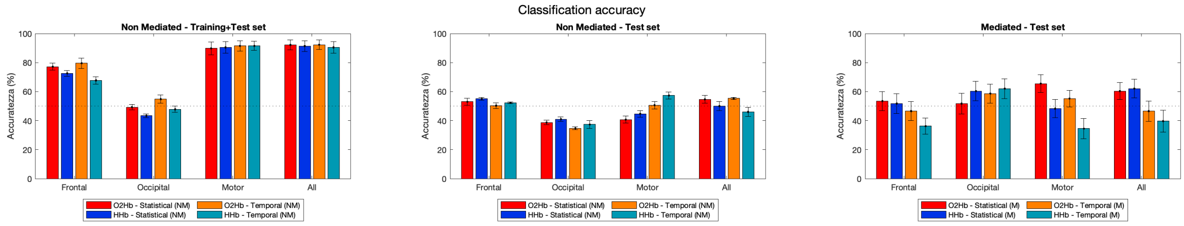
Figure 2: Topographical mapping of p-values, for each channel in all our classifiers

The null hypothesis was never rejected except for ch. FPZ-FP2, OZ-O1, C5-FC5 (in HHb, Non Mediated, with statistical features); ch. C6-C4 (in O_2Hb, Non Mediated, with temporal features), ch. OZ-O1, FC3-FC1 (in HHb, Non Mediated, with temporal features); ch. FC3-C3 (in O_2Hb, Non Mediated, with temporal features); ch. OZ-O1, OZ-O2 (in O_2Hb, Non Mediated, with statistical features).
These results show that both approaches are not truly consistent when evaluated across subjects. Consequently, even though it is used very frequently in the literature, this may turn out not to be the correct approach for training a classifier.

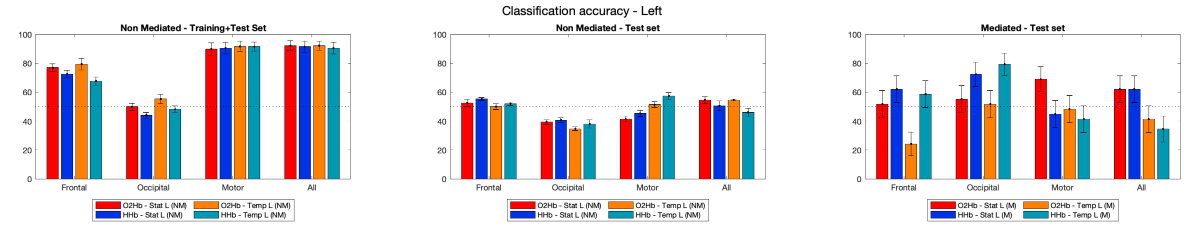
3 Results

The overall performances of our classifiers are shown in Figure 3. At a glance, it can be noticed that accuracy values simultaneously estimated on both Training and Test Set show that our classifier correctly learns from the Training Set, while the same parameters evaluated only on Test Set show that the classifier does not have the ability to generalize well.
Going further into detail, the classification accuracy oscillates around 50 % when evaluated on the Test Set. This result is not encouraging for our study as accuracies above 80 % would be needed to validate the hypothesis that NIRS signal can be used to distinguish between left and right task on its own.

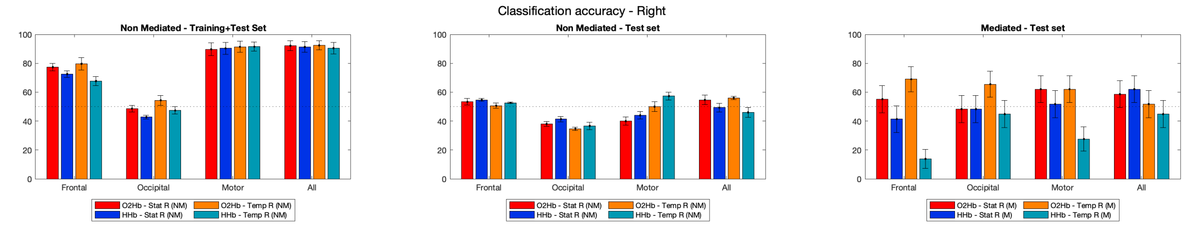
We noticed that we achieved lower results in the Non Mediated case than in the Mediated classifiers. This is possibily due to the fact that, by averaging the epochs, the signal variability is reduced. Although, an averaged approach would take us away from the hypothesis of using our classifier for a everyday use BCI, as it would require a long processing time before recognising a task.
We also tried to discern the classification consistency by studying the accuracy of our trained classifiers only for the recognition of right and left tasks. As can be seen comparing the results showed in figs. 3b and 3c, the differences in accuracy between the two cases are negligible. The non consistency of the implemented approaches is also confirmed by the statistical test performed with MANOVA and described in paragraph 2.3.



(a) Classification accuracy values analizing all tasks (Left+Right)



(b) Classification accuracy values analizing only Left tasks



(c) Classification accuracy values analizing only Right tasks

Figure 3: Every image has three histograms corresponding to accuracies of the three analized cases: Non Mediated case evaluated on the whole dataset, Non Mediated case evaluated on Test Set and Mediated case evaluated on Test Set. The results are showed for both HHb and O_2Hb. Each bar represents the accuracy of a classifier trained and tested on frontal, occipital, motor or all channels, with temporal (Temp) or statistical (Stat) features.

In literature, relevant results have been obtained in distinguishing between Motor Imagery and Motor Execution or between Motor Task and Rest [3], while little has been done in distinguishing between a left and a right Motor Imagery task.

A group of researchers, that used the same dataset as this project, seemed to have reached high performances in classifying left and right tasks [5]. Their approach was similar to our temporal approach for the non-averaged signal, except for extracting the features by averaging within 1 s windows. However, their high performances were calculated considering the accuracy of the classifier on both training and test set. As shown in fig. 3a, our classifier also reached high performances when tested on all the dataset, but this was not the purpose of our study. For our purpose, only the performances on Test Set should be considered. Unfortunately, none of the classifiers with temporal features reached performances that could reasonably justify a clinical application.

Similarly, a moving time window method was applied on signals coming from the same dataset as our study [6]. In this research, for NIRS signals, only 19 and 15 subject out of 29 achieved an acceptable classification for HHb and O_2Hb, respectively. The averaged accuracy between all subjects was lower than 60 %, perfectly fitting with our results. In this study, in addition, an hybrid approach was performed by combining NIRS signals with EEG. The hybrid approach showed higher performances, demonstrating that EEG is capable of compensating the low performances obtained considering NIRS signal on its own.

Other studies performed Motor Imagery classification with a SVM classifier by using some of the statistical features that we extracted [3], [4]. They were able to reach high performances in distinguishing between Motor Imagery and Motor Execution and between Motor Imagery/Execution and Rest. Since their classifier wasn't tested only on left or right tasks, their accuracy results cannot be compared to ours.

4 Conclusions

Considering the results showed in the previous paragraph, we can assert that an inter-subject approach is not sufficient to validate the use of NIRS signal alone for task side classification. Following other studies ([2], [3], [4], [6]), we tested both the use of temporal and statistical features, with the aim to compare them and to discriminate the best approach. Unfortunately, none of these methods was successful. Even when we tried to be more selective, by training our classifiers on restricted area of the brain (temporal, occipital or motor area), we didn't reach significant improvements.

We believe that the reason of reaching such low performances is due to the excessive variability of the NIRS signals between different subjects. Many previous studies have reported high performances when NIRS was coupled with other techniques (i.e. EEG) or when the Motor Imagery task was compared with Motor Execution or Rest task, but none of them achieved good performances in distinguishing between two controlateral Motor Imagery tasks.

Following our findings, we can assert that NIRS signal in Motor Imagery is extremely different between different subjects and it depends on too many factors that can't be taken into account and generalized (i.e. the ability of a subject to learn the task or its experience in performing it). Without any measurable feedback, it is not possible to know the exact instant in which the subject starts the imagery task on each trial. Moreover, the duration of the window in which he/she imagine the task on different trials may be different too, creating considerable difficulties for the subsequent signal recognition algorithm [11]. The Multivariate analysis of variance (MANOVA) performed on all features for each channels, confirmed that they were not statistically independent and, though, not sufficient to discriminate between such fine tasks.

As already discussed in section 3, similar studies with the same dataset have been carried out with an approach that consider each subject singularly (intra-subject approach) in order to train the classifier on the subject characteristics. This approach, although, only worked for almost 58 % of the subjects, confirming the hypothesis NIRS signal alone is not reliable for this kind of classification.

In addition to these considerations, as NIRS response arises between 10 s and 20 s after the task, its use would require a long time before being able to effectively detect the subject's intentions and send the command to a prosthesis. For these reasons we would exclude an application of NIRS signals for everyday use BCIs. In any case, it could be interesting to investigate its use for telerehabilitation in order to monitor changes in brain activities. Its low cost and portability, indeed, make it eligible for this purpose.

Bibliography and references

[1] N. Iso, T. Moriuchi, A. Sagari, *et al.*, "Monitoring Local Regional Hemodynamic Signal Changes during Motor Execution and Motor Imagery Using Near-Infrared Spectroscopy," *Frontiers in Physiology*, vol. 6, p. 416, 2016, ISSN: 1664-042X. DOI: 10.3389/fphys.2015.00416.

[2] P. Pinti, I. Tachtsidis, A. Hamilton, *et al.*, "The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience," *Annals of the New York Academy of Sciences*, vol. 1464, no. 1, pp. 5–29, 2020. DOI: 10.1111/nyas.13948.

[3] S. B. Erdoĝan, E. Özsarfati, B. Dilek, K. S. Kadak, L. Hanoĝlu, and A. Akın, "Classification of motor imagery and execution signals with population-level feature sets: Implications for probe design in fNIRS based BCI," *Journal of Neural Engineering*, vol. 16, no. 2, p. 026 029, Feb. 2019. DOI: 10.1088/1741-2552/aafdca.

[4] N. K. Qureshi, N. Naseer, F. M. Noori, H. Nazeer, R. A. Khan, and S. Saleem, "Enhancing classification performance of functional near-infrared spectroscopy- brain–computer interface using adaptive estimation of general linear model coefficients," *Frontiers in Neurorobotics*, vol. 11, p. 33, 2017, ISSN: 1662-5218. DOI: 10.3389/fnbot.2017.00033.

[5] A. H. Moslehi and T. C. Davies, "Comparison of classification accuracies between different brain areas during a two-class motor imagery in a fnirs based bci," vol. 2021-May, IEEE Computer Society, May 2021, pp. 702–705, ISBN: 9781728143378. DOI: 10.1109/NER49283.2021.9441376.

[6] J. Shin, A. von Lühmann, B. Blankertz, *et al.*, "Open Access Dataset for EEG+NIRS Single-Trial Classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1735–1745, 2017. DOI: 10.1109/TNSRE.2016.2628057.

[7] *MATLAB version 9.11.0.1809720 (R2021b)*, The Mathworks, Inc., Natick, Massachusetts, 2021.

[8] V. Martínez-Cagigal, *Topographic EEG/MEG plot*, en, 2021. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/72729-topographic-eeg-meg-plot.

[9] *Tabulated Molar Extinction Coefficient for Hemoglobin in Water.* [Online]. Available: https://omlc.org/spectra/hemoglobin/summary.html.

[10] I. Oshina and J. Spigulis, "Beer–Lambert law for optical tissue diagnostics: current state of the art and the main limitations," *Journal of Biomedical Optics*, vol. 26, no. 10, pp. 1–17, 2021. DOI: 10.1117/1.JBO.26.10.100901.

[11] R. Sitaram, H. Zhang, C. Guan, *et al.*, "Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface," *NeuroImage*, vol. 34, 4 2007, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2006.11.005.