Assessing Impact of Channel Selection on Decoding of Motor and Cognitive Imagery from MEG Data

Sujit Roy [1], Dheeraj Rathee [2], Anirban Chowdhury [2], Karl McCreadie [1], Girijesh Prasad [1]

[1] Intelligent Systems Research Centre, School of Computing & Intelligent Systems, Ulster University, Derry ~ Londonderry, N. Ireland, UK. [2] School of Computer Science and Electronic Engineering, University of Essex, UK.

E-mail: Roy-s2@ulster.ac.uk

August 2019

Abstract.
Objective: Magnetoencephalography (MEG) based Brain-Computer Interface (BCI) involves a large number of sensors allowing better spatiotemporal resolution for assessing brain activity patterns. There have been many efforts to develop BCI using MEG with high accuracy, though an increase in the number of channels means an increase in computational complexity. However, not all sensors necessarily contribute significantly to an increase in classification accuracy and specifically in the case of MEG-based BCI no channel selection methodology has been performed. Therefore, this study investigates the effect of channel selection on the performance of MEG-based BCI.
Approach: MEG data were recorded for two sessions from 15 healthy participants performing motor imagery, cognitive imagery and a mixed imagery task pair using a unique paradigm. Performance of four state-of-the-art channel selection methods (i.e. Class-Correlation (CC), ReliefF (RF), Random Forest (RandF), and Infinite Latent Feature Selection (ILFS) were applied across six binary tasks in three different frequency bands) was evaluated in this study on two state-of-the-art features i.e. bandpower and CSP.
Main results: All four methods provided a statistically significant increase in classification accuracy (CA) compared to a baseline method using all gradiometer sensors, i.e. 204 channels with band-power features from alpha (8-12Hz), beta (13-30Hz), or broadband (α+β) (8-30Hz). It is also observed that the alpha frequency band performed better than the beta and broadband frequency bands. The performance of the beta band gave the lowest CA compared with the other two bands. Channel selection improved accuracy irrespective of feature types. Moreover, all the methods reduced the number of channels significantly, from 204 to a range of 1-25, using bandpower as a feature and from 15-105 for CSP. The optimal channel number also varied not only in each session but also for each participant. Reducing the number of channels will help to decrease the computation cost and maintain numerical stability in cases of low trial numbers.
Significance: The study showed significant improvement in performance of MEG-BCI with channel selection irrespective of feature type and hence can be successfully applied for BCI applications.

Keywords: MEG, BCI, Bandpower, CSP, Channel selection.
Submitted to: *J. Neural Eng.*

1. Introduction

Motor disabilities and severe neurological injury require an extra measure in rehabilitation for active and effective environmental interaction. Motor Imagery (MI) practice through brain-computer interface (BCI) has been found to be useful as a therapeutic substitute for standard rehabilitation practices for post-stroke patients [1, 2], and an alternative approach for interaction with the environment [3, 4, 5, 6, 7] for people with severe movement disability. For post-stroke rehabilitation, the patient is required to voluntarily practice activities of daily living (ADL) with very high focus [2, 8, 9, 10]. Advancements in BCI based technologies have shown promising results in terms of focused interaction for stroke patients [11]. Current BCI systems may use magnetoencephalography (MEG), electroencephalography (EEG), functional magnetic resonance imaging (fMRI) or electrocorticography (ECoG) approaches for mapping brain responses [7, 12, 13, 14, 15, 16, 17]. Although the majority of the available BCI research is focused on EEG, MEG may provide better performance due to its higher signal-to-noise (SNR) ratio and spatio-temporal resolution compared to EEG. Moreover, unlike EEG, MEG sensors are placed in a dedicated helmet rather than physically placed on subjects' scalp resulting in significant signal attenuation [18].

One of the key challenges of MEG-based BCI systems is their low accuracy [19, 20, 21]. Previous studies have focused on either development of novel feature extraction methods or improvement of current classification methods. Signal processing methods such as channel selection have been completely ignored in the case of MEG. It has been observed that signal pre-processing methods can improve the performance of an EEG based BCI system [22]. There is a substantial literature supporting the effectiveness of channel selection methods with EEG-based BCIs and it is thus intuitive to explore their effectiveness with MEG-based systems.

MEG systems typically have a large number of channels for very high spatial resolutions e.g. 306 for Elekta Neuromag Triux system. It is known that classification performance of a BCI system is dependent upon data pre-processing, feature extraction, and the use of an appropriate classifier but by selecting optimum channels, classification performance can be improved further. Provided with a large number of MEG channels, the extracted features can outnumber the trials resulting in an over-parameterised classification scenario. Dimensionality reduction can help to select the most important features without affecting performance [18]. It is however observed that the use of a large number of channels can affect the performance in a negative way, as those channels that are not contributing to feature separability can make the feature set noisier.

Prasad et al. [2] presented promising results with 5 stroke patients in their neuro-feedback based BCI which facilitated motor recovery with moderate BCI accuracy. Neurorehabilitation, using MEG-based BCI by extraction of relevant information from brain activity, is a challenge. Selecting 204 gradiometers provides a higher signal-to-noise ratio (SNR) as compared to 102 magnetometers, as gradiometers are more sensitive to rate of change in cortical activations nearer to scalp. Hence, it would be more appropriate in this instance to choose gradiometers instead of magnetometers for an MI-BCI application. This still leaves a huge number of channels which will result in higher computation irrespective of their positive or negative contribution to the performance of a BCI system.

Common spatial pattern (CSP) and its extended algorithms like Sparse CSP, have been used for dimensionality reduction and show that the channels can be selected based on large CSP vector coefficients [23, 24, 25] whilst maintaining a sufficient level of accuracy. Further, different CSP methods have been implemented in an attempt to increase the accuracy [19]. A channel selection method was presented by He et al. [26] based on the Bhattacharya bound CSP for classification of MI-related signals. It considers the Bhattacharyya bound as an index and progressively searches for optimized channel combinations. A 95% classification accuracy was later claimed with an average of 33 channels, which was a higher accuracy than that of any other channel selection method. Zhang et al. [27] used ReliefF method for EEG sensor selection for emotion classification and reported approx 9% improvement by using 30 EEG channels. Roy et al. [18] presented promising results using correlation and ReleifF based channel selection in MEG. A maximum increase of 24.22% was observed for cross-validation performance. For a review of other methods implemented in the field of EEG motor imagery, readers can refer to Alotaiby et. al and Lotte et al. [28, 29].

In this study, our previous work [18] has been extended, by implementing four state-of-the-art feature selection methods i.e. class correlation [30], ReliefF [31, 32], random forest feature ranking [33], and infinite latent feature selection [34] and evaluating their performances using a four-class MEG MI BCI dataset. These methods have been used widely for image and time-series datasets. To the best of authors' knowledge, this is the first attempt to evaluate the effect of channel selection process on performance of MEG-based MI-BCI system.

The remainder of this paper proceeds as follows: Section 2 provides details of the acquired data sets,

experimental paradigm, signal processing pipeline, and channel selection methods. Next, Section 3 presents the performance analysis.The outcomes of the study are discussed in Section 4, and Section 5 summarises the findings of this study.

2. Materials and Methods

2.1. Experiment and Data Description

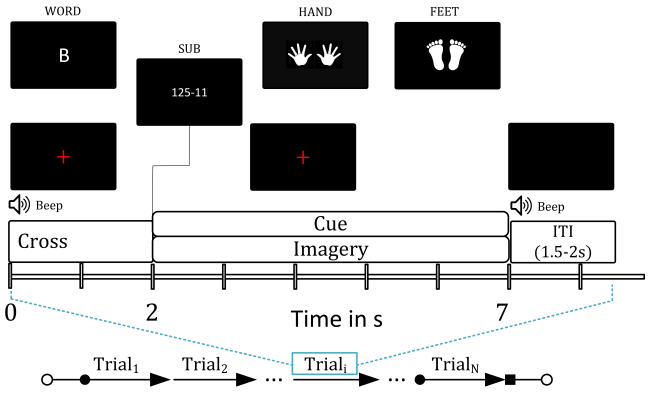

Figure 1. Timing diagram of MEG-BCI paradigm.

For this study, an MEG dataset of 15 healthy participants was acquired using a typical MEG-based experimental paradigm which can be used for BCI as well. The participants included 12 males and 3 females with a mean age of 29.3 years ± 5.96, with 13 right-handed and 2 left-handed as per self-report. The participants signed a written consent form before starting the experiments and the experimental protocol was approved by the ethics committee of Ulster University. MEG data were acquired over 2 sessions (each session on different days) using the same experimental paradigm. Figure 1 presents the timing diagram of the MEG-based experimental paradigm. Each trial starts with a rest period of 2 s followed by 5 s of imagery task period. The cue remained visible during the imagery task period. A randomly selected inter-trial-interval (ITI) of $1.5-2$ s was presented after the imagery task period. Participants were seated on a comfortable chair approximately 80 cm away from the projector screen. Elekta Neuromag Triux system was used for recording the MEG data at a sampling rate of 1kHz.

The experimental paradigm was designed to cover motor imagery (MI) tasks and cognitive imagery tasks. The dataset includes four mental imagery tasks: both hand movement, both feet movement, subtraction, and word generation. During the MI-related tasks, participants imagined movement of both hands/both foot when the cue appeared at the screen. Similarly, for cognitive imagery tasks, participants either subtracted two numbers presented as cues or generated words related to an English language letter presented as a cue. Each session consisted of 50 trials for each of the imagery tasks with a total of 200 trials.

2.2. Data Processing

The MEG dataset was acquired from all 306 sensors (204 gradiometers and 102 magnetometers). However, for this study, the raw data from only 204 gradiometers were used before being band-passed into 3 frequency bands, i.e. alpha(α) band (8-12 Hz), beta(β) band (13-30 Hz) and alpha and beta combined ($\alpha + \beta$) band (8-30 Hz) using a two pass Butterworth filter. Selecting 204 gradiometers provides a higher signal-to-noise ratio (SNR) as compared to 102 magnetometers, as gradiometers are more sensitive to changes in cortical activations. Bad channels were discarded due to extraneous noise, interference or artefacts. For feature extraction, data were down-sampled from 1 kHz to 500 Hz. After pre-processing, 3.5 s of the data segment related to imagery activity was selected from each trial (0.5 s after the cue onwards) and signal power was estimated for each MEG gradiometer channel for each task i.e. hand, feet, math and word generation in each band separately. For the selection of best channels, four ranking based methods were used, i.e. Class-correlation (CC), ReliefF (RF), Infinite Latent Feature Selection (ILFS) and Random Forest (RandF). These methods were considered for the evaluation of the effect of channel selection in MEG in the three frequency bands. Furthermore, a 10-fold cross validation classification accuracy (CA) was estimated for six binary classification tasks, i.e. hand versus feet (H-F), hand versus word (H-W), hand versus math (H-M), feet versus word (F-W), feet versus math (F-M), and word versus math (W-M) using a linear discriminant analysis (LDA) classifier (Figure 2). The performance of each channel selection method was compared with the baseline condition, i.e. 204 gradiometer channels, in all three conditions.

2.3. Channel Selection

Feature selection, or feature reduction, is a process of selecting a subset of relevant features based on positively contributing criteria. This process aims towards significantly reducing the noise created by relatively less important features and counter the curse of dimensionality. In this study, overall signal band-power was estimated for each MEG channel separately and the following four methods were applied to find the best set of channels for each binary combination of classes.

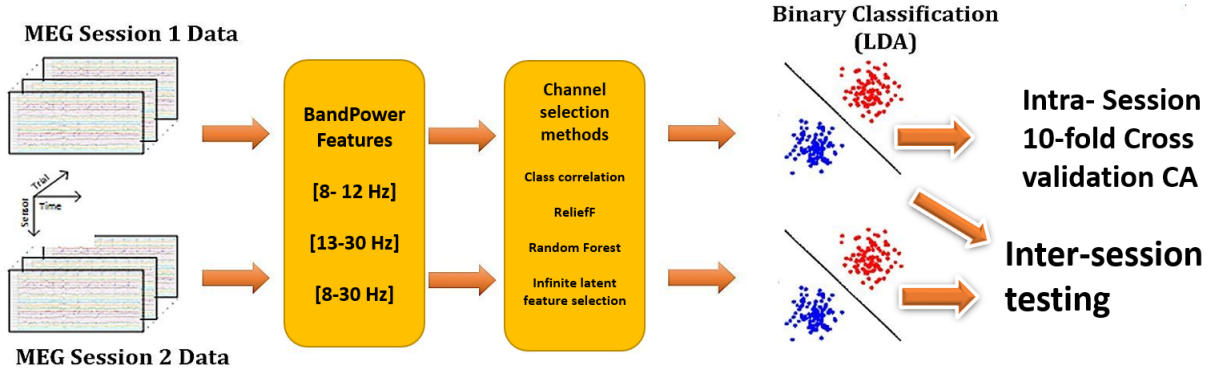

Figure 2. Schematic diagram providing the details of signal processing pipeline using bandpower feature for classification.

2.3.1. Class-Correlation Method (CC): The Pearson correlation coefficient is used to determine the statistical relationship between two random variables A and B. The values of the coefficients range from -1 to 1 representing no relation to direct relation i.e. it is a measure of the linear dependence of the two random variables [30].

$$\rho(A,B) = \frac{1}{N-1}\sum_{i=1}^{N}\frac{(A_i-\mu_A)(B_i-\mu_B)}{\sigma_A.\sigma_A} \quad (1)$$

where μ_A is the mean and σ_A is the standard deviation of A. Similarly, μ_B is the mean and σ_B is the standard deviation of B. A represents the bandpower feature vector and B represented labels of corresponding class , i.e. 1, -1. The values of the correlation coefficients were calculated by creating a dummy label corresponding to the observations of each of the features. The dummy class label had a numeric value which indicated (L=1) for class 1 and (L=−1) class 2. As per the coefficients, the channels were ranked in decreasing order keeping the highest correlated channel at the top and were stored for further evaluation. The computational complexity of CC method is $\mathcal{O}(n)$ and running time is 0.10 sec.

Data: For each training instance, obtains vectors of attribute values and corresponding class value
Result: For computing the vector W of estimations of the qualities of attributes
initialise the weights: W[A] ← 0 ;
for $i \leftarrow 1$ **to** m **do**
  randomly select an instance R_i;
  find k nearest hits H_j ;
  for *each class* $C \neq class(R_i)$ **do**
    from class C find k nearest misses $M_j(C)$;
  end
end
for $A \leftarrow 1$ **to** a **do**
  $$W[A] = W[A] - \sum_{j=1}^{k}\frac{diff(A,R_i,H_j)}{m.k} + \sum_{C\neq class(R_i)}\frac{\frac{P(C)}{1-P(class(R_i))}\sum_{j=1}^{k} diff(A,R_i,M_j(C))}{m.k}$$
end

Algorithm 1: Pseudocode for ReliefF algorithm

2.3.2. ReliefF Method (RF): The ReliefF (RF) algorithm is an extension of Relief algorithm proposed by Kira et al. [31, 32]. The RF algorithm differs from Relief algorithm on the basis of the selection of nearest hits (same class) and misses (other class) and also on the update of attributes [32, 35]. The RF method firstly initialises a random instance R_i and then searches for K nearest hits (H_j) and k nearest misses($M_j(C)$). The weights $W[A]$ are updated based on the values of R_i, H_j and $M_j(C)$ as shown in Algorithm 1. Using the RF method feature importance along with weights was found by 2 inputs, i.e. feature vector and their corresponding labels of 1 or -1 and was stored in an array after sorting according to their weights for further evaluation. The size of the nearest neighbours (i.e. k) was set to 25. The computational complexity of ReliefF method is $\mathcal{O}(iSnC)$ and running time is 0.47 sec, S is the number of samples, n is the number of initial features, i is the number of iterations in the case of iterative algorithms, and C is the number of classes.

2.3.3. Infinite Latent Feature Selection (ILFS): This feature selection method was proposed by Roffo et al. [34] where a training set is represented by a feature distribution set $P = \{P_1, P_2, ..., P_m\}$. Each n x 1 vector P_i, is considered as the distribution of values assumed by the ith feature concerning n samples to build an undirected graph G, where nodes correspond to features and edges model relationships among pair

of nodes. Edges are represented weights as a_{ij} which is an element of A, 1≤i, j≤n, and which model the pairwise relationship between features, assuming that feature x_i and x_j are good candidates. Weights can be associated with a binary function of graph nodes;

$$a_{ij} = \phi(\vec{x_i}, \vec{x_j}) \tag{2}$$

where $\phi(.,.)$ is a real valued potential function learned by the proposed approach in a probabilistic latent semantic analysis (PLSA)-inspired framework. Considering a weighted graph G, ILFS represents a subset of features as path connecting them. Similarly, like other methods, a feature vector of 2 classes were fed along with corresponding labels using the method developed by Roffo et al. [34, 36, 37]. The computational complexity of ILFS method is $\mathcal{O}(n^{2.37} + iN + S + C)$ and running time is 0.26 s. Here S is the number of samples, N is the number of initial features, i is the number of iterations in the case of iterative algorithms, and C is the number of classes.

2.3.4. Random Forest Based Ranking (RandF): This is a permutation technique presented by Breiman to measure the importance of features in the prediction [33] referred to as an out-of-bag (OOB) importance score. At each node t in a decision tree, a split is determined by the decrease in node impurity $\Delta R(t)$. The node impurity R(t) is the gini index. If a subset in node t contains samples from c classes, Gini(t) is defined as,

$$Gini(t) = R(t) = 1 - \sum_{j=1}^{c} \hat{a}_j^2 \tag{3}$$

where $\hat{a}_j^2$ is the relative frequency of class j in t. After splitting into two child node t_1 and t_2 with sample size N_1 (t) and N_2 (t), the gini index of split data is defined as:

$$Gini_{split}(t) = \frac{N_1(t)}{N_t} Gini(t_1) + \frac{N_2(t)}{N_t} Gini(t_2) \tag{4}$$

The feature providing the smallest $Gini_{split}$ (t) is chosen to split the node. The importance score of feature vector X_j in a single decision tree T_k is

$$Ik(X_j) = \sum_{t \in T_k} \Delta R(X_j, t) \tag{5}$$

where I is the importance of feature X of class j and t represents node. The same is computed over all k trees in a random forest, defined as

$$I(X_j) = \frac{1}{k} \sum_{k=1}^{k} I_k(x_j) \tag{6}$$

Feature vector of 2 classes was used to create an ensemble along with their corresponding labels. The minimum number of observations per tree leaf was set to 5. The time complexity for constructing a complete decision tree is $\mathcal{O}(a * nlog(n))$, where n is the number of records and a is the number of attributes. While constructing the Random forest, it is required to define the number of trees needed to build (ntree) and number of the attributes wished to sample at each node (mtry). Since only mtry variables will be used at each node the complexity to build one tree would be $\mathcal{O}(mtry * nlog(n))$. Hence, for building a random forest with ntree, the complexity would be $\mathcal{O}(ntree * mtry * nlog(n))$ and running time is 1.81 s in our case. This is the worst case scenario, i.e., assuming the depth of the tree is going to be $\mathcal{O}(log(n))$.

The methods discussed above have been used for ranking of channels for session one and session two data individually. But to calculate the final set of channels, a forward elimination approach is used. The forward elimination approach helps to select a subset of channels that positively contribute towards accuracy from the set of all ranked channels. Let X be a set of 204 MEG channels. After the ranking of channels, the top-ranked channel is selected and added to an empty set X_ranked which is used for calculating CA. If the second-ranked channel positively contributes to the first ranked channel, then it is added to the set Y else the next channel is tested for its accuracy contribution. The process terminates after the addition of a channel that does not contribute positively to CA. At the end of the process, Y has only positively contributing channels. The classifier used in this instance is LDA.

2.4. Common Spatial Patterns (CSP)

CSP is an efficient tool to analyze multichannel data such as EEG/MEG for binary classification and provides a supervised method for decomposition of signals parameterized by a matrix $W \in R^{\{C \times C\}}$ (C: number of channels). The matrix is used to project the original sensor space E into the surrogate sensor space, using eq.(7)

$$Z = WE \tag{7}$$

where, $E \in R^{\{C \times T\}}$ is the MEG measurement of a single-trial and T is the number of samples per channel. W is the CSP projection matrix. The rows of W are the spatial filters and the columns are the common spatial patterns. A small number of spatially filtered signals are used as features for classification purposes. These are generally the first and last m rows of Z, i.e. Z_t, where $t \in \{1, ..., 2m\}$. In our case, m= 1, which implies that the first and last component were considered. The feature vector is derived from Z_t by

eq.(8) as,

$$x_t = log\left(\frac{var(Z_t)}{\sum_{t=1}^{2m} var(Z_t)}\right) \quad (8)$$

After the temporal filtering in mu([8-12] Hz) and beta([13-30] Hz) bands and spatial filtering as shown in Figure 3, and finally taking the log variance using (8), we obtain the feature vector.

MATLAB 2018b was used for creating scripts for building learning models and evaluating their performance. For statistical analysis, paired-sample t-test was used. The system used Windows 10 with an i7 8th gen processor, and an NVidia RTX2080 Ti.

3. Results

3.1. Performance comparison using a band power feature

Figs 4, 5, 6 show bar plots of the mean CA obtained using 10-fold cross-validation of individual sessions under the five experimental conditions (i.e. baseline, CC, RF, RandF and ILFS) and six binary classification tasks for α, β and α+β frequency bands. For all three frequency bands, CC, RF, RandF, and ILFS provided statistically significant improvement ($p < 0.05$) as compared to the baseline in terms of CA for all the six binary classification tasks. Moreover, for all combinations, the mixed imagery task pairs (H-W, H-M, F-W & F-M) provided higher separability as compared to the H-F and W-M task pairs.

In reference to Figure 4, the RandF method performed better than ILFS with an increased performance of 1.82% on average across all subjects over 6 binary classification tasks. RandF provided a statistically significant improvement over ILFS in H-F, H-M, W-M, F-W and W-M task pairs ($p < 0.05$). RandF did not provide a statistically significant improvement over other methods. RF and CC provided statistical significant improvement over Randf in just two task pairs i.e. for F-W and F-M respectively. The overall mean CA across subjects using RandF is 81.11% (±6.02), ILFS is 79.30% (±6.51), CC is 81.72% (±6.25), and RF is 81.14% (±6.22) for session 1 whereas baseline was 65.32% (±8.09). Similarly, for session 2 mean CA across subjects were 81.87% (±6.59) for RandF, 79.57% (±7.30) for ILFS, 81.99% (±6.25) for CC, and 81.24% (±6.86) for RF for the α frequency band.

Figure 5 shows the mean CA obtained using the β frequency band (12-30 Hz). This figure shows the same behavior as Figure 4, where RandF has performed better than other channel selection methods. However, a point worth highlighting is the drop in accuracy compared to the α frequency band. The overall mean accuracy for session 1 using RandF is 75.65% (±5.59), ILFS is 73.33%(±6.09), RF is 74.02%(±5.77), CC is 72.40%(±5.85) and for baseline is 59.88% (±5.72). Similarly, for session 2 mean CA is 75.70%(±5.74) for RandF, 74.83%(±6.67) for ILFS, 75.28%(±6.05) for RF, 72.93%(±6.05) for CC and 59.88%(±5.72) for baseline.

Figure 6 shows the mean CA obtained using α+β frequency band (8-30 Hz). RandF and RF have both performed better than any other channel selection method. The overall mean accuracy for session 1 using RandF is 79.99% (±6.36), ILFS is 78.08%(±7.15), RF is 78.93%(±5.92), CC is 77.28%(±7.20) whilst for baseline is 64.31% (±7.41). Similarly, for session 2 mean CA is 79.58%(±7.22) for RandF, 77.68%(±7.39) for ILFS, 79.04%(±7.35) for RF, 78.32%(±7.83) for CC and 65.84%(±8.07) for baseline. CA is significantly higher in the α band as well as the α+β band compared to the β band alone. However, after performing channel selection the grand mean accuracy of the α band at 79.30% (±6.51) is higher than in the α+β band at 78.08% (±7.15).

Figure 7 shows the CA in the three bands when the classifier was trained on session 1 data and tested on session 2. For this experiment, the channels selected for session 1 using each ranking method, were considered for evaluation in session 2. Since the session 2 is unseen data, the accuracies drop as compared to the cross-validation CAs shown in Fig 4, 5 and 6. However, overall the α band (64.24% ± 8.34) performed better than the β (57.88% ± 7.34) and α+β (63.67% ± 7.85) bands. Similarly, in the cross-session evaluation, the H-W group showed higher accuracy than the motor imagery (H-F) group.

The effect of the addition of the top channels can be seen on CA as it starts to fall after a certain number of channels (Figure 8). The result was calculated on session 1 data for each participant using the motor imagery paradigm (H-F) in the α frequency band. As observed, the CA tends to rise up to a certain number of channels, i.e. 15 and then starts decreasing with further addition of channels. A few channels can also be observed to contribute negatively before reaching the peak accuracy. Therefore, by using a forward elimination approach the CA accuracy was calculated and the maximum CA was achieved with just nine channels (Figure 8(b)).

Table 1 gives an overview of the total number of channels involved for achieving the maximum CA for session 1 (S01) and session 2 (S02) based on the RandF method for the α+β frequency band. The number of channels contributing to a maximum CA using forward elimination using all the 4 ranking methods in the respective bands, ranged from 1 to 23. It is to be

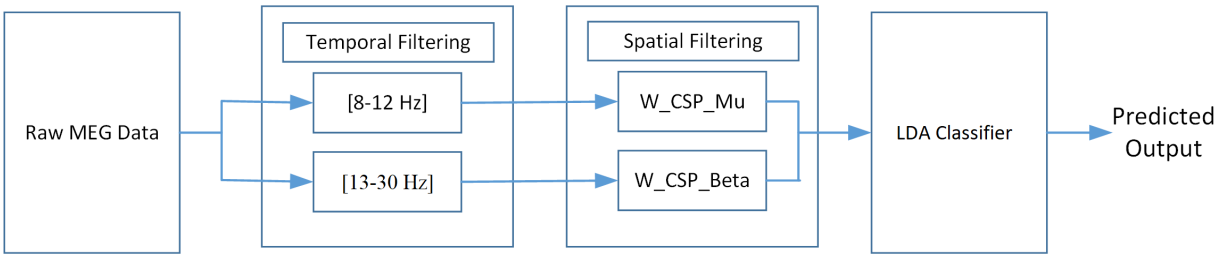

Figure 3. The data processing pipeline using CSP. The raw MEG signals are first passed through bandpass filters in mu (8-12 Hz) and beta (13-30) Hz. Then these bandpass filtered signals are spatially filtered with their respective CSP weight matrices. After spatial filtering the feature vector is formed and subsequently classified by the LDA classifier. The CSP weight matrices and the LDA classifier was trained using the session 1 data.

noted that the mean of the number of selected channels across the subjects for session 1 and session 2 looks more or less similar, although there were variations in the number of selected channels for individual subjects between session 1 and session 2..

3.2. Performance evaluation using a CSP feature

Figure 9 shows the classification performance after channel selection using common spatial patterns (CSP) in all the three bands i.e. α, β, and $\alpha+\beta$. Firstly, all the channels were ranked using RandF method with data from session 1. Then the ranked channels were added block by block starting from 15 in intervals of 10 till 125. Then the ordered channel blocks were used to train a classifier on session 1 data and test on session 2 data. This figure contains the global mean of all the binary classification tasks. A steady rise in accuracy can clearly be observed before either saturation or drop. Going by the trendlines shown in Figure 9, for the band $\alpha+\beta$, the CA hits a plateau at NoC=75 with 72%, for α the CA hits the plateau at NoC=85 with 69.3%, and for beta it is NoC=95 near the plateau where the CA is 65.7%. It is to be noted that the performance of the beta band is much lower than the other two bands. This can be due to the fact that the event related desynchronization is more consistent during inter-trial and inter-session transitions in alpha frequency range between 8 and 12 Hz. This was found in our previous study on EEG-EMG correlation where the desynchronization in 8-12 Hz gave more consistent correlation with the EMG activity than for beta band [38]. In the same study, ERD distribution for alpha band showed significantly less inter-trial variability than for beta band from the data across 8 healthy participants. This could be a possible reason why alpha band performed better than beta band. Consequently, alpha+beta band also performed better than beta because the addition of alpha band led to more consistent pattern in the sensori-motor rhythm than for beta band only. Suppression of alpha band power around 10 Hz is also a popular marker for movement planning, execution and imagery than beta band [39]. These findings indicates that the alpha band may have more impact on classification performance in the context of motor imagery. However beta band is more related to the longitudinal changes in motor behaviour, as revealed by a study on post stroke rehabilitation [40]. CSP is a spatial filtering technique which needs larger number of channels to optimize the CSP projection matrix. The CSP projection matrix transforms the signals in such a way that it maximizes the variance of one class and minimizes the variance of the other class [41, 42]. Providing larger number of channels to the bandpower features may overfit the data due to high dimensionality as they do not have the ability to maximize the discrimination between the two classes unlike CSP. Hence, a feature vector with lesser dimension works better in case of bandpower. On the other hand, if we feed the CSP with a larger number of channels it can optimize the CSP projection matrix very well which could lead to higher discriminability between the two classes. However, it is also true that the performance of CSP saturates and drops beyond a certain number of channels (as evident in Figure 9) which may be due to the inclusion of lower ranking MEG channels as which contribute very little or no useful information but primarily noise.

As can be seen from Figure 9, the performance of the $\alpha+\beta$ band is clearly higher than the alpha or beta band in the case of CSP features with channel selection, hence we further show the comparison between the CA at optimal channel set and at all channels in Figure 10. Hand vs. feet shows a significant ($p < 0.05$) increase in performance of motor imagery task when the optimal channel set was chosen as compared to selection of all the channels. The mean accuracy by all subjects is 70.67%($\pm$12.69) at optimal channel set, while the mean CA is 55.8%($\pm$13.44) when all the channels are considered. The mean of the optimal channel set length contributing maximum accuracy in hand vs. feet is 64.33 (Table 2). In hand vs. word the mean CA is 81.80%($\pm$13.63) and mean number of channels contributing to this accuracy is 47. For

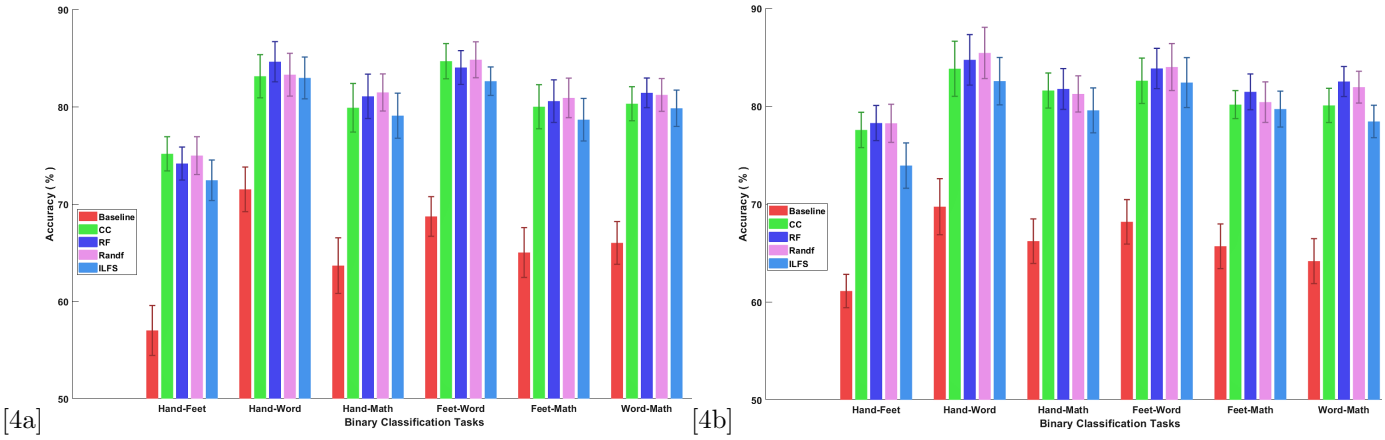

Figure 4. Mean classification accuracies obtained with the five experimental conditions (baseline, CC, RF, RandF and ILFS) and six binary classification tasks for the α **frequency band** (8-12 Hz) for **session 1 (4a) and session 2 (4b)** using 10-fold cross-validation.

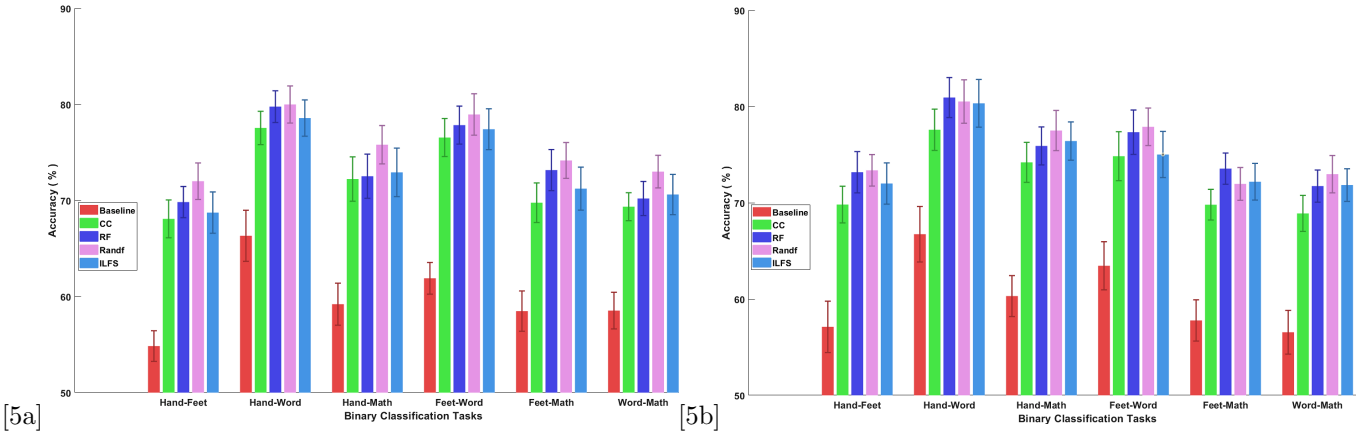

Figure 5. Mean classification accuracies obtained with the five experimental conditions (baseline, CC, RF, RandF and ILFS) and six binary classification tasks for β **frequency band** (12-30 Hz) for **session 1 (5a) and session 2 (5b)** using 10-fold cross-validation.

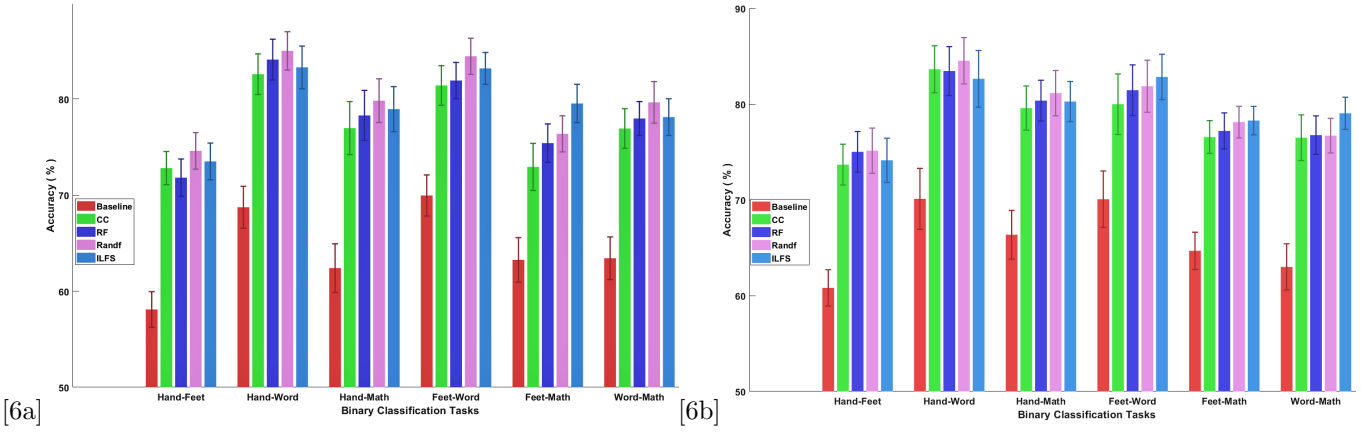

Figure 6. Mean classification accuracies obtained with the five experimental conditions (baseline, CC, RF, RandF and ILFS) and six binary classification tasks for α **and** β **frequency band** (8-30 Hz) for **session 1 data (6a) and session 2 (6b)** using 10-fold cross-validation.

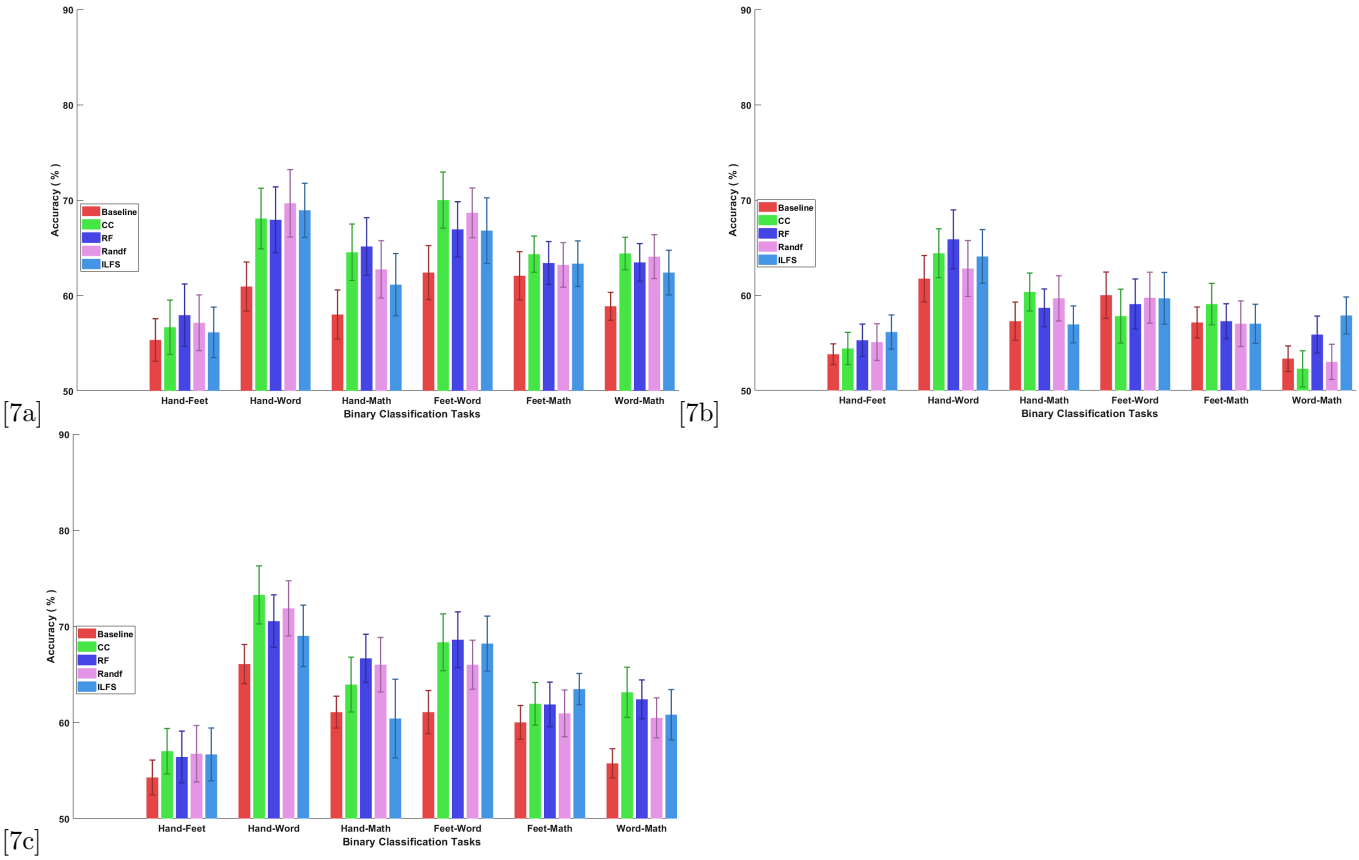

Figure 7. Mean classification accuracies (CAs) obtained with the five experimental conditions (baseline, CC, RF, RandF and ILFS) and six binary classification tasks for α **(8-12Hz),** β **(13-30 Hz) and** α **and** β **frequency band (8-30 Hz)**, a, b, and c respectively, for a classifier **trained on session 1 data and tested on session 2 data**.

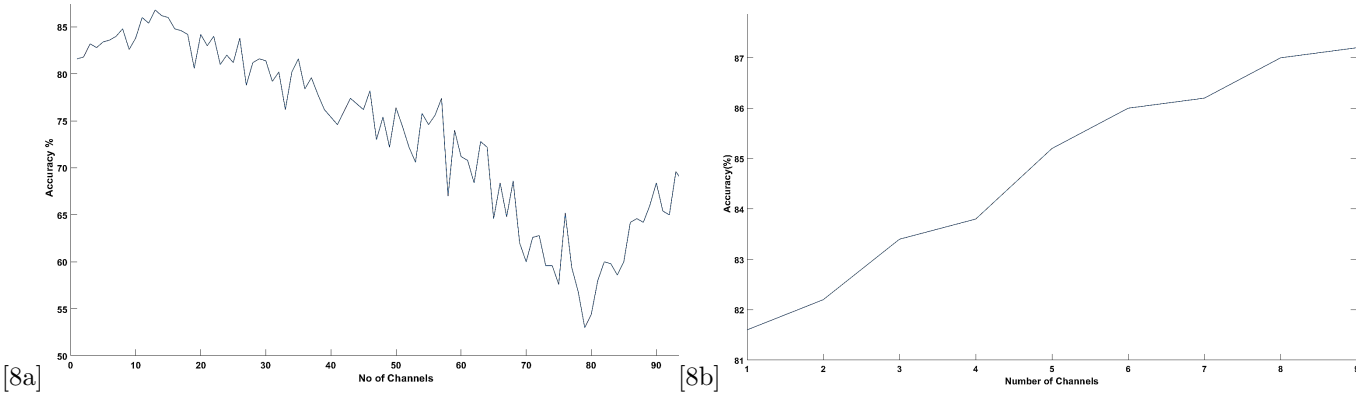

Figure 8. 8a and 8b show the classification accuracy for a subject in a single session for the hand vs feet category calculated from the sequential addition of ranked channels and channels obtained by forward elimination method respectively using RandF method for motor imagery.

Table 1. Number of channels contributing to maximum accuracy using RandF method for session 1 and 2 data in $\alpha + \beta$ frequency band using bandpower feature.

Participants	**Hand vs Foot**		**Hand vs Word**		**Hand vs Math**		**Feet vs Word**		**Feet vs Math**		**Math vs Word**	
	*S*01	*S*02	*S*01	*S*02	*S*01	*S*02	*S*01	*S*02	*S*01	*S*02	*S*01	*S*02
P01	7	7	14	5	16	5	9	7	14	19	11	12
P02	13	11	14	4	23	12	15	12	6	15	9	9
P03	10	8	10	13	10	12	9	9	8	8	14	13
P04	14	11	5	13	8	14	14	16	17	1	12	10
P05	9	9	14	9	7	21	5	6	8	6	16	9
P06	14	6	24	15	11	15	18	17	14	9	18	6
P07	16	6	19	11	14	10	11	10	6	9	14	6
P08	13	12	13	14	9	14	13	10	8	12	10	21
P09	22	12	18	9	13	13	16	9	13	2	13	9
P10	4	13	9	16	10	12	15	18	10	16	18	6
P11	16	11	9	12	14	14	19	18	14	7	12	14
P12	9	19	2	12	3	7	17	12	7	12	9	6
P13	13	11	16	7	10	13	10	11	3	8	5	4
P14	7	5	12	8	15	14	10	5	6	20	7	2
P15	11	18	16	8	9	11	16	18	8	13	11	11
Mean	**11.87**	**10.6**	**13**	**10.4**	**11.47**	**12.47**	**13.13**	**11.87**	**9.47**	**10.47**	**11.93**	**9.2**

hand vs. math mean CA is 74.20%(±14.4), and mean number of selected channels is 61.67. For feet vs. word, feet vs. math and word vs math, mean CAs are 76.47%(±15.28), 75.07%(±12.19) and 74.20%(±9.93) respectively. Clearly mixed imagery has performed better than motor imagery.

Table 2 describes the maximum CA attained by each subject using the respective number of channels (NoC) in a binary classification task in $\alpha + \beta$ band using CSP. The channels were selected from session 1 data and tested in the same order on session 2.

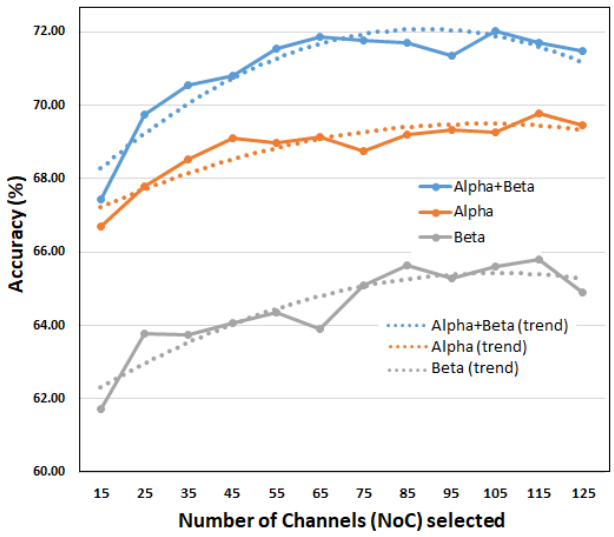

Figure 9. Mean between-sessions accuracy averaged over all classification tasks classification performance with the channels selected using CSP in all the three bands.

3.3. Selected Channels

Figure 11 displays the contributing channels giving maximum accuracy in binary tasks classification in $\alpha + \beta$ frequency band (8-30 Hz) using RandF method. There are three images for every task, i.e. Left Hemisphere (LH) view, Right Hemisphere (RH) view , and Top View. For generating Figure 11 fieldtrip brain template was used along with SPM toolbox. The channel locations on the template are matched to the list of optimal channels obtained from the analysis via MEG channel labels and these channels are shown as markers on the head template. The plot shows the approximate location of channels with respect to the helmet. Figure 11 displays the common contributing channels for subject's binary classification tasks in $\alpha+\beta$ frequency band. The channels which are contributing positively and are common to minimum of 3 subjects are plotted. First, the channels were ranked by RandF feature importance method and then they were selected using forward elimination approach. The first triplet represents selected channels in the case of Hand vs Feet task using the left hemisphere view, right hemisphere view and top view respectively. The total number of contributing channels are shown in Table 1 for all the subjects and for all binary classes. The red dot represents the common channel in session 1 and the green diamond represents common channels for session 2. The channels can be observed in motor areas for hands and feet. Some shift in channel location can also be observed. The next triplet represents the Top, LH, and RH view of the hand vs word imagery task, respectively. Channels in motor imagery are selected in frontal, parietal lobe, and occipital lobe. The next triplet presents the channels for hand vs math tasks.

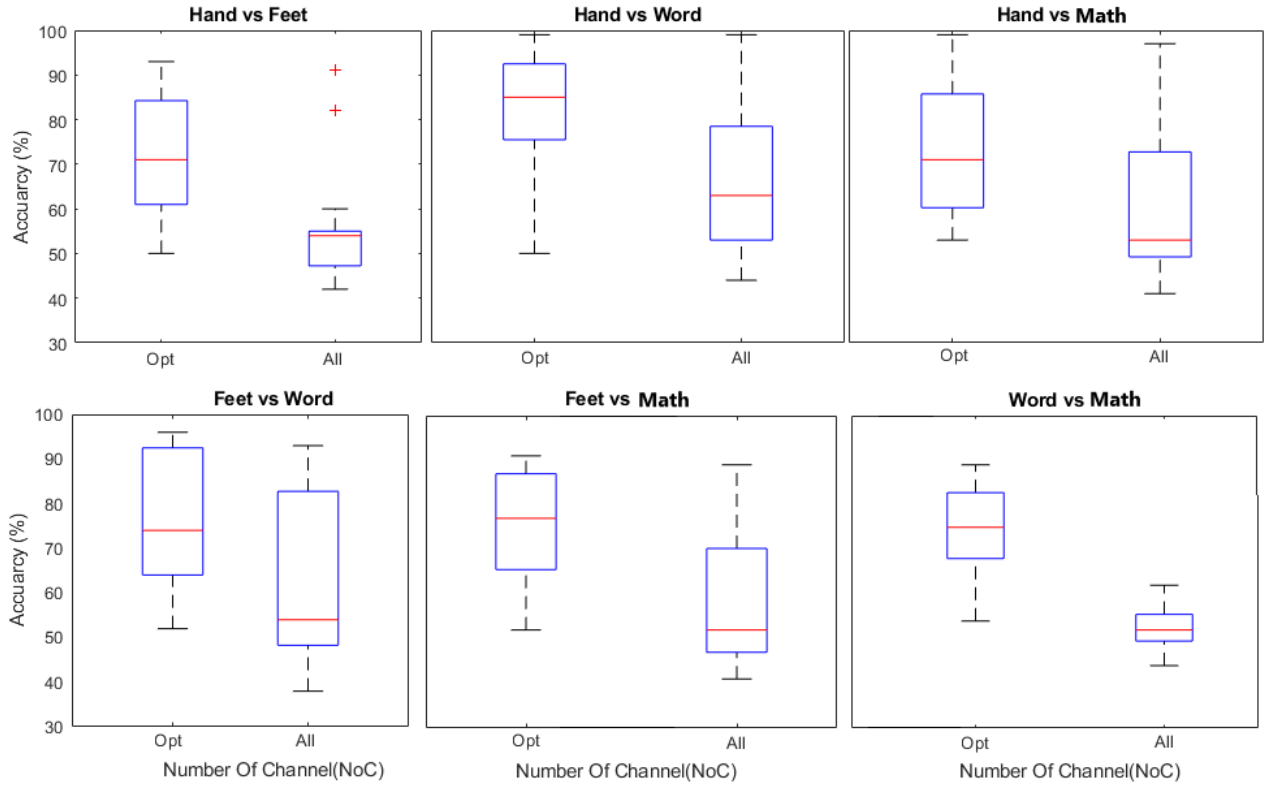

Figure 10. Classification accuracy comparison after channel selection in $\alpha+\beta$ band vs all the channels when CSP was used. The classification accuracy is obtained after training on session 1 and testing on session 2. NoC : number of channel, opt: optimum

Table 2. Number of channels contributing to maximum accuracy using RandF method for model trained on session 1 and tested on session 2 data in $\alpha + \beta$ frequency band using CSP.

Participants	**Hand vs Foot**		**Hand vs Word**		**Hand vs Math**		**Feet vs Word**		**Feet vs Math**		**Math vs Word**	
	NoC	*Acc*(%)	*NoC*	*Acc*(%)	*NoC*	*Acc*(%)	*NoC*	*Acc*(%)	*NoC*	*Acc*(%)	*NoC*	*Acc*(%)
P01	55	58	25	50	45	53	15	52	85	67	55	55
P02	95	50	25	93	25	70	15	93	75	79	55	77
P03	55	93	55	99	35	98	65	93	35	91	55	84
P04	55	82	35	90	105	57	25	71	55	62	65	73
P05	65	85	35	97	35	99	75	91	65	89	35	83
P06	105	72	15	77	35	80	15	63	55	65	35	66
P07	75	61	65	81	65	64	85	96	55	75	25	71
P08	45	62	15	81	65	79	85	85	55	88	75	67
P09	15	58	35	75	75	59	75	58	55	52	25	82
P10	65	86	65	85	75	82	45	74	35	84	65	74
P11	85	71	85	89	75	71	45	67	75	81	65	83
P12	65	61	55	61	105	70	55	52	95	56	35	54
P13	75	86	65	91	85	88	45	87	105	77	75	80
P14	85	61	85	65	65	56	75	71	95	70	25	75
P15	25	74	45	93	35	87	95	94	35	90	85	89
Mean	**64.33**	**70.67**	**47.00**	**81.80**	**61.67**	**74.20**	**54.33**	**76.47**	**65.00**	**75.07**	**51.67**	**74.20**

Channels are primarily in motor areas. The next triplet is for the feet vs word classification task. Again the majority of channels are in the motor areas. The next triplet is for word vs math imagery tasks where channels are evenly distributed over all areas of the brain. The last triplet represents contributing channels for feet vs math. Channels in the motor areas as well as the occipital lobe can be observed. It is worth noticing that in all the categories, channels in the occipital region are selected for motor imagery classification. As the experimental paradigm involved presentation of a cue (an image) to the participants, the visual cue may result in positive contributions from the occipital region channels due to visual display/processing.

4. Discussion

This paper provided an empirical assessment of the impact of channel selection on MEG based BCI performance. MEG-based BCI recordings are conducted using spatially fixed sensor array, and since participants' heads are positioned inside helmet, which is of not exactly same as a participant's head size, it is possible to have slight head movements, which is unintentional. The head movements are not wanted during any MEG experiment (participants were advised to keep their head still during the experiment), but there are unintentional head movements and also different positioning in the different sessions causes change in activated sensors. Unintentional head movements and variable positioning across sessions may thus cause non-stationarity in the acquired MEG data and may result in change of optimal selection of channels across different sessions (for the same participant) and across different participants. Any movement in the initial head positions results in lower accuracy. There is a statistically significant effect of head movement on the dipole reconstruction as reported by Stokes et al. [43], which can also result in lower accuracy from session-to-session transfer. Also, the number of channels in MEG is much higher compared to EEG. Fewer trials in MEG based BCI demands dimensionality reduction, hence there is a need for the selection of a minimum number of sensors which contributes positively in the classification of different mental tasks. Different participants have different head sizes and it would be difficult to acquire MRI of all subjects, thus a head template provided by fieldtrip has been used for this study. This paper provides an approach to improve classification accuracy in real-time with the existing MEG system by accounting for low accuracy due to head movement and head position in the helmet. The performance is evaluated with binary CAs across four different classes. Channel ranking was performed using state-of-the-art feature selection methods i.e. CC [30], RF [31], RandF [33], and ILFS [34] followed by the selection of a set of best channels based on improvement of CA using a forward elimination method. The impact of channel selection was studied with two feature types i.e. power in various frequency bands (α, β, and $\alpha + \beta$) and CSP. Furthermore, the analysis was performed within-session (10-fold CV) and across-session conditions (training on S01 and testing on S02). The study provided several outcomes. Firstly, all four methods significantly improved the performance as compared to a baseline method wherein all channels were included in both within-session and cross-session conditions. Secondly, the RandF method provided more stability and overall better CA than the rest of the ranking methods i.e. CC, ILFS, and RF. Thirdly, MEG-based imagery BCI performed better with mix-imagery classification tasks (i.e. combinations of CI and MI classes e.g. H-W, H-M, and F-M) as compared to motor-imagery tasks (i.e. H-F). Fourthly, the MEG-BCI system performed better for the within-session condition as compared to the across-session condition. Lastly, the number of best channels selected varies a lot both across sessions and subjects. Moreover, the list of best channels for a particular binary classification task changes across sessions and the pattern is consistent across all classification comparisons.

Using bandpower as a feature and the RandF method for channel selection, on average the accuracy improved by 15.79% using the α band, 15.77% using β band and 15.67% using $\alpha + \beta$ band in 10-fold cross validation over the corresponding accuracies obtained with all the channels in the respective bands. There was a marginal increase in CA in the α band compared to α+β band but there was no statistically significant improvement overall. It was also observed that ranking and selecting positively contributing channels helped to improve the accuracy. From the results it is evident that the performance of CSP for between-session accuracy is higher than the bandpower based approach and the improvement is also statistically significant (p <0.05, paired-sample t-test). The power of CSP to minimize the inter-session inconsistencies in feature distribution is well evident in previous literature [44, 45].However, the use of CSP in this study is two-fold: first of all using two different features (bandpower and CSP) we showed that the application of channel selection algorithm is feature agnostic. This means that if we use channel selection on top of any feature extraction technique popular in BCI domain we should get higher performance than without channel selection; the second objective was to enhance the between-session accuracy to a level sufficient for practical uses such as issuing neurofeedback for rehabilitation purposes. Due to the presence of a large number

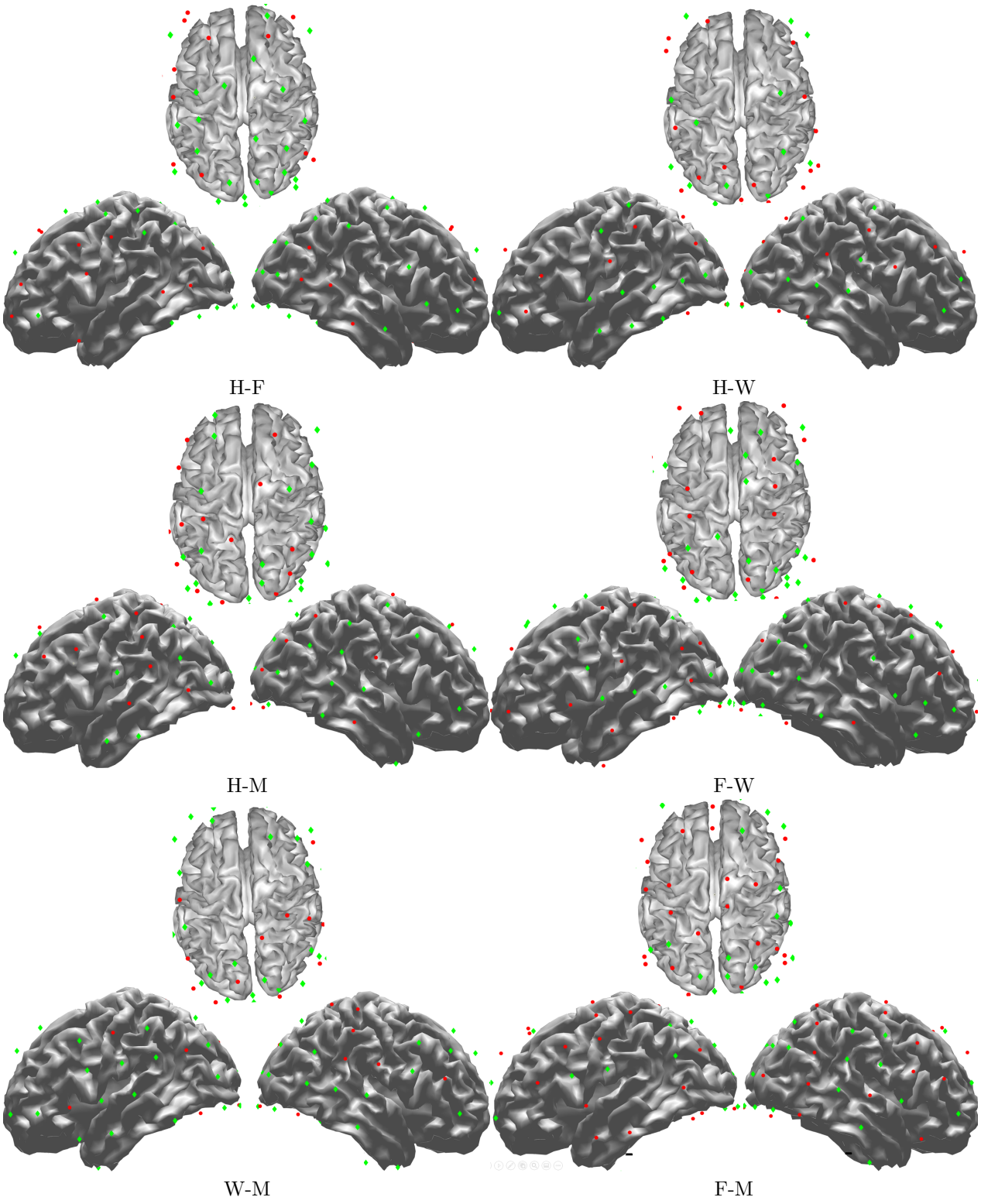

Figure 11. Plot of channels common for minimum of 3 participants for six binary classification imagery tasks for 8-30 Hz using RandF method. • represents channels in session 01 and & ♦ represents selected channel in session 2.

of channels in MEG data acquisition it is hard to avoid overfitting even after using CSP. Therefore, further channels selection was used to reduce the number of channels, possibly close to a typical EEG data acquisition settings using the forward elimination procedure.

When the classifier was trained on session 1 data and tested on session 2 data there was a sharp decrease in CA with or without channel selection. However, an overall increase in CA of 4.64% was observed in the α band including all experimental conditions. With regards to binary classification an increase in CA was observed of 1.8% (H-F), 8.73% (H-W), 4.73% (H-M), 6.27% (F-W), 1.13% (F-M) and 5.20% (W-M) in the α band ($p < 0.05$). There was no statistically significant improvement in H-F and F-M binary classification categories. By examining the top-ranked channels in Table 1, it is clear that the numbers of channels are different between session 1 and session 2. But if the mean across all subjects is observed, it will be noticed that there is not a statistically significant difference overall. Taking the same participant, and the same activity in two different sessions, a clear reduction in the channel number is seen. Also if Figure 11 is observed, a substantial shift in the selected channels' pattern can be seen in H-F(LH). A similar pattern can be observed in W-M(LH). Unintentional head movements and variable positioning across sessions may cause non-stationarity in the acquired MEG data and may result in change of optimal selection of channels across different sessions (for the same participant) and across different participants. Since, this conclusion is drawn on a low number of trials, so the study provides a foundational base for classification methodology for data with a low trial count.

It is to be noted that positively contributing channels appear over all brain areas (Figure 11) but there are some definite changes in selected optimal channels in cases of mixed imagery and motor imagery. The plots, show the channels common to 3 or more subjects which might include variability among participants. This can be one of the plausible causes of channels spread over multiple brain areas. The density of channels is however higher in the functional brain area of specific activity. For example the density of channels in motor task is more for motor area compared to other brain regions. Additionally, it is now widely known the brain connectivity networks can be altered while performing imagery tasks which might contribute towards involvement of channels from other brain regions. Several studies indicated significant changes within the fronto-parietal brain networks during cognitive and motor tasks. Thus, some additional channels may appear due to brain region connectivity changes contributing positively for classification [46, 47]. The contribution of frontal and occipital areas during hand vs feet classification is likely to be because the tasks involve imagination of different real life actions which may activate the frontal area while occipital channels may be relevant due to activation of visual processing of different visual cues [48, 49]. The contribution of fronto-temporal reason for feet-vs-word task classification is likely to be because tasks involving word generation and motor action may activate Broca's areas which is language center in human brain [50]. An interesting application for the outcome of this study would be for a real-time BCI. For instance, it may be possible to design a machine learning model that determines the best channels during the initial trials of an experiment, as part of calibration. Over the complete session, this should result in higher classification accuracy. It was also observed that channels contributing positively to an MI-based classification remain almost the same for good participants. Also, it can be observed that mixed imagery classification performance is significantly higher than motor imagery; this observation can be used to design an enhanced paradigm.

Using the forward elimination approach, in some cases, there were just two channels contributing positively which would make the application of CSP techniques problematic as CSP can only be used on three or more channels. So, in this case CSP channels were ranked using a RandF method and then a minimum of the first 15 channels were selected to calculate accuracy. Then a block of 10 channels was added to see the change in performance. It was noticed that the accuracy improved up to 75-85 channels overall and then plateaued or dropped as seen in Figure 9. Now, as per the previous literature, researchers have used only sensori-motor area channels to reduce the total number of channels for motor-imagery classification. But even when these channels are selected manually, the number is still in the region of 80. Channels selected in the frontal and occipital lobes can also be observed. Similarly for cognitive tasks, a higher number of channels must be used for classification. Thus, reliance upon the classical brain areas may not yield the best results. Using feature importance helps in improving the classification performance with a minimum number of channels. As the trend can be observed from Figure 8, by addition of channels the CA increases and after a certain point it either falls or saturates. Though channel selection helps to improve accuracy, head movement needs to be accounted for. Channel selection not only reduces the dimensionality of data but also helps in online BCI as the computational cost will be lower. Using channel selection, the number of channels was reduced from 204 to a range of 1-25 using bandpower and 15-105 using

CSP. This method is also helpful in handling data if the ratio of trials to number of channels is low, as in this case, it was 50:204.

From the results it can be concluded that mixed imagery (i.e. H-W, H-M, F-W or F-M) has shown better classification accuracy than motor imagery tasks. For making an effective BCI a blend of channel selection and mixed imagery can be used. In order to incorporate channel selection in a real-time MEG BCI implementation, the minimum number of trials needed to perform a numerically stable optimal channel selection was investigated. It was observed that the order of the ranked channels remains almost the same after 15 trials. Thus in practice, an optimal channel selection can be implemented in two ways for a real-time BCI using MEG: (1) a classifier can be trained on session 1 data and selected channels from session 1 can be used for session 2 classification using bandpower and/or CSP; (2) channels can be selected, based on the first few (mixed) trials (e.g. 15) based on bandpower or the first n number of channels can be considered for CSP, where n will be decided by observing data from same session.Using the second method, improved performance can be observed in the same session.

5. Conclusion

This paper analysed the effect of channel selection in improving classification performance for MEG for the first time. It has been observed, that using positively contributing channels reduces the channel count dramatically while the classification accuracy can be significantly improved both in the case of bandpower and CSP features, which was consistent across all the 6 binary classification tasks. However, CSP requires more channels to perform better than the bandpower feature. There was a statistically significant increase ($p < 0.05$) in performance in all cases. Additionally, it is also observed that mixed imagery (e.g. H-W) performed significantly better than pure MI (e.g. H-F) and CI tasks (e.g. W-M), which can be helpful in designing a paradigm where performance is the priority.

Acknowledgment

This work is supported by the Department of Science and Technology (DST), India and UK India Education and Research Initiative (UKIERI) Thematic Partnership project, 'Advancing MEG based Brain-Computer Interface Supported Upper Limb Post-Stroke Rehabilitation' (DST-UKIERI-2016-17-0128). G.P. is also supported by the Northern Ireland Functional Brain Mapping Facility project (1303/101154803), funded by InvestNI and Ulster University.

References

[1] Leigh R Hochberg, Mijail D Serruya, Gerhard M Friehs, Jon A Mukand, Maryam Saleh, Abraham H Caplan, Almut Branner, David Chen, Richard D Penn, and John P Donoghue. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164, 2006.

[2] Girijesh Prasad, Pawel Herman, Damien Coyle, Suzanne McDonough, and Jacqueline Crosbie. Applying a brain-computer interface to support motor imagery practice in people with stroke for upper limb recovery: a feasibility study. *Journal of neuroengineering and rehabilitation*, 7(1):60, 2010.

[3] Bin He, Shangkai Gao, Han Yuan, and Jonathan R Wolpaw. Brain–computer interfaces. In *Neural Engineering*, pages 87–151. Springer, 2013.

[4] Han Yuan and Bin He. Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives. *IEEE Transactions on Biomedical Engineering*, 61(5):1425–1435, 2014.

[5] Pieter-Jan Kindermans, Hannes Verschore, and Benjamin Schrauwen. A unified probabilistic approach to improve spelling in an event-related potential-based brain–computer interface. *IEEE Transactions on Biomedical Engineering*, 60(10):2696–2705, 2013.

[6] C-C Postelnicu and Doru Talaba. P300-based brain-neuronal computer interaction for spelling applications. *IEEE transactions on biomedical engineering*, 60(2):534–543, 2012.

[7] Wojciech Samek, Frank C Meinecke, and Klaus-Robert Müller. Transferring subspaces between subjects in brain–computer interfacing. *IEEE Transactions on Biomedical Engineering*, 60(8):2289–2298, 2013.

[8] M Felix Orlando, Himanshu Akolkar, Ashish Dutta, Anupam Saxena, and Laxmidhar Behera. Optimal design and control of a hand exoskeleton. pages 72–77, 2010.

[9] J Farmer, XV Zhao, H Van Praag, K Wodtke, FH Gage, and BR Christie. Effects of voluntary exercise on synaptic plasticity and gene expression in the dentate gyrus of adult male sprague–dawley rats in vivo. *Neuroscience*, 124(1):71–79, 2004.

[10] Niels Birbaumer and Leonardo G Cohen. Brain–computer interfaces: communication and restoration of movement in paralysis. *The Journal of physiology*, 579(3):621–636, 2007.

[11] Anirban Chowdhury, Yogesh Kumar Meena, Haider Raza, Braj Bhushan, Ashwani Kumar Uttam, Nirmal Pandey, Adnan Ariz Hashmi, Alok Bajpai, Ashish Dutta, and Girijesh Prasad. Active physical practice followed by mental practice using BCI-driven hand exoskeleton: a pilot trial for clinical effectiveness and usability. *IEEE journal of biomedical and health informatics*, 22(6):1786–1795, 2018.

[12] Benjamin Blankertz, Gabriel Curio, and Klaus-Robert Müller. Classifying single trial EEG: Towards brain computer interfacing. In *Advances in neural information processing systems*, pages 157–164, 2002.

[13] Sujit Roy, Karl McCreadie, and Girijesh Prasad. Can a single model deep learning approach enhance classification accuracy of an EEG-based brain-computer interface? In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 1317–1321. IEEE, 2019.

[14] Peter T Lin, Kartikeya Sharma, Tom Holroyd, Harsha Battapady, Ding-Yu Fei, Ou Bai, and Fr Signorelli. A high performance MEG based BCI using single trial detection of human movement intention. *Functional brain mapping and the endeavor to understand the working brain. InTech*, pages 17–36, 2013.

[15] Stephen T Foldes, Douglas J Weber, and Jennifer L Collinger. MEG-based neurofeedback for hand rehabilitation. *Journal of neuroengineering and rehabilitation*, 12(1):1–9, 2015.

[16] Franca Tecchio, Filippo Zappasodi, Mario Tombini, Antonio Oliviero, Patrizio Pasqualetti, Fabrizio Vernieri, Matilde Ercolani, Vittorio Pizzella, and Paolo Maria Rossini. Brain plasticity in recovery from stroke: an MEG assessment. *Neuroimage*, 32(3):1326–1334, 2006.

[17] Jürgen Mellinger, Gerwin Schalk, Christoph Braun, Hubert Preissl, Wolfgang Rosenstiel, Niels Birbaumer, and Andrea Kübler. An MEG-based brain–computer interface (BCI). *Neuroimage*, 36(3):581–593, 2007.

[18] Sujit Roy, Dheeraj Rathee, Karl McCreadie, and Girijesh Prasad. Channel selection improves MEG-based brain-computer interface. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 295–298. IEEE, 2019.

[19] Hanna-Leena Halme and Lauri Parkkonen. Comparing features for classification of MEG responses to motor imagery. *PloS one*, 11(12):e0168766, 2016.

[20] Ethan Buch, Cornelia Weber, Leonardo G Cohen, Christoph Braun, Michael A Dimyan, Tyler Ard, Jurgen Mellinger, Andrea Caria, Surjo Soekadar, Alissa Fourkas, et al. Think to move: a neuromagnetic brain-computer interface (BCI) system for chronic stroke. *Stroke*, 39(3):910–917, 2008.

[21] Ryohei Fukuma, Takufumi Yanagisawa, Youichi Saitoh, Koichi Hosomi, Haruhiko Kishima, Takeshi Shimizu, Hisato Sugata, Hiroshi Yokoi, Masayuki Hirata, Yukiyasu Kamitani, et al. Real-time control of a neuroprosthetic hand by magnetoencephalographic signals from paralysed patients. volume 6, pages 1–14. Nature Publishing Group, 2016.

[22] Dheeraj Rathee, Haider Raza, Girijesh Prasad, and Hubert Cecotti. Current source density estimation enhances the performance of motor-imagery-related brain–computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(12):2461–2471, 2017.

[23] Yijun Wang, Shangkai Gao, and Xiaornog Gao. Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 5392–5395. IEEE, 2006.

[24] Mahnaz Arvaneh, Cuntai Guan, Kai Keng Ang, and Chai Quek. Optimizing the channel selection and classification accuracy in EEG-based BCI. *IEEE Transactions on Biomedical Engineering*, 58(6):1865–1873, 2011.

[25] Claudia Sannelli, Thorsten Dickhaus, Sebastian Halder, Eva-Maria Hammer, Klaus-Robert Müller, and Benjamin Blankertz. On optimal channel configurations for SMR-based brain–computer interfaces. *Brain topography*, 23(2):186–193, 2010.

[26] Lin He, Zhuliang Yu, Zhenghui Gu, and Yuanqing Li. Bhattacharyya bound based channel selection for classification of motor imageries in EEG signals. In *2009 Chinese Control and Decision Conference*, pages 2353–2356. IEEE, 2009.

[27] Jianhai Zhang, Ming Chen, Shaokai Zhao, Sanqing Hu, Zhiguo Shi, and Yu Cao. Relieff-based EEG sensor selection methods for emotion recognition. *Sensors*, 16(10):1558, 2016.

[28] Turky Alotaiby, Fathi E Abd El-Samie, Saleh A Alshebeili, and Ishtiaq Ahmad. A review of channel selection algorithms for EEG signal processing. *EURASIP Journal on Advances in Signal Processing*, 2015(1):66, 2015.

[29] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005, 2018.

[30] Mark A Hall. Correlation-based feature selection of discrete and numeric class machine learning. 2000.

[31] Kenji Kira, Larry A Rendell, et al. The feature selection problem: Traditional methods and a new algorithm. In *Aaai*, volume 2, pages 129–134, 1992.

[32] Marko Robnik-Šikonja and Igor Kononenko. Theoretical and empirical analysis of relieff and rrelieff. *Machine learning*, 53(1-2):23–69, 2003.

[33] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.

[34] Giorgio Roffo, Simone Melzi, Umberto Castellani, and Alessandro Vinciarelli. Infinite latent feature selection: A probabilistic latent graph-based ranking approach. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1398–1406, 2017.

[35] Igor Kononenko. Estimating attributes: analysis and extensions of RELIEF. In *European conference on machine learning*, pages 171–182. Springer, 1994.

[36] Giorgio Roffo, Simone Melzi, and Marco Cristani. Infinite feature selection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4202–4210, 2015.

[37] Giorgio Roffo. Ranking to learn and learning to rank: On the role of ranking in pattern recognition applications. *arXiv preprint arXiv:1706.05933*, 2017.

[38] Anirban Chowdhury, Haider Raza, Yogesh Kumar Meena, Ashish Dutta, and Girijesh Prasad. An eeg-emg correlation-based brain-computer interface for hand orthosis supported neuro-rehabilitation. *Journal of Neuroscience Methods*, 312:1 – 11, 2019.

[39] Siyang Yin, Yuelu Liu, and Mingzhou Ding. Amplitude of sensorimotor mu rhythm is correlated with bold from multiple brain regions: A simultaneous eeg-fmri study. *Frontiers in Human Neuroscience*, 10:364, 2016.

[40] D. Rathee, A. Chowdhury, Y. K. Meena, A. Dutta, S. McDonough, and G. Prasad. Brain–machine interface-driven post-stroke upper-limb functional recovery correlates with beta-band mediated cortical networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(5):1020–1031, 2019.

[41] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek. Optimizing spatial filters by minimizing within-class dissimilarities in electroencephalogram-based brain–computer interface. *IEEE Transactions on Neural Networks and Learning Systems*, 24(4):610–619, 2013.

[42] F. Lotte and C. Guan. Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms. *IEEE Transactions on Biomedical Engineering*, 58(2):355–362, 2011.

[43] Arjen Stolk, Ana Todorovic, Jan-Mathijs Schoffelen, and Robert Oostenveld. Online and offline tools for head movement compensation in MEG. *Neuroimage*, 68:39–48, 2013.

[44] H. Raza et al. Adaptive learning with covariate shift-detection for motor imagery-based brain–computer interface. *Soft Computing*, 20:3085–3096, 2016.

[45] A. Chowdhury, H. Raza, Y. K. Meena, A. Dutta, and G. Prasad. Online covariate shift detection-based adaptive brain–computer interface to trigger hand exoskeleton feedback for neuro-rehabilitation. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4):1070–1080, 2018.

[46] Christian Grefkes, Simon B Eickhoff, Dennis A Nowak, Manuel Dafotakis, and Gereon R Fink. Dynamic intra- and interhemispheric interactions during unilateral and bilateral hand movements assessed with fmri and dcm.

Neuroimage, 41(4):1382–1394, 2008.

[47] Qing Gao, Xujun Duan, and Huafu Chen. Evaluation of effective connectivity of motor areas during motor imagery and execution using conditional granger causality. *Neuroimage*, 54(2):1280–1288, 2011.

[48] Jennifer A Stevens. Interference effects demonstrate distinct roles for visual and motor imagery during the mental representation of human action. *Cognition*, 95(3):329–350, 2005.

[49] Ana Solodkin, Petr Hlustik, E Elinor Chen, and Steven L Small. Fine modulation in network activation during motor execution and motor imagery. *Cerebral cortex*, 14(11):1246–1255, 2004.

[50] Byron Bernal, Alfredo Ardila, and Monica Rosselli. Broca's area network in language function: a pooling-data connectivity study. *Frontiers in psychology*, 6:687, 2015.