A graph-theoretic sensor-selection scheme for covariance-based Motor Imagery (MI) decoding

Kostas Georgiadis
AIIA lab, Informatics dept., AUTH
Information Technologies Institute (ITI), CERTH
Thessaloniki, Greece
georgiaki@csd.auth.gr

Dimitrios A. Adamos
Department of Computing, Imperial College London, London, UK
NeuroInformatics.GRoup, AUTH
School of Music Studies, AUTH
Thessaloniki, Greece
d.adamos@ieee.org

Spiros Nikolopoulos
Information Technologies Institute (ITI), CERTH
Thessaloniki, Greece
nikolopo@iti.gr

Nikos Laskaris
AIIA lab, Informatics dept., AUTH
NeuroInformatics.GRoup, AUTH
Thessaloniki, Greece
laskaris@csd.auth.gr

Ioannis Kompatsiaris
Information Technologies Institute (ITI), CERTH
Thessaloniki, Greece
ikom@iti.gr

***Abstract*— Optimal sensor selection is an issue of paramount importance in brain decoding. When associated with estimates of covariance, its implications concern not only classification accuracy, but also computational efficiency. However, very few attempts have been made so far, since it constitutes a challenging mathematical problem. Herein, we propose an efficient heuristic scheme that combines discriminative learning (from a small training dataset of labelled trials) with unsupervised learning (the automated detection of sensors that collectively maximize the trial discriminability of the induced Covariance structure). The approach is motivated from a complex network modelling perspective. Its efficacy and efficiency are demonstrated experimentally, based on BCI-competition datasets concerning MI-tasks, and compared against popular techniques in the field.**

Keywords—discriminative learning, complex network modelling, graph clustering.

I. INTRODUCTION

Brain Computer Interfaces (BCIs) constitute a promising neuroengineering technology as they provide an alternative communication pathway, by converting brain signals to machine commands [1]. BCIs can be operated solely by brain commands and do not require any peripheral nerve activity. Hence, they were originally opted to assist people with partial or complete loss of their fine motor skills by increasing the individuals' independence or aiding in their rehabilitation process [2]. Besides their initial motivation, BCIs have also found application in other activities such as gaming or mental workload monitoring [3], [4]. In respect, electroencephalography (EEG) appears as one of the most popular options for registering the brain activity due to its low cost, non-invasiveness, high-temporal resolution and easy adaptation to non-clinical settings.

BCIs built upon motor imagery (MI) activity receive increasing attention, since they can operate in asynchronous mode (i.e. as self-paced BCIs). They are based on changes in sensorimotor rhythms (SMR) that become detectable when someone either plans or executes a movement. Originally, these brain activity modifications were sought in the form of an initial power decrease in μ-band (desynchronization) followed by a power increase in β-band (synchronization) once the MI task was completed [5]. An alternative, well-established, MI decoding scheme is the data-learning technique of common spatial filters (CSP) [6] that aims in the maximal discrimination between two classes of brain activity (e.g. imagination of hand vs. foot movement). More recent approaches include decoders based on descriptors form Graph Signal Processing domain [7], [8], Riemannian geometry [9] and the field of complex networks [10].

Almost all the aforementioned decoding approaches require several sensors to perform sufficiently, which hinders the ability of a BCI to operate promptly. Aiming to tackle this limitation, several research groups have employed different strategies in order to reduce the required number of sensors while preserving high classification performance. The most prominent sensor selection approaches encompass the fundamentals of the CSP method using either spatial (e.g. [11] and [12]) or spatiotemporal filters (e.g. [13]-[15]). Despite the abundance of sensor selection approaches, the identification of an optimal sensor set regarding the registered covariance structure has received only very limited attention so far. To the best of our knowledge, the problem has been addressed in [16] for a more general BCI-setting and in [17-18] for MI-decoding.

We herein introduce a methodology for crafting a robust and parsimonious covariance-based MI decoder from an extended array of sensors. It is a hybrid-learning scheme that combines discriminative learning (an initial feature ranking step) with unsupervised learning (the detection of most informative subarray of sensors). It relies on the complex network modelling of brain signals recorded from the whole sensor array and exploits the graph-clustering technique of Dominant-Set [19] for detecting the subset of sensors with covariance structure best discriminating between MI-tasks (see Fig.1). Considering that a given covariance matrix encapsulates the pairwise relations among all variates (in particular, a spatial covariance matrix incorporates the functional covariation between signals recorded at distinct sites), it can be thought as representing a fully connected graph. Hence, sensor selection can be cast as a subgraph detection problem. Following this line of thought, we commence by building a connectivity graph that incorporates all available sensors as nodes and has a structure that directly reflects the ability of recording sites, taken in pairs, to discriminate between MI-tasks. At this first stage, a small-set of training trials is necessary for estimating the class-separability of each entry in the full covariance matrix. Next, the Dominant-Set algorithm is invoked in order to detect the

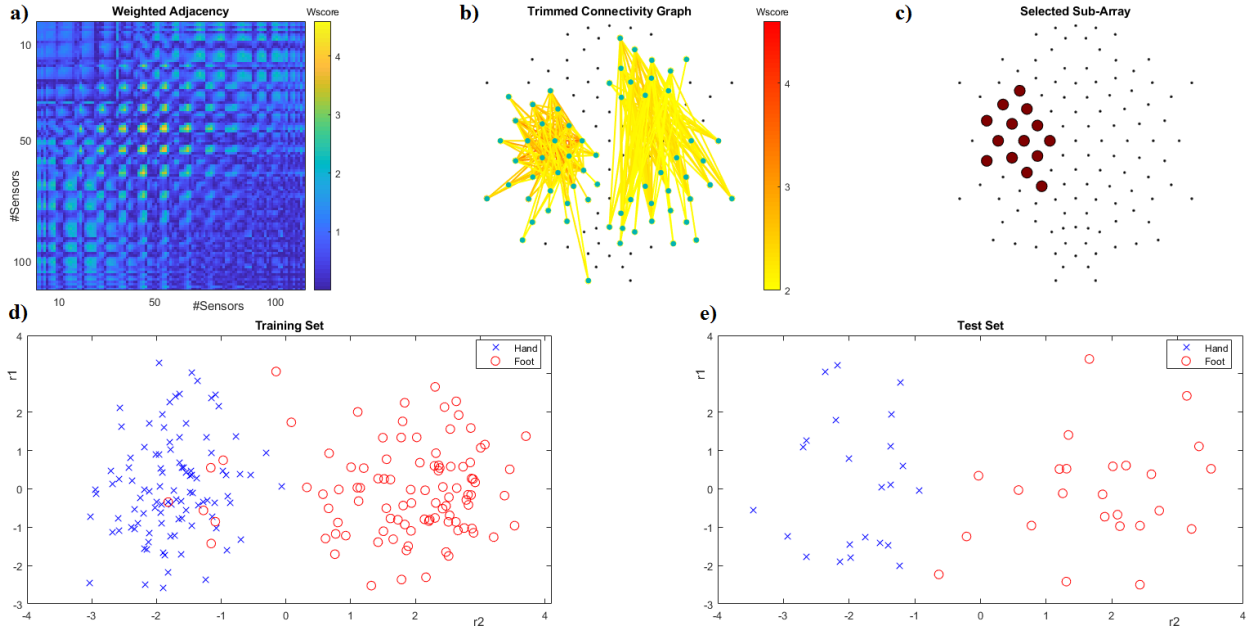

Figure 1 The Dominant-Set detection approach to sensor selection, exemplified for A2 subject. a) The Weighted Adjacency matrix derived from the trial-based covariances in the training set: each (i,j) entry reflects the Wilcoxon score from contrasting the corresponding cov(x_i,x_j) between MI-tasks. b) The associated connectivity graph, drawn over the sensor-array, after a simple sparsification step (during which only edges with Wscore>2 have been kept). c) The Dominant-Set of nodes detected within the initial, fully-connected, graph. d) MDS embedding of the training trials based on DomSetCov representation. e) The corresponding MDS map of test trials.

most compact subgraph among the emerged network of relations. Since edge-strength encodes the covariates' discriminability, the delineation of the Dominant-Set is equivalent with mining a subset of sensors with highly informative covariance structure. Finally, the MI-decoding is realized, based on the reduced covariance matrices by means of a support vector machine (SVM) operating within an appropriate Riemannian geometry framework [9].

The proposed decoding scheme, denoted hereafter as SVM+$^{\text{DomSet}}$Cov is validated using trials of electroencephalographic activity arising from two distinct mental tasks, provided within the two datasets of BCI competition III and its performance is directly compared against popular alternative schemes.

II. Methodology

A. *A graph-representation of discriminability based on estimates of functional covariation.*

Sample covariance matrices (SCMs) are known to provide information-rich descriptions of multichannel EEG signals, sufficient for the discrimination between specific mental tasks [21]. For the clarity of the presentation, we consider here a two-class scenario and that a training set of representative trials is available along with the associated class labels. The SCM for a given single trial $X_i \in \mathbb{R}^{S\times T}, i = 1,2,\dots N_{trials}$, with S and T denoting the number of sensors and samples in time respectively, is calculated as $\boldsymbol{C}_i = X_i X_i^T/(T-1)$ (or through an equivalent regularized estimator). This simple computation leads to an $S \times S$ representation for each trial that will be wrapped in one of the two sets of covariance matrices (i.e. $\boldsymbol{COV}^A$ or $\boldsymbol{COV}^B$), according to the trial's label $y_i \in \{0,1\}$. The covariation patterns, conveyed by these trial-based SCMs, are treated as feature-vectors and a *feature-ranking* algorithm is employed to evaluate the separability among the two brain-states (mental tasks) regarding each entry in the formed SCMs. In particular, the non-parametric Wilcoxon test is applied $S(S-1)/2$ times along every dimension in the vectorized instantiations of the SCMs. The obtained results are reshaped into a single symmetric $S \times S$ matrix $\mathbf{W}$, directly reflecting the discriminability of the entries in the original (full-sized) SCM representation. This step is exemplified via the visualization in Fig.1a, where the matrix of feature-ranking scores has been color-coded (high W_{score} corresponds to high separability), providing some evidence about the block-structure of the underlying measurements. Adopting, the perspective of complex network modelling, the matrix $\mathbf{W}$ is treated as the adjacency matrix of a graph that involves all sensors and characterized via a connectivity that reflects the covariations facilitating mental-task discrimination. This concept is further illustrated is Fig.1b, where the connectivity pattern has been sparsified via a simple edge-filtration step.

B. *Sensor Selection by means of Dominant Set detection*

With the purpose of identifying a small-sized (sub)array of sensors that is characterized by the most discriminant covariance structure, we scrutinize the connectivity pattern in $\mathbf{W}$ by means of Dominant-Set algorithm. This is an efficient graph-clustering algorithm, that operates via simple operations on the weighted-adjacency matrix of a given graph and isolates the most cohesive subgraph.

Formally, the problem is formulated as maximizing the following objective function of cohesiveness:

$$F(Z) = Z^T \boldsymbol{W} Z \qquad (1)$$

subject to $Z \in \Delta$, where $\Delta = \{Z \in \mathbb{R}^S : z_i \geq 0, \forall i \text{ and } \sum_{i=1}^{S} z_i = 1\}$. The solution is based on the approach of replicator dynamics, as it is described in [19] and implemented in the Matlab-code provided in [20]. The

algorithm converges fast to a vector Z* that contains the memberships of nodes in the Dominant-Set. In our case, where **W** contains the discriminability of the SCM-entries, the detected subgraph can be thought as the functional module of the underlying complex brain network that is expected to result in a reduced in size but highly discriminative covariance pattern. Returning to the previous example, we illustrate in Fig.1c topographically the solution that corresponds to a cohesiveness-level F(Z*) of 2.62, when the cohesiveness-level of the overall graph is only 0.77.

C. Classification and Validation

The representation of trials based on the parsimonious description of covariance, as constrained within the selected (sub)array of sensors, is by definition symmetric positive definite (SPD) considering that the post-stimulus activity is lengthy enough to ensure the full rank property of the covariance matrix. According to differential geometry, SPD matrices are not embedded in a vector space, but reside on a Riemannian manifold, denoted as $Sym^{+}_{S_{RED}}$, that defines a hypercone in the Euclidean space [9]. The inner product between two points, $\boldsymbol{A} = \boldsymbol{C_i}$ and $\boldsymbol{B} = \boldsymbol{C_j}$, laying on the tangent space of $Sym^{+}_{S_{RED}}$ at point $\boldsymbol{P}$ can be readily estimated using the Affine Invariant Riemannian Metric (AIRM) [22] as follows:

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle_{\boldsymbol{P}} \triangleq Trace(\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}^{-1}\boldsymbol{B}) \quad (3)$$

Since the estimation of the inner product between two points is ensured, SVMs that have proved to provide robust results regarding the classification of brain activity [23], can be adopted for carrying out the discrimination between the two mental tasks based on the reduced covariance representation. A typical choice of $\boldsymbol{P}$ for this task, is the *geodesic mean* of covariance matrices of the training set.

III. RESULTS

A. Dataset Description and Performance Evaluation

For the experiments conducted in this study, two publicly available datasets from previous BCI competitions consisting of EEG recordings during MI tasks were used. The first one was the Dataset IVa of BCI competition III [24], denoted hereafter as *Dataset IVa*, and consisted of the electroencephalographic activity of five subjects (coded as A1, A2, …, A5) performing movement imagination of either their right hand or their foot. This dataset accommodates the issue of sensor selection since EEG is registered with numerous sensors (i.e. 118) at a sampling frequency of 100 Hz. All subjects participated in a total of 280 MI trials, equally distributed among the two classes, with duration of 3.5 sec. The train/test set split was different among subjects, with the training set encompassing 168, 224, 84, 56 and 28 trials for subjects A1, A2, A3, A4 and A5 respectively, with the remainder of the trials comprising the test set.

An additional dataset from the same competition, denoted as *Dataset IIIa* [25], was also considered for the further evaluation of our approach. In the recordings of this dataset, the task for the three participants, denoted as B1, B2 and B3, was to perform left hand, right hand, foot or tongue MI tasks and their brain activity was registered via 60 sensors with a sampling frequency of 250 Hz. In our study only trials concerning the left and right hand MI tasks were selected, resulting in a total number of trials equal to 180 (for the first subject) and to 120 (for the second and third subject). Trials were equally distributed among both train/test set and the two classes.

Since both datasets had a predefined sequence of train/test trials by the organizers of the BCI competition, the validation scheme for this study was tailored accordingly. This was dictated by the need to cast the performance of our decoding scheme in a way to be directly comparable with the results reported in previous works.

B. ^{DomSet}Cov-based decoding of MI activity

Working in a personalized fashion, each trial of both train and test set was first band-pass filtered in the 8-30Hz and SCMs were then derived based on the post-stimulus brain activity from a 2-second interval (starting 0.5 second after the visual trigger for movement initiation). Finally, the Dominant-Set of sensors was identified for each subject independently, based on the trials in the training set.

The discriminability of the derived representation and its generalizability are demonstrated, respectively, via Fig.1d and Fig.1e, where the training and test trials of subject A2 have been visualized as 2D point-clouds based on the technique of multidimensional scaling (MDS). MDS is a distance preserving visualization that acted on the geodesic distances between all the reduced SCMs in the ensemble of training(test)set and resulted to a scatterplot reflecting the geometrical relationships on the Riemannian manifold. The inter-covariance distance was induced by (3) as $\delta(\boldsymbol{C}_i, \boldsymbol{C}_j) = \left\| logm(\boldsymbol{C}_i^{-1/2}\boldsymbol{C}_j\boldsymbol{C}_i^{-1/2}) \right\|_F$, where $logm(.)$ is the log-matrix operator and $\|.\|_F$ denotes the Frobenius norm of the matrix [17]. The clear separation seen in the MDS space aligns well with the high classification performance seen for this subject in Table I.

Table I refers to *Dataset IVa* and provides a direct comparison, in terms of classification accuracy among the introduced decoder (i.e. SVM+^{DomSet}Cov), the SVM decoder that worked on the full-sized covariance matrix, the standard CSP technique, two popular CSP alternatives for sensor selection (WTRCSP and SRCSP [11]) and a recently introduced sensor selection scheme that uses Riemannian distance [17], denoted hereafter as *Riedist*. It is evident that the proposed decoder significantly outperforms the CSP approach, providing an improvement of roughly 15%. Additionally, SVM+^{DomSet}Cov demonstrates higher classification performance compared to the CSP-based sensor selection algorithms, with a 6% improvement relative to the WTRCSP technique and a 3% improvement relative to the SRCSP technique. Moreover, the ^{DomSet}Cov representation compares favorably against both the full-sized SCM and the *Riedist* technique, with a 12% and 4% performance improvement respectively.

Finally, it is important to note here that the decoder's performance is being reached with a Dominant-Set consisting of 15 (out of 118) sensors on average, which is equivalent to gathering information from 1 sensor out of 10 (i.e. 13%). Also, the selection (an example can be seen in Fig.1c) can be characterized as neurophysiologically meaningful, since the selected sensors lie over the sensorimotor area that is highly activated when a mental task is performed. This proves not only the effectiveness of our approach but also indicates its low computational cost, an essential feature for all asynchronous BCIs.

TABLE I. CLASSIFICATION ACCURACY (%) IN THE "HAND VS FOOT" TASK

Subj ID	Dataset IVa					
	^{DomSet}Cov (*#sensors*)	*Cov*	*CSP*	*WTRCSP [11]*	*SRCSP [11]*	*Riedist [17]*
A1	**74.1** *(13)*	68.8	66.1	69.6	72.3	74.1
A2	**98.2** *(10)*	96.4	96.4	**98.2**	96.4	**98.2**
A3	**68.4***(15)*	54.6	47.5	54.6	60.2	59.2
A4	**80.4** *(18)*	75.4	71.9	71.9	77.7	77.7
A5	**89.3** *(19)*	53.6	49.6	86.9	86.5	80.6
aver.	**82.1** *(15)*	69.8	66.3	76.2	78.6	78.0

Similar trends were observed in *Dataset IIIa* (refer to Table II), where an average improvement between 3% and 9% is achieved, with the Dominant-Set being comprised of 10 (out of 60) sensors on average. It is important to note here that a comparison between the proposed decoder and *Riedist* technique was not made for this dataset, since the results provided by the authors in [17] concern only *Dataset IVa*.

IV. DISCUSSION

A graph-theoretic algorithm for crafting robust and parsimonious covariance-based MI decoders for multichannel EEG signals has been presented. Our approach was introduced in the setting of binary classification. However, its modification to accommodate more classes is, in principle, feasible and remains to be thoroughly investigated, with one solution being a "one vs all" approach. Among the most prominent advantages are the fast execution during the learning stage (sensor selection is accomplished swiftly) and the lower computational cost (with respect to the full-sized covariances) of the decoder during operation. An additional, less obvious, advantage is the robustness of the approach to both bad-sensors and intense artifact contamination at particular sites (like blinks over frontal regions). The discriminative learning during the training stage ensures that such sensors are excluded from the Dominant-Set. A future extension of this work will focus on crafting decoders for self-paced BCIs. For instance, a two-stage decision system can be built as proposed in our previous work [26]. More specifically, such a system will include a "brain switch" based on the covariance structure enhancing the separability between rest and MI-segments followed by an MI-classifier based on the covariance structure enhancing the discrimination among distinct tasks.

ACKNOWLEDGMENT

This work is part of project MAMEM that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644780 and project Evision that has been co‐financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code:T1EDK-02454).

TABLE II. CLASSIFICATION ACCURACY (%) IN THE "LEFT HAND VS RIGHT HAND" TASK

Subj ID	Dataset IIIa				
	^{DomSet}Cov (*#sensors*)	*Cov*	*CSP*	*WTRCSP [11]*	*SRCSP [11]*
B1	**98.9** (*11*)	95.6	95.6	98.9	96.9
B2	**75.0** *(11)*	63.3	61.7	71.7	53.3
B3	**96.7** *(9)*	93.3	93.4	93.3	93.3
aver.	**90.2** (*10.3*)	84.1	83.5	87.3	81.1

REFERENCES

[1] M.A. Lebedev, and M.A. Nicolelis, "Brain-machine interfaces: From basic science to neuroprostheses and neurorehabilitation." Physiological reviews 97, no. 2, pp. 767-837, Mar 2017.

[2] S. Moghimi, A. Kushki, A.M. Guerguerian, and T. Chau, "A review of EEG-based brain-computer interfaces as access pathways for individuals with severe disabilities." Assistive Technology 25, no. 2, pp. 99-110, 2013.

[3] H, Gurkok, A. Nijholt, and M. Poel, "Brain-computer interface games: Towards a framework." Handbook of Digital Games and Entertainment Technologies, pp. 133-150, 2017.

[4] M.V. Kosti, K. Georgiadis, D.A. Adamos, N. Laskaris, D. Spinellis, and L. Angelis, "Towards an affordable brain computer interface for the assessment of programmers' mental workload." International Journal of Human-Computer Studies 115, pp. 52-66, July 2018.

[5] G. Pfurtscheller, and F.H. Lopes Da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles." Clinical neurophysiology 110, no. 11, pp. 1842-1857, Nov 1999.

[6] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement." IEEE transactions on rehabilitation engineering 8, no. 4, pp. 441-446, Dec 2000.

[7] K. Georgiadis, N. Laskaris, S. Nikolopoulos, and I. Kompatsiaris. "Connectivity steered graph Fourier transform for motor imagery BCI decoding." Journal of neural engineering 16, no. 5 (2019): 056021.

[8] K. Georgiadis, N. Laskaris, S. Nikolopoulos, D.A. Adamos, and I. Kompatsiaris. "Using Discriminative Lasso to Detect a Graph Fourier Transform (GFT) Subspace for robust decoding in Motor Imagery BCI." In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 6167-6171. IEEE, 2019.

[9] F. Yger, M. Berar, and F. Lotte. "Riemannian approaches in brain-computer interfaces: a review." IEEE Transactions on Neural Systems and Rehabilitation Engineering 25, no. 10 (2016): 1753-1762.

[10] F.D.V. Fallani, and D.S. Bassett, "Network neuroscience for optimizing brain-computer interfaces." Physics of life reviews, Jan 2019.

[11] F. Lotte, and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms." IEEE Transactions on biomedical Engineering 58, no. 2, pp. 355-362, Feb 2011.

[12] B. Yang, H. Li, Q. Wang, and Y. Zhang. "Subject-based feature extraction by using fisher WPD-CSP in brain–computer interfaces." Computer methods and programs in biomedicine 129 (2016): 21-28.

[13] Y. Zhang, C.S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI." IEEE transactions on cybernetics 99, pp 1-11, June 2018.

[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), pp. 2390-2397. IEEE, 2008.

[15] F. Qi, W. Wu, Z. L. Yu, Z. Gu, Z. Wen, T. Yu, and Y. Li. "Spatiotemporal-Filtering-Based Channel Selection for Single-Trial EEG Classification." IEEE Transactions on Cybernetics (2020).

[16] F. Kalaganis, N. Laskaris, E. Chatzilari, S. Nikolopoulos, and I. Kompatsiaris. "A Riemannian geometry approach to reduced and

discriminative covariance estimation in Brain Computer Interfaces." IEEE Transactions on Biomedical Engineering (2019).

[17] A. Barachant and S. Bonnet. "Channel selection procedure using Riemannian distance for BCI applications." In 2011 5th International IEEE/EMBS Conference on Neural Engineering, pp. 348-351. IEEE, 2011.

[18] K. Sadatnejad, A. Roc, L. Pillette, A. Appriou, T. Monseigne, and F. Lotte. " Channel Selection over Riemannian Manifold with Non-Stationarity Consideration for Brain-Computer Interface Applications." In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1364-1368. IEEE, 2020.

[19] M. Pavan, and M. Pelillo. "Dominant sets and pairwise clustering." IEEE transactions on pattern analysis and machine intelligence 29, no. 1 (2006): 167-172.

[20] D. A. Adamos, N. A. Laskaris, and S. Micheloyannis. "Harnessing functional segregation across brain rhythms as a means to detect EEG oscillatory multiplexing during music listening." Journal of neural engineering 15, no. 3 (2018): 036012.

[21] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller. "Optimizing spatial filters for robust EEG single-trial analysis." IEEE Signal processing magazine 25, no. 1 (2007): 41-56.

[22] X. Pennec, P. Fillard, and N. Ayache. "A Riemannian framework for tensor computing." International Journal of computer vision 66, no. 1 (2006): 41-66.

[23] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update." Journal of neural engineering 15, no. 3, 031005, April 2018.

[24] B. Blankertz, K.R. Muller, D.J. Krusienski, G. Schalk, J.R. Wolpaw, A. Schlogl, G. Pfurtscheller, J.R. Millan, M. Schroder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems." IEEE transactions on neural systems and rehabilitation engineering 14, no. 2, pp. 153-159, June 2006.

[25] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller. "Characterization of four-class motor imagery EEG data for the BCI-competition 2005." Journal of neural engineering 2, no. 4 (2005): L14.

[26] K. Georgiadis, N. Laskaris, S. Nikolopoulos, and I. Kompatsiaris, "Exploiting the heightened phase synchrony in patients with neuromuscular disease for the establishment of efficient motor imagery BCIs." Journal of neuroengineering and rehabilitation 15, no. 1, 90, Oct 2018.