Channel Selection for Optimal EEG Measurement in Motor Imagery-Based Brain-Computer Interfaces

International Journal of Neural Systems, Vol. 0, No. 0 (2020) 1–16

Channel selection for optimal EEG measurement in motor imagery-based brain-computer interfaces

PASQUALE ARPAIA
Department of Electrical Engineering and Information Technology (DIETI), Universita' degli Studi di Napoli Federico II, Naples, Italy.
Augmented Reality for Health Monitoring Laboratory (ARHeMLab).
E-mail: pasquale.arpaia@unina.it

FRANCESCO DONNARUMMA
Institute of Cognitive Sciences and Technologies, National Research Council (ISTC-CNR), Rome, Italy.
Augmented Reality for Health Monitoring Laboratory (ARHeMLab).

ANTONIO ESPOSITO
Department of Electronics and Telecommunications (DET), Politecnico di Torino, Turin, Italy.
Augmented Reality for Health Monitoring Laboratory (ARHeMLab).

MARCO PARVIS
Department of Electronics and Telecommunications (DET), Politecnico di Torino, Turin, Italy.

A method for selecting electroencephalographic (EEG) signals in motor imagery-based brain-computer interfaces (MI-BCI) is proposed for enhancing the online interoperability and portability of BCI systems, as well as user comfort. The attempt is also to reduce variability and noise of MI-BCI, which could be affected by a large number of EEG channels. The relation between selected channels and MI-BCI performance is therefore analyzed. The proposed method is able to select acquisition channels common to all subjects, while achieving a performance compatible with the use of all the channels. Results are reported with reference to a standard benchmark dataset, the BCI competition IV dataset 2a. They prove that a performance compatible with the best state-of-the-art approaches can be achieved, while adopting a significantly smaller number of channels, both in two and in four tasks classification. In particular, classification accuracy is about 77% - 83% in binary classification with down to 6 EEG channels, and above 60% for the four-classes case when 10 channels are employed. This gives a contribution in optimizing the EEG measurement while developing non-invasive and wearable MI-based brain-computer interfaces.

Keywords: EEG channels selection; EEG channel reduction; motor imagery; brain-computer interface.

1. Introduction

A brain-computer interface (BCI) provides an alternative communication channel by directly reading the brain activity and associating a meaning.[1–4] Although invasive measurements guarantee better signal-to-noise ratio,[5,6] entering the scalp is not suitable for many applications, especially for people with no impairments. Meanwhile, among the mostly adopted non-invasive measurements, electroencephalography (EEG) allows wearable, portable, and inexpensive acquisition of the electrical brain signals with the electrodes placed along the user's scalp.[7] A BCI based on sensorimotor rhythms is known as "motor imagery" (MI) BCI.[8,9] These rhythms can be measured by EEG. Other-

wise, magnetoencephalography (MEG), functional near-infrared spectroscopy (fNIRS), and hybrid approaches are also used.[10] Motor imagery refers to imagining a specific movement without executing it. Several researches have shown MI-BCI suitability in different applications, from communication and control to rehabilitation.[11,12] Unfortunately, it is more affected by inter-subject and inter-session variability than other brain activities.[13,14] Nonetheless, this kind of BCI is easy to use and comfortable because it relies on endogenous brain potentials, unlike other BCI paradigms requiring, for example, visual stimulation.[15–18] MI can be detected either synchronously (cue-based), or asynchronously (no cue is provided). Therefore, a MI-BCI is a suitable candidate for a wearable system in daily-life applications of communication and control. EEG channels selection allows to minimize the number of electrodes while guaranteeing acceptable performance, wearability, portability, and easiness to use.

Several literature studies concern the classification of motor imagery tasks. Most of them consider the discrimination between two classes, while fewer discriminate more classes.[19–24] This is confirmed by a recent review[25] that highlights classification accuracies up to 80% - 90% with down to 4 EEG channels for two-classes problems.[26–28] In an another review,[29] main finding about channel selection algorithms for motor imagery are reported. Some results are recalled in the following, while the reader is addressed to the bibliography for further details. For instance, a research of 2015[30] reports the results obtained with varying the number of EEG channels (14, 22, and 29). The number of selected channels and the accuracies crucially depend on the dataset taken into account and on the tasks to classify. This justifies the need to establish one or more benchmark datasets when validating a method by comparing its results to literature. With particular reference to the dataset 2a of BCI competition IV [31], considered as a benchmark in the present work, several approaches are reported in literature.[22,32–36] In Arvaneh et al.[27] two implementations of a sparse common spatial pattern (SCSP) are proposed for a two class discrimination: left hand versus right hand imagery. By selecting 13 channels, the classification accuracy was 81%, while it was 79% when reducing them to 8. However, different channels were selected for each subject. Other works report an a-priori channel selection.[37] As an example, the electrodes C3, Cz, and C4 of the 10-20 system were simply selected because they are related to the motor area of the scalp.[38] This approach was criticized for the poor performance, even in case of time-frequency optimization.[27,39] Even single-channels approaches were proposed with the aim to build utmost wearable and portable BCIs [40,41]. For instance, in,[41] the results in terms of cross-validation accuracy on training data seemed promising for 3 randomly chosen subjects. Nonetheless, these preliminary results were not confirmed when considering more data.

In this paper, a method for selecting and validating EEG channels is proposed for BCI design by exploiting the relation between channels and performance. Notably, the proposed method progressively selects EEG channels that are common to all subjects by considering their contribution to classification accuracy. Four motor imagery tasks are taken into account. The results show that accuracies compatible with the best approaches in literature can be achieved with a smaller number of channels, both in the two-classes and four-classes discrimination. Therefore, in the following, Section 2 introduces the proposed channel selection approach. Then, the results are reported in Section 3 and extensively discussed in Section 4. Conclusions follow in Section 5.

2. Proposal

In designing a BCI, the selection of a minimum number of meaningful EEG channels is a crucial task for enhancing the wearability and portability of the system, as well as to optimize the system performance by eliminating noisy channels or reducing over-fitting.[25] Notably, for MI-based BCI, the contribution of each channel to the recognition of a motor behavior is to be estimated. A method for selecting and validating EEG channels is proposed by exploiting the signal processing approach of the BCI competition IV winners with respect to the dataset 2a.[32] The proposed method adds a non-uniform embedding strategy[42] to evaluate the contribution of the candidate channels to the final performance. This proposal is a wrapper technique, because, in contrast with filtering approaches, the classifier is also involved.[25] The proposed method carries out a progressive selection of EEG channels in order to highlight the trade-off between classification performance and number of channels.

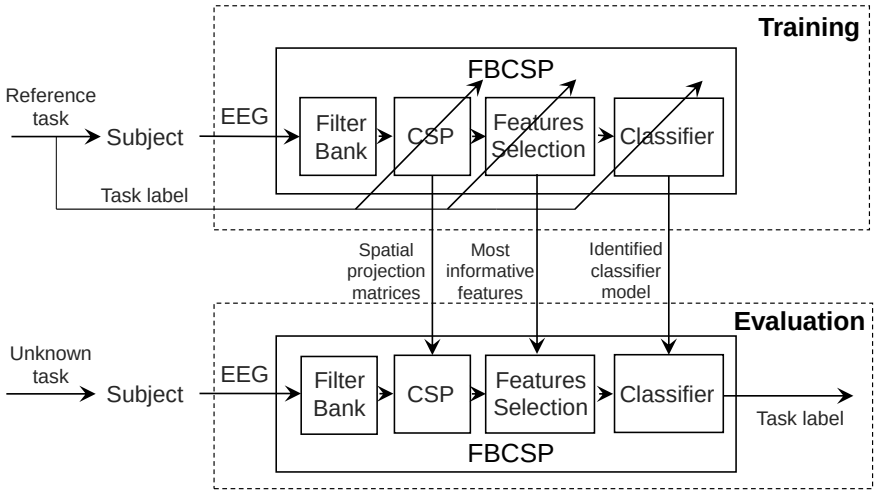

Figure 1: Signal processing approach proposed by the winners of the BCI competition IV, also known as "filter bank common spatial pattern", or FBCSP.[32]

2.1. *Channel selection*

A sequential forward selection (SFS) strategy[25,43] is adopted for an iterative selection of the best-performing channels. In the first iteration, a single channel is used for motor imagery classification. All available channels are tried one-by-one, and the channel leading to the best performance is selected. Similarly, in the second iteration, two channels are used. The first one (equal for all subjects) is the best of the previous iteration, while the second one is searched by scanning across the remaining channels and computing the performance of the resulting double-channel system. The next iterations follow the same principle, until the maximum number of available channels is reached. For each guess of the channel set, the classification performance is assessed by considering the mean classification accuracy μ and the standard deviation σ among all subjects. The best performance is determined as the one associated with the maximum value of $\mu-\sigma$. The aim is to select common channels for all subjects while minimizing the performance variance among them.

The channel selection exploits a 6-fold cross-validation. The folds contain the same amount of data and they are balanced with respect to the MI tasks. Hence, the available data are split $n_{cv} = 6$ times into 5 folds for training and 1 fold for evaluation. For each split, a processing based on the BCI competition IV winning approach[32] is exploited. Cross-validation accuracy is obtained for every subject by averaging the classification accuracies of the different splits. Then, μ and σ are the mean and standard deviation of such accuracies. The chosen signal processing approach is based on four main blocks, constituting both the training phase and the evaluation phase for each cross-validation split:

- a filter bank (FB),
- a common spatial pattern (CSP) filter,
- features selection with "mutual information-based best individual feature" (MIBIF),
- and a classifier.

In the training phase, the CSP, the MIBIF, and the classifier are trained with labeled EEG data. Then, during the evaluation phase, further EEG trials are classified by assigning a suitable label to each of them. The classification accuracy is assessed by comparing the obtained labels with the actual ones.

The described algorithm is represented in Fig. 1. Notably, for the channel selection, data from a first measurement session are taken into account, while data from a second session are exploited afterwards for validation. In both cases, these data can be arranged in an array with dimensions $n_{ch} \times n_{sub} \times n_{tr} \times s$, where n_{ch} is the number of channels, n_{sub} the number of subjects, n_{tr} the number of trials, and s the number of samples for each signal. The pseudo-code of the selection procedure is detailed in Alg. 1. Moreover, each block of the training and evaluation phases is explained in detail in the next subsections. For the purpose of illustrating the signal processing approach, the arrays X_T and X_E are considered

as the training and evaluation data of a single subject, respectively. Therefore, the dimension 'n_{sub}' is mainly ignored in presenting the blocks more easily. Nevertheless, it is easy to extend the reasoning to multiple subjects, as the proposed method requires.

Algorithm 1 Channel selection(X, y)

Require: selection data array X,
selection data labels y.
Ensure: sorted channels array CH,
(μ, σ) associated with channel.

$CH, M, \Sigma = [\text{“}empty'']$
for $iteration = 1 : n_{ch}$ **do**
  for $ch = 1 : n_{ch} \wedge ch \notin CH$ **do**
    $CH_{test} = [CH, ch]$,
    $X' = X(CH_{test})$
    for $sub = 1 : n_{sub}$ **do**
      split X' *in* (X_T,X_E) n_{cv} *times,*
      split y *in* (y_T,y_E) *accordingly.*
      for $i = 1 : n_{cv}$ **do**

block B1 - filter bank
        $X_T^f = ChebychevII(X_T, band_f)$
        $X_E^f = ChebychevII(X_E, band_f)$
block B2 - common spatial pattern
        $W_p = findMatrices(X_T^f, y_T)$
        $\xi_T = features(X_T^f, W_p)$
        $\xi_E = features(X_E^f, W_p)$
block B3 - features selection
        $I_p = findBestFeatures(\xi_T, y_T)$
        $\xi_T^r = selectFeatures(\xi_T, I_p)$
        $\xi_E^r = selectFeatures(\xi_E, I_p)$
block B4 - classifier
        1) Training with $(\xi_T^r, y_T) \rightarrow clf_p$
        2) $\mu(i, sub, ch) = clf_p(\xi_E^r, y_E)$

      end for*:cross-validation*
      $\mu_{cv}(sub, ch) = \frac{\sum_i \mu(i,sub,ch)}{n_{cv}}$
    end for*:subjects*
    $\mu = \frac{\sum_{sub} \mu_{cv}(sub,ch)}{n_{sub}}$
    $\sigma = \sqrt{\frac{\sum_{sub}[\mu_{cv}(sub,ch)-\mu]}{n_{sub}-1}}$
  end for*:test channel*
  select ch_{opt} associated with maximum $\mu - \sigma$
  $CH = [CH, ch_{opt}]$
  $M = [M, \mu_{opt}]$, $\Sigma = [\Sigma, \sigma_{opt}]$
end for*:iterations*

2.1.1. *Filter bank*

Type–II Chebyshev filters are employed for the filter bank. Each filter has a 4 Hz-wide pass band, and consecutive filters have a 2 Hz overlap, resulting in 17 bands from 4 Hz to 40 Hz (4-8, 6-10, 8-12, ..., 36-40). Later, the features selector chooses the best bands subject-by-subject. This also means that the band choice is age-related. However, an age-dependent baseline could be also adopted in future, and the 4-40 Hz range extended. The attenuation in the stop band was set to -40 dB, while the order was set to 10, so to achieve a -200 dB/decade slope in the frequency response. By filtering each signal of X_T, a 4D array is derived. This array is referred to as X_T^f, and its dimensions are $n_{ch} \times n_{tr} \times s \times f$, with $f = 17$ in the current implementation. The same filtering is also applied to each signal of X_E to obtain X_E^f.

2.1.2. *Spatial filtering*

The CSP spatial filtering is applied to filtered data by means of projection matrices derived in the training phase. It has been demonstrated that spatial filters for multi-channel EEG effectively extract discriminatory information from two populations of single-trial EEGs.[44] Hence, the CSP typically applies to binary classification, but multi-class extensions are possible.[45] Among the various possibilities, the one-versus-rest (OVR) approach is adopted here.[32] This means that, since the CSP can discriminate between two classes, each class is considered against the remaining ones.

The computation of the projection matrices from training data is conducted as follows. A projection matrix exists for each filter band and for each class, i.e. W_p^i associated with the $i-th$ band and class p. In the new space, the first m CSP components have maximum variance associated with class p and minimum variance associated with the remaining classes. Simultaneously, the last m components have minimum variance associated with class p and maximum variance for the others. The calculation steps are here reported for the sake of an easy reproducibility of the algorithm, while the reader is addressed to the literature for further details.[32,45]

To simplify the notation, the filter band is fixed, but the following applies to all bands. Each filtered trial of the training data belonging to class p, namely $\chi_{j,p} = \chi_{j,p}^i$, is involved in the calculation of the class

p covariance matrix

$$K_p = \frac{1}{n_{tr,p}} \sum_{j=1}^{n_{tr,p}} \frac{\chi_{j,p}\chi_{j,p}{}'}{trace\left(\chi_{j,p}\chi_{j,p}{}'\right)}, \tag{1}$$

where $n_{tr,p}$ is the number of training trials belonging to class p, and the dimensions of the matrix $\chi_{j,p}$ are $n_{ch} \times s$. The composite covariance matrix is then obtained by summing the covariance matrix of each class, namely

$$K = \sum_p K_p. \tag{2}$$

The complete projection matrix W_p^c is computed by solving the eigenvalue decomposition problem

$$K_p W_p^c = K W_p^c \Lambda_p, \tag{3}$$

where Λ_p is a diagonal matrix made of eigenvalues. In the present implementation, the eigenvalues are sorted in ascending order in Λ_p, and the eigenvectors constituting W_p^c are sorted accordingly. Finally, the projection matrix W_p^i is obtained by considering the first m and the last m columns of the complete projection matrix, where $m = 2$ is empirically set.[32] Ultimately, $f = 17$ projection matrices are obtained for each class p.

The matrices W_p^i transform each $n_{ch} \times s$ subarray of X_T or X_E, corresponding to a single trial and a single band, into a new space associated with the class p. In applying the CSP, each signal is projected in all the possible class-related spaces. Since the class of such signals is in principle unknown during this step, the adopted notation will be χ_j^i, i.e. the subscript p cannot be applied. Clearly, each signal must be described with features corresponding to all possible classes. The signal features are derived from the covariance matrix

$$C_{j,p}^i = W_p^{i\prime} \chi_j^i \chi_j^{i\prime} W_p^i, \tag{4}$$

where χ_j^i is the $n_{ch} \times s$ filtered signal to project, and W' denotes the transposed matrix. It is to remark that, while the computation of projection matrices discussed above is only carried out during the training phase, the CSP projection is conducted both in the training and evaluation phases. From the matrix $C_{j,p}^i$ the features are obtained as

$$f_{j,p}^i = \log\left[\frac{diag(C_{j,p}^i)}{trace(C_{j,p}^i)}\right]. \tag{5}$$

The $f_{j,p}^i$ describe the $j-th$ trial with respect to the class p and the $i-th$ band. Since the projection matrix W_p^i has dimensions $n_{ch} \times 2m$, the covariance matrix $C_{j,p}^i$ is a square matrix with dimensions $2m \times 2m$ and the features for each trial are $2m$ per band. By concatenating the features of all the bands, each trial j is described with $2mf$ features with respect to the class p. By bringing together all the trails, two features matrices can be created per each class, $\xi_{T,p}$ and $\xi_{E,p}$. They have dimensions $n_{tr} \times 2mf$, and they are associated with training data and evaluation data, respectively. Note that the computation explained in this subsection in the general multi-class case can be obviously applied to binary classification problems.

2.1.3. *Features selection*

At this step, the features matrix $\xi_{T,p}$ (or the $\xi_{E,p}$) is processed. As abovementioned, its dimensions are $n_{tr} \times 2mf$, and the selection reduces the columns to I_p features. Hence, the subset of features will be different for each class p. In particular, for this selection, the mutual information (MIBIF) between a feature and the class p is calculated for each feature, and the features with the highest mutual information are selected. The features matrices associated with the labeled training data, i.e. $\xi_{T,p}$, are employed to train the selection process, while the same features are then selected both from $\xi_{T,p}$ and $\xi_{E,p}$.[32]

The number of selected features depends on the fact that the CSP features are paired. Therefore, if one element of the pair is selected, the other has to be selected too. If the two features of a pair are both selected as features with high mutual information, the total number of features will then be smaller. Thus, the outputs of this third step are reduced feature matrices $\xi_{T,p}^r$ (or $\xi_{E,p}^r$), whose dimensions are $n_{tr} \times I_p$, with I_p ranging from k to $2k$. In the proposed implementation, $k = 5$ was empirically chosen.

2.1.4. *Classification*

The proposed channel selection method consists of determining, among the available channels describing the observed processes, the most significant in the sense of predictive information. In doing that, the iterative procedure is defined for selecting a minimum number of channels that can effectively classify the target behaviors. A new candidate channel is added to the set if the related classification performance is

the best among the other possible channel combinations. Thus, in the final processing step, the features obtained from the MIBIF must be classified.

In choosing the most suitable classifier, state-of-the-art solutions were compared, i.e. a k nearest neighbors (kNN) classifier,[46] a Naïve Bayesian Parzen Window (NBPW),[32] and a support vector machine (SVM)[47]. The best classifier was then chosen according to experimental results. This are basically binary classifiers, but an OVR approach was again considered for the multi-class extension: each class is discriminated against the remaining ones, and then the most probable class is assigned to each trial thanks to the score associated with each binary classification. The classifier is trained with the features matrix $\xi^r_{T,p}$ and its associated labels. Then, the features arrays associated with each unlabeled trial are extracted from $\xi^r_{E,p}$ for classification, and the derived label are compared with the real ones to calculate the classification accuracy.

The channel selection procedure is performed by exploiting the classification step. In particular, the resulting accuracies are averaged among the different splits of the cross-validation. Finally, in each iteration of the channel selection, the mean μ and standard deviation σ of the accuracies, calculated among the subjects, are considered in choosing the channel with optimal $\mu - \sigma$.

2.2. *Channel validation*

The result of channel selection is a sequence of channels sorted according to their predictive information. In designing a BCI system, these results are useful in choosing a combination of $n_{sel} = 1, ..., n_{ch}$ channels. Given the number of channels, the proposed method also points out an expected classification performance in terms of (μ, σ). Next, a subset of the data from the first measurement session can be extracted by considering the n_{sel} identified channels. The BCI design is finalised by employing these data subset for the CSP, MIBIF, and classifier training. The final system is thus defined by the CSP projection matrices W_p, the features I_p to select according to the MIBIF, the trained classifier clf_p, and, of course, the EEG channels $CH(1 : n_{sel})$.

A testing phase can be carried out in order to validate such a design. This procedure is described in Alg. 2. In this case, an array X_v containing validation data is taken into account for the evaluation phase. The data in X_v must be independent from the data from the first measurement session contained in the array X. Thus, the former is usually associated with a further measurement session. In general, the number of trials and even the number of samples could be different between X_v and X, while the available EEG channels must be the same.

Algorithm 2 Validation(X_v, y_v, CH, W_p, I_p, clf_p)

Require: validation data array X_v,
validation data labels y_v,
n_{sel} channels selected from CH.
CSP projection matrices W_p,
features to select I_p,
trained classifiers clf_p *for* $p = 1, ..., classes$,
Ensure: system performance: (μ, σ).

1: $X'_v = X_v(CH(1 : n_{sel}))$
block B1 - filter bank
2: $X^f_v = ChebychevII(X'_v, band_f)$
block B2 - common spatial pattern
3: $\xi_v = features(X^f_v, W_p)$
block B3 - features selection
4: $\xi^r_v = selectFeatures(\xi_v, I_p)$
block B4 - classification
5: $\mu(i, sub, ch) = clf_p(\xi^r_v, y_v)$

It is worth noting that the system performance is assessed by considering the mean performance among subjects, but the algorithm is trained subject by subject. Therefore, the procedure of Alg. 2 applies to each subject, while the mean classification accuracy among subjects must be ultimately considered for validation since the claim is to select subject-independent channels.

3. Results

In this section, the results achieved in applying the proposed method are reported. The analyses were conducted on signals extracted from the dataset 2a of BCI competition IV. This is briefly described in the next subsection, while further details can be found in the bibliography.[31] The scripts for data processing were implemented in MATLAB®.

3.1. *Dataset*

The dataset 2a of BCI competition IV contains data related to 9 subjects. For every subject, signals were

recorded in 2 sessions held on different days, referred to as "session T" and "session E". Each session consisted of up to 3 runs related to eye movements, and 6 runs related to motor imagery separated by short breaks. The runs related to eye movements are ignored in the present work. The remaining 6 runs consist of 48 trials each. The subjects could perform four motor imagery tasks, namely imagining the movement of left hand, right hand, both feet, or tongue. A cue-based paradigm presented the sequence of tasks to perform in random order but balancing the tasks per run. Each trial contains data from 3 EOG channels and 22 EEG channels. Only the EEG channels were considered, and the 3 s time windows related to the motor imagery were extracted. These signals were recorded with Ag/AgCl wet electrodes placed according to the 10-20 system[37] to cover the motor cortex. As shown in Fig. 2, reference and ground electrodes are placed on the left and right ear lobes, respectively. The sampling rate was 250 Sa/s, a bandpass-filter was applied with band 0.5 Hz - 100 Hz, and a 50 Hz notch filter was also enabled to suppress line noise. Artefacts were identified by visual inspection and trials containing them are marked within the dataset. However, they were not removed from the dataset.

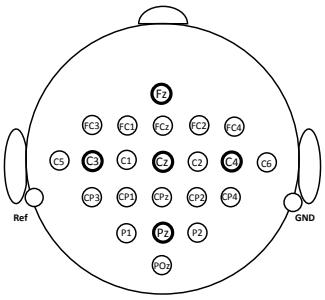

Figure 2: EEG electrodes employed during the data acquisition, placed according to the 10-20 standard.[31,37]

3.2. *Optimal Classifier*

Preliminary tests were performed to choose the classifier to employ in Alg. 1 and Alg. 2. According to previous section, the investigated classifiers where a kNN, an NBPW, and an SVM, as well as their multi-class extensions (OVR approach). Tab. 1 shows the results in terms of mean accuracy and standard deviation among the 9 subjects. For each subject, and for each classifier, the accuracy was calculated with a 6-fold cross-validation on data from the session T, thus choosing the optimal classifier regardless of data from session E used afterward for validation. All the 22 EEG channels were taken into account for this step. Results show comparable performances on the classifiers. Nevertheless, it is possible to argue that NBPW is to prefer. To this aim, paired t-tests[48] were performed to compare first NBPW with kNN, and then NBPW with SVM. Tests were conducted for each row of Tab. 1 with the null hypothesis that

$$\mu_N \leq \mu_{K/S}, \tag{6}$$

where μ_N and $\mu_{K/S}$ are the mean accuracies for the NBPW and kNN/SVM, respectively. Rejecting the null hypothesis would suggest that the NBPW is better than the competing classifier. The level of significance for this one-sided test was fixed at $\alpha = 5\%$. The null hypothesis was rejected only in a single case ('left hand vs tongue') in comparing the NBPW with the kNN, thus suggesting that the NBPW is slightly better (p-value = 0.0145). Instead, no evidence showed a difference between the NBPW and the SVM. However, it is worth mentioning that, in training the classifiers, hyperparameters tuning was carried out for SVM and kNN, while this was not needed for NBPW. This suggests that NBPW for the implemented processing.

	ACCURACY (%)		
CLASSES / CLASSIFIERS	kNN	NBPW	SVM
left hand vs right hand	73 ± 19	74 ± 20	74 ± 20
left hand vs feet	82 ± 13	81 ± 13	82 ± 12
left hand vs tongue	82 ± 13	82 ± 13	81 ± 13
right hand vs feet	81 ± 15	81 ± 15	81 ± 15
right hand vs tongue	83 ± 13	84 ± 13	83 ± 14
feet vs tongue	75 ± 14	75 ± 13	76 ± 13
four classes	63 ± 20	63 ± 19	63 ± 19

Table 1: Comparison between different classifiers, carried out by taking into account all EEG channels. Mean and standard deviation of the cross-validation accuracy are considered among the 9 subjects.

3.3. *Classification accuracy*

The iterative selection method was implemented both for the discrimination among the six possible pairs of classes, and for the discrimination among the four classes. Channel selection was conducted by considering a cross-validation on data from session T. The results in terms of mean cross-validation accuracy among the 9 subjects are reported with a blue line in Fig. 3 and Fig. 4, for the binary cases and the multi-class case respectively. The x-axis reports

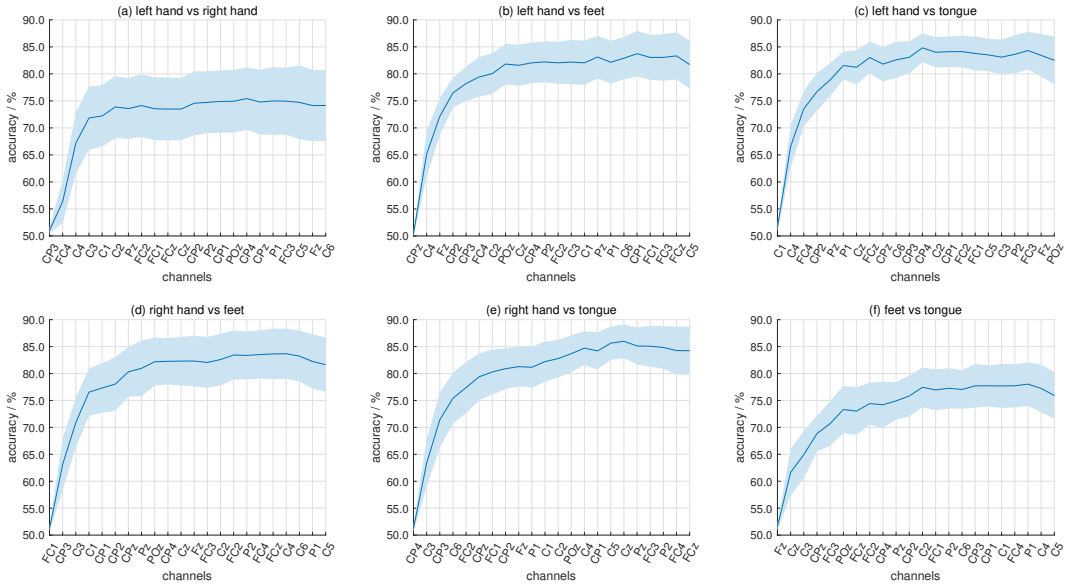

Figure 3: Mean classification accuracy and standard deviation of the mean in cross-validation for progressively selected channels and for each pair of classes.

the progressively selected channels, while the y-axis reports the classification accuracy μ in percentage. Actually, aiming at comparing Fig. 3 and Fig. 4 directly, the y-axes should report normalized performance metrics such as the kappa coefficient.[49] However, classification accuracies are commonly reported in literature. Therefore, the values reported on the y-axes are such that there is a one-by-one correspondence between Fig. 3(a)-(f) and Fig. 4, i.e. they correspond to the same kappa coefficients. These plots show the trade-off between the number of selected channels and the classification accuracy. Depending on the considered tasks, an acceptable performance can be also achieved with only 4 channels. This is especially true for binary classification problems, while more channels are required in the four-task classification. The area defined by the standard deviation of the mean accuracy is also reported in light blue. This is calculated as the standard deviation among subjects divided by the square root of the number of subjects, i.e. $\sigma_\mu = \sigma/\sqrt{n_{sub}}$. It should also be noticed that maximum accuracy is sometimes reached with less than 22 channels. In the present case, those values are merely a random occurrence because those values are not significantly different from the accuracy at 22 (from the statistical point of view).

In a further step, the sequences of selected channels were validated by employing the independent data from the session E. The mean accuracy among the 9 subjects and its standard deviation are plotted in Fig. 5 and Fig. 6. For each plot, the channels on the x-axis correspond to the respective sequence found during the selection step, while the accuracy values on the y-axes were chosen as explained before. In detail, the validation starts by considering the first channel of the proper sequence, and then progressively the other channels are added as they were found within the selection procedure.

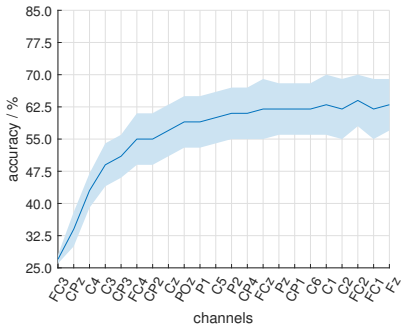

Figure 4: Mean classification accuracy and standard deviation of the mean in cross-validation for progressively selected channels in the multi-class problem.

Validation results were analyzed with paired t-tests[48] in order to compare the accuracy at 22 chan-

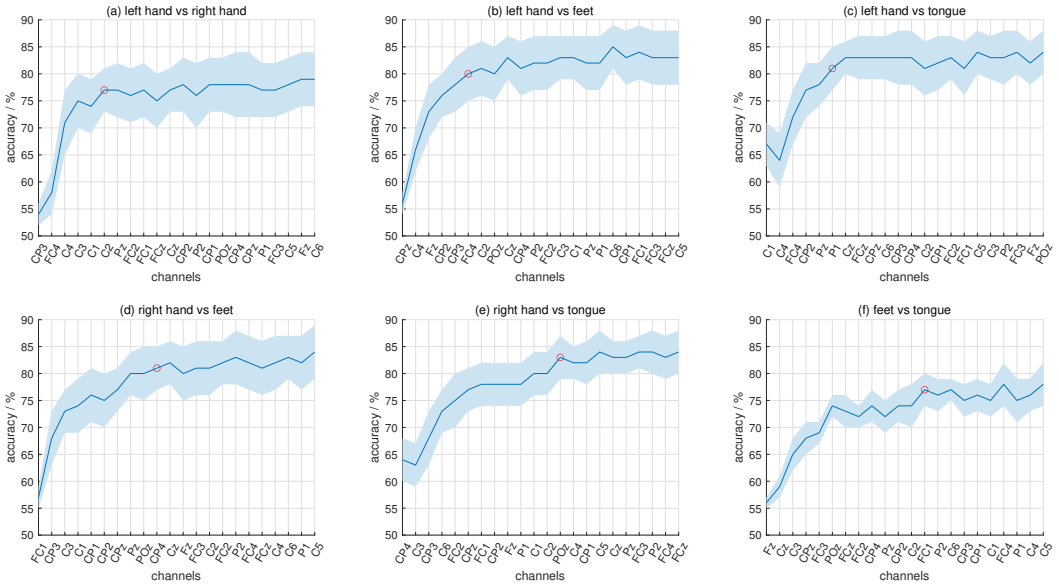

Figure 5: Mean classification accuracy and standard deviation of the mean obtained in validating the respective channels sequence.

nels with the one at a reduced number of channels. The null hypothesis was

$$\mu_{22} \leq \mu_{red}, \tag{7}$$

where μ_{22} and μ_{red} are the mean accuracies for the 22 channels and a reduced set of channels, respectively. Rejecting the null hypothesis would suggest that the classification performance at 22 channels is significantly better. For this test, the level of significance was set to $\alpha = 5\%$. Nevertheless, failing to reject would not mean that the null hypothesis is necessarily true. Therefore, the probability β of a false positive was also taken into account. In particular, $\beta \leq 5\%$ was considered as a reasonable risk of accepting a false positive. In accordance with the figures, these tests pointed out that performances are significantly worse when 3-5 channels are considered, while they become acceptable with at least 6-13 channels (depending on the considered tasks). In detail, the acceptable minimum number of channels is reported in Fig. 5 and Fig. 6 with a red circle. To sum up these results, the classification performances achieved with the reduced number of channels are also reported in Tab. 2. Notably, this table reports the standard deviation σ, instead of the standard deviation of the mean $\sigma/\sqrt{n_{sub}}$, since this is common in literature.

Finally, to better understand where the most predictive information is located on the scalp, Fig. 7 and Fig. 8 show contour plots in which the i-th channel is weighted according to

$$w(i) = 1 - \frac{i-1}{n_{ch}}, \tag{8}$$

where $n_{ch} = 22$ in the present case. Thus, $w(i)$ is a weight between 1 and $1/n_{ch}$ assigned according to the selection order of channels per each sequence.

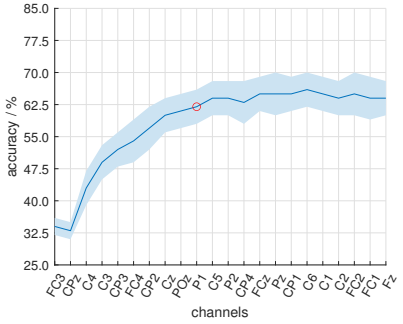

Figure 6: Mean classification accuracy and standard deviation of the mean obtained in validating the channels sequence.

Eq. 8 assigns a decreasing importance to the progressively selected channels and it helps the discussion, reported in the following section, by highlighting the most informative brain areas.

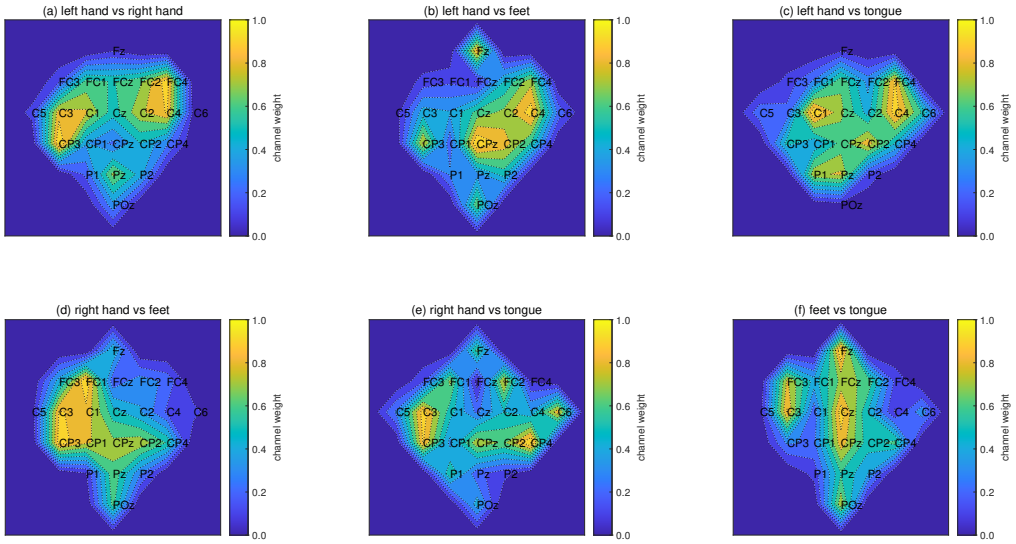

Figure 7: Location of the most predictive information on the scalp for each pair of classes

	ACCURACY (%)	
TASKS / CLASSIFIERS	reduced ch.	22 channels
left hand vs right hand (6)	77 ± 12	79 ± 15
left hand vs feet (6)	80 ± 15	83 ± 15
left hand vs tongue (6)	81 ± 12	84 ± 12
right hand vs feet (10)	81 ± 12	84 ± 15
right hand vs tongue (13)	83 ± 12	84 ± 12
feet vs tongue (13)	77 ± 9	78 ± 12
four classes (10)	62 ± 12	64 ± 12

Table 2: Mean and standard deviation of the classification accuracy obtained during validation, for both the minimum number of channels (reported for each row) and 22 channels.

4. Discussion

The main idea of the work is to reduce the number of EEG electrodes in designing a wearable BCI device for daily-life applications. The proposed method allows to select a reduced set of EEG channels while keeping good performance. The classification accuracies decrease of few percentage points with respect to 22 channels when more than 3-5 channels are selected (depending on the considered tasks), but statistical tests demonstrated that the performance differences were not significant. In particular, the classification accuracy is kept around 77% - 83% for 2 tasks, and it is kept around 62% for 4 tasks classification. In some cases, the channel reduction also leads to a slight improvement (decrease) of the standard deviation. Further validation tests were also carried out by inverting the data from session T and session E of the benchmark dataset, and then by considering data from BCI competition III dataset 3a.[50] Statistical tests confirmed that classification performance was significantly different when the channels set is reduced down to the first 3-5, while for more channels the performance was acceptable. Nonetheless, some limitations emerged.

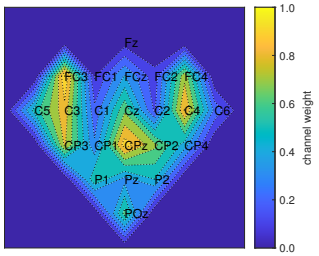

Figure 8: Location of the most predictive information on the scalp for the multi-class problem

4.1. *Results significance*

Firstly, if data from session E is employed for training, data from session T for evaluation, and the same channel sequence of Fig. 3 and Fig. 4 are respectively used for each task combination, there is

an accuracy diminishing of 1%-5%. This diminishing holds for both the reduced and non-reduced sets. Therefore, the channel selection is still effective because the performances for reduced and non-reduced sets result not statistically different. Meanwhile, the processing approach should be improved with respect to the training data sensitivity. In a second step, data from BCI competition III was exploited, with particular reference to the dataset 3a. This dataset contains data from 3 subjects, and, by adopting the FBCSP processing approach, the classification accuracy for the maximum number of channels (60) resulted above 80% for 2 tasks, and above 70% for 4 tasks. In aiming at using the reduced sets of channels as before, it is to point out that the signals in the dataset 3a were not acquired at the exact same positions as dataset 2a. Hence, the channels closest to the respective standard positions adopted in the BCI competition IV were selected. The accuracies for the respective reduced sets of channels resulted between 78% and 92% for in case of 2 tasks, and equal to 72% with 10 channels employed in the 4 tasks case. It is to notice that the four mental tasks involved in the dataset 3a are quite compatible with the ones of the benchmark dataset, but a single foot is considered instead of both feet. In spite of this, satisfying performance was achieved. Also note that the considered mental tasks are intuitively associated with commands such as 'left', 'right', 'down', and 'up'.

The discussed results are in general comparable or improve the findings in the recent literature. Actually, only a single work recently validated a channel reduction procedure on the dataset 2a of BCI competition IV.[51] In that work, the authors considered one binary case, namely right hand versus left hand, and they obtained a mean classification accuracy equal to 76%, among 8 subjects out of 9, by selecting different channels for each subject (6−15 channels). In another work on the same dataset, all the 4 tasks are considered, but in this case a channel selection procedure was not proposed. Instead, 9 channels were manually selected, and the resulting accuracy equaled 50%.[52] Several other works propose channel selection or reduction procedures working on different datasets. The method suggests by Parashiva et al.[53] in discriminating between left and right hand reduces the channels number from 31 to 13 obtaining the 74% of accuracy with a subject-dependent selection, while the achieved accuracy is 66% when channels are reduced to 11 with a subject-independent approach. Feng et al.[54] found the minimum number of channels depending on the considered frequency band. Only two classes were discriminated, and two different datasets were taken into account. For the first dataset, the channels were reduced from 118 to 30, and the resulting accuracy is 82%. Instead, for the second dataset, the channels are reduced from 59 to 24 and the resulting accuracy is 77%. Although they start from a number of channels that is greater than the one considered in the present manuscript, the reduced set basically contains channels located in the sensorimotor areas of the scalp. Therefore, the findings are somehow compatible with the results presented above.

4.2. *Possible improvements*

Considering the widely spread datasets from BCI competitions is indeed useful in comparing the proposed method with other works in literature. However, the classification accuracies for the general population might be lower because the best subjects could be selected for a competition. Notably, it was shown that about 7% of subjects from general population give lower performance in two-class motor imagery.[55] This consideration supports the choice of the dataset 2a of BCI competition IV as benchmark, because a relatively large number of subjects was involved. On the other side, the results indicate that improvements are still needed for practical applications. The first limiting factors are indeed the poor signal-to-noise ratio associated with non-invasive techniques like EEG. Interestingly, the present FBCSP approach proved quite robust to artifacts, because excluding marked artefacts from the dataset was not enhancing performance. Ultimately, results suggest that future works should focus on EEG non-stationary. This is particularly evident from validation results, where the employment of data from a different session introduced performance fluctuations. Nonetheless, in the present case, figures suggest that there is no significant difference. For instance, in Fig. 6, the accuracy diminishes when the second channel is added to the first, but a t-test can demonstrate that this difference is not statistically significant.

Indeed, the adopted architecture is modular and can be furtherly refined and explored, especially in

dealing with EEG non-stationarity. For instance, the adoption of a convolutional neural network for extracting and classifying CSP features was recently proposed to capture temporal changes in the EEG signals.[56] Such a framework seems promising to mitigate the effects of inter-subjects and inter-sessions variability. One possible improvement of the proposed method, which can be addressed in future works, is an efficient treatment of online signals when dealing with a very large number of starting available channels. In this case, the framework could be integrated with a preliminary reduction analysis step based on techniques of sparse dictionary learning, e.g. sPCA, sparse coding, and related approaches (see[57,58]). An efficiency trade-off could guide the tuning of such a pre-analysis step. Finally, the iterative channel selection phase could be also enhanced. In detail, though the proposed method selects the best possible channel for each iteration, figures show that there is an accuracy diminishing for some selection steps, while this increases again if one or more other channels are added. This evidence suggests that channels could be correlated, and correlation-based selection could improve the results.[59] A reasonable target will be to overcome 90% accuracy for two-tasks discrimination, and 85% for four-tasks discrimination (both would correspond to a 0.8 kappa coefficient). In the latter case, further studies will probably be needed also for the multi-class processing approach.

4.3. *Channels significance*

Literature reports that the left hand generally shows bilateral activation, while the right hand mainly shows contralateral activation schemes, i.e. the left hemisphere of motor area is typically involved. Then, the feet-related area is located inside the interhemispheric fissure of the sensorimotor cortex, and feet show a strong bilateral activation when they are considered together.[60,61] Finally, patterns and motor area of the tongue are concentrated in the bilateral premotor cortex that covers the central sulcus and in the right putamen. Apparently, this should limit the capability of recording tongue-related activity with surface electrodes. However, it was demonstrated[62] that tongue movement imagery, as well as foot-related imagery, enhances the rhythms in the neighboring cortical areas, and by consequence this allows the possibility to discriminate the tongue imagery itself. In this regard, Fig. 7 tries to highlight the areas of the scalp were the predictive information is concentrated per each binary classification problem, while Fig. 8 highlights the areas for the four classes problem. Indeed, there is a correspondence between these brain areas and the literature recalled above. Notably, the importance of the median line activities in Fig. 7 is minimized when feet and/or tongue imagery is involved, in accordance with the considerations about the specific task-related activity. This supports the idea that the channels identified for a reduced set have a meaningful neurological counterpart. It is also interesting to note that, by considering the weights associated with the channels per each classes pair, the electrodes that most likely covered the field potentials are "CP3", "C3", "CPz", "CP2", "C1", "Cz", "CP4", and "C4". They are directly related to the somatosensory and motor areas,[63] and most of them are also among the first channels of the multi-class sequence.

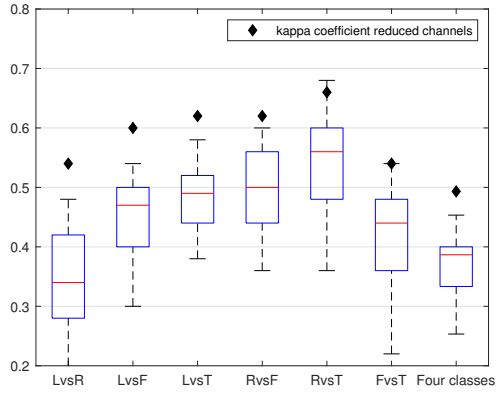

Figure 9: Comparison between the performance achieved with a reduced number of EEG channels and thirty tests with random channels.

A final test was carried out to prove the robustness of the presented results. In this test, random channel sequences were selected for thirty times. The number of random channels goes from 4 to 13, i.e. a number of channels for which statistical tests showed no evidence of significant difference with the 22 channels case. The aim was to show that it is not enough to select any set with a reduced number of channels, but that the proposed method allows to identify an optimal set. The proof of that is shown in the box plot of Fig. 9, where each distribution of the kappa

coefficients of the "random channels tests" are plotted against the ones corresponding to the reduced channel set reported in Tab. 2. Kappa coefficients were used in spite of accuracies in order to compare the 2 task cases with the 4 task case. It can be seen that the values found with the proposed procedure (black diamonds in Fig. 9) are always in the upper part of the respective distribution. Interestingly, the proposed channel selection finds an optimal set without testing all possible combinations of channels. Therefore, such a procedure could be also exploited in online channel selection due to its computational efficiency.

5. Conclusions

This paper proposes and validates a method aimed at EEG channel selection applied to the analysis of motor imagery signals. The selected channels are subject-independent because they are chosen according to the performance of all subjects. The effectiveness of the proposal is validated on a benchmark dataset that is largely employed in recent works. It has been shown that it is possible to reduce the number of EEG channels down to 6 (depending on the considered task) while keeping accuracy around 77% - 83% in binary classification, and above 60% for the four-classes case with a 10 channels sequence. The results have been extensively discussed, and they provide a valuable support to the design of wearable BCI systems for applications in daily life.

This work also addresses some future developments. Indeed, the main principles of this work can be extended to other datasets and BCI paradigms different from motor imagery. Then, further analyses could be conducted and the present results should be compared to a subject-dependent selection, in aiming to analyse the performance difference. Finally, resulting accuracies could be enhanced by better managing EEG non-stationarity, and by attempting different multi-class approaches. These developments are foreseen for future studies in designing wearable MI-BCI systems, and the proposed channel selection could be also exploitable in online channel selection due to its computational efficiency.

Acknowledgements

The authors thank Angela Natalizio for here indispensable support in the design and execution of MATLAB trials, and for the interesting discussion about the paper drafting. The authors also thank the "Excellence Department project" (LD n. 232/2016) whose support is gratefully acknowledged.

Bibliography

1. A. Ortiz-Rosario and H. Adeli, "Brain-computer interface technologies: from signal to action," *Reviews in the Neurosciences*, vol. 24, no. 5, pp. 537–552, 2013.
2. A. Ortiz-Rosario, I. Berrios-Torres, H. Adeli, and J. A. Buford, "Combined corticospinal and reticulospinal effects on upper limb muscles," *Neuroscience letters*, vol. 561, pp. 30–34, 2014.
3. A. Burns, H. Adeli, and J. A. Buford, "Brain–computer interface after nervous system injury," *The Neuroscientist*, vol. 20, no. 6, pp. 639–651, 2014.
4. X. Chen, B. Zhao, Y. Wang, S. Xu, and X. Gao, "Control of a 7-DOF robotic arm system with an SSVEP-based BCI," *International journal of neural systems*, vol. 28, no. 08, p. 1850018, 2018.
5. F. Xu, W. Zhou, Y. Zhen, Q. Yuan, and Q. Wu, "Using fractal and local binary pattern features for classification of ECOG motor imagery tasks obtained from the right brain hemisphere," *International journal of neural systems*, vol. 26, no. 06, p. 1650022, 2016.
6. I. Rembado, E. Castagnola, L. Turella, T. Ius, R. Budai, A. Ansaldo, G. N. Angotzi, F. Debertoldi, D. Ricci, M. Skrap, *et al.*, "Independent component decomposition of human somatosensory evoked potentials recorded by micro-electrocorticography," *International journal of neural systems*, vol. 27, no. 04, p. 1650052, 2017.
7. M. Teplan *et al.*, "Fundamentals of EEG measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.
8. M. Hamedi, S.-H. Salleh, and A. M. Noor, "Electroencephalographic motor imagery brain connectivity analysis for BCI: a review," *Neural computation*, vol. 28, no. 6, pp. 999–1041, 2016.
9. J. A. Barios, S. Ezquerro, A. Bertomeu-Motos, M. Nann, F. J. Badesa, E. Fernandez, S. R. Soekadar, and N. Garcia-Aracil, "Synchronization of Slow Cortical Rhythms During Motor Imagery-Based Brain-Machine Interface Control," *International journal of neural systems*, vol. 29, no. 5, pp. 1850045–1850045, 2019.
10. M.-C. Corsi, M. Chavez, D. Schwartz, L. Hugueville, A. N. Khambhati, D. S. Bassett, and F. de Vico Fallani, "Integrating EEG and meg signals to improve motor imagery classification in brain–computer interface," *International journal of neural systems*, vol. 29, no. 01, p. 1850014, 2019.
11. S. De Vries and T. Mulder, "Motor imagery and stroke rehabilitation: a critical discussion," *Journal*

of rehabilitation medicine, vol. 39, no. 1, pp. 5–13, 2007.
12. R. Ron-Angevin, F. Velasco-Álvarez, Á. Fernández-Rodríguez, A. Díaz-Estrella, M. J. Blanca-Mena, and F. J. Vizcaíno-Martín, "Brain-Computer Interface application: auditory serial interface to control a two-class motor-imagery-based wheelchair," *Journal of neuroengineering and rehabilitation*, vol. 14, no. 1, p. 49, 2017.
13. P. Shenoy, M. Krauledat, B. Blankertz, R. P. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *Journal of neural engineering*, vol. 3, no. 1, p. R13, 2006.
14. B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive berlin brain–computer interface: fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.
15. C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, and G. Edlinger, "How many people are able to control a P300-based brain–computer interface (BCI)?," *Neuroscience letters*, vol. 462, no. 1, pp. 94–98, 2009.
16. S. Ajami, A. Mahnam, and V. Abootalebi, "Development of a practical high frequency brain–computer interface based on steady-state visual evoked potentials using a single channel of EEG," *Biocybernetics and Biomedical Engineering*, vol. 38, no. 1, pp. 106–114, 2018.
17. Y. Jiao, Y. Zhang, Y. Wang, B. Wang, J. Jin, and X. Wang, "A novel multilayer correlation maximization model for improving CCA-based frequency recognition in SSVEP brain–computer interface," *International journal of neural systems*, vol. 28, no. 04, p. 1750039, 2018.
18. C. Yang, X. Han, Y. Wang, R. Saab, S. Gao, and X. Gao, "A dynamic window recognition algorithm for SSVEP-based brain–computer interfaces using a spatio-temporal equalizer," *International journal of neural systems*, vol. 28, no. 10, p. 1850028, 2018.
19. J. Meng, L. Yao, X. Sheng, D. Zhang, and X. Zhu, "Simultaneously optimizing spatial spectral features based on mutual information for EEG classification," *IEEE transactions on biomedical engineering*, vol. 62, no. 1, pp. 227–240, 2014.
20. Y. Zhang, Y. Wang, J. Jin, and X. Wang, "Sparse Bayesian learning for obtaining sparsity of EEG frequency bands based feature vectors in motor imagery classification," *International journal of neural systems*, vol. 27, no. 02, p. 1650032, 2017.
21. R. Liu, Y. Wang, G. I. Newman, N. V. Thakor, and S. Ying, "EEG classification with a sequential decision-making method in motor imagery BCI," *International journal of neural systems*, vol. 27, no. 08, p. 1750046, 2017.
22. B. E. Olivas-Padilla and M. I. Chacon-Murguia, "Classification of multiple motor imagery using deep convolutional neural networks and spatial filters," *Applied Soft Computing*, vol. 75, pp. 461–472, 2019.
23. C. H. Nguyen, G. K. Karavas, and P. Artemiadis, "Adaptive multi-degree of freedom Brain Computer Interface using online feedback: Towards novel methods and metrics of mutual adaptation between humans and machines for BCI," *PloS one*, vol. 14, no. 3, p. e0212620, 2019.
24. P. Gaur, K. McCreadie, R. B. Pachori, H. Wang, and G. Prasad, "Tangent Space Features-Based Transfer Learning Classification Model for Two-Class Motor Imagery Brain–Computer Interface," *International journal of neural systems*, vol. 29, no. 10, p. 1950025, 2019.
25. M. Z. Baig, N. Aslam, and H. P. Shum, "Filtering techniques for channel selection in motor imagery EEG applications: a survey," *Artificial intelligence review*, vol. 53, pp. 1207–1232, 2020.
26. X. Yong, R. K. Ward, and G. E. Birch, "Sparse spatial filter optimization for eeg channel reduction in brain-computer interface," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 417–420, IEEE, 2008.
27. M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Optimizing the channel selection and classification accuracy in EEG-based BCI," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 6, pp. 1865–1873, 2011.
28. H. V. Shenoy and A. P. Vinod, "An iterative optimization technique for robust channel selection in motor imagery based brain computer interface," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1858–1863, IEEE, 2014.
29. T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, p. 66, 2015.
30. H. Shan, H. Xu, S. Zhu, and B. He, "A novel channel selection method for optimal classification in different motor imagery BCI paradigms," *Biomedical engineering online*, vol. 14, no. 1, p. 93, 2015.
31. C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI Competition 2008–Graz data set A," *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, vol. 16, 2008.
32. K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers in neuroscience*, vol. 6, p. 39, 2012.
33. L. F. Nicolas-Alonso, R. Corralejo, J. Gomez-Pilar, D. Álvarez, and R. Hornero, "Adaptive stacked generalization for multiclass motor imagery-based brain computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 4, pp. 702–712, 2015.
34. P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "A

multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry," *Expert Systems with Applications*, vol. 95, pp. 201–211, 2018.
35. Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE transactions on cybernetics*, vol. 49, no. 9, pp. 3322–3332, 2018.
36. A. Singh, S. Lal, and H. W. Guesgen, "Reduce calibration time in motor imagery using spatially regularized symmetric positives-definite matrices based classification," *Sensors*, vol. 19, no. 2, p. 379, 2019.
37. G. H. Klem, H. O. Lüders, H. Jasper, C. Elger, *et al.*, "The ten-twenty electrode system of the International Federation," *Electroencephalogr Clin Neurophysiol*, vol. 52, no. 3, pp. 3–6, 1999.
38. Y. Yang, S. Chevallier, J. Wiart, and I. Bloch, "Subject-specific time-frequency selection for multi-class motor imagery-based BCIs using few Laplacian EEG channels," *Biomedical Signal Processing and Control*, vol. 38, pp. 302–311, 2017.
39. Y. Yang, I. Bloch, S. Chevallier, and J. Wiart, "Subject-specific channel selection using time information for motor imagery brain–computer interfaces," *Cognitive Computation*, vol. 8, no. 3, pp. 505–518, 2016.
40. S. Ge, R. Wang, and D. Yu, "Classification of four-class motor imagery employing single-channel electroencephalography," *PloS one*, vol. 9, no. 6, p. e98019, 2014.
41. L. Angrisani, P. Arpaia, F. Donnarumma, A. Esposito, N. Moccaldi, and M. Parvis, "Metrological performance of a single-channel Brain-Computer Interface based on Motor Imagery," in *2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pp. 1–5, IEEE, 2019.
42. L. Faes, G. Nollo, and A. Porta, "Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique," *Physical Review E*, vol. 83, no. 5, p. 051112, 2011.
43. A. W. Whitney, "A direct method of nonparametric measurement selection," *IEEE Transactions on Computers*, vol. 100, no. 9, pp. 1100–1103, 1971.
44. J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical neurophysiology*, vol. 110, no. 5, pp. 787–798, 1999.
45. D. Thiyam and E. Rajkumar, "Common Spatial Pattern Algorithm Based Signal Processing Techniques for Classification of Motor Imagery Movements: A Mini Review," *IJCTA*, vol. 9, no. 36, pp. 53–65, 2016.
46. B. V. Dasarathy, "Nearest neighbor (NN) norms: NN pattern classification techniques," *IEEE Computer Society Tutorial*, 1991.
47. C.-W. Hsu, C.-C. Chang, C.-J. Lin, *et al.*, "A practical guide to support vector classification," 2003.
48. D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*. John Wiley & Sons, 2010.
49. A. Schlogl, J. Kronegg, J. Huggins, and S. Mason, "19 evaluation criteria for bci research," *Toward brain-computer interfacing*, 2007.
50. "BCI competition III, data sets IIIa: motor imagery, multi-class," 2006.
51. A. Ghaemi, E. Rashedi, A. M. Pourrahimi, M. Kamandar, and F. Rahdari, "Automatic channel selection in EEG signals for classification of left or right hand movement in Brain Computer Interfaces using improved binary gravitation search algorithm," *Biomedical Signal Processing and Control*, vol. 33, pp. 109–118, 2017.
52. S. Saha, K. I. U. Ahmed, R. Mostafa, L. Hadjileontiadis, and A. Khandoker, "Evidence of variabilities in EEG dynamics during motor imagery-based multiclass brain–computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 2, pp. 371–382, 2017.
53. P. K. Parashiva and A. Vinod, "A New Channel Selection Method using Autoencoder for Motor Imagery based Brain Computer Interface," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 3641–3646, IEEE, 2019.
54. J. K. Feng, J. Jin, I. Daly, J. Zhou, Y. Niu, X. Wang, and A. Cichocki, "An optimized channel selection method based on multifrequency CSP-rank for motor imagery-based BCI system," *Computational Intelligence and Neuroscience*, vol. 2019, 2019.
55. C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller, "How many people are able to operate an EEG-based brain-computer interface (BCI)?," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 11, no. 2, pp. 145–147, 2003.
56. X. Zhu, P. Li, C. Li, D. Yao, R. Zhang, and P. Xu, "Separated channel convolutional neural network to realize the training free motor imagery BCI systems," *Biomedical Signal Processing and Control*, vol. 49, pp. 396–403, 2019.
57. I. Jolliffe, *Principal component analysis*. Springer, 2011.
58. R. Prevete, F. Donnarumma, A. D'Avella, and G. Pezzulo, "Evidence for sparse synergies in grasping actions," *Scientific reports*, vol. 8, no. 1, pp. 1–16, 2018.
59. J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Networks*, vol. 118, pp. 262–270, 2019.
60. H. H. Ehrsson, S. Geyer, and E. Naito, "Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations," *Journal of neurophysiology*, vol. 90, no. 5, pp. 3304–3316, 2003.
61. A. M. Batula, J. A. Mark, Y. E. Kim, and H. Ayaz,

"Comparison of brain activation during motor imagery and motor movement using fNIRS," *Computational intelligence and neuroscience*, vol. 2017, 2017.

62. G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. Da Silva, "Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.
63. G. D. Schott, "Penfield's homunculus: a note on cerebral cartography," *Journal of neurology, neurosurgery, and psychiatry*, vol. 56, no. 4, p. 329, 1993.