Original Article

A Dimensionality Reduction Approach for Motor Imagery Brain–Computer Interface Using Functional Clustering and Graph Signal Processing

Mohammad Davood Khalili, Vahid Abootalebi, Hamid Saeedi-Sourck

Department of Electrical Engineering, Yazd University, Yazd, Iran

Abstract

Background: This paper introduces an approach for dimensionality reduction and classification of electroencephalogram signals in motor imagery brain–computer interface (MI-BCI) systems. **Materials and Methods:** The proposed Kron-reduced generic learning regularization with differential evolution (K-GLR-DE) framework leverages graph signal processing (GSP) with a meta-heuristic optimizer, integrating functional clustering, Kron reduction, regularized common spatial patterns with generic learning (GLRCSP), and differential evolution (DE). Brain graphs are constructed within a structural–functional framework, where edge weights are defined based on geometric distances and correlations. Graph's dimensionality reduction is achieved by applying physiological regions of interest (ROIs) and Kron reduction to preserve essential topological-spectral features. Feature extraction is performed using graph total variation and GLRCSP, followed by DE-based feature selection. **Results:** The approach was evaluated on BCI Competition III Dataset IVa and the PhysioNet eegmmidb dataset. The support vector machine with a radial basis function (SVM-RBF) classifier achieved superior performance, yielding a mean accuracy of 96.46% ± 0.81% on BCIC III-IVa. **Conclusions:** The proposed K-GLR-DE method demonstrates significant performance in MI-BCI classification across various training conditions, including scenarios with small and limited training sets.

Keywords: *Brain-computer interface (BCI), electroencephalography (EEG), graph signal processing (GSP), Kron reduction, motor cortex*

Submitted: 20-Dec-2024 **Revised:** 09-May-2025 **Accepted:** 21-Jul-2025 **Published:** 02-Feb-2026

Introduction

A brain–computer interface (BCI) serves as a communication system that bypasses traditional output pathways of the brain, specifically peripheral nerves and muscles.[1] Fundamentally, a BCI empowers subjects to communicate and control using their brain signals.[2-4] Electroencephalogram (EEG) is the preferred choice for BCI applications due to its cost-effectiveness, non-invasiveness, high temporal resolution, and user-friendly nature compared to alternative methods.[5] EEG-based noninvasive BCI systems are typically organized into five primary categories: slow cortical potentials, motor imagery (MI), P300 event-related potential (P300 ERP), steady-state visual evoked potential, and error-related negative evoked potential.[6] While most BCI systems require signals from multiple locations across the scalp to optimize their performance,[7] excessive EEG channels may introduce redundancy and potentially undermine BCI performance.[8,9] Thus, strategically selecting an adequate number of channels that balance accuracy and practicality becomes crucial.[7]

In MI-BCI, subjects mentally simulate precise motor actions without physically executing the movements. Subsequently, the BCI system discerns the neural activity associated with this mental simulation and encodes the EEG signal. This process helps subjects acquire the neural patterns associated with specific movements, enabling them to learn the corresponding physical action more effectively.

Advances in data science have increased the interest in complex signals with irregular structures.[10] High-dimensional data can live on weighted graph vertices and show the geometric structure of data in various applications, such as social networks, energy systems, transportation, sensors, and

How to cite this article: Khalili MD, Abootalebi V, Saeedi-Sourck H. A dimensionality reduction approach for motor imagery brain–computer interface using functional clustering and graph signal processing. J Med Sign Sens 2026;16:6.

Address for correspondence:
Dr. Vahid Abootalebi, Department of Electrical Engineering, Yazd University, Yazd, Iran.
E-mail: abootalebi@yazd.ac.ir

Access this article online

Website: www.jmssjournal.net

DOI: 10.4103/jmss.jmss_88_24

Quick Response Code:

neural networks.[11] Graph signal processing (GSP) is an innovative tool for analyzing data on complex and irregular structures, with particular applications in brain signals.[12] GSP enhances signal processing through a graph-based perspective integrated with classical signal processing, providing enhanced insight into signal connectivity effects.[13] GSP encompasses methods such as graph Fourier transform (GFT),[12-14] graph shift,[12,13] graph filter,[11,13,15] and spectral graph wavelet transform[15-17] for processing graph signals.

Numerous studies within brain signal analysis have employed conventional machine learning techniques like principal component analysis (PCA), independent component analysis (ICA), and support vector machine (SVM). Recent studies have utilized graph-based frameworks and GSP perspectives, enhancing our understanding of brain phenomena and interconnectivity among brain regions. These investigations have provided insights into brain signal classification,[18,19] diagnosis of brain diseases,[20-23] BCI,[24-26] and graph frequency analysis of brain signals.[27]

In recent years, the field of BCI has seen numerous research efforts focused on reducing the dimensions of EEG data using GSP techniques. Tanaka *et al.*[24] pursued the dimensionality reduction of sample covariance matrices (SCM) in BCI applications. They created geometric graphs based on the electrode distribution on the head and studied structural connections using geographical distance measures. To estimate SCM accurately in a lower-dimensional space, they used a method called tangent space mapping (TSM) and reduced dimensions using GFT.[24] Kalantar *et al.*[25] built geometric graphs to study the brain's structural and functional connections. They combined geographical distance and Pearson correlation among EEG electrodes to derive edge weights for the BCI application. They employed GFT, TSM, and PCA for dimension reduction, feature extraction, and feature reduction, respectively.[25] In another study, Kalantar *et al.*[26] developed hybrid graphs for BCI applications. They performed graph clustering and selected the top two clusters using linear discriminant analysis (LDA). They extracted features using common spatial patterns (CSP) and then inputted these features into LDA, quadratic discriminant analysis (QDA), and logistic regression (LR) classifiers. In our prior work, we constructed structural–functional graphs similar to[25] but with different normalization and optimization coefficients.[28] We efficiently reduced the dimensionality of these graphs through a fusion of Kron reduction and GFT. We conducted statistical feature extraction using the Ledoit–Wolf shrinkage estimator and TSM on EEG data from superior vertices. We achieved dimension reduction of these features using PCA and an optimization technique known as differential evolution (DE).

Prior studies[24-26,28] have revealed various challenges that hinder improving BCI application outcomes. The primary contributions of this paper to MI-BCI are as follows:

1. Efficient dimension reduction of extensive EEG data within a complex brain graph framework
2. Excellent channel subset selection for high performance
3. Adaptive processing approach for varied dataset sizes, with adjustable parameters for improved decoding accuracy.

These contributions collectively advance the state-of-the-art in MI-BCI research, providing valuable insights. This paper addresses the contributions mentioned above and meticulously analyzes the corresponding results. We recommend a strategic fusion of Kron reduction and functional clustering in the first and second contributions. In this way, the dimensions of EEG data are adeptly reduced by leveraging the perspective of GSP. In addition, we employ the DE algorithm to discern and selectively employ superior features. We introduce an approach that seamlessly amalgamates Kron reduction and regularized CSP with generic learning (GLRCSP) for feature extraction and DE for feature selection in the third contribution. This methodological synthesis operates through diverse stages where we optimize and fine-tune parameters tailored to each subject, ensuring accuracy and efficacy in our analysis.

The rest of this paper is organized as follows: Section 2 provides an overview of the necessary background and preliminaries for our work, including the dataset description, the basic concepts and definitions of GSP and brain graph, and the existing methods of Kron reduction, RCSP, and DE. Section 3 describes the main contribution of this paper, which is our proposed method, Kron-reduced generic learning regularization with differential evolution (K-GLR-DE). Section 4 presents the experimental results of applying our framework to classify different cognitive states based on the MI-EEG data. The section also compares the performance of our framework with other state-of-the-art methods and discusses the advantages and limitations of our approach. Section 5 summarizes the key findings and contributions made in this paper. It also outlines potential avenues for further research in this field.

Materials and Methods

Dataset

The Intelligent Data Analysis Group[29] provided dataset IVa from BCI competition III. This publicly available dataset includes EEG signals recorded during MI tasks involving the right hand and right foot. EEG data were collected from five healthy subjects using 118 electrodes based on the extended international 10–20 system. The signals were filtered within the range of (0.05–200 Hz) and digitized at a sampling rate of 1000 Hz. The dataset comprises data from the initial four sessions conducted without any feedback. It encompasses two types of visual stimulation: (1) target

letters presented behind a fixation cross, potentially inducing slight target-correlated eye movements and (2) targets displayed as randomly moving objects, potentially causing target-uncorrelated eye movements. Visual cues were presented for 3.5 s during each trial. Each subject performed 280 MI trials involving either the right hand (R) or foot (F) MI tasks. The presentation of target cues was randomly interrupted by resting intervals lasting from 1.75 to 2.25 s. This strategic interruption prevents the formation of temporal dependencies, thereby ensuring the independence of each trial.[29,30]

This paper incorporates training trials for all subjects, utilizing three modes: limited (60 trials), small (100 trials), and conventional (200 trials). In addition, a specific training trial mode was formulated to align with the BCI Competition III-dataset IVa [as presented in Table 1] to ensure unbiased comparability with approaches employed in this competition.

Graph signal processing

A weighted graph G is represented as G = (V, E, **W**), where V, E, and **W** denote vertices, edges, and weighted adjacency matrix, respectively. The set V = {1,2,...,N} represents vertices, and E is a set of edges with tuples (i, j) that reflects the connectivity between the i^{th} and j^{th} nodes. We focus on connected and undirected graphs without self-loops or multiple edges. The Laplacian and normalized Laplacian matrices are defined as $\mathbf{L} = \mathbf{D}-\mathbf{W}$, and $\mathbf{L}_n = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$, respectively. Here, **W** is the weighted adjacency matrix, and **D** is the diagonal degree matrix with diagonal elements $[D]_{i,j} = \sum_{l=1}^{N}[W]_{i,l}$.

In the case of an undirected graph, **L** is a real symmetric positive semi-definite matrix with nonnegative eigenvalues $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_N$, and the corresponding real-valued orthonormal eigenvectors $\{u_i\}_{i=1}^{N}$. Therefore, **L** can be decomposed as $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{\Lambda}$ = diag $\{\lambda_l\}$ and the columns of **U** are $\{u_i\}_{i=1}^{N}$. In this context, $\mathbf{x} = [x(1),\cdots,x(N)]$ represents a graph signal, where sample x(m) is associated with the m^{th} node. Furthermore, GFT of a given signal **x** over an undirected graph is defined as $\hat{\mathbf{x}} = \mathbf{U}^T\mathbf{x}$.[11] Equivalently, the inverse GFT is represented as $\mathbf{x} = \mathbf{U}\hat{\mathbf{x}}$.

Total variation (TV) is a widely used graph measure that quantifies signal smoothness concerning the graph. $TV_G(\mathbf{x})$ represents the extent of signal changes within the graph, where smaller values indicate slower changes.[12] The TV of a graph signal **x** is defined as:[31]

$$TV_G(\mathbf{x}) = \frac{1}{\|\mathbf{x}\|_2^2}\left\|\mathbf{x} - \frac{1}{|\lambda_{max}|}\mathbf{W}\mathbf{x}\right\|_2^2. \quad (1)$$

Using (1), we calculate the sum of graph signal variations at each time instant for all time samples in all trials. Here, **x** represents the vector of values obtained from all vertices at a specific time, **W** denotes the weighted adjacency matrix, and λ_{max} corresponds to the largest-magnitude eigenvalue of **W**.

Table 1: Train and test trials in BCI competition III-dataset IVa

Subjects	*aa*	*al*	*av*	*aw*	*ay*
Train trials	168	224	84	56	28
Test trials	112	56	196	224	252

Brain graph

Feature extraction, dimension reduction, and classification of brain signals form fundamental aspects in the design of a BCI system. In this study, we employ a graph-based approach to process brain signals. This approach leverages the rich information of the brain graphs. The brain graphs are defined by vertices and weighted edges, which represent the interconnections among various brain regions. By applying GSP techniques to the signal on the brain graph, we extract relevant features and reduce data dimensionality. GSP-based approaches exhibit remarkable versatility in brain signal processing, as demonstrated by their diverse applications, including: (1) dimension reduction of EEG signals,[24-26,28] (2) inference of brain graph topology,[10,27] (3) spectral analysis of brain signals using wavelet transform,[17,32] (4) analysis of dynamic brain connectivity,[17,33] and (5) classification of ERPs.[27,34] By considering the intricate interplay between brain signals and the underlying graph, GSP enables the integration of physiological and anatomical information. This integration enhances our understanding of brain function and facilitates the development of more robust BCI systems. In this study, EEG signals are assigned to the vertices of the brain graph, contributing to the construction of a comprehensive, large-scale brain graph.

Kron reduction

Kron reduction is a highly notable graph simplification method in the vertex domain that has been applied in recent years for simplifying electrical circuits,[35] power systems,[35,36] and graph structures.[37] Given a weighted and undirected graph G = (V, E,**W**) with vertices V, if **L** represents the Laplacian matrix, and V_1 is a subset containing at least two vertices from V, the reduced graph by the Kron reduction can be defined as $G^{Kron} = \{V_1, E^{Kron}, \mathbf{W}^{Kron}\}$. Furthermore, based on[37] and Schur complement theorem,[38] considering V_1 as the selected vertices and its complement, V_1^c as the deleted vertices, the reduced Laplacian matrix obtained through Kron reduction and the Schur complement of the deleted vertices block from the original Laplacian matrix can be defined as:

$$\mathbf{L}^{Kron} = K(\mathbf{L}^{\text{Original}}, V_1) = \mathbf{L}_{V_1,V_1} - \mathbf{L}_{V_1,V_1^c}\mathbf{L}^{-1}_{V_1^c,V_1^c}\mathbf{L}_{V_1^c,V_1}. \quad (2)$$

The $K(\mathbf{L}^{\text{Original}}, V_1)$ represents the Kron reduction operator. The Kron reduction process eliminates the non-selected vertices V_1^c, preserving graph connectivity, spectral fidelity, and structural–topological characteristics over V_1. In (2),

$\mathbf{L}_{A,B}$ denotes the submatrix $|\mathbf{A}| \times |\mathbf{B}|$, encompassing all elements of matrix **L** with row indices in block **A** and column indices in block **B**. Moreover, the first term represents the Laplacian of the selected vertices, while the second term indicates the impact of the Laplacian of the deleted vertices on the Laplacian of the selected vertices. One of the key advantages of Kron reduction is its ability to utilize weights and information from deleted vertices to construct the final Laplacian matrix for the selected vertices. In line with the reduced Laplacian matrix obtained through Kron reduction, the weight of the new graph edges is defined as:[36,37]

$$\mathbf{W}_{ij}^{\text{Kron}} = \begin{cases} -\mathbf{L}_{ij}^{\text{Kron}}; i \neq j \\ 0; i = j \end{cases} \quad (3)$$

Common spatial patterns and regularized common spatial patterns

The CSP approach is a widely used technique for EEG feature extraction. This method discriminates EEG signals or features into two distinct patterns. Each pattern corresponds to a specific class. This discrimination is achieved by simultaneously diagonalizing two real symmetric matrices. The classes are distinguished based on their variance, with one exhibiting the highest and the other the lowest. Studies have demonstrated that spatial filters significantly enhance spatial resolution and increase signal-to-noise-ratio (SNR) in EEG signals.[39,40] The effectiveness of the CSP approach requires the availability of a clean and sufficient EEG training set for estimating the spatial covariance matrix for each class. Otherwise, the resulting covariance matrices may poorly represent the mental states and lead to ineffective spatial filters.[41] Regularized CSP (RCSP) methods have emerged to address these drawbacks. As proposed in,[41] regularizing CSP by incorporating prior information proves beneficial. As a result, a proper estimation for the covariance matrix is made by integrating prior knowledge and regularization terms as:[41]

$$\tilde{\mathbf{C}}_{\text{i}} = (1-\gamma)\hat{\mathbf{C}}_{\text{i}} + \gamma\mathbf{I}, \quad (4)$$

where

$$\hat{\mathbf{C}}_{\text{i}} = (1-\beta)s_i\mathbf{C}_i + \beta\mathbf{G}_i. \quad (5)$$

In (4) and (5), $\mathbf{C}_i$ is the initial spatial covariance matrix for class i and $\tilde{\mathbf{C}}_i$ is its regularized estimation. Furthermore, **I** represents the identity matrix, and s_i is a scaling parameter. The regularization process involves two parameters, γ and β, both constrained within the range [0,1]. Additionally, $\mathbf{G}_i$ represents the generic covariance matrix. The parameter γ is employed to shrink the $\mathbf{C}_i$ estimate towards I, effectively mitigating the estimation bias stemming from small training data. Conversely, the parameter β facilitates the shrinkage of the $\mathbf{C}_i$ estimate toward $\mathbf{G}_i$, enhancing the stability of the estimation. $\mathbf{G}_i$ signifies the ideal covariance matrix configuration corresponding to each distinct mental state. The acquisition of $\mathbf{G}_i$ involves utilizing signals from diverse subjects who participated same experiments.[41] These methods require training the spatial filters by replacing $\mathbf{C}_1$ and $\mathbf{C}_2$ in the conventional CSP algorithm with their regularized counterparts, denoted as $\tilde{\mathbf{C}}_1$ and $\tilde{\mathbf{C}}_2$. A collection of RCSP methods has been developed by $\mathbf{G}_i$ and using adjustable parameters.

Generic Learning Regularized CSP (GLRCSP) leverages covariance matrix regularization using cross-participant data. This approach incorporates regularization terms, β and γ, which steer the covariance matrix toward both **I** and $\mathbf{G}_i$. This dual effect enhances estimation by mitigating bias (γ) and reinforcing stability (β). $\mathbf{G}_i$ is computed through covariance matrices of other subjects as:

$$\mathbf{G}_i = s_i \sum_{j \in \Omega} \mathbf{C}_i^j, \quad (6)$$

where

$$s_i = \frac{1}{(1-\beta)M_{C_i} + \beta \sum_{j \in \Omega} M_{C_i^j}}. \quad (7)$$

In (6) and (7), s_i is the scaling parameter, Ω denotes the dataset of subjects, $\boldsymbol{C}_i^j$ signifies the spatial covariance matrix for class i and subject j, and M_C represents the number of trials utilized to compute the covariance matrix **C**. GLRCSP demonstrates two important advantages over conventional RCSP. First, it maintains high classification performance even when training data is limited, due to the effect of the regularization parameter γ. Second, by integrating population-level trends through β, it enhances the stability and generalizability of the spatial filters across subjects.

Differential evolution

Meta-heuristic methods provide a strategic response to addressing the dimensionality challenges by intelligently selecting features based on evaluations of classification performance. These algorithms bear significant promise, having demonstrated success in effectively navigating expansive feature spaces to uncover optimal solutions across diverse applications. Moreover, they have proven effective in feature selection and dimensionality reduction.[42]

The DE method emerges as a prominent contender among evolutionary optimization techniques. This method, characterized by its straightforward yet robust population-based random search algorithm, has demonstrated effective application in addressing optimization challenges spanning diverse fields of basic sciences and engineering. DE offers an array of strategies to generate trial vectors, each with distinct suitability for addressing specific problems.[43] Three pivotal control parameters – population size (n), mutation scale factor (F), and crossover rate (CR) – are integral components of this method, profoundly influencing the effectiveness of the DE approach. As an evolutionary process, the DE algorithm focuses on a population of size n

denoted as $\mathbf{X}_{i,G} = \left\{\mathbf{x}_{i,G}^{1}, \mathbf{x}_{i,G}^{2}, \ldots, \mathbf{x}_{i,G}^{D}\right\} (i = 1, \ldots, n)$ and encodes potential solutions. This population consists of n-dimensional parametric D vectors known as individuals and converge toward the globally optimal direction.[44,45] Analogous to the genetic algorithm, the DE optimization method employs vector-based search techniques involving crossover and mutation operations. DE boasts two primary strengths: high convergence ability and global optimization capability, while overcoming local optimization traps.

Proposed Method: K-GLR-DE

Machine learning techniques are widely recognized as a fundamental approach for MI processing. In this MI-BCI study, we integrate classical signal processing with GSP techniques to finely construct and reduce the brain graph, selecting optimal features derived from the graph. This section describes the stages and procedures employed for preprocessing, classical signal processing, and GSP. Figure 1 illustrates the block diagram of our proposed method, referred to as Kron-reduced generic learning regularization with differential evolution (K-GLR-DE).

Preprocessing

We employ a fifth-order Butterworth bandpass filter within the range of 8–30 Hz, guided by previous studies,[26,46] to preserve the alpha (8-13 Hz) and beta (13-30 Hz) frequency bands. Reliable studies have demonstrated the significant importance of these two frequency bands in classifying MI-BCI tasks.[46,47] As proposed by,[26] we employ the weighted moving average (WMA) method to account for the subjects' responses to each stimulus. By analyzing various WMA functions in this application, we find that the square function yields lower reconstruction error and superior results compared to linear and binomial functions.

Numerous studies on BCI systems, specifically those using the IVa dataset, including,[24-26,48] have employed downsampling as an effective preprocessing technique. A fixed downsampling factor of 10 is used, reducing the sampling frequency from 1000 Hz to 100 Hz. To extract MI activity, similar to,[24-26] we analyze the time interval of 0.5 to 4 s following the visual stimulus in each trial of the dataset.

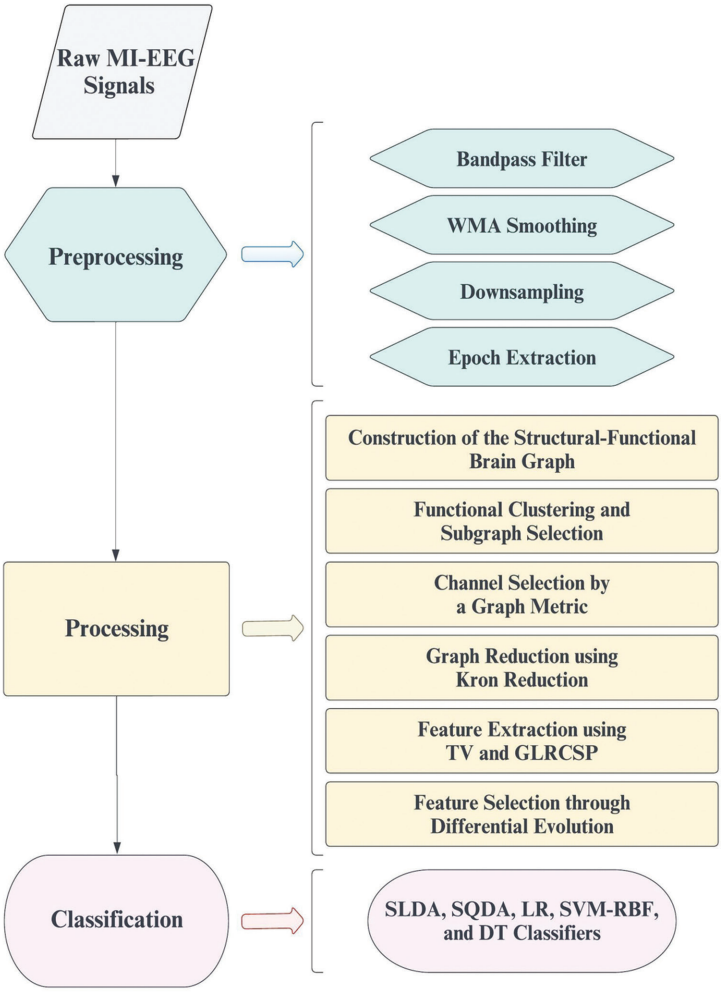


Figure 1: Block diagram of the proposed method, K-GLR-DE

Processing

In the BCI system, EEG signals are processed to extract the essential control commands. This paper employs classical signal processing and GSP as the primary approaches. In this section, the processing steps of the proposed approach are explained: graph construction, graph reduction, feature extraction, feature selection, and classification.

Construction of the structural-functional brain graphs

As described in Section 2.3, the construction of graphs, including brain graphs, necessitates the identification of two fundamental components: vertices and weighted edges. Expanding on our previous study,[28] this paper investigates the utilization of EEG signals on structural-functional brain graphs (SFG brain graphs). The vertices in our brain network correspond to the EEG electrodes, resulting in 118 vertices. The weights of brain graph edges are determined by measures of structural connectivity through geometric distance and functional connectivity through correlation, following a methodology akin to.[28] Within the SFG brain graphs, edge weights are determined as the product of two exponential expressions, representing the normalized squares of the Euclidean distance and the Pearson correlation. The weight matrix $\mathbf{W}_{SFG}$ for $p = 1,...,N$ and $q = 1,...,N$ is defined as:

$$W_{SFG}(p,q) = \exp(-\frac{D^2(p,q)}{2\sigma_d^2}).\exp(-\frac{\left(1-\left\|\rho(p,q)\right\|\right)^2}{2\sigma_\rho^2}), \quad (8)$$

Where $D(p,q) < k$ represents an element from the Euclidean distance matrix, and $\rho\ (p,q)$ represents an element from the correlation matrix. In (8), σ_d and σ_ρ denote the normalization coefficients for the squared Euclidean distance and squared Pearson correlation between vertices, respectively. The value of k represents the threshold for the Euclidean distance between two vertices. If this distance falls below the threshold k the weight of the edge connecting the two electrodes is calculated using (8); otherwise, the weight is set to 0. In this study, we adopt the empirical method described in[28] to select the normalizing coefficients σ_d and σ_ρ for each subject based on a trial-and-error approach and evaluation of decoding performance.

These coefficients are selected from three predefined strategies: the constant value in,[24] the variance of the matrix elements, and the maximum of the matrix elements – based on decoding performance. The threshold value k is set to 1 (according to the head radius), as recommended by.[24] Consequently, using (8), we compute the weighted adjacency matrix ($\mathbf{W}_{118\times118}$) for the SFG brain graphs. The corresponding graph Laplacian matrix ($\mathbf{L}_{118\times118}$) is then obtained and used in the subsequent steps. These graphs are extracted individually for each subject but jointly for the two MI classes.

Functional clustering and subgraph selection

To identify the active brain regions during hand-foot MI tasks, we consider the specific physiology of the case and apply a functional clustering method to the EEG electrodes. Functional clustering is performed based on the Brodmann area (BA) atlas, which provides anatomical priors corresponding to all EEG channels in the current dataset. Channels refer to the 118 EEG electrodes in this dataset; regions of interest (ROIs) correspond to neuroanatomically defined motor-relevant brain areas (motor-relevant Brodmann areas); clusters are functional groupings of EEG channels derived from electrode-to-ROI mapping.

Functional clustering was guided by Brodmann areas, which served to assign EEG electrodes to physiologically relevant regions. Specifically, EEG channels were mapped to the nearest cortical Brodmann areas using a scalp-to-cortex projection model that accounts for volume conduction and head geometry.[49,50] We consider the somatosensory cortical areas responsible for the MI tasks to improve classification performance.[49,50] Figure 2a displays the twenty-four ROIs for various types of MI tasks (hand-foot MI tasks: 18 ROIs). Figure 2b illustrates the functional clustering of EEG electrodes based on bilateral MI-related ROIs, encompassing both the left and right hemispheres. The details for all 24 MI-related ROIs can be found in Table 2.[49,50] The number of electrodes included in each subgraph is as follows:

- Cluster 1: (1-13),
- Cluster 2: (17-19, 25-28, 35-37, 45-46),
- Cluster 3: (14-16, 23-24, 31, 33-34),
- Cluster 4: (20-22, 29-30, 38-39, 41),
- Cluster 5: (32, 42, 50, 59, 67-68),
- Cluster 6: (40, 49, 58, 66, 76-77),
- Cluster 7: (43-44, 51-52, 60-61),
- Cluster 8: (47-48, 56-57, 64-65),
- Cluster 9: (53-55, 62-63, 71-73, 81-82, 90-92, 99-100),
- Cluster 10: (69-70, 78-80, 86-89, 97-98),
- Cluster 11: (74-75, 83-85, 93-96, 101-102),
- Cluster 12: (103-118).

The analysis was conducted within 18 specific regions, which related to the hand and foot tasks, including pairs

Table 2: 24 regions of interest in motor imagery BCI

ROI number	ROI name
ROI (1+13)	SAC
ROI (2+14)	Primary foot somatosensory area (S1F)
ROI (3+15)	Primary hand somatosensory area (S1H)
ROI (4+16)	Secondary somatosensory area (S2)
ROI (5+17)	CMA
ROI (6+18)	Primary foot motor area (M1F)
ROI (7+19)	Primary hand motor area (M1H)
ROI (8+20)	Primary lip motor area (M1L)
ROI (9+21)	SMA
ROI (10+22)	pSMA
ROI (11+23)	Dorsal premotor cortex (PMd)
ROI (12+24)	Ventral premotor cortex (PMv)

ROI – Regions of interest; SAC – Somatosensory association cortex; CMA – Cingulate motor area; SMA – Supplementary motor area; pSMA – Presupplementary motor area

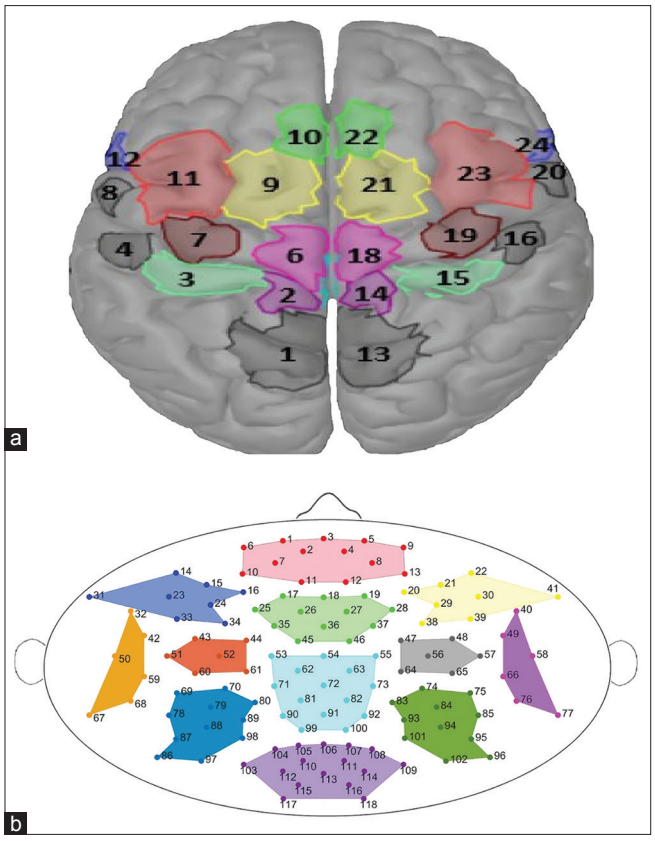


Figure 2: (a) 24 ROIs for MI tasks,[50] (b) Functional clustering of 118 EEG electrodes based on specific ROIs for hand–foot MI tasks

such as {R2+R14, R3+R15, R5+R17, R6+R18, R7+R19, R9+R21, R10+R22, R11+R23, R12+R24}. As depicted in Figure 2, these 18 ROIs are associated with eight subgraphs, specifically {C2,C3,C4,C7,C8,C9,C10,C11}. In this way, we identify significant physiological areas in MI application and conduct preliminary channel selection by prioritizing more important subgraphs, resulting in the selection of 77 out of 118 vertices.

The mapping of EEG electrodes to Brodmann-area-based ROIs was conducted using a standard scalp-to-cortex projection model aligned to the MNI template. Each electrode was associated with the nearest cortical Brodmann area using methods informed by canonical source modeling and neuroanatomical atlases.[49,50] This mapping accounts for volume conduction effects and provides a physiologically grounded electrode-to-ROI assignment strategy. Although this template-based approach does not fully eliminate inter-subject anatomical variability, the individualized graph construction improves robustness to such variability.

Channel selection by a graph metric

Channel selection plays a crucial role in brain signal processing as it reduces computational complexity. Choosing the appropriate channels in EEG processing is analogous to selecting important vertices in the brain graph from the perspective of GSP. Centrality metrics are a set of criteria in network science that examines the relative importance of a vertex or an edge concerning the structure and function of the network. Therefore, centralities are influential criteria for selecting vertices.[51] In this step, we employ a widely adopted and straightforward metric known as the weighted degree to select 24 significant vertices (~20% of total vertices) from the eight related brain subgraphs, which are then used in the Kron reduction process to simplify the graph. According to neurophysiological characteristics,[41,50,52] MI-EEG of the right hand and right foot occurs more frequently in the brain's left hemisphere than in the right hemisphere. Based on our experimental observations, we have found that assigning greater weight to the left hemisphere of the brain in this context produces superior outcomes. Instead of selecting the three top vertices from each of the eight subgraphs, we choose four superior vertices from three left brain subgraphs, three superior vertices from two middle brain subgraphs, and two superior vertices from three right brain subgraphs, resulting in 24 superior vertices in the brain graph.

Graph reduction using Kron reduction

This study addresses the challenge of dimension reduction in large EEG data through Kron reduction. To reduce the vertex dimension in the brain graph, we employ a combination of selecting significant vertices based on degree centrality and utilizing Kron reduction. By using degree centrality and Kron reduction, as presented in (2) and (3), we identify 24 superior vertices from eight related subgraphs. We reduce the number of vertices in eight subgraphs from 77 to 24 and the size of the Laplacian matrix from (77×77) to (24×24) in the vertex domain. This approach enables meaningful sampling in the vertex domain by merging the selection of the top 24 vertices from the initial 77 vertices and applying Kron reduction to the Laplacian matrix. Figure 3a illustrates an example of an EEG signal on the SFG brain graph in the vertex domain for subject "aa," while Figure 3b demonstrates the reduced SFG brain graph achieved through weighted degree and Kron reduction for subject "aa."

An ablation study was conducted to compare classification accuracy across different graph reduction levels (top 12, 16, 20, 24, 28, 32 vertices). The 24-vertex configuration yielded the highest average decoding performance and was therefore adopted [Supplementary Table 1 in Supplementary Material]. To validate that the essential spectral properties of the graph are preserved, we computed and compared the first 24 eigenvalues of the original and Kron-reduced Laplacians. The resulting distributions showed high similarity in spectral shape and smoothness, consistent with prior findings that Kron reduction retains spectral and energy profiles[36,37] [Supplementary Figure 1 and Supplementary Table 2 in Supplementary Material].

Feature extraction using TV_G and GLRCSP

We used TV_G (**x**) as GSP-based feature vectors. After Kron

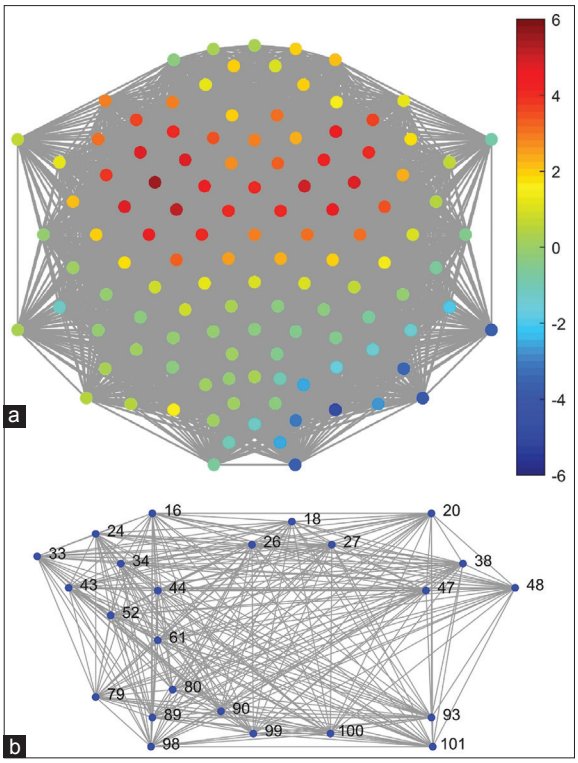


Figure 3: (a) Example of an EEG signal on the original SFG brain graph for subject "aa" in the vertex domain, (b) Example of the reduced SFG brain graph for subject "aa" (24 significant vertices) achieved through weighted degree and Kron reduction in the vertex domain

reduction, matrix **W** has dimensions of (24×24) in (1). We apply the collected data separately for the 24 vertices of eight subgraphs as described in (1). Consequently, the feature vectors for each subgraph and each trial will have dimensions of (350×1). Inspired by findings from,[41] we employ the GLRCSP method to regularize spatial filters in our study. CSP-based feature extraction was conducted by inputting EEG data from eight subgraphs into GLRCSP with m spatial filters. Similar to the methodology outlined in, [41] we fine-tuned the hyperparameters to attain maximum accuracy using LDA and 10-fold cross-validation (CV) on the training dataset. The values assigned to β and γ are selected from the set {0,0.1,0.2,...,0.9}.

Our feature extraction process comprised two distinctive categories: 1) TV_G from eight subgraphs, resulting in dimensions of 8×350 per trial and 2) The ratio of the variance of each GLCRSP output to the total variance of all GLRCSP outputs, yielding dimensions of m×1 per trial, with $m \in \{2,4\}$ spatial filters applied. These categories collectively represent a set of graph-based features in the vertex domain and statistical features. We integrated feature vectors derived from these two categories and subsequently performed feature selection using DE.

Feature selection through differential evolution

As described in Section 3.2.5, the number of features per trial is too large for the number of trials. However, this count proves excessive for robust classification within our dataset comprising 280 trials, potentially triggering the curse of dimensionality. In response, we leverage an approach that optimizes feature selection and effectively reduces feature dimensions. This involves the application of an approximate optimization technique known as DE. Similar to,[43,44] the optimization objective function $f(.)$ equates to the percentage of classification accuracy for training data. The training process incorporates $f(.)$ to maximize classification accuracy for the training data.

Although other feature selection methods, such as PCA[24,28] and Mutual Information, were considered, they did not yield superior performance in our prior experiments. PCA, while effective for variance preservation, does not optimize for class discriminability, often leading to suboptimal classification results. Mutual Information-based ranking, although supervised, was observed to suffer from instability due to data sparsity and feature redundancy in high-dimensional EEG spaces. In contrast, DE directly optimizes classification accuracy, allowing the discovery of synergistic feature subsets that best separate classes. Empirical tests conducted on the same MI-BCI dataset showed that DE improved classification accuracy by 4.93% over PCA.[28] Further experiments in this study [Supplementary Table 3 in Supplementary Material] demonstrated that DE consistently outperformed PCA and Mutual Information by margins of 6.10% and 4.21% on average across subjects.

A noteworthy advantage of the DE algorithm lies in its utilization of differential vectors, capturing the disparity between responses and incorporating valuable insights from excluded individuals in the population. In this study, we have employed a common variant of the DE algorithm. This variant uses three random vectors in the mutation process and incorporates a binomial crossover mechanism. To configure the DE algorithm for feature selection, we explored multiple candidate values for population size, number of generations, mutation factor, and crossover rate based on prior studies.[43,44] The final configuration (Population Size = 50, Generations = 400, F = 0.75, CR = 0.7) was chosen based on classification accuracy and runtime feasibility. In addition, we conducted a parameter sensitivity analysis [Supplementary Table 4 in Supplementary Material] by varying one DE parameter at a time and averaging classification accuracy across all subjects. This demonstrated the performance of our K-GLR-DE method across DE configurations. Feature selection via the DE algorithm encompasses two key components: first, the construction of a feature subset through DE, and second, using of a classification algorithm to assess the validity of the chosen subset. Ultimately, the DE algorithm's output yields a selection of the top 10 features, which are then employed for classification. Table 3 presents the parameters employed for the DE algorithm.

Although DE, being population-based, is computationally more intensive than greedy methods, the optimization is performed only during the offline training phase. Importantly, DE imposes no additional runtime cost during the testing phase. During the testing process, feature selection requires a few seconds per subject on standard hardware (Intel Core i7-6700HQ, 16GB RAM), as it reduces to indexing the pre-optimized 10-feature subset identified during training.

Classification

Following the feature selection process, we proceed to the classification stage. In MI-BCI processing, our primary objective is to assign samples to their classes accurately. To assess the efficacy of our proposed approach (K-GLR-DE), we employ the features identified by the DE algorithm as input for five classifiers: shrinkage LDA (SLDA), shrinkage QDA (SQDA), LR, SVM with radial basis function (SVM-RBF), and decision tree (DT). The validation process is carried out using a 10 × 10-fold cross-validation method.

The SLDA classifier, an extension of LDA, incorporates regularization through the shrinkage of class-related covariance matrices.[53] This regularization is crucial because covariance matrices estimated from small datasets often exhibit extreme eigenvalues, leading to inaccurate covariance estimates. This challenge is mitigated by applying regularization to covariance matrices $\mathbf{\Sigma}$, resulting in $\hat{\mathbf{\Sigma}}(\lambda) = (1-\lambda)\mathbf{\Sigma} + \lambda\mathbf{I}$, where $\mathbf{I}$ represents the identity matrix and λ is the regularization parameter. Notably, analytical solutions exist for the automatically selecting the optimal λ value. In ERP-based BCI[53] and oscillatory activity BCI,[54] the resulting SLDA classifier has demonstrated superiority over the conventional LDA classifier. Similarly, the SQDA classifier, a variant of QDA, is also subject to regularization for enhanced performance.

For the SVM-RBF classifier, we performed a subject-specific grid search for the regularization parameter (C) and kernel bandwidth (σ), using 10-fold cross-validation accuracy on the training folds as the selection criterion. The parameter search space was: $C \in \{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$, $\sigma \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$. The tuning was conducted independently for each subject and strictly within the training partitions of the 10-fold cross-validation to prevent any form of data leakage or test set contamination. The final selected SVM-RBF hyperparameters are listed in Supplementary Table 5 in the Supplementary Material.

Table 3: Parameters used in differential evolution for selecting the optimal set of 10-features

DE parameters	Value
Population size	50
Number of generations	400
Crossover type	Binomial
Crossover rate (CR)	0.70
Mutation type	Three vectors (DE/rand/1)
Mutation scale factor (F)	0.75

DE – Differential evolution; CR – Crossover rate

Results and Discussion

This section provides a detailed examination of our results under various experimental conditions, supported by both quantitative metrics and visual illustrations (Tables and Figures). The analysis is organized into four parts:

a. We compare the performance of the proposed approach with the GDR-BCI method[26] and assess the impact of Kron reduction under three training conditions: limited, small, and conventional
b. We examine the effect of applying Kron reduction within the training setup of the BCI Competition III Dataset IVa
c. We investigate the combined effect of Kron reduction and DE, and benchmark our approach against two relevant alternatives: KG-DE[28] and CSP-DE.[44] This subsection also includes feature importance analysis and statistical significance testing
d. Finally, we evaluate the generalizability of K-GLR-DE using a second independent dataset, PhysioNet-eegmmidb, a publicly available EEG repository with 64-channel recordings (10–10 international system) from 109 participants. Subjects S001 to S020 were selected for this evaluation.

We further provide a detailed analysis of the classification accuracies achieved by the five classifiers (introduced in Section 3.3) under varying conditions, offering insights into the separability of the two-class MI tasks. All parameters and hyperparameters used in both GLR-DE and K-GLR-DE methods across the different experimental scenarios were exclusively derived from the training data and applied directly – without any post-hoc tuning – to the test data.

Impact of Kron reduction on limited, small, and conventional training trials

Given the significance of data size in classification, we thoroughly analyze the effects of Kron reduction across three distinct training data size modes and two spatial filter modes within this section.

Comparative analysis of GLR-DE and GDR-BCI[26] results

Table 4 presents the mean outcomes for GLR-DE and the mean results of GDR-BCI's best fold across all five subjects, employing LDA, QDA, and LR classifiers in six distinct modes. Upon reviewing Table 4, it is evident that the average accuracy of LDA, QDA, and LR has shown improvement in this study compared to the average accuracy of the best fold for these three classifiers in[26] across all six modes. In addition, our experiments show that the standard deviation of GLR-DE accuracy is lower than that

of GDR-BCI accuracy, signifying the enhanced stability of our recommended approach over the GDR-BCI method.[26]

Limited (60) training trials with 2 and 4 GLRCSP spatial filters

Table 5 presents results for (2-60) GLR-DE and (2-60) K-GLR-DE, while Figure 4 illustrates the impact of Kron reduction in (2-60). Similarly, Table 6 depicts findings for (4-60) GLR-DE and (4-60) K-GLR-DE, highlighting Kron reduction efficacy in the (4-60) mode through Figure 5.

Applying the K-GLR-DE approach to 2-GLRCSP and limited (60) training trials yielded average accuracy improvements of 2.04%, 2.19%, 2.21%, 1.94%, and 2.18% for the SLDA, SQDA, LR, SVM-RBF, and DT classifiers compared to GLR-DE, respectively. Similarly, with 4-GLRCSP and limited (60) training trials, applying the K-GLR-DE approach led to average accuracy enhancements of 3.22%, 3.21%, 3.12%, 3.26%, and 3.11% for the same classifiers compared to GLR-DE, respectively. As a result, the Kron reduction applied to the GLR-DE approach increased the average accuracy of all five classifiers in both (2-60) and (4-60) scenarios.

Small (100) training trials with 2 and 4 GLRCSP spatial filters

Table 7 presents the outcomes of (2-100) GLR-DE and (2-100) K-GLR-DE, respectively, while Figure 6 illustrates the impact of Kron reduction within the (2-100) context. Similarly, Table 8 details the findings of (4-100) GLR-DE and (4-100) K-GLR-DE, respectively, showcasing the effectiveness of Kron reduction in the (4-100) mode as illustrated in Figure 7.

In the (2-100) scenario, the K-GLR-DE approach yielded mean improvements of 2.84%, 3.29%, 3.47%, 3.88%, and 3.49% in the performance of SLDA, SQDA, LR, SVM-RBF, and DT classifiers, respectively, compared to the GLR-DE approach. Correspondingly, in the (4-100) scenario, the K-GLR-DE approach led to average enhancements of 3.31%, 2.58%, 1.45%, 2.09%, and 2.06% in the same classifiers, compared to the GLR-DE approach. Consequently, the application of the K-GLR-DE approach led to heightened mean accuracy across

Table 4: Comparative analysis of GLR-DE and GDR-BCI[26] results across six different modes

Scenarios*	LDA	LDA[26]	QDA	QDA[26]	LR	LR[26]
2–60	72.27±1.16	64.73	71.36±0.72	63.36	**73.00±1.39**	64.55
4–60	78.91±0.83	74.36	79.09±1.24	70.91	**79.45±0.99**	73.91
2–100	82.89±0.91	76.44	81.78±0.82	78.00	**82.89±1.54**	76.89
4–100	85.44±1.07	78.22	85.33±1.28	78.00	**86.78±1.07**	78.89
2–200	87.25±1.85	80.50	87.00±1.43	80.25	**87.75±2.40**	80.75
4–200	88.25±1.43	80.25	88.50±1.63	80.75	**89.00±1.05**	81.00

*The numbers in the Scenarios column represent the number of spatial filters and the number of training trials, respectively. LDA – Linear discriminant analysis; QDA – Quadratic discriminant analysis; LR – Logistic regression

Table 5: Classification accuracy (mean±standard deviation) with two spatial filters and 60 training trials employing GLR-DE and K-GLR-DE

(2–60) GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	68.27±1.40	68.36±0.98	67.45±1.80	**72.64±2.04**	71.00±0.88
al	83.27±0.78	82.00±0.74	82.45±1.46	**85.73±1.48**	84.91±1.22
av	67.91±3.54	67.82±1.83	67.45±4.15	**71.91±3.02**	70.00±2.44
aw	71.82±2.04	71.09±0.98	72.27±2.04	**76.27±2.13**	74.09±1.62
ay	75.73±1.98	76.00±1.28	75.45±2.14	**79.64±2.48**	78.45±2.10
Mean±SD	73.40±1.87	73.05±1.16	73.01±2.32	**77.24±2.23**	75.69±1.65
(2–60) K-GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	71.00±1.76	70.73±1.57	70.45±1.69	**75.00±1.88**	73.18±1.79
al	84.27±0.71	84.00±0.76	84.09±1.37	**86.91±1.26**	86.09±0.59
av	70.27±3.25	70.36±1.82	69.91±3.28	**73.73±2.58**	73.18±2.03
aw	74.91±2.33	74.09±0.88	74.91±0.92	**79.55±2.04**	77.36±1.39
ay	76.73±1.56	77.00±0.93	76.73±2.01	**80.73±2.36**	79.55±1.09
Mean±SD	75.44±1.92	75.24±1.19	75.22±1.85	**79.18±2.02**	77.87±1.38

SD – Standard deviation; LR – Logistic regression; DT – Decision tree; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; GLR-DE – Generic learning regularization with differential evolution; SVM-RBF – Support vector machine with a radial basis function

all five classifications for both the (2-100) and (4-100) cases, surpassing the performance of the GLR-DE approach.

Conventional (200) training trials with 2 and 4 GLRCSP spatial filters

Table 9 provides the outcomes for (2-200) GLR-DE and (2-200) K-GLR-DE, respectively, while Figure 8 illustrates the impact of Kron reduction in this context. Similarly, Table 10 presents the results of (4-200) GLR-DE and (4-200) K-GLR-DE, respectively, showcasing the effectiveness of Kron reduction in the (4-200) mode, as demonstrated in Figure 9.

Within the context of 2-GLRCSP mode and conventional training data (200), the mean accuracy of all subjects in SLDA, SQDA, LR, SVM-RBF, and DT classifiers using the K-GLR-DE approach has exhibited increments of 2.00%, 1.85%, 1.45%, 1.35%, and 1.05%, respectively, when compared to the GLR-DE approach. Similarly, in the scenario involving 4-GLRCSP mode and conventional training data (200), the K-GLR-DE approach has led to average enhancements of 2.00%, 2.00%, 2.75%, 2.35%, and 2.50% across the same classifiers, as opposed to the GLR-DE approach. Because of applying Kron reduction, we observed an enhancement in the mean accuracy for all five classifiers, evident in both the (2-200) and (4-200) scenarios.

Notably, the investigation conducted in this study reveals that the average accuracy of SLDA and SQDA adaptive classifiers consistently surpasses that of the fixed LDA and QDA cases across all modes. This includes scenarios with limited, small, and conventional training datasets and configurations involving two and four spatial filters.

In summary, the data suggests a positive correlation between the number of training trials and the accuracy of SVM-RBF within both the 2-GLRCSP and 4-GLRCSP modes for both GLR-DE and K-GLR-DE methodologies. In the Supplementary Material, Supplementary Figure 2a and b illustrate the variability in the average accuracy of this study's top-performing classifier, SVM-RBF, about the number of training trials. This variability is demonstrated for both GLR-DE and K-GLR-DE methodologies within the 2-GLRCSP mode. Similarly, Supplementary Figure 3a and b in the Supplementary Material depict the changes in the average accuracy of SVM-RBF as the number of training trials varies. This comparison is made between GLR-DE and K-GLR-DE methodologies within the 4-GLRCSP mode.

Impact of Kron reduction on training trials of BCI Competition III-dataset IVa

We have implemented experimental conditions akin to those used in the BCI Competition III-Dataset IVa, enabling us to conduct a comparative analysis of the results. In the 4-GLRCSP mode, aligned with the number of training trials in BCI Competition III, the GLR-DE approach yields average accuracies of 85.26%, 86.00%,

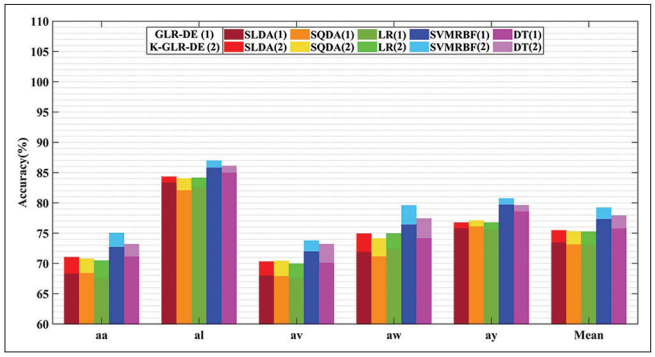


Figure 4: Effect of Kron reduction on classification accuracy in (2-60) GLR-DE and (2-60) K-GLR-DE

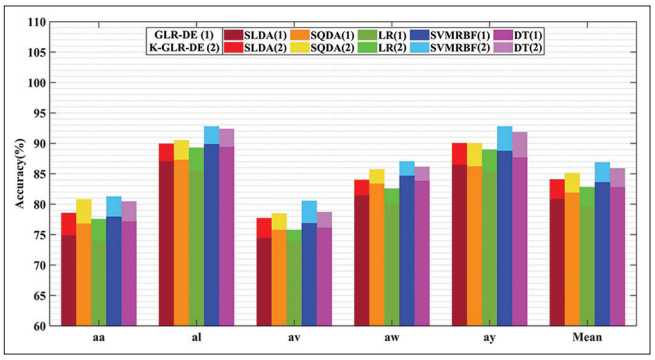


Figure 5: Effect of Kron reduction on classification accuracy in (4-60) GLR-DE and (4-60) K-GLR-DE

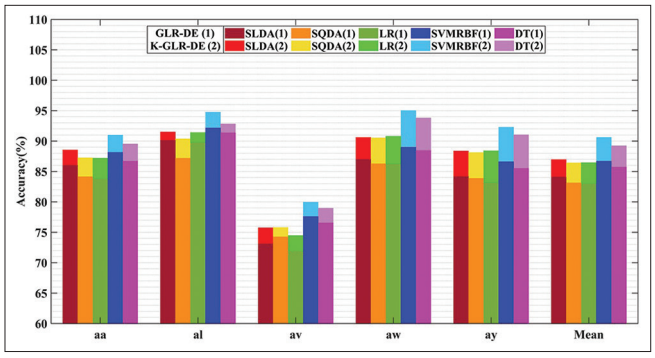


Figure 6: Effect of Kron reduction on classification accuracy in (2-100) GLR-DE and (2-100) K-GLR-DE

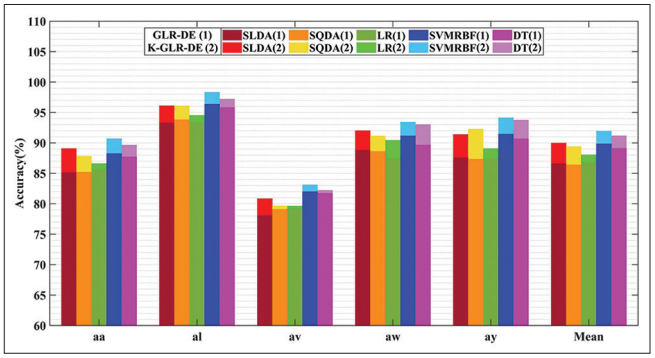


Figure 7: Effect of Kron reduction on classification accuracy in (4-100) GLR-DE and (4-100) K-GLR-DE

Table 6: Classification accuracy (mean±standard deviation) with four spatial filters and 60 training trials employing GLR-DE and K-GLR-DE

(4–60) GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	74.82±3.55	76.73±3.24	73.82±3.66	**77.91±2.65**	77.09±2.23
al	87.00±0.70	87.27±1.33	85.36±1.54	**89.82±2.01**	89.36±2.13
av	74.36±4.24	75.73±4.18	73.82±4.41	**76.82±3.60**	76.00±2.84
aw	81.36±2.51	83.27±2.49	80.00±2.67	**84.64±2.84**	83.73±2.72
ay	86.36±0.97	86.18±2.17	85.18±1.94	**88.73±2.25**	87.64±1.64
Mean±SD	80.78±2.39	81.84±2.68	79.64±2.84	**83.58±2.67**	82.76±2.31
(4–60) K-GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	78.55±2.87	80.73±3.10	77.45±3.12	**81.18±2.76**	80.45±2.90
al	89.91±1.00	90.45±1.55	89.18±1.65	**92.73±1.74**	92.36±1.91
av	77.64±3.55	78.45±3.50	75.73±3.94	**80.55±3.26**	78.64±3.07
aw	83.91±2.04	85.64±2.55	82.55±2.22	**87.00±2.52**	86.09±2.57
ay	90.00±1.39	90.00±2.14	88.91±2.04	**92.73±2.52**	91.82±2.28
Mean±SD	84.00±2.17	85.05±2.57	82.76±2.59	**86.84±2.56**	85.87±2.55

SD – Standard deviation; LR – Logistic regression; DT – Decision tree; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; GLR-DE – Generic learning regularization with differential evolution; SVM-RBF – Support vector machine with a radial basis function

Table 7: Classification accuracy (mean±standard deviation) with two spatial filters and 100 training trials employing GLR-DE and K-GLR-DE

(2–100) GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	86.00±1.30	84.11±1.76	83.78±1.50	**88.11±2.17**	86.67±1.40
al	90.11±1.24	87.11±1.58	89.67±1.27	**92.11±0.74**	91.33±1.10
av	73.11±1.97	74.22±1.87	71.89±1.69	**77.56±2.13**	76.56±1.64
aw	87.00±1.17	86.22±1.34	86.22±1.52	**88.89±1.59**	88.44±1.36
ay	84.11±1.12	83.78±0.95	83.11±0.85	**86.56±1.38**	85.44±0.76
Mean±SD	84.07±1.36	83.09±1.50	82.93±1.37	**86.65±1.60**	85.69±1.25
(2–100) K-GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	88.56±1.63	87.22±2.14	87.11±1.85	**90.89±2.45**	89.44±1.63
al	91.44±1.52	90.33±1.94	91.33±1.59	**94.67±0.96**	92.78±1.33
av	75.67±2.37	75.78±2.26	74.44±2.06	**79.89±2.41**	78.89±1.90
aw	90.56±1.67	90.44±1.67	90.78±1.49	**95.00±1.62**	93.78±1.02
ay	88.33±1.48	88.11±1.24	88.33±1.18	**92.22±1.85**	91.00±1.60
Mean±SD	86.91±1.73	86.38±1.85	86.40±1.63	**90.53±1.86**	89.18±1.50

SD – Standard deviation; LR – Logistic regression; DT – Decision tree; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; GLR-DE – Generic learning regularization with differential evolution; SVM-RBF – Support vector machine with a radial basis function

84.00%, 88.29%, and 87.60% for SLDA, SQDA, LR, SVM-RBF, and DT classifiers, respectively [Table 11 and Figure 10].

Similarly, under the same settings, the K-GLR-DE approach produces average accuracies of 88.86%, 88.64%, 87.07%, 91.62%, and 90.43% for the classifiers above [Table 12 and Figure 11]. This demonstrates that incorporating Kron reduction enhanced the average accuracy across all classifiers. Moreover, the average performance of the K-GLR-DE method with SVM-RBF is noteworthy. Compared to the top three groups in the BCI Competition III-Dataset IVa, which achieved average accuracies of 94.17%, 85.12%, and 83.45%,[29] the K-GLR-DE results demonstrate competitiveness [Refer to Table 12 and Figure 11 for further details].

As illustrated in Figures 10 and 11, it is clear that the standard deviation for the results of subject "av" exceeds that of other subjects. Conversely, the standard deviation for "al" and "aw" is lower, indicating higher stability in the data results for "al" and "aw". This trend is also apparent in Tables 5-10.

Tables 11 and 12 provide additional insights, showing that the classification results for "al" and "aw" surpass those of other

Table 8: Classification accuracy (mean±standard deviation) with four spatial filters and 100 training trials employing GLR-DE and K-GLR-DE

(4–100) GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	85.11±2.64	85.22±1.96	85.56±1.86	**88.22±2.43**	87.67±1.66
al	93.22±1.37	93.78±0.52	93.33±0.90	**96.33±1.80**	95.78±1.42
av	78.00±3.48	79.00±2.98	79.11±2.74	**82.00±2.60**	81.67±2.54
aw	88.78±2.87	88.56±1.29	87.33±0.98	**91.11±2.23**	89.56±1.58
ay	87.56±1.83	87.33±1.90	87.33±2.12	**91.44±1.84**	90.67±1.80
Mean±SD	86.53±2.44	86.78±1.73	86.53±1.72	**89.82±2.18**	89.07±1.80
(4–100) K-GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	89.00±2.83	87.78±2.16	86.56±2.36	**90.67±2.21**	89.56±2.16
al	96.11±1.57	96.00±0.71	94.44±1.11	**98.33±1.26**	97.22±1.43
av	80.78±3.18	79.67±3.18	79.56±2.94	**83.11±3.43**	82.22±3.12
aw	92.00±2.04	91.11±1.79	90.33±1.18	**93.33±2.29**	93.00±2.04
ay	91.33±3.07	92.22±1.90	89.00±2.32	**94.11±2.60**	93.67±2.34
Mean±SD	89.84±2.54	89.36±1.95	87.98±1.98	**91.91±2.36**	91.13±2.22

SD – Standard deviation; LR – Logistic regression; DT – Decision tree; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; GLR-DE – Generic learning regularization with differential evolution; SVM-RBF – Support vector machine with a radial basis function

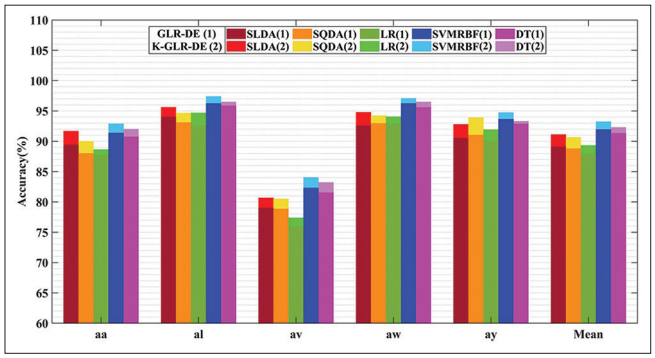


Figure 8: Effect of Kron reduction on classification accuracy in (2-200) GLR-DE and (2-200) K-GLR-DE

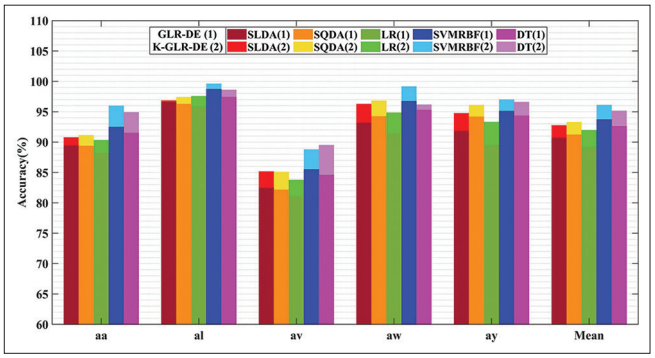


Figure 9: Effect of Kron reduction on classification accuracy in (4-200) GLR-DE and (4-200) K-GLR-DE

subjects, despite undergoing a similar number of test trials as in BCI Competition III. This superior performance could be due to the data recording process. Specifically, "al" and "aw" participated in two sessions each of task 1 (target-correlated eye movement) and task 2 (target-uncorrelated eye movement), whereas other subjects participated in one session of task 1 and three sessions of task 2. This observation is consistent with findings from previous studies.

Synergistic impact of Kron reduction and differential evolution algorithm

In this section, we examine the combined impact of Kron reduction and the DE algorithm on K-GLR-DE, with the goal of comparing the findings to those of the two prior studies.[28,44] The outcomes of the K-GLR-DE approach, evaluated through a 10 × 10-fold cross-validation across all 280 trials encompassing different individuals and classifiers, are documented in Table 13. Table 13 highlights that the K-GLR-DE approach coupled with SVM-RBF classification yields the highest average accuracy across all subjects. It is worth noting that in Table 13, the number of selected features for all subjects in[44] exceeds the ten features selected in this study and our previous work.[28] In the K-GLR-DE methodology, specifically adapted for the MI-BCI system, all processing steps, including structural-functional graph construction, functional clustering, Kron reduction, and DE optimization for feature selection, were performed strictly within the training folds of cross-validation. That is, for each fold in the trial-wise 10 × 10-fold CV setup, only the training data were used for model fitting and parameter tuning. The trained model was then evaluated on the corresponding test fold, with no information from the test fold used during training or feature selection. This strict separation ensured that no train-test data leakage occurred during any stage of the process.

The average performance of SVM-RBF using the K-GLR-DE approach demonstrated an increase of 0.44% compared to the CSP-DE method[44] under the same

Table 9: Classification accuracy (mean±standard deviation) with two spatial filters and 200 training trials employing GLR-DE and K-GLR-DE

(2–200) GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	89.25±1.46	88.00±1.50	87.75±1.46	**91.25±2.06**	90.75±1.94
al	94.00±1.37	93.00±1.50	92.50±0.79	**96.25±0.79**	95.75±1.50
av	79.00±1.58	78.75±1.94	76.00±1.50	**82.25±2.44**	81.50±2.06
aw	92.50±0.79	93.00±1.00	92.75±0.50	**96.25±1.37**	95.50±1.00
ay	90.50±1.46	91.00±1.46	90.00±1.22	**93.50±2.00**	92.75±2.00
Mean±SD	89.05±1.33	88.75±1.48	87.80±1.09	**91.90±1.73**	91.25±1.70
(2–200) K-GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	91.75±1.00	90.00±2.50	88.50±2.00	**92.75±2.00**	92.00±1.00
al	95.50±1.00	94.50±1.50	94.75±0.50	**97.25±0.94**	96.50±1.22
av	80.50±1.22	80.50±1.58	77.25±1.46	**84.00±2.06**	83.25±1.37
aw	94.75±0.50	94.25±0.61	94.00±0.50	**97.00±0.61**	96.50±0.50
ay	92.75±0.50	93.75±2.50	91.75±1.00	**94.75±2.00**	93.25±1.00
Mean±SD	91.05±0.84	90.60±1.74	89.25±1.09	**93.25±1.52**	92.30±1.02

SD – Standard deviation; LR – Logistic regression; DT – Decision tree; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; GLR-DE – Generic learning regularization with differential evolution; SVM-RBF – Support vector machine with a radial basis function

Table 10: Classification accuracy (mean±standard deviation) with four spatial filters and 200 training trials employing GLR-DE and K-GLR-DE

(4–200) GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	89.50±1.00	89.25±0.79	88.00±1.22	**92.50±1.22**	91.50±1.00
al	96.50±0.50	96.25±0.79	95.75±1.00	**98.75±0.79**	97.25±0.50
av	82.50±1.46	82.00±1.22	81.00±1.58	**85.50±1.22**	84.50±2.21
aw	93.25±0.79	94.25±0.79	91.50±1.46	**96.75±0.61**	95.25±1.58
ay	91.75±1.37	94.25±1.00	89.50±1.37	**95.00±0.94**	94.25±1.00
Mean±SD	90.70±1.02	91.20±0.92	89.15±1.33	**93.70±0.96**	92.55±1.26
(4–200) K-GLR-DE					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	90.75±1.58	91.00±2.00	90.25±2.00	**96.00±1.46**	94.75±1.22
al	96.75±0.61	97.25±1.22	97.50±1.00	**99.50±0.61**	98.50±0.50
av	85.00±2.26	85.00±2.33	83.75±2.00	**88.75±1.70**	89.50±1.50
aw	96.25±1.58	96.75±1.00	94.75±0.50	**99.00±0.50**	96.00±0.50
ay	94.75±2.00	96.00±0.94	93.25±2.00	**97.00±1.50**	96.50±1.22
Mean±SD	92.70±1.61	93.20±1.50	91.90±1.50	**96.05±1.15**	95.05±0.99

SD – Standard deviation; LR – Logistic regression; DT – Decision tree; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; GLR-DE – Generic learning regularization with differential evolution; SVM-RBF – Support vector machine with a radial basis function

10 × 10-fold cross-validation scenario. Furthermore, the increase was notably substantial at 0.96% when compared to the KG-DE approach.[28] For a more intuitive understanding of the differences in outcomes, Figure 12 presents box plots of the statistical summaries of classification accuracy (mean, median, first and third quartiles, maximum, and minimum). Figure 12 illustrates the results of the K-GLR-DE method using 10 × 10-fold cross-validation. In Figure 12(a), the box plots represent an overview of the average results attained by the 4-K-GLR-DE method across all subjects and various classifiers. In Figure 12(b), the box plots depict the average results achieved by the 4-K-GLR-DE method for the best classifier (SVM-RBF) and different subjects. Notably, a pattern consistent with previous studies on this dataset emerges: subject "al" shows the highest classifier performance, while subject "av" exhibits the lowest.

To examine the contributions of each core component – Kron reduction and DE – we conducted a controlled ablation study, reported in Supplementary Table 6. This study compares classification performance across four configurations: (i) baseline 4-GLR, (ii) Kron-reduced 4-K-GLR, (iii) DE-enhanced 4-GLR-DE, and (iv) the full pipeline 4-K-GLR-DE. Each component

Table 11: Classification accuracy (mean±standard deviation) for 4-GLR-DE approach using training trials of BCI Competition III-dataset IVa

4-GLR-DE (BCI competition III-dataset IVa)					
Subjects*	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa (112)	87.86±2.50	90.36±2.58	86.25±2.65	**90.89±2.99**	89.64±2.71
al (56)	94.64±1.13	96.79±1.34	94.64±1.96	**98.57±1.33**	98.21±1.60
av (224)	76.25±2.92	76.70±2.90	74.82±2.81	**78.57±3.35**	78.04±3.23
aw (196)	91.33±2.25	91.94±2.22	89.69±2.47	**93.98±2.62**	93.67±2.62
ay (252)	85.32±2.66	85.32±2.38	84.37±2.57	**89.05±2.95**	88.10±2.88
Mean±SD	85.26±2.51	86.00±2.44	84.00±2.58	**88.29±2.88**	87.60±2.80

*The numbers in the Subjects column indicate the number of test trials in BCI Competition III dataset IVa. SD – Standard deviation; BCI – Brain-computer interface; LR – Logistic regression; DT – Decision tree; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; GLR-DE – Generic learning regularization with differential evolution; SVM-RBF – Support vector machine with a radial basis function

Table 12: Classification accuracy (mean±standard deviation) for 4-K-GLR-DE approach using training trials of BCI Competition III-dataset IVa

4-K-GLR-DE (BCI competition III-dataset IVa)					
Subjects*	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa (112)	88.93±1.84	91.43±1.96	87.32±1.85	**91.79±2.19**	90.18±2.15
al (56)	98.21±1.60	97.50±1.82	97.86±2.08	**99.64±0.71**	98.93±1.42
av (224)	78.93±2.04	79.20±2.11	77.50±2.16	**81.25±2.36**	80.27±2.45
aw (196)	94.90±1.67	93.67±1.90	92.35±1.88	**98.27±1.96**	97.55±2.07
ay (252)	90.87±1.69	89.92±1.92	88.97±1.69	**93.81±1.98**	92.14±2.04
Mean±SD	88.86±1.79	88.64±1.96	87.07±1.91	**91.62±2.02**	90.43±2.13

*The numbers in the Subjects column indicate the number of test trials in BCI Competition III dataset IVa. SD – Standard deviation; BCI – Brain-computer interface; LR – Logistic regression; DT – Decision tree; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; GLR-DE – Generic learning regularization with differential evolution; SVM-RBF – Support vector machine with a radial basis function

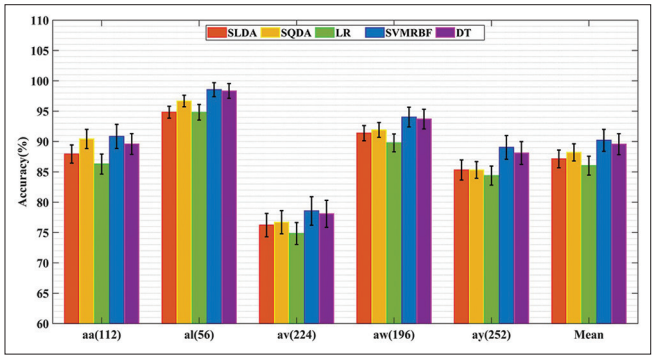


Figure 10: Effect of Kron reduction on classification accuracy (mean ± SD) for 4-GLR-DE approach using training trials of BCI Competition III-dataset IVa

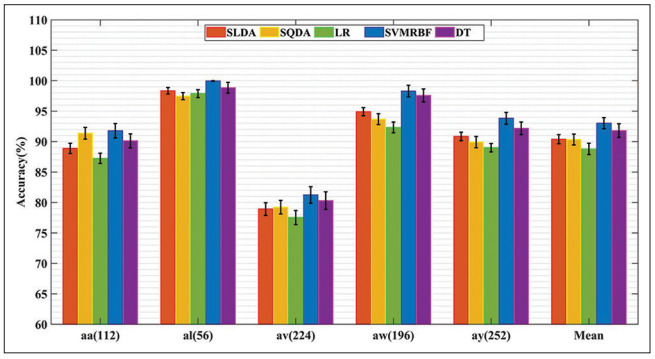


Figure 11: Effect of Kron reduction on classification accuracy (mean ± SD) for 4-K-GLR-DE approach using training trials of BCI Competition III-dataset IVa

independently improves accuracy over the base model, while the full integration leads to the highest performance (96.46% ± 0.81%). These results demonstrate the individual and combined benefits of graph reduction and evolutionary feature selection.

To assess the contribution of each feature category – graph-based total variation (TV_G) and statistical variance-based GLRCSP – we conducted an ablation-style feature importance analysis. Using the same settings as Table 13 (4 spatial filters, SVM-RBF, 10 × 10-fold CV), we separately evaluated the classification performance using TV_G-only or GLRCSP-only features. As shown in Supplementary Table 7, the results confirm that while GLRCSP features provide strong discriminative power due to spatial filtering, TV_G features from eight subgraphs also provide valuable structural information. Their combination leads to the best performance across subjects, demonstrating the complementary nature of structural and statistical-spatial features in the K-GLR-DE framework.

In addition to the classification accuracy results presented in Table 13, we provide three complementary evaluation metrics – sensitivity, specificity, and precision – to enable a

more comprehensive assessment of the 4-K-GLR-DE model under the 10 × 10-fold cross-validation protocol. These results are reported in Supplementary Tables 8-10 of the Supplementary Material. To further enhance interpretability and offer detailed insight into the model's performance across different subjects and classifiers, we have also included extensive visualizations: confusion matrices [Supplementary Figures 4-8] and ROC curves [Supplementary Figures 9 and 10] are presented in the Supplementary Material.

We recommend using simple yet effective classifiers, including SLDA, SQDA, and DT, which deliver promising results. Both our research findings and[55] support this recommendation. Notably, SLDA and SQDA stand out because of their absence of hyperparameters, which enhances their user-friendliness. However, it's worth noting that the linear nature of LDA may yield suboptimal results with extensive training data. Conversely, DT, a non-linear classifier, excels across various training set sizes. In terms of computational efficiency, DT, similar to SLDA and SQDA, is a fast and effective method. In addition, our investigation notes that within the K-GLR-DE context, the SVM and DT classifiers exhibit superior accuracy compared to the SLDA, SQDA, and LR classifiers.

To evaluate the statistical robustness of the observed performance improvements across methods, we conducted a one-way ANOVA and Bonferroni-corrected paired *t*-tests on the 10 × 10-fold CV results per subject. ANOVA results confirmed significant differences ($p < 0.001$) across methods for all five subjects. Pairwise *t*-tests revealed that K-GLR-DE significantly outperformed all other methods, with adjusted p values < 0.00167. In addition, Cohen's d effect sizes showed large improvements for K-GLR-DE. The ANOVA and pairwise *t*-test results are provided in Supplementary Tables 11 and 12 and illustrated in Supplementary Figures 11 and 12.

Generalization to an independent dataset

While the primary evaluation was conducted on BCI Competition III Dataset IVa (Right hand vs. Foot) to ensure comparability with prior work,[24-26,28,44] we further validated the generalizability of the proposed K-GLR-DE model on an independent dataset, PhysioNet-eegmmidb (Right fist vs. Left fist). For this evaluation, we selected Subjects S001–S020, focusing on right-fist vs left-fist MI tasks. The same preprocessing pipeline, functional clustering, Kron reduction, DE-based feature selection, and SVM-RBF classification strategy were employed as described earlier.

Using four spatial filters and 10 × 10-fold CV, the model achieved an average classification accuracy of 84.12% ± 2.51% across these 20 subjects. Detailed results are provided in Supplementary Table 13. These findings confirm the robustness and transferability of the proposed method to different recording conditions and MI-EEG systems.

Conclusions

This paper introduces the K-GLR-DE approach, designed for dimensionality reduction and classification of EEG data in MI-BCI systems. Key components of the method include subgraph selection based on physiological ROIs, dimensionality reduction of the graph Laplacian matrix using Kron reduction, feature extraction via GSP-based and CSP-based features, and feature selection through the DE optimization algorithm.

Adding GSP-based features to conventional CSP-based ones significantly enhances MI classification performance. By selecting optimal features from a combined pool of graph-based total variations and CSP-based statistical features using DE optimization, the K-GLR-DE approach achieves notable improvements in classification accuracy. Specifically, this method combines Kron reduction with physiology-graph-based channel selection, excelling in scenarios with limited and small training trials. When paired with SVM-RBF and DT classifiers, the proposed approach achieves comparable accuracy with the top performers in BCI Competition III-Dataset IVa, including an average of 91.62% ± 2.02% with four spatial filters. Notably, within the K-GLR-DE framework, the SVM-RBF classifier achieves an average accuracy of 96.46% ± 0.81% – outperforming the CSP-DE method[44] by 0.44% and the KG-DE method[28] by 0.96%.

Table 13: Classification accuracy (mean±standard deviation) in 4-K GLR-DE approach with 10×10-fold cross-validation compared to KG-DE[28] and CSP-DE[44]

	Accuracy in 4-K-GLR-DE (10×10-fold cross-validation)						
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**	**SVM-RBF[28]**	**SVM[44]**
aa	90.89±0.84	91.21±0.96	90.39±0.85	**96.11±1.20**	93.79±1.15	94.2±1.36	95.8±2.40
al	96.89±0.53	97.61±0.58	97.57±0.65	**99.71±0.14**	98.71±0.88	99.3±0.61	98.8±0.79
av	85.21±1.04	84.79±1.11	83.29±1.16	**89.89±1.36**	89.50±1.45	88.5±2.43	89.8±3.36
aw	95.61±0.67	97.21±0.90	96.39±0.88	**99.39±0.39**	96.43±1.07	98.6±0.94	99.2±0.72
ay	95.89±0.69	97.11±0.92	93.50±0.69	**97.21±0.98**	96.89±1.04	96.9±1.03	96.5±1.11
Mean±SD	92.90±0.75	93.59±0.89	92.23±0.85	**96.46±0.81**	95.06±1.12	95.5±1.27	96.02±1.68

LR – Logistic regression; DT – Decision tree; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; GLR-DE – Generic learning regularization with differential evolution; SVM-RBF – Support vector machine with a radial basis function; SD – Standard deviation

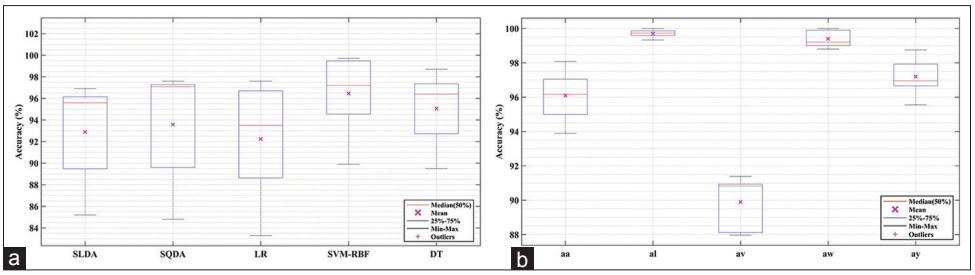


Figure 12: 4-K-GLR-DE approach with 10 × 10-fold cross-validation. (a) Accuracy of classifiers, and (b) Accuracy of subjects

The K-GLR-DE framework shows strong potential for advancing MI-BCI applications. Future work may explore its adaptation to other BCI paradigms and brain signal types, investigate optimal combinations of GSP-based, CSP-based, and alternative feature sets, and develop more efficient graph simplification techniques. Identifying optimal edge reduction strategies can further reduce computational overhead and enhance model efficiency. In addition, employing more advanced optimization strategies for feature selection and reduction may further improve effectiveness and computational scalability. These directions offer promising avenues for the continued development of robust and efficient MI-BCI systems.

Financial support and sponsorship

Nil.

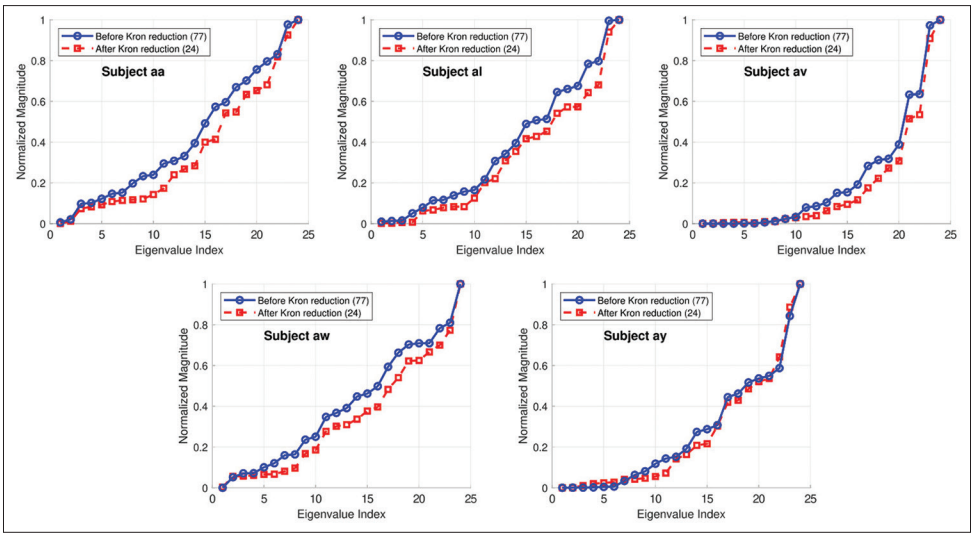
Conflicts of interest

There are no conflicts of interest.

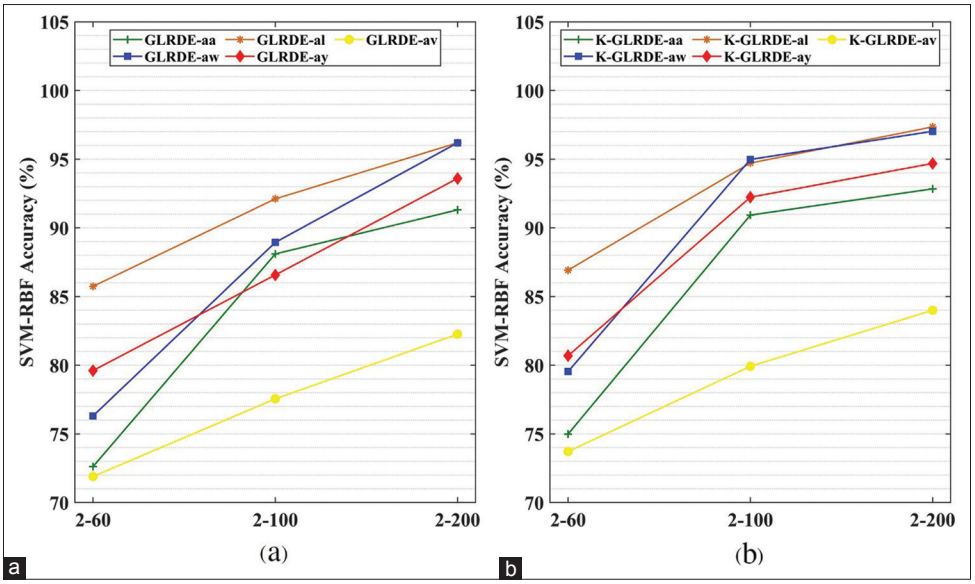
References

1. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. Clin Neurophysiol 2002;113:767-91.
2. Wolpaw JR, McFarland DJ, Vaughan TM. Brain-computer interface research at the Wadsworth center. IEEE Trans Rehabil Eng 2000;8:222-6.
3. Curran EA, Stokes MJ. Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems. Brain Cogn 2003;51:326-36.
4. Birbaumer N. Brain-computer-interface research: Coming of age. Clin Neurophysiol 2006;117:479-83.
5. Zander TO, Kothe C. Towards passive brain-computer interfaces: Applying brain-computer interface technology to human-machine systems in general. J Neural Eng 2011;8:1-5.
6. Chaudhary U, Birbaumer N, Ramos-Murguialday A. Brain-computer interfaces for communication and rehabilitation. Nat Rev Neurol 2016;12:513-25.
7. Blankertz B, Losch F, Krauledat M, Dornhege G, Curio G, Müller KR. The Berlin brain – Computer interface: Accurate performance from first-session in BCI-naïve subjects. IEEE Trans Biomed Eng 2008;55:2452-62.
8. Lal TN, Schröder M, Hinterberger T, Weston J, Bogdan M, Birbaumer N, *et al*. Support vector channel selection in BCI. IEEE Trans Biomed Eng 2004;51:1003-10.
9. Popescu F, Fazli S, Badower Y, Blankertz B, Müller KR. Single trial classification of motor imagination using 6 dry EEG electrodes. PLoS One 2007;2:e637.
10. Mateos G, Segarra S, Marques AG, Ribeiro A. Connecting the dots: Identifying network structure via graph signal processing. IEEE Signal Process Mag 2019;36:16-43.
11. Shuman DI, Narang SK, Frossard P, Ortega A, Vandergheynst P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Process Mag 2013;30:83-98.
12. Ortega A, Frossard P, Kovacevic J, Moura JM, Vandergheynst P. Graph signal processing: Overview, challenges, and applications. Proc IEEE 2018;106:808-28.
13. Sandryhaila A, Moura JM. Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure. IEEE Signal Process Mag 2014;31:80-90.
14. Medaglia JD, Huang W, Karuza EA, Kelkar A, Thompson-Schill SL, Ribeiro A, *et al*. Functional alignment with anatomical networks is associated with cognitive flexibility. Nat Hum Behav 2018;2:156-64.
15. Aviyente S, Villafañe-Delgado M. Graph signal processing on neuronal networks. In: Djurić PM, Richard C, editors. Cooperative and Graph Signal Processing: Principles and Applications. 1st ed. London, United Kingdom: Academic Press, an Imprint of Elsevier; 2018. p. 799-816.
16. Hammond DK, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory. Appl Comput Harmon Anal 2011;30:129-50.
17. Leonardi N, Van De Ville D. Tight wavelet frames on multislice graphs. IEEE Trans Signal Process 2013;61:3357-67.
18. Rui L, Nejati H, Cheung NM. Dimensionality Reduction of Brain Imaging Data Using Graph Signal Processing. In: 2016 Proc Int Conf Image Proc. Phoenix, USA; 2016:p. 1329-33.
19. Hu C, Sepulcre J, Johnson KA, Fakhri GE, Lu YM, Li Q. Matched signal detection on graphs: Theory and application to brain imaging data classification. Neuroimage 2016;125:587-600.
20. Stam CJ, Reijneveld JC. Graph theoretical analysis of complex networks in the brain. Nonlinear Biomed Phys 2007;1:19.
21. delEtoile J, Adeli H. Graph theory and brain connectivity in Alzheimer's disease. Neuroscientist 2017;23:616-26.
22. Kabbara A, Eid H, El Falou W, Khalil M, Wendling F, Hassan M. Reduced integration and improved segregation of functional brain networks in Alzheimer's disease. J Neural Eng 2018;15:1-13.
23. Dimitriadis SI. Graph signal processing of low and high-order dynamic functional connectivity networks using EEG resting-state for schizophrenia: A whole brain breakdown. BioRxiv 2019:1-26.
24. Tanaka T, Uehara T, Tanaka Y. Dimensionality Reduction of Sample Covariance Matrices by Graph Fourier Transform for Motor Imagery Brain-Machine Interface. In: 2016 IEEE Stat Signal Processing Workshop. IEEE. Palma de Mallorca, Spain;

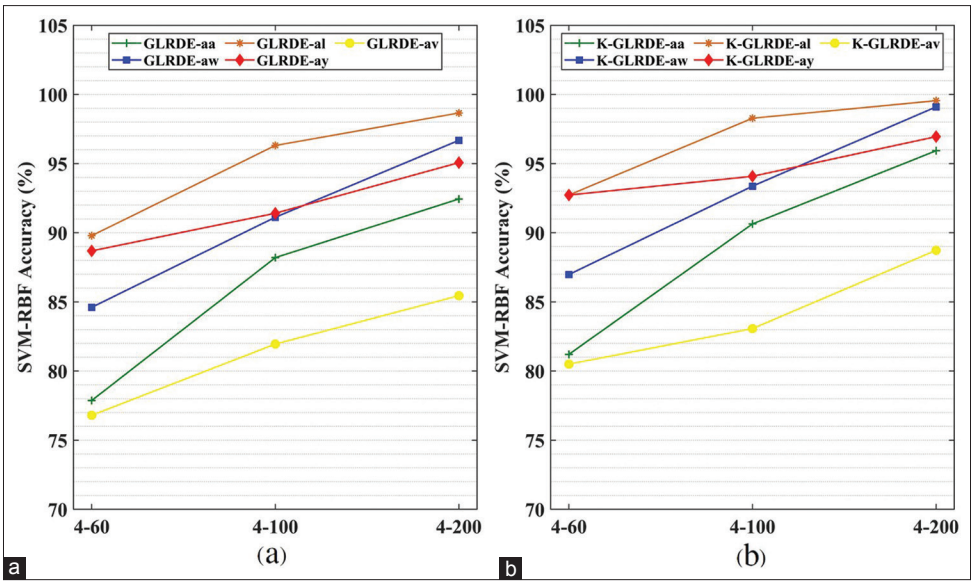
2016. p. 1-5.
25. Kalantar G, Sadreazami H, Mohammadi A, Asif A. Adaptive Dimensionality Reduction Method Using Graph-Based Spectral Decomposition for Motor Imagery-Based Brain-Computer Interfaces. IEEE Glob Conf Signal Inf Process. 2017; 990-4.
26. Kalantar G, Mohammadi A. Graph-based dimensionality reduction of EEG signals via functional clustering and total variation measure for BCI systems. Annu Int Conf IEEE Eng Med Biol Soc 2018;2018:4603-6.
27. Huang W, Goldsberry L, Wymbs NF, Grafton ST, Bassett DS, Ribeiro A. Graph frequency analysis of brain signals. IEEE J Sel Top Signal Process 2016;10:1189-203.
28. Khalili MD, Abootalebi V, Saeedi-Sourck H. Dimensionality reduction for motor imagery BCI system using kron reduction, graph Fourier transform and differential evolution. Iran J Biomed Eng 2022;16:75-94. ([Published in Persian]). Available from: https://www.ijbme.org/article_252164.html?lang=en. [Last accessed on 2025 Apr 04].
29. Müller KR, Blankertz B. Berlin Institute of Technology, Charitex-University Medicine Berlin. Available from: https://www.bbci.de/competition/iii/desc_IVa.html. [Last accessed on 2024 Jun 06].
30. Blankertz B, Müller KR, Krusienski DJ, Schalk G, Wolpaw JR, Schlögl A, *et al*. The BCI competition. III: Validating alternative approaches to actual BCI problems. IEEE Trans Neural Syst Rehabil Eng 2006;14:153-9.
31. Sandryhaila A, Moura JM. Classification Via Regularization on Graphs. IEEE Glob Conf Signal Inf Process. 2013; 495-8.
32. Behjat H, Leonardi N, Sörnmo L, Van De Ville D. Anatomically-adapted graph wavelets for improved group-level fMRI activation mapping. Neuroimage 2015;123:185-99.
33. Yu Q, Du Y, Chen J, Sui J, Adali T, Pearlson GD, *et al.* Application of graph theory to assess static and dynamic brain connectivity: Approaches for building brain graphs. Proc IEEE 2018;106:886-906.
34. Moran TP, Bernat EM, Aviyente S, Schroder HS, Moser JS. Sending mixed signals: Worry is associated with enhanced initial error processing but reduced call for subsequent cognitive control. Soc Cogn Affect Neurosci 2015;10:1548-56.
35. Caliskan SY, Tabuada P. Towards Kron reduction of generalized electrical networks. Automatica 2014;50:2586-90.
36. Dörfler F, Bullo F. Kron reduction of graphs with applications to electrical networks. IEEE Trans Circuits Syst I Regul Pap 2013;60:150-63.
37. Shuman DI, Faraji MJ, Vandergheynst P. A multiscale pyramid transform for graph signals. IEEE Trans Signal Process 2016;64:2119-34.
38. Zhang F. The Schur Complement and Its Applications. Vol. 4. New York, USA: Springer Science and Business Media; 2006.
39. Soong AC, Koles ZJ. Principal-component localization of the sources of the background EEG. IEEE Trans Biomed Eng 1995;42:59-67.
40. Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process Mag 2008;25:41-56.
41. Lotte F, Guan C. Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms. IEEE Trans Biomed Eng 2011;58:355-62.
42. Padfield N, Zabalza J, Zhao H, Masero V, Ren J. EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges. Sensors (Basel) 2019;19:1423.
43. Khushaba RN, Al-Ani A, Al-Jumaily A. Feature subset selection using differential evolution and a statistical repair mechanism. Expert Syst Appl 2011;38:11515-26.
44. Baig MZ, Aslam N, Shum HP, Zhang L. Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG. Expert Syst Appl 2017;90:184-95.
45. Qin AK, Huang VL, Suganthan PN. Differential evolution algorithm with strategy adaptation for global numerical optimization. IEEE Trans Evol Comput 2009;13:398-417.
46. Ramoser H, Müller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Rehabil Eng 2000;8:441-6.
47. Pfurtscheller G, Neuper C, Flotzinger D, Pregenzer M. EEG-based discrimination between imagination of right and left hand movement. Electroencephalogr Clin Neurophysiol 1997;103:642-51.
48. Blankertz B, Curio G, Müller KR. Classifying Single Trial EEG: Towards Brain Computer Interfacing. Adv Neural Inf Process Syst. 2001;14:157-64.
49. Xygonakis I, Athanasiou A, Pandria N, Kugiumtzis D, Bamidis PD. Decoding motor imagery through common spatial pattern filters at the EEG source space. Comput Intell Neurosci 2018;2018:1-10.
50. Yazici M, Ulutas M, Okuyan M. A comprehensive sLORETA study on the contribution of cortical somatomotor regions to motor imagery. Brain Sci 2019;9:372.
51. van den Heuvel MP, Sporns O. Network hubs in the human brain. Trends Cogn Sci 2013;17:683-96.
52. Park Y, Chung W. Optimal channel selection using correlation coefficient for CSP based EEG classification. IEEE Access 2020;8:111514-21.
53. Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. Single-trial analysis and classification of ERP components – A tutorial. Neuroimage 2011;56:814-25.
54. Lotte F. Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces. Proc IEEE 2015;103:871-90.
55. Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, *et al*. A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update. J Neural Eng 2018;15:1-28.



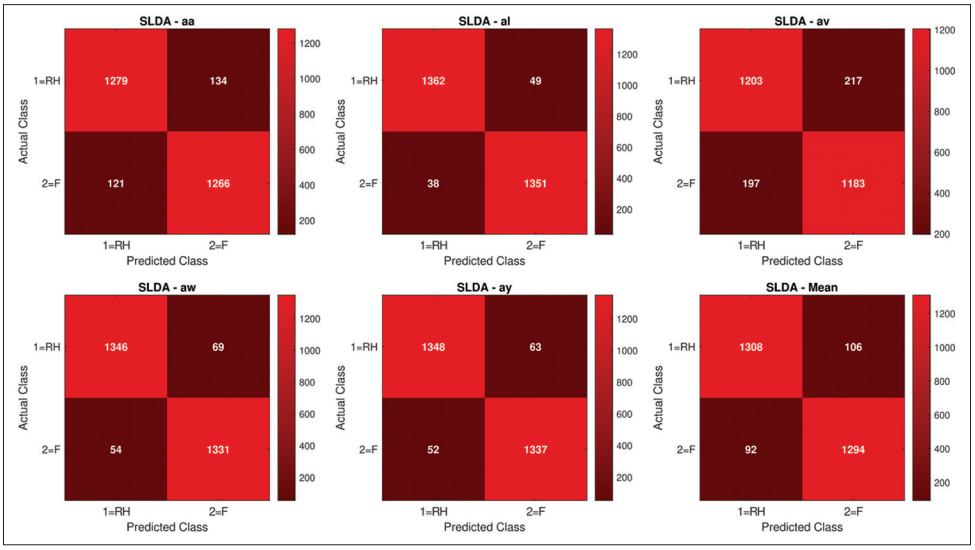
Supplementary Figure 1: Eigenvalue distributions of 77-vertex vs. Kron-reduced (24-vertex) graphs Laplacians across all five subjects



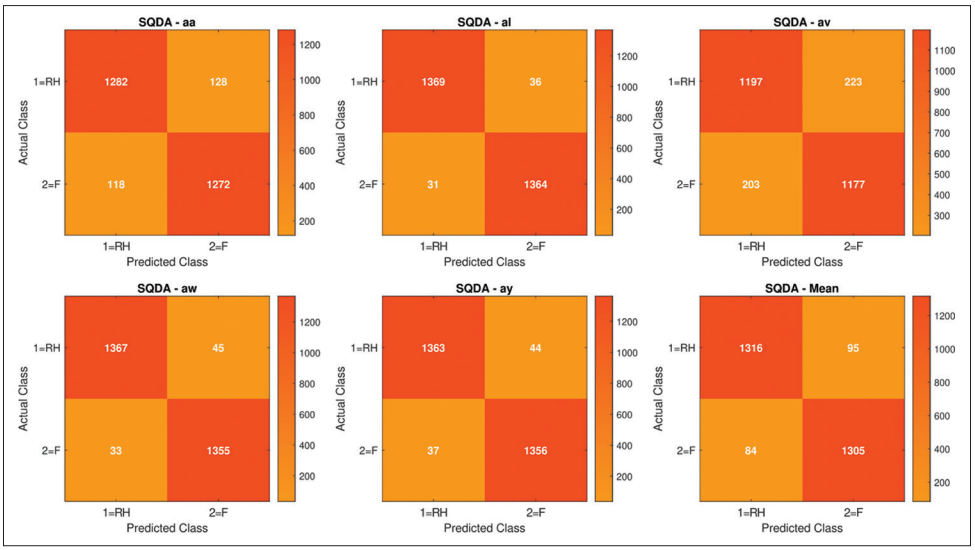
Supplementary Figure 2: Classification performance trends across varying training trials (60, 100, 200) using two spatial filters in: (a) GLR-DE and (b) K-GLR-DE approaches



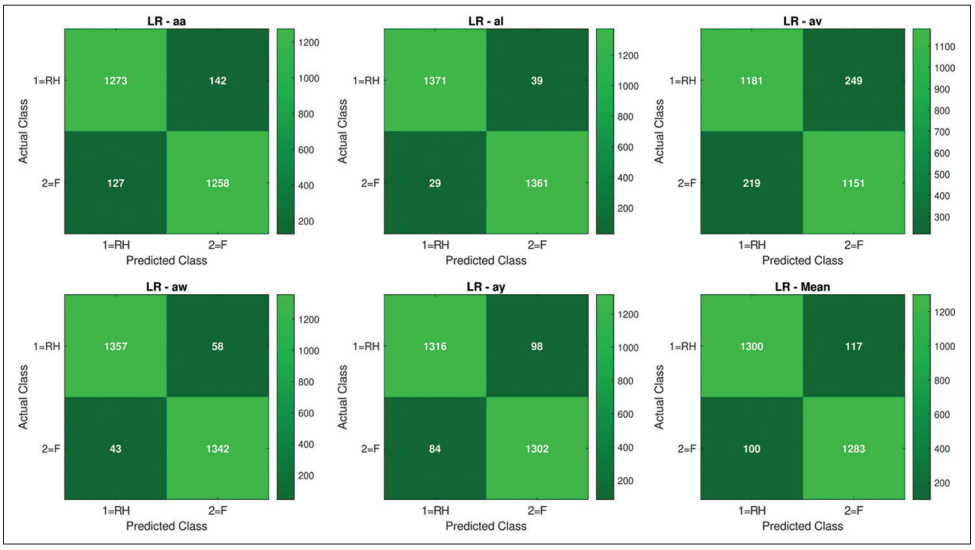
Supplementary Figure 3: Classification performance trends across varying training trials (60, 100, 200) using four spatial filters in: (a) GLR-DE and (b) K-GLR-DE approaches



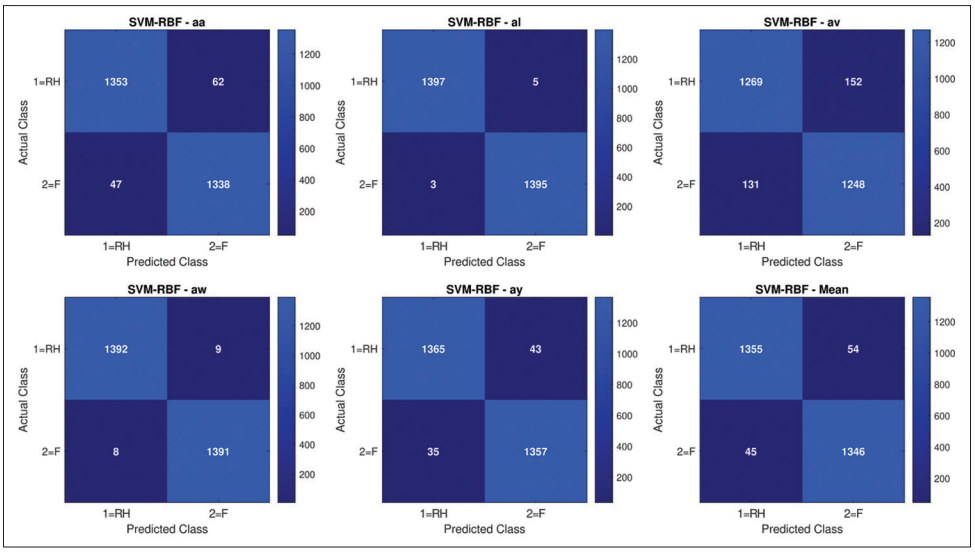
Supplementary Figure 4: Confusion matrices for the SLDA classifier across all five subjects (aa, al, av, aw, ay) and their mean. Results obtained using the best setting: 4 spatial filters and 10 × 10-fold CV



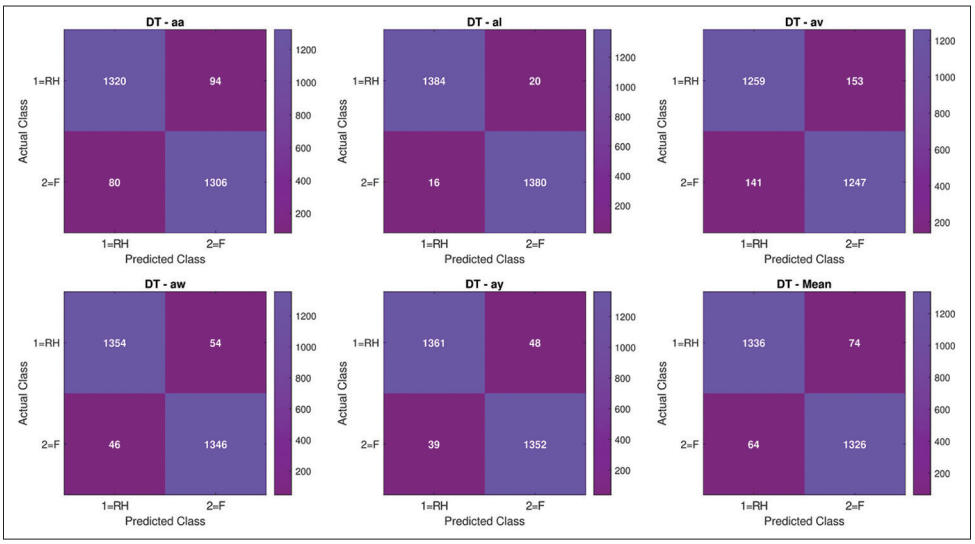
Supplementary Figure 5: Confusion matrices for the SQDA classifier across all five subjects (aa, al, av, aw, ay) and their mean. Results obtained using the best setting: 4 spatial filters and 10 × 10-fold CV



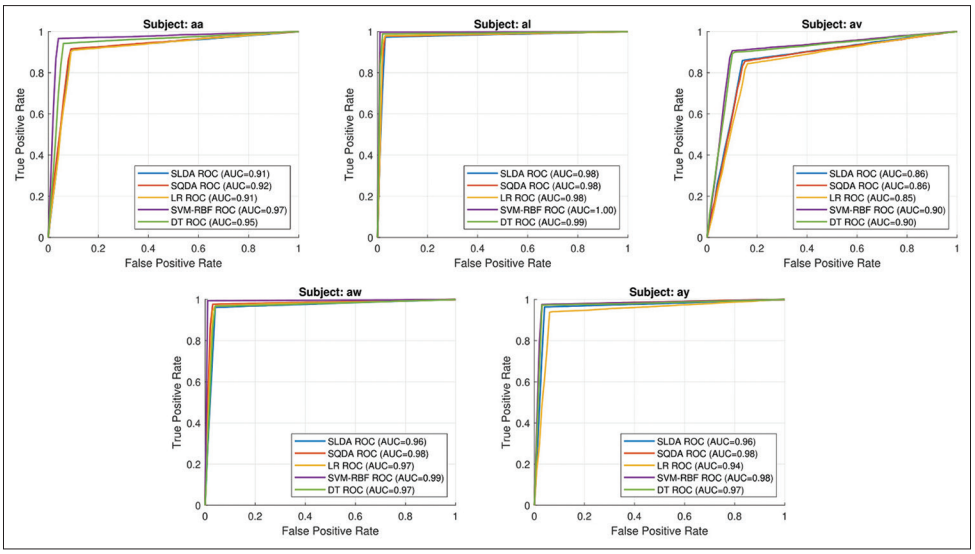
Supplementary Figure 6: Confusion matrices for the LR classifier across all five subjects (aa, al, av, aw, ay) and their mean. Results obtained using the best setting: 4 spatial filters and 10 × 10-fold CV



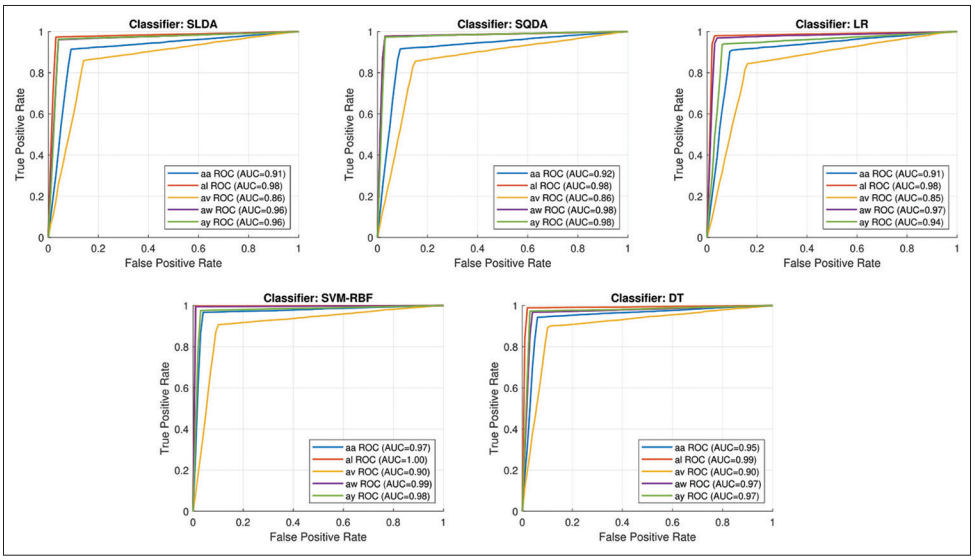
Supplementary Figure 7: Confusion matrices for the SVM-RBF classifier across all five subjects (aa, al, av, aw, ay) and their mean. Results obtained using the best setting: 4 spatial filters and 10 × 10-fold CV



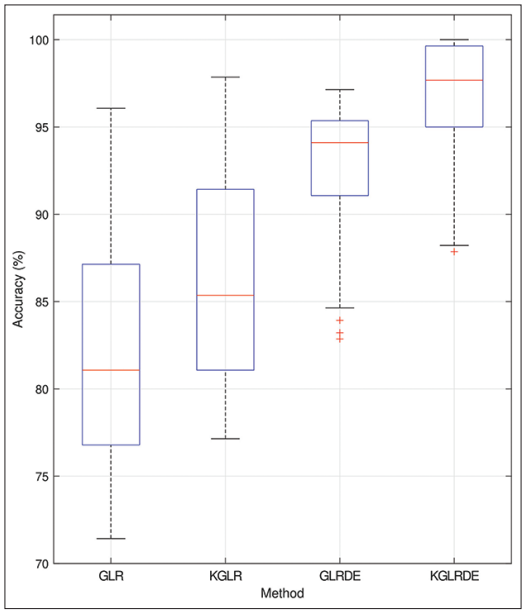
Supplementary Figure 8: Confusion matrices for the DT classifier across all five subjects (aa, al, av, aw, ay) and their mean. Results obtained using the best setting: 4 spatial filters and 10 × 10-fold CV



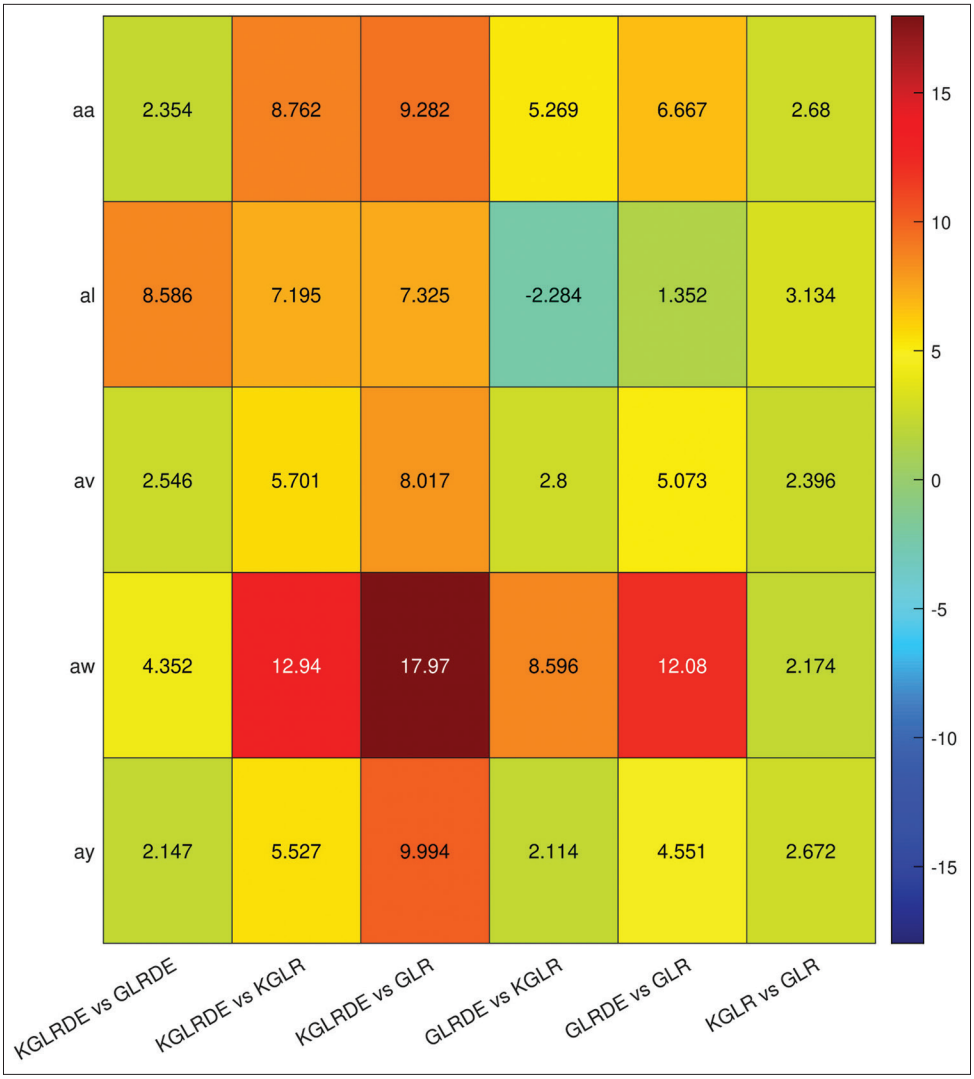
Supplementary Figure 9: ROC curves for all five subjects (aa, al, av, aw, ay) across the five classifiers. Each subplot represents one subject, showing ROC curves over classifiers. (Using the best setting: 4 spatial filters, 10 × 10-fold CV)



Supplementary Figure 10: ROC curves for all five classifiers (SLDA, SQDA, LR, SVM-RBF, DT) across the five subjects. Each subplot represents one classifier, showing ROC curves over subjects. (Using the best setting: 4 spatial filters, 10 × 10-fold CV)



Supplementary Figure 11: Classification accuracy boxplots for the four methods (GLR, K-GLR, GLR-DE, and K-GLR-DE) using the SVM-RBF classifier with 10 × 10-fold CV



Supplementary Figure 12: Pairwise effect sizes (Cohen's d) among the four methods (GLR, K-GLR, GLR-DE, and K-GLR-DE) using the SVM-RBF classifier with 10 × 10-fold CV

Supplementary Table 1: Classification accuracy (mean±standard deviation) across different numbers of selected vertices in 4-K-GLR-DE approach with 10×10-fold cross-validation

Subject	Top 12-SVM	Top 16-SVM	Top 20-SVM	Top 24-SVM	Top 28-SVM	Top 32-SVM
aa	85.82±0.55	88.54±0.35	92.29±0.92	**96.11±1.20**	94.46±0.51	93.11±0.78
al	89.79±0.45	91.75±0.60	95.39±0.33	**99.71±0.14**	97.82±0.51	96.04±0.55
av	82.32±0.48	83.79±0.56	86.54±0.58	**89.89±1.36**	87.36±0.56	86.89±0.58
aw	88.04±0.50	90.04±0.46	94.64±0.62	**99.39±0.39**	96.00±0.85	94.39±0.63
ay	86.00±0.72	88.93±0.64	93.46±0.84	**97.21±0.98**	95.07±0.74	93.86±0.88
Mean±SD	86.39±0.54	88.60±0.52	92.46±0.66	**96.46±0.81**	94.14±0.63	92.86±0.68

SD – Standard deviation; SVM – Support vector machine

Supplementary Table 2: Quantitative spectral preservation metrics for Kron reduction across all five subjects

Subject	Relative spectral distance (L2 norm)	Spectral correlation coefficient	Spectral energy retention ratio
aa	0.1540	0.9895	0.7942
al	0.1359	0.9942	0.7929
av	0.1523	0.9916	0.8091
aw	0.1439	0.9920	0.7891
ay	0.0908	0.9934	0.9773
Mean±SD	0.1354±0.0259	0.9921±0.0018	0.8325±0.0813

SD – Standard deviation

Supplementary Table 3: Classification accuracy (mean±standard deviation) in K-GLR approach using different feature selection methods with the best setting. (4 spatial filters, SVM-RBF classifier, and 10×10-fold cross-validation)

Subject	K-GLR + DE	K-GLR + PCA	K-GLR + mutual information
aa	**96.11±1.20**	90.54±2.54	92.36±2.14
al	**99.71±0.14**	93.00±1.00	95.32±0.99
av	**89.89±1.36**	82.96±3.46	85.64±2.71
aw	**99.39±0.39**	93.46±1.50	94.64±1.73
ay	**97.21±0.98**	91.82±1.73	93.29±1.51
Mean±SD	**96.46±0.81**	90.36±2.05	92.25±1.82

SD – Standard deviation, GLR-DE – Generic learning regularization with differential evolution, PCA – Principal component analysis

Supplementary Table 4: Sensitivity analysis of differential evolution, averaged across five subjects using the best setting. (4 spatial filters, SVM-RBF classifier, and 10×10-fold cross-validation)

DE parameter	Values (+other defaults)	Accuracy (mean±SD)
Population size	25	92.89±1.51
	50 (default)	96.46±0.81
	75	96.79±1.02
	100	96.68±1.48
Number of generations	200	94.54±1.35
	400 (default)	96.46±0.81
	600	96.36±1.49
Mutation factor (F)	0.5	91.86±1.57
	0.75 (default)	96.46±0.81
	1.0	95.11±1.18
CR	0.3	91.21±1.25
	0.5	95.46±1.07
	0.7 (default)	96.46±0.81
	0.9	95.29±0.98

DE – Differential evolution; CR – Crossover rate; SD – Standard deviation

Supplementary Table 5: Final SVM-RBF hyperparameters selected via 10 fold CV

Subject	Selected C	Selected σ
aa	10	0.01
al	100	0.001
av	1	0.1
aw	10	0.01
ay	100	0.01

Supplementary Table 6: Classification accuracy (mean±standard deviation) in 4-GLR, 4-K-GLR, 4-GLR-DE, and 4-K-GLR-DE (using 4 spatial filters, SVM-RBF classifier, and 10×10-fold CV)

Subjects	4-GLR	4-K-GLR	4-GLR-DE	4-K-GLR-DE
aa	81.21±1.93	85.54±1.22	92.86±1.54	**96.11±1.20**
al	94.32±1.03	96.89±0.53	95.50±0.68	**99.71±0.14**
av	74.61±2.33	79.89±2.08	85.57±1.98	**89.89±1.36**
aw	76.93±1.73	80.96±1.98	95.29±1.28	**99.39±0.39**
ay	86.14±1.22	89.93±1.59	93.79±2.04	**97.21±0.98**
Mean±SD	82.64±1.65	86.64±1.48	92.61±1.50	**96.46±0.81**

GLR – Generic learning regularization; DE – Differential evolution; SD – Standard deviation

Supplementary Table 7: Feature importance analysis

Subjects	TV_G only 1	GLRCSP only 1 (4GLR)	TV_G only 2 (K + TV_G + DE)	GLRCSP only 2 (K +4GLR)	TV_G + GLRCSP (K +4GLR + TV_G + DE)
aa	72.46±2.59	78.39±1.95	87.07±1.28	83.18±1.24	**96.11±1.20**
al	85.96±1.65	91.86±1.59	94.39±0.73	94.96±0.46	**99.71±0.14**
av	66.14±2.92	72.25±2.50	80.89±1.57	76.50±1.48	**89.89±1.36**
aw	68.71±2.32	74.21±1.99	87.54±0.93	79.18±0.56	**99.39±0.39**
ay	77.79±1.87	83.82±1.77	89.75±1.09	87.79±0.98	**97.21±0.98**
Mean±SD	74.21±2.27	80.11±1.96	87.93±1.12	84.32±0.94	**96.46±0.81**

GLR – Generic learning regularization; DE – Differential evolution; SD – Standard deviation; TV_G - Graph-based Total variation; CSP – Common spatial patterns

Supplementary Table 8: Classification sensitivity (mean±standard deviation) in 4-K-GLR-DE approach with 10×10-fold CV

Sensitivity in 4-K-GLR-DE (10×10-fold CV)					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	91.28±1.44	91.51±2.12	90.83±2.42	**96.61±1.56**	94.23±1.36
al	97.26±0.66	97.77±0.74	97.91±0.67	**99.79±0.32**	98.85±0.81
av	85.72±2.37	85.29±1.56	84.01±2.47	**90.50±2.37**	89.84±1.93
aw	96.10±0.57	97.62±0.59	96.90±0.69	**99.43±0.53**	96.70±0.66
ay	96.26±1.65	97.34±1.61	93.94±2.92	**97.49±1.56**	97.20±1.29
Mean±SD	93.32±1.34	93.91±1.32	92.72±1.83	**96.76±1.27**	95.36±1.21

GLR – Generic learning regularization; DE – Differential evolution; SD – Standard deviation; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; SVM-RBF – Support vector machine with a radial basis function; CV – Cross-validation; LR – Logistic regression; DT – Decision tree

Supplementary Table 9: Classification specificity (mean±standard deviation) in 4-K-GLR-DE approach with 10×10-fold CV

Specificity in 4-K-GLR-DE (10×10-fold CV)					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	90.52±1.43	90.92±1.54	89.96±1.34	**95.61±2.42**	93.35±2.24
al	96.53±1.63	97.44±1.75	97.23±1.51	**99.64±0.36**	98.58±0.73
av	84.72±1.92	84.30±2.52	82.59±2.25	**89.30±2.35**	89.16±1.37
aw	95.12±0.79	96.81±1.18	95.90±1.72	**99.36±0.50**	96.16±0.54
ay	95.54±1.29	96.87±2.53	93.07±3.06	**96.95±2.69**	96.59±1.36
Mean±SD	92.49±1.41	93.27±1.90	91.75±1.98	**96.17±1.63**	94.77±1.25

GLR – Generic learning regularization; DE – Differential evolution; SD – Standard deviation; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; SVM-RBF – Support vector machine with a radial basis function; CV – Cross-validation; LR – Logistic regression; DT – Decision tree

Supplementary Table 10: Classification precision (mean±standard deviation) in 4-K-GLR-DE approach with 10×10-fold CV

Precision in 4-K-GLR-DE (10×10-fold CV)					
Subjects	**SLDA**	**SQDA**	**LR**	**SVM-RBF**	**DT**
aa	90.43±2.43	90.86±1.56	89.86±1.71	**95.57±1.64**	93.29±1.46
al	96.50±0.64	97.43±1.62	97.21±1.44	**99.64±0.49**	98.57±0.64
av	84.50±1.66	84.07±2.56	82.21±1.77	**89.14±2.77**	89.07±2.44
aw	95.07±1.10	96.79±1.36	95.85±1.80	**99.36±0.22**	96.14±0.67
ay	95.50±1.65	96.86±2.19	93.00±2.79	**96.93±1.50**	96.57±2.19
Mean±SD	92.40±1.50	93.20±1.86	91.63±1.90	**96.13±1.32**	94.73±1.48

GLR – Generic learning regularization; DE – Differential evolution; SD – Standard deviation; SLDA – Shrinkage linear discriminant analysis; SQDA – Shrinkage quadratic discriminant analysis; SVM-RBF – Support vector machine with a radial basis function; CV – Cross- validation; LR – Logistic regression; DT – Decision tree

Supplementary Table 11: One-way ANOVA results (*F*-statistics, *p*-value) across all subjects

One-way ANOVA		
Subject	***F*-statistic**	***p***
aa	204.1812	1.1907e-22
al	118.0465	1.0980e-18
av	114.5841	1.7823e-18
aw	545.3551	4.7795e-30
ay	100.5728	1.4633e-17

Supplementary Table 12: Paired *t*-test results (*t*-value, *p*-value) with Bonferroni correction across all subjects

Paired *t*-tests (Bonferroni corrected, significance level=0.01)			
Comparison	**|*t*|-value**	***p*-value (raw)**	***p*-value (adjusted)**
Subject aa			
KGLR versus GLR	17.526	2.9003e-08	1.7402e-07
GLRDE versus GLR	81.500	3.1919e-14	1.9151e-13
GLRDE versus KGLR	55.007	1.0910e-12	6.5458e-12
KGLRDE versus GLR	62.472	3.4803e-13	2.0882e-12
KGLRDE versus KGLR	96.889	6.7408e-15	4.0445e-14
KGLRDE versus GLRDE	22.365	3.3811e-09	2.0287e-08
Subject al			
KGLR versus GLR	15.429	8.8306e-08	5.2983e-07
GLRDE versus GLR	8.3373	1.5890e-05	9.5343e-05
GLRDE versus KGLR	21.726	4.3695e-09	2.6217e-08
KGLRDE versus GLR	18.356	1.9316e-08	1.1590e-07
KGLRDE versus KGLR	20.867	6.2436e-09	3.7462e-08
KGLRDE versus GLRDE	23.043	2.5947e-09	1.5568e-08
Subject av			
KGLR versus GLR	41.224	1.4491e-11	8.6946e-11
GLRDE versus GLR	59.327	5.5344e-13	3.3207e-12
GLRDE versus KGLR	68.143	1.5946e-13	9.5675e-13
KGLRDE versus GLR	42.895	1.0151e-11	6.0905e-11
KGLRDE versus KGLR	39.165	2.2928e-11	1.3757e-10
KGLRDE versus GLRDE	20.013	9.0283e-09	5.4170e-08
Subject aw			
KGLR versus GLR	26.717	6.9747e-10	4.1848e-09
GLRDE versus GLR	120.410	9.5457e-16	5.7274e-15
GLRDE versus KGLR	59.485	5.4041e-13	3.2424e-12
KGLRDE versus GLR	51.377	2.0134e-12	1.2080e-11
KGLRDE versus KGLR	35.476	5.5561e-11	3.3337e-10
KGLRDE versus GLRDE	14.026	2.0179e-07	1.2107e-06
Subject ay			
KGLR versus GLR	28.557	3.8541e-10	2.3125e-09
GLRDE versus GLR	27.526	5.3490e-10	3.2094e-09
GLRDE versus KGLR	23.143	2.4975e-09	1.4985e-08
KGLRDE versus GLR	120.060	9.7971e-16	5.8782e-15
KGLRDE versus KGLR	35.101	6.1109e-11	3.6666e-10
KGLRDE versus GLRDE	9.798	4.2397e-06	2.5438e-05

GLR – Generic learning regularization; DE – Differential evolution; K - Kron reduction

Supplementary Table 13: Classification accuracy of the K-GLR-DE method for motor imagery tasks (right fist vs. left fist) across 20 subjects (S001–S020) from the PhysioNet-eegmmidb dataset (Using four spatial filters, SVM-RBF classifier, and 10×10-fold CV)

Subject	K-GLR-DE accuracy (mean±SD)	Subject	K-GLR-DE accuracy (mean±SD)
S001	85.00±4.77	S011	87.62±2.19
S002	84.52±4.24	S012	82.38±2.56
S003	85.71±2.24	S013	79.29±1.96
S004	83.81±3.51	S014	81.19±1.35
S005	81.66±3.38	S015	80.71±2.37
S006	83.33±2.51	S016	86.66±2.01
S007	84.05±3.73	S017	87.38±1.61
S008	85.95±1.76	S018	85.48±2.08
S009	86.43±1.96	S019	77.86±1.96
S010	86.19±1.51	S020	87.14±2.56
		Mean±SD	**84.12±2.51**

GLR – Generic learning regularization; DE – Differential evolution; SD – Standard deviation; K - Kron reduction