# Brain-Computer Interface using temporal- spectral features and neural network classifier

Moran Cerf [1] and Gan Wang [2]

[1]Northwestern University
[2]Affiliation not available

October 30, 2023

Posted on 1 Feb 2020 — CC-BY 4.0 — https://doi.org/10.36227/techrxiv.19069529.v1 — e-Prints posted on TechRxiv are preliminary reports that are not peer reviewed. They should not b...

### Abstract

We show that by extracting temporal and spectral features from EEG signal and, following, using neural network to classify those features, one can significantly improve the performance of Brain-Computer Interfaces (BCIs) in predicting which motor movement was imagined by a subject. Our movement prediction algorithm uses Sequential Backward Selection technique to jointly select the temporal and spectral features, and a radial basis function neural network for the classification. The method shows an average performance increase of 5.96% compared to state-of-the-art benchmark algorithms. Using two popular public datasets, our algorithm reaches 91.73% accuracy (compared to an average benchmark of 81.10%) on the first dataset, and 88.78% (average benchmark: 82.76%) on the second dataset. Given the high variability within- and across-subjects in EEG-based motion decoding, we suggest that using features from multiple modalities along with neural network feature selection and classification protocol is likely to increase BCI performance across various tasks.

# Brain-Computer Interface using temporal-spectral features and neural network classifier

Gan Wang, and Moran Cerf

***Abstract*—We show that by extracting temporal and spectral features from EEG signal and, following, using neural network to classify those features, one can significantly improve the performance of Brain-Computer Interfaces (BCIs) in predicting which motor movement was imagined by a subject. Our movement prediction algorithm uses Sequential Backward Selection technique to jointly select the temporal and spectral features, and a radial basis function neural network for the classification. The method shows an average performance increase of 5.96% compared to state-of-the-art benchmark algorithms. Using two popular public datasets, our algorithm reaches 91.73% accuracy (compared to an average benchmark of 81.10%) on the first dataset, and 88.78% (average benchmark: 82.76%) on the second dataset. Given the high variability within- and across-subjects in EEG-based motion decoding, we suggest that using features from multiple modalities along with neural network feature selection and classification protocol is likely to increase BCI performance across various tasks.**

***Index Terms:*** **[Science–general, Neuroscience], [Brain-Computer Interface], [Common Spatial Patterns], [EEG]**

Submitted on January 22, 2022. The authors received no funding or support for this work. The authors declare no conflict of interest in the work. The data and analyses codes (in *Matlab*) are available on www.morancerf.com/publications. GW conducted the analyses. GW and MC wrote the manuscript and generated the final output.

## I. Introduction

BRAIN-Computer Interfaces (BCI) act as a link between neural activity and machine operations. The BCI extracts data from electrodes or sensors acquiring neural signals and translates those data into digital code[1]. Application of BCI range from ones focused on improved health outcomes (i.e., rehabilitation of impaired motor function[2], restoring sensory functions[3], or interpreting thoughts from individuals who cannot otherwise communicate them[5]), enhanced control of devices (i.e., operating heavy machinery, flying drones[6], or driving[7]), and even recreational uses (i.e., gaming[8]). Invasive BCIs that access raw neural signals have recently shown high accuracy in interpreting human/animal intentions, actions, and imagery[3,5]. Similarly, non-invasive tools such as electroencephalography (EEG) have demonstrated high performance in interpreting thoughts in real-time. For example, interpreting imagined motor action – a commonly used task for evaluating BCIs – has shown accuracies ranging between 70-85% in recent works[9].

BCIs based on Motor Imagination (MI) task, where a participant imagines an action repeatedly over multiple trials, typically aim to identify the action class (i.e., clenching of the fist) by neural signatures such as significant power changes in the alpha and beta rhythm in sensory-motor regions. Given that non-invasive signals generated by EEG are often contaminated by artifacts derived from eye movement or muscle movement, a typical EEG-based BCI requires larger training data and isolated trials to accurately identify the action type. The repeated trials enable the averaging of the event-related signals and the extraction of a synchronized clean input. Variance across individual subjects, electrode montages, experimental sessions, and trials adds difficulty to the accurate interpretation of the signal.

Given the challenges in BCI development using the noisy inputs, numerous methods have been proposed to improve the performance of EEG-based BCIs. The suggested methods often focus on isolation of either the temporal or spectral components of the signal. Algorithms based on spectral feature selection are more prominent in the BCI arsenal since the time course of event-related synchronization (or de-synchronization) vary

G. Wang is with the School of Electrical Engineering, Soochow University, Soochow, China (e-mail: 296309197@qq.com). M. Cerf is with the Interdepartmental Neuroscience Program, Northwestern University, Evanston, IL 60208 USA (e-mail: moran@morancerf.com).

heavily among subjects during motor movement (or imagined movement) tasks[3-4].

Within the feature selection arsenal for BCI signals, Common Spatial Patterns (CSP) algorithms are dominant[12]. These algorithms seek to find an optimal spatial filter that distinguishes one brain state from another. In EEG, the performance of CSPs is highly sensitive to the choice of frequency bands, making the decision on which filter to use heavily dependent on the recording configuration. To afford some generalization, variants of CSP that use narrower frequency bands (termed: sub-band CSP; SBCSP)[13] and Filter Banks (FBCSP)[14] were proposed. These variants show increased performance for BCIs, yet are still scarce.

In addition to the extended frequency bands and filters, recent attempts to include temporal signals in the BCI emerged in the form of the Temporally Constrained Group Spatial Patterns (TSGSP) algorithms[5,11]. These algorithms optimally select the CSP features by considering different temporal windows for signal extraction derived from multiple-task learnings. That is, instead of collapsing all the trials within one MI class (i.e., all left-hand movement trials) various MI tasks are combined to suggest the ideal CSP for that individual. The TSGSP algorithm uses Support Vector Machines (SVM) for the classification of new trials to their corresponding action class. The inclusion of temporal data was shown to improve the performance of CSP-based BCIs[10,11,15].

Neural network based classifiers that often show superiority in data-rich non-linear classification tasks such as MI BCI were recently suggested as a potential additional improvement for the CSP algorithms[12]. Specifically, the usage of Sequential Backward Floating Selection method along with a radial basis function neural network for optimal CSP features selection was suggested as a potential superior algorithm for BCIs[12]. Here, we attempt to implement and test those suggested improvements for MI BCI.

Namely, we introduce a number of additions to the BCI motor classification algorithms arsenal. First, we incorporate both temporal and spectral features in the MI BCI. Second, we use sub-bands rather than typical frequency bands (i.e., alpha, beta). Third, we separate the feature selection process from the following feature classification process. Fourth, we incorporate the suggested radial basis function neural network (rather than SVM) for the motor classification.

Inspired by the above literature, we combine the successful Sequential Backward Selection method with CSP features for the temporal/spectral feature selection.

We suggest that our enhanced feature set and selection optimization process should improve EEG-based BCIs across multiple tasks because it is less prone to individual subject's signal variations. We demonstrate the effectiveness of our methods on popular public datasets and compare out outcomes to the current state-of-the-art BCI benchmark methods.

## II. Methods

### A. Data

Two popular BCI datasets were used for the algorithm testing:

#### 1) Dataset 1

BCI competition IV, dataset 2a, which contains 22-channel EEG data recorded from 9 healthy subjects (A01-A09) participating in different MI tasks. In each task subjects were asked to imagine movement of the left hand, right hand, feet, and tongue. There were 72 trials for each of the 4 classes of movement. The EEG signals were sampled at 250Hz and bandpass filtered between 0.5-100Hz (with a 50Hz notch filter). We used the data from the left- and right-hand imagery tasks alone to align with the second dataset and some of the benchmark algorithms that focused solely on those movement classes.

#### 2) Dataset 2

BCI competition IV, dataset 2b, which contains 3-channel EEG data recorded from 9 different subjects (B01-B09) participating in two MI tasks. The experimental protocol was identical to dataset 1 other than the fact that subjects only imaged movements of the left-hand and right-hand and had 80 trials for each class. For each subject, a separate training and testing set were available. The EEG signals were sampled at 250Hz and bandpass filtered between 1-50Hz (with 50Hz notch filter).

See [16] for additional details on the two datasets.

### B. Feature extraction

#### 1) Preprocessing

Raw EEG signals were filtered between 4-40Hz with fifth-order Butterworth filter. For each trial, we used samples between 500-4,500ms from the trial onset in the analyses. The first 500ms were excluded, in alignment with the instructions of the BCI IV competition winners, as the imagined actions onset deviated across trials due to response times.

#### 2) Feature selection

The neural signals were divided to 5 overlapping 2-second windows with a step size of 500ms. This ensured temporal generalizability within a trial. Following, we filtered the data along 17 overlapping frequency bands ranging from 4-40Hz with a 2Hz step. Finally, we identified the common spatial filter[12] that maximized the variance within a single class (i.e., across all left-hand trials) and minimized the variance across classes (i.e., between left-hand and right-hand trials).

The data for a single trial were represented as a matrix, $X \in R^{N \cdot T}$ (with $N$ reflecting the number of channels, and $T$ the time) whose normalized covariance matrix, $C$, is:

$$C = \frac{XX^T}{trace(XX^T)} \tag{1}$$

Averaging across all trials within a class yielded a matrix, $C_t$, ($t$ indicating the class type).

The spatial covariance was calculated by averaging all covariance matrices:

$$C_c = \overline{C_{left-hand}} + \overline{C_{right-hand}} \quad (2)$$

The $C_c$ matrix was white transformed:

$$C_{C} = U_C \lambda_C U_C^T \quad (3)$$

with $U_C$ the eigenvector matrix and $\lambda_C$ the eigenvectors.

Defining $P$ as:

$$P = \sqrt{\lambda_C^{-1}} U_C^T \quad (4)$$

we transformed the individual class matrices to:

$$S_{left-hand} = P\overline{C_{left-hand}}P^T \quad (5)$$
$$S_{right-hand} = P\overline{C_{right-hand}}P^T \quad (6)$$

such that the $S_t$ matrices have the same eigenvectors.

Given that $S_t$ could be represented as $B\lambda_t B^T$ with $B$ the eigenvectors matrix and $\lambda_t$ the eigenvalues:

$$S_t = B\lambda_t B^T \quad (7)$$

we derived the projection matrix, *W*:

$$W = B^T P \quad (8)$$

Thus, the EEG data were projected to a matrix, *Z*:

$$Z = W^T X \quad (9)$$

where the columns of *Z* corresponded to the data's spatial source distribution vectors. The vectors maximized the variance across classes and corresponded to the maximum eigenvalues ($\lambda_{left-hand}$ and $\lambda_{right-hand}$).

Finally, the classification features were represented by:

$$f_p = \log\left(\frac{var(Z_p)}{\sum_{i=1}^{n} var(Z_i)}\right) \quad (10)$$

where $Z_p$ are the common spatial patterns ($p$ = *1-N*).

Typically, a subset of *Z* (first and last *m* rows) is used in further analyses.

A Sequential Backward Selection (SBS; see [17]) was used to reduce the initial 85-feature set (17 frequency bands *x* 5 time-windows) from each individual trials to as little as 8 features.

Finally, we used a radial basis function neural network[12] to classify the MI signals within each trial.

#### 3) *Analyses*

We compared our classification algorithm to the state-of-the-art algorithms available from BCI competition IV:

For dataset 1 we compared our performance to (see results in **table 01**):

a. Weighted Overlap Add Common Spatial Patterns (WOLA-CSP[18])
b. Subject Specific Multivariate Empirical Mode Decomposition Based Filtering (SS-MEMDBF[19])
c. Regularized Minimum Distance to Riemannian Mean (R-MDRM[20])
d. Spatial Regularized Minimum Distance to Riemannian Mean (SR-MDRM[21])
e. Temporally Constrained Sparse Group Spatial Patterns (TSGSP[11])

For dataset 2 we compared our results to (see results in **table 02**):

a. Robust Support Matrix Machine (RSMM[22])
b. Dynamic Joint Domain Adaptation (DJDA[23])
c. Frequential Deep Belief Network (FDBN[24])
d. Random Forest Dynamic Frequency Feature Selection (RF-DFFS[25])
e. Temporally-constrained Sparse Group Spatial Patterns (TSGSP[11])
f. Wavelet Package Decomposition Spatio-Temporal Discrepancy Feature (WPD-STDF[26])
g. Central Distance Loss Convolutional Neural Network (CD-CNN[27])

Additionally, we implemented a version of the Sequential Backward Selection Filter Bank Common Spatial Patterns (SBS-FBCSP) algorithm, which is an adapted version of the Sub-Band Common Spatial Patterns with Sequential Backwards Floating Selection (SBCSP-SBFS) proposed in [12]. The SBS-FBCSP algorithm resembles our suggested method in that it, too, uses sub-bands and common spatial patterns feature selection. This contender algorithm did not use temporal features and was limited to 12 overlapping frequency bands (4-30Hz).

Given that dataset 1 had 22 channels, we varied the parameter *m* from 1-11 (with *2m* options yielding up to 22 features in each trial). For dataset 2 we varied *m* from 1-3 yielding up to 6 features. For each subject, we used the first 3 sessions ([72/80] trials *x* 2 classes *x* 3 sessions) as training set and the remaining trials for testing.

## III. Results

### A. *Performance*

Our algorithm, a Sequential Backward Selection with Temporal Filter Bank Common Spatial Patterns (SBS-TFBCSP), significantly outperformed the average (81.10±1.58%) of all other algorithms (T(8)=2.998, p=0.017; t-test) and each of those algorithms individually when tested on the first dataset (**table 01**).

The algorithm also outperformed the contender leading algorithm (SBS-FBCSP) by 8.60% (T(8)=2.850, p=0.021, t-test).

TABLE I
PERFORMANCE COMPARISON WITH DATASET 1

| Method | Year | A01 | A02 | A03 | A04 | A05 | A06 | A07 | A08 | A09 | *Mean ± std* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **WOLA-CSP** | **2018** | 86.81 | 63.19 | 94.44 | 68.75 | 56.25 | 69.44 | 78.47 | 97.91 | 93.75 | **78.85±15.15** |
| **SS-MEMDBF** | **2018** | 91.49 | 60.56 | 94.16 | 76.16 | 58.52 | 68.52 | 78.57 | 97.01 | 93.85 | **79.93±14.14** |
| **MDRM** | **2019** | 91.61 | 63.28 | 97.20 | 72.91 | 64.08 | 69.71 | 81.25 | 96.52 | 92.30 | **80.98±13.86** |
| **SR-MDRM** | **2019** | 90.21 | 63.28 | 96.55 | 76.38 | 65.49 | 69.01 | 81.94 | 95.14 | 93.01 | **81.22±12.43** |
| **TSGSP** | **2018** | 87.00 | 64.70 | 93.80 | 74.30 | 90.40 | 63.90 | 91.40 | 95.80 | 81.30 | **82.51±12.23** |
| **SBS-FBCSP** | **2019** | 85.71 | 73.21 | 94.64 | 82.14 | 78.57 | 71.43 | 78.57 | 98.21 | 85.71 | **83.13±9.02** |
| *Ours* | *2022* | *91.07* | *85.71* | *92.86* | *87.50* | *83.93* | *89.29* | *92.86* | *92.86* | *96.43* | ***90.28±4.01*** |

Using the second dataset, our algorithm again significantly outperformed the average (82.76±3.08%) of all other algorithms by 2.69% (T(8)=3.228, p=0.012; t-test) and each of those algorithms individually (**table 02**). Comparing the performance to the leading state-of-the-art contender algorithm (CD-CNN), we see a 3.32% increase (T(8)=1.314, p=0.225, t-test).

TABLE II
PERFORMANCE COMPARISON WITH DATASET 2

| Method | Year | B01 | B02 | B03 | B04 | B05 | B06 | B07 | B08 | B09 | *Mean ± std* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **SBS-FBCSP** | 2019 | 75.00 | 63.16 | 62.50 | 93.42 | 88.82 | 78.95 | 76.32 | 80.92 | 81.58 | **77.85±10.30** |
| **RSMM** | 2016 | 72.50 | 56.43 | 55.63 | 97.19 | 88.44 | 78.75 | 77.50 | 91.88 | 83.44 | **77.97±13.73** |
| **DJDA** | 2021 | 83.44 | 58.57 | 59.06 | 98.13 | 96.56 | 84.38 | 86.25 | 92.81 | 87.81 | **83.00±14.64** |
| **FDBN** | 2016 | 81.00 | 65.00 | 66.00 | 98.00 | 93.00 | 88.00 | 82.00 | 94.00 | 91.00 | **84.22±11.93** |
| **RF-DFFS** | 2016 | 73.24 | 67.48 | 63.01 | 97.40 | 95.49 | 86.66 | 84.68 | 95.93 | 96.21 | **84.06±12.31** |
| **TSGSP** | 2018 | 84.00 | 62.60 | 56.30 | 99.40 | 94.80 | 83.80 | 94.10 | 93.30 | 90.10 | **84.26±15.01** |
| **WPD-STDF** | 2019 | 69.50 | 64.00 | 86.50 | 96.00 | 94.00 | 87.00 | 83.00 | 95.50 | 92.00 | **85.28±10.81** |
| **CD-CNN** | 2021 | 79.69 | 60.71 | 82.19 | 96.87 | 94.37 | 89.37 | 82.19 | 93.75 | 90.00 | **85.46±10.44** |
| *Ours* | *2022* | *89.47* | *75.69* | *75.69* | *100* | *92.76* | *87.84* | *83.78* | *94.08* | *95.39* | ***88.30±8.52*** |

Focusing on the similar SBS-FBCSP algorithm, we note that the contender algorithm shows a notable drop in performance in classifying dataset 2. While the key difference between our algorithm and the SBS-FBCSP is the feature selection method, the inclusion of temporal features in the common spatial patterns; **figure 01**) seems to notably contribute to the performance increase. This expansion of the frequency range used by the radial basis function neural network increases the feature selection granularity. As an intuition for the advantage of the method with respect to the feature selection, we show an example from one of the highest performing subjects (A01, chosen arbitrarily; **figure 01**). We also note that one of the dominant features selected was in a frequency band above 30Hz (34-38Hz) which would be excluded in the typical SBCSP-SBFS implementations[12] since it is not typically associated with MI.

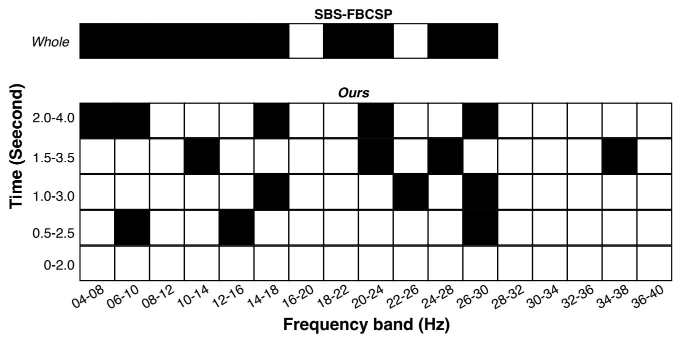



Fig. 1. Illustration of the feature selection difference between the SBS-FBCSP (top rectangle) algorithm and our algorithm (bottom rectangle) that includes the temporal features. The data (taken from subject A01, arbitrarily chosen) highlights the 10 features selected (black squares) in the contender algorithm, and the 15 features selected in our algorithm (the performance with 15 features was optimal, hence the selected number; all options between 8-85 were evaluated).

## *B. Parameter sensitivity*

Given that the performance of our proposed methods heavily depends on the selection of the *m* parameter, we tested our results with all *m* values relevant to dataset 1 (**table 03**) and dataset 2 (**table 04**) to demonstrate the robustness of the method. While, indeed, the choice of *m* affects the algorithm performance across subjects, the average impact of the selection on dataset 1 was 4.25±1.18% with the highest drop in performance yielding 83.76% accuracy — still placing the algorithm above its peers. For dataset 2, the performance change with different *m* values averaged 1.73±1.10%, with the bigger performance drop yielding 86.57% accuracy. The lowest performance is still higher than peer algorithms' accuracy. Combined, these results suggest that the method is robust and maintains its efficiency irrespective of the parameter choice.

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT VALUES OF *m* FOR OUR METHOD, USING DATASET 1

| *m* | A01 | A02 | A03 | A04 | A05 | A06 | A07 | A08 | A09 |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 85.71 | 80.36 | 80.36 | 87.50 | 83.93 | 85.71 | 92.86 | 91.07 | 83.93 |
| **2** | 78.57 | 78.57 | 92.86 | 80.36 | 83.93 | 89.29 | 91.07 | 89.29 | 91.07 |
| **3** | 83.93 | 85.71 | 85.71 | 76.79 | 73.21 | 82.14 | 85.71 | 91.07 | 92.86 |
| **4** | 83.93 | 83.93 | 89.29 | 78.57 | 76.76 | 76.79 | 82.14 | 89.29 | 92.86 |
| **5** | 83.93 | 76.76 | 92.86 | 73.21 | 80.36 | 83.93 | 76.79 | 92.86 | 96.43 |
| **6** | 91.07 | 80.36 | 87.50 | 73.21 | 80.36 | 85.71 | 83.93 | 92.86 | 94.64 |
| **7** | 82.14 | 80.36 | 89.29 | 76.79 | 76.79 | 89.29 | 78.57 | 92.86 | 91.07 |
| **8** | 83.93 | 76.79 | 91.07 | 71.43 | 82.14 | 78.57 | 76.79 | 92.86 | 89.29 |
| **9** | 87.50 | 69.64 | 91.07 | 73.21 | 76.79 | 75 | 76.79 | 89.29 | 83.93 |
| **10** | 82.14 | 76.76 | 83.93 | 73.21 | 80.36 | 76.79 | 83.93 | 83.93 | 87.50 |
| **11** | 83.93 | 76.76 | 91.07 | 69.64 | 82.14 | 76.79 | 71.43 | 89.29 | 89.29 |
| *STD* | 3.17 | 4.26 | 3.93 | 5.01 | 3.41 | 5.28 | 6.52 | 2.68 | 4.02 |

TABLE IV
PERFORMANCE COMPARISON OF DIFFERENT VALUES OF *m* FOR OUR METHOD, USING DATASET 2

| *m* | B01 | B02 | B03 | B04 | B05 | B06 | B07 | B08 | B09 |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 86.81 | 74.31 | 75.69 | 99.32 | 91.22 | 86.49 | 80.41 | 89.47 | 92.76 |
| **2** | 89.47 | 75.66 | 70.39 | 100 | 92.76 | 84.87 | 81.58 | 91.45 | 94.08 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **3** | 81.94 | 75.69 | 70.83 | 100 | 91.22 | 87.84 | 83.78 | 94.08 | 95.39 |
| *STD* | 3.81 | 0.78 | 2.94 | 0.39 | 0.88 | 1.48 | 1.71 | 2.31 | 1.31 |

### *C. Comparison of computational time*

Finally, to demonstrate that the new method is useful for BCI implementation, we tested its computational efficiency. As BCIs require not only high decoding accuracy but also relatively fast parsing of the intended motion, a speedy classification is important. We used a 2.67GHz i5-M480 processor with 4Gb RAM for the analyses.

Runtime profiling of the algorithm took 366.91±51.29 seconds for the entire assessment. While this is nearly 2.8 orders of magnitude longer than the contender algorithm (SBS-FBCSP) which took only 8.05±3.02 seconds) this test compares both the feature selection/validation and classification. As the feature selection is only required for the model training, a comparison of the online classification alone shows that our algorithm is on par with competing algorithms that report their computational efficiency[11]. Namely, it is within 3 seconds from the SBS-FBCSP algorithm (*n.s.*). Together with the improved classification accuracy, we argue, the sacrifice in computational efficiency still renders our method ideal for BCI applications.

## IV. Discussion

### *A. Performance comparison with other methods*

We compare the performance of a novel neural decoding algorithm, which use both temporal and spectral EEG signals for MI. Our algorithm shows increased accuracy of 3.32-8.60% above benchmark algorithms (**tables 01-02**). Investigating individual cases (rather than solely the aggregate performance) shows that the algorithm outperforms all benchmark comparisons across 8 out of 18 individuals tested. The highest improvement above the average benchmark for a single individual reaches 25%. Comparing our method to an algorithm that uses similar routines (SBS-FBCSP) shows an average increased performance of 11% (8.60% for dataset 1, and 13.42% for dataset 2). As the main difference in our method and the SBS-FBCSP algorithm is the addition of temporal signals to the feature set, we suggest that these features capture information that controls for the variance within trials of a single individual and therefore increase the performance.

Given that our method relies on the choice of a parameter, $m$, we tested the algorithm's robustness to the parameter selection and show that the results remain consistent (**tables 03-04)**.

### *B. Prior works*

Our method is not the first to consider the temporal structure of EEG signals during MI task classification. For example, previous work[15] has combined temporally constrained group LASSO with Convolutional Neural Network aiming at interpretating the mechanism of the EEGNet[28] model. Similarly, a framework for time frequency CSP smoothing was recently implemented to improve EEG decoding performance through ensemble learning[29]. Both those methods focused on selecting the features of the CSP by ranked weight. Our method both incorporates the temporal features, as well as uses a neural network based feature selection strategies for the classification.

A different approach that showed highly promising results in MI classification focused on the selection of the relevant channels subset from a single individual's data. This method – Spatio-Temporal filtering-based Channel Selection (STECS) proves superior to state-of-the-art methods that use the entire EEG dataset with about half the channel count. In BCIs that rely on high channel numbers this method is likely to outperform many of the methods we evaluated here[30].

Neural networks classifiers were suggested to outperform SVM classifiers[12] and, together, with the additional features seem to drive the notable performance increase.

Additionally, as our method separated the feature selection process from the following classification task, we suggest that this two-stage process, which enabled the reduction of features number, contributed to the performance increase. Indeed, recent work using EEG-based BCI — in the domain of emotional memory recall — used similar two-stage process to, first, determine a suitable features set and train a model, and following perform the real-time classification, and has shown remarkable results[31].

### *C. Limitations*

The proposed MI decoding method suffers from a number of limitations that are driven by the addition of the temporal components. First, the method requires a-priory intuition about the data in order to accurately choose the EEG temporal segments. To prove the method's superiority in datasets where no prior knowledge is available it would be useful to test either arbitrary datasets, or randomly selected temporal windows. If the method proves superior even with those selections, it will be regarded more robust.

Second, our method is orders of magnitude slower in its initial computation time. This means that usage of the method for BCIs that continuously update the feature set would either be challenging or require extensive computational resources. To overcome this challenge, one should investigate whether smaller time-window sizes (presumably yielding faster processing) could produce higher performance. Shorter time-window that maintain the high performance would elevate the usefulness of the algorithm.

Third, it is not clear whether the method would easily generalize to BCI tasks outside of MI ones. Specifically, because MI tasks are less likely to show the types of noise that pollutes active movement tasks, the fact that our method shows superiority in one domain does not guarantee its success in others.

### *D. Future directions*

Accordingly, two research venues that directly extend our work are suggested: i) enhancing the features selection granularity (while attempting to maintain the feature-classification performance), and ii) generalizing the temporal features classification process. Specifically, as EEG and other biological signals are heavily dependent on temporal dynamics, usage of feature selection process with tools such as the recently proposed attention guided neural networks[32] may improve the ability to extract the appropriate features without a-priori knowledge on the data. This would make the algorithm generalizable to other BCI inputs.

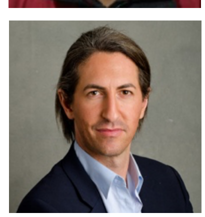
Further, as the majority of the benchmark algorithms we compared used neural networks for the full classification process (thereby effectively using all the available features without pre-selection) we suggest that amending the benchmark algorithms to incorporate the two-step selection-classification process may increase the performance of all the benchmark methods.

It has not escaped our notice that as SVMs were previously shown to be superior with respect to feature classification (whereas deep learning networks were shown to be superior in BCI feature selection[15,23,24,27]) a combination of both methods might improve our algorithm further and allow it to generalize to tasks outside of motor imagination or control (i.e., non-verbal communication, language decoding, or parsing of thoughts).

## References

[1] S. Bulárka and A. Gontean, "Brain-computer interface review," in 2016 12th IEEE International Symposium on Electronics and Telecommunications (ISETC), 2016, pp. 219–222.
[2] G. Courtine, S. Micera, J. DiGiovanna, and J. del R Millán, "Brain–machine interface: closer to therapeutic reality?," Lancet, vol. 381, no. 9866, pp. 515–517, 2013.
[3] L. R. Hochberg et al., "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," Nature, vol. 485, no. 7398, pp. 372–375, 2012.
[4] R. A. Andersen, T. Aflalo, and S. Kellis, "From thought to action: The brain–machine interface in posterior parietal cortex," Proc. Natl. Acad. Sci., vol. 116, no. 52, pp. 26274–26279, 2019.
[5] M. Cerf et al., "On-line, voluntary control of human temporal lobe neurons," Nature, vol. 467, no. 7319, pp. 1104–1108, 2010, doi: 10.1038/nature09510.
[6] A. Chiuzbaian, J. Jakobsen, and S. Puthusserypady, "Mind controlled drone: An innovative multiclass SSVEP based brain computer interface," in 2019 7th International Winter Conference on Brain-Computer Interface (BCI), 2019, pp. 1–5.
[7] M. Nader, I. Jacyna-Golda, S. Nader, and K. Nehring, "Using BCI and EEG to process and analyze driver's brain activity signals during VR simulation," Transport, vol. 60, no. 4, pp. 137–153.
[8] P. Kotler, Consumer neuroscience. MIT Press, 2017.
[9] S. Y. Gordleeva et al., "Exoskeleton control system based on motor-imaginary brain–computer interface," Mod. Technol. Med., vol. 9, no. 3 (eng), 2017.
[10] S. Sakhavi, C. Guan, and S. Yan, "Learning temporal information for brain-computer interface using convolutional neural networks," IEEE Trans. neural networks Learn. Syst., vol. 29, no. 11, pp. 5619–5629, 2018.
[11] 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 Trans. Cybern., vol. 49, no. 9, pp. 3322–3332, 2018.
[12] M. H. Bhatti et al., "Soft computing-based EEG classification by optimal feature selection and neural networks," IEEE Trans. Ind. Informatics, vol. 15, no. 10, pp. 5747–5754, 2019.
[13] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in 2007 3rd International IEEE/EMBS Conference on Neural Engineering, 2007, pp. 204–207.
[14] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), 2008, pp. 2390–2397.
[15] X. Deng, B. Zhang, N. Yu, K. Liu, and K. Sun, "Advanced TSGL-EEGNet for Motor Imagery EEG-Based Brain-Computer Interfaces," IEEE Access, vol. 9, pp. 25118–25130, 2021.
[16] R. Leeb, C. Brunner, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI Competition 2008–Graz data set B," Graz Univ. Technol. Austria, pp. 1–6, 2008.
[17] A. Pasyuk, E. Semenov, and D. Tyuhtyaev, "Feature Selection in the Classification of Network Traffic Flows," in 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon), 2019, pp. 1–5.
[18] K. Belwafi, O. Romain, S. Gannouni, F. Ghaffari, R. Djemal, and B. Ouni, "An embedded implementation based on adaptive filter bank for brain–computer interface systems," J. Neurosci. Methods, vol. 305, pp. 1–16, 2018.
[19] 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 Syst. Appl., vol. 95, pp. 201–211, 2018.
[20] A. Singh, S. Lal, and H. W. Guesgen, "Small sample motor imagery classification using regularized Riemannian features," IEEE Access, vol. 7, pp. 46858–46869, 2019.
[21] 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.
[22] Q. Zheng, F. Zhu, and P.-A. Heng, "Robust support matrix machine for single trial EEG classification," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 26, no. 3, pp. 551–562, 2018.
[23] X. Hong et al., "Dynamic Joint Domain Adaptation Network for Motor Imagery Classification," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 29, pp. 556–565, 2021.
[24] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," IEEE Trans. neural Syst. Rehabil. Eng., vol. 25, no. 6, pp. 566–576, 2016.
[25] J. Luo, Z. Feng, J. Zhang, and N. Lu, "Dynamic frequency feature selection based approach for classification of motor imageries," Comput. Biol. Med., vol. 75, pp. 45–53, 2016.
[26] J. Luo, Z. Feng, and N. Lu, "Spatio-temporal discrepancy feature for classification of motor imageries," Biomed. Signal Process. Control, vol. 47, pp. 137–144, 2019.
[27] L. Yang, Y. Song, K. Ma, and L. Xie, "Motor imagery EEG decoding method based on a discriminative feature learning strategy," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 29, pp. 368–379, 2021.
[28] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," J. Neural Eng., vol. 15, no. 5, p. 56013, 2018.
[29] Y. Miao et al., "Learning Common Time-Frequency-Spatial Patterns for Motor Imagery Classification," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 29, pp. 699–707, 2021.
[30] Q. Feifei, W. Wu, Z. Liang Yu, Z. Gu, Z. Wen, T. Yu, and Y. Li. "Spatiotemporal-filtering-based channel selection for single-trial EEG classification." IEEE Trans. Cybern., vol. 51, no. 2, pp/ 558-567, 2020. D. Iacoviello, A. Petracca, M. Spezialetti, and G. Placidi, "A Classification Algorithm for Electroencephalography Signals by Self-Induced Emotional Stimuli," IEEE Trans. CYbern., vol. 46, no. 12, pp. 3171-3180, 2016.
[31] L. Sun, W. Shao, D. Zhang, and M. Liu, "Anatomical attention guided deep networks for ROI segmentation of brain MR images," IEEE Trans. Med. Imaging, vol. 39, no. 6, pp. 2000–2012, 2019.

**Gan Wang** is a graduate student at Soochow University. His research interests include machine learning, deep learning, signal processing for biomedical engineering and applications of EEG data analyses to neural problems.



**Moran Cerf** (PhD) is a professor of neuroscience and business at Northwestern University. He received a PhD in computational neuroscience at the California Institute of Technology, Pasadena, CA (2009), an MA in philosophy (2001) and a BSc in physics (2000) at the Tel-Aviv University, Tel-Aviv, Israel. Prof. Cerf holds numerous patents, and over 70 academic papers. His work was featured in popular outlets such as CNN, BBC, Times, Forbes and hundred others. Much of his work was made public through his popular

talks at large-scale events such as TED, TEDx, The World Economic Forum and numerous others, where he garnered millions of followers. Recently, he was named one of the "40 leading professors under 40".

Prior to his academic career, Prof. Cerf worked in the cybersecurity industry and was on the board of leading tech companies.