Computers in Biology and Medicine 153 (2023) 106498

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A dynamical graph-based feature extraction approach to enhance mental task classification in brain–computer interfaces

Shaotong Zhu [a], Sarah Ismail Hosni [b], Xiaofei Huang [a], Michael Wan [a], Seyyed Bahram Borgheai [b], John McLinden [b], Yalda Shahriari [b], Sarah Ostadabbas [a,*]

[a] The Department of Electrical and Computer Engineering, Northeastern University, 360 Huntington Ave, Boston, MA 02115, USA
[b] The Electrical, Computer, and Biomedical Engineering Department, University of Rhode Island, Kingston, RI 02881, USA

ARTICLE INFO

Keywords:
Brain–computer interfaces (BCI)
Graph theory
Feature selection
Mental task classification

ABSTRACT

Graph theoretic approaches in analyzing spatiotemporal dynamics of brain activities are under-studied but could be very promising directions in developing effective brain–computer interfaces (BCIs). Many existing BCI systems use electroencephalogram (EEG) signals to record and decode human neural activities noninvasively. Often, however, the features extracted from the EEG signals ignore the topological information hidden in the EEG temporal dynamics. Moreover, existing graph theoretic approaches are mostly used to reveal the topological patterns of brain functional networks based on synchronization between signals from distinctive spatial regions, instead of interdependence between states at different timestamps. In this study, we present a robust fold-wise hyperparameter optimization framework utilizing a series of conventional graph-based measurements combined with spectral graph features and investigate its discriminative performance on classification of a designed mental task in 6 participants with amyotrophic lateral sclerosis (ALS). Across all of our participants, we reached an average accuracy of 71.1% ± 4.5% for mental task classification by combining the global graph-based measurements and the spectral graph features, higher than the conventional non-graph based feature performance (67.1% ± 7.5%). Compared to using either one of the graphic features (66.3% ± 6.5% for the eigenvalues and 65.9% ± 5.2% for the global graph features), our feature combination strategy shows considerable improvement in both accuracy and robustness performance. Our results indicate the feasibility and advantage of the presented fold-wise optimization framework utilizing graph-based features in BCI systems targeted at end-users.

1. Introduction

Decoding the neural responses for brain–computer interface (BCI)-based communication is of substantial importance for improving the quality of life for people with severe disabilities such as later-stage amyotrophic lateral sclerosis (ALS) patients who gradually lose all their voluntary muscle control.

This is especially important for ALS patients in the later stages of the disease, when eye gaze control begins to fail, impacting the efficacy of communication using visual BCIs [1,2] and other methods dependent on eye gaze, such as eye trackers [3]. Current augmentative and alternative communication (AAC) technologies for individuals in the later stages of ALS are insufficient, although previous BCI studies have proposed multiple means to augment established visual paradigms [2, 4].

From a topological perspective, mapping the human brain connectivity using powerful tools based on graph theory could shed a light on the nature of interactions between different brain regions. These approaches have gained plenty of attention in BCI research for representing brain networks of different mental tasks [5,6].

Since the late 1990s when the small-world-ness and scale-free network models [7,8] were proposed, graph theory has been increasingly used to investigate the topological patterns of the brain networks to discover meaningful functional brain activities. In general, the graph theory approach considers the sources of signal sequences as nodes and the measurement of sequences' interdependence as the edge between nodes. These inherent topological patterns of brain activities are represented via the adjacency matrix that enables the characterization of the statistical interdependence of physiological time series between each pair of brain regions, which cannot be achieved by simple linear methods. Typical graph measurements can be calculated from the adjacency matrix for the quantitative investigation of the brain connectivity network depending on the mental tasks and brain signal modalities.

* Corresponding author.
E-mail address: ostadabbas@ece.neu.edu (S. Ostadabbas).

https://doi.org/10.1016/j.compbiomed.2022.106498
Received 29 August 2022; Received in revised form 8 December 2022; Accepted 27 December 2022
Available online 30 December 2022

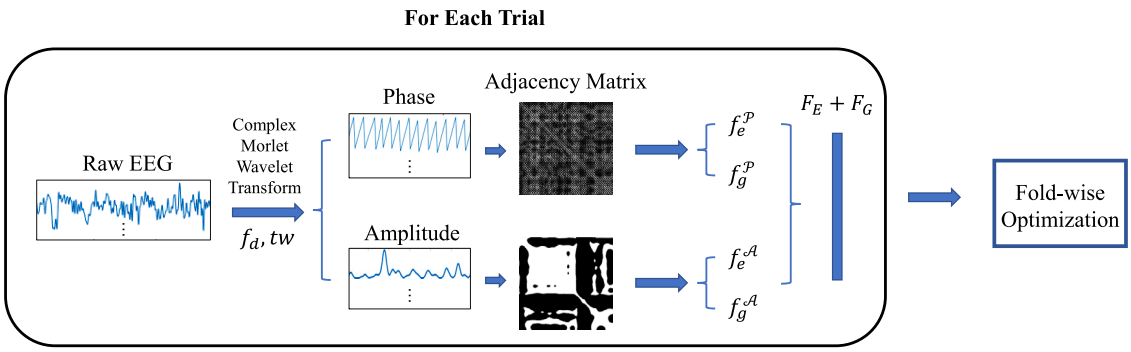

Fig. 1. An overview of the proposed graph-based feature extraction approach. Each trial of a given subject consists of a 6-s EEG data frame bandpassed through 0.5–30 Hz. The input EEG frame is prepossessed under the given time window tw and downsampling factor f_d, then transform into corresponding instantaneous phase and amplitude data frames based on which the adjacency matrices are calculated. $f_e^{\mathcal{P}}$ and $f_e^{\mathcal{A}}$ are the thresholded eigenvalues extracted from the phase-based graph and the amplitude-based graph noted as F_E. $f_g^{\mathcal{P}}$ and $f_g^{\mathcal{A}}$ are the global graph features from the corresponding graphs above, noted as F_G. The features from all trials are then used for the classification with fold-wise optimization.

The essence of any BCI system lies in the interpretation of dynamic cognitive processes to specific the corresponding brain activities. Therefore, a high temporal resolution technique to capture the dynamics of the brain network is necessary. A modern electroencephalogram (EEG) system can capture the temporal dynamics of brain activity on the sub-second scale, which enables it to reflect rapid changes within the brain network. EEG is capable of capturing rich temporal information that aids identification of the directions of the flow of information among different brain regions. Furthermore, EEG systems are noninvasive, portable, wireless, and easy to use, making them attractive and applicable to neuroergonomics studies [9]. There have been ongoing efforts to exploit graph theory in the field of BCI system development for a great variety of cognitive [10] and physical tasks [11] with promising results. However, current graph-theoretical approaches for BCI systems are mostly based on functional connectivity features, which have been limited to quantifying the neural dynamics on a spatial level using dependence between individual electrodes. The pipeline of generating graph-based features for the BCI system generally considers the sources of the temporal signal sequence (such as EEG electrodes) as the nodes. Edges are represented by the statistical measures of association, including anatomical, functional, or effective relationships between brain regions [12,13]. The temporal dynamics encoded in these complex signals are not fully explored using these graph-based approaches. The spatial graph strategy generates the graph via a temporally independent statistical method that neglects the temporal interdependence existing in the signal sequences.

To capture the advantages of the graph theory while incorporating the essence of the brain's temporal dynamics in the model, we present a spectral graph theoretic method, which is guided by the EEG temporal dynamic characteristics while performing mental calculations [14] and considers the timestamps as graph nodes, and the radial based distance between the values of all EEG channels at given two timestamps as the edge between two nodes. Two types of graph-based measures were extracted and combined after the graph was formed. Global graph measures, such as characteristic path length and clustering coefficient, and spectral features including eigenvalues of the adjacency matrix were extracted and combined to reveal the temporal synchronization via graph structural properties. In sum, this paper presents the following contributions:

- Introducing a new perspective for graph generation in which temporal dynamics within amplitude and phase instantaneous signals are used to build the corresponding graph;
- Proposing a novel combination of graph measurements and eigenvalue features from EEG data towards developing robust mental task classification in BCI systems;
- Presenting a fold-wise parameter optimization scheme that selects the optimal hyperparameters for each training process to account for the EEG nonstationarity and neural variability observed within subjects.
- Detailedly evaluating the proposed pipeline with ablation studies on involved parameters and classifiers to demonstrate its reasonability.

2. Related work

In this section, we summarize the existing works that utilize graph theory in the fields involving functional brain connectivity analysis. Then, we discuss the research gap that current graph-based approaches rarely utilize the temporal dynamic patterns in the brain signals, which is addressed by our proposed graph-based feature extraction approach.

2.1. Spatial graph theory for functional brain connectivity analysis

Graph theory has been applied in multiple types of EEG connectivity related studies to reveal the latent functional network patterns between signals from the topological perspective, which cannot be achieved via simple linear methods. The general pipeline for processing EEG signals via graph theory is as follows [15]: first, define the nodes of graphs as the EEG electrodes, whereas the edges represent statistical measures of association, such as phase-locking value (PLV) [16], partial directed coherence (PDC) [17], or phase lag index (PLI) [18]. Then, form the connectivity matrix using the states of edges and convert the connectivity matrix into a binarized adjacency matrix via a chosen threshold value. Finally, estimate the typical graph measurements including global measurements such as characteristic path length, or nodal measurements such as local efficiency. A comprehensive review of common measurements is provided in [19,20].

The application of this pipeline in the quantification of EEG data has gained much attention in the clinical field and in neuroscience for the characterization and diagnosis of brain disorders. Hann et al. utilized the synchronization likelihood between EEG signals from pairs of electrodes as the graph edges and found the clustering coefficient decreased in the lower alpha and beta bands ($p < 0.001$), and the characteristic path length decreased in the lower alpha and gamma bands ($p < 0.05$) for the patients with Alzheimer's disease compared to controls [21]. Wang et al. investigated the resting neural networks of hemianopia stroke patients using graphs constructed with phase synchronization indexes and found the left primary visual cortex of patients tends to be less active than that of healthy subjects [22]. Ponten and others [23] studied EEG recordings from patients with seizures and reported that absences are characterized by an increase in synchronization and that the functional network topology changed

towards a more ordered pattern compared to pre-ictal network configuration. All these studies revealed informative spatial topological patterns for the corresponding investigation topic by applying the graph theory which considers the electrodes as nodes and the time series correlation between two different nodes as edges.

The graph theoretic approaches for functional brain connectivity can also play a role in engineering applications aiming to characterize different brain states or recognize mental intentions from EEG data, which aligns with the essence of BCI systems. Many efforts have been made to apply graph theory to the BCI field. Daly et al. [24] achieved the first application of the graph theoretic approach in the motor imagery (MI) based BCI system. The author assessed the discrimination ability of mean clustering coefficients to differentiate between tapping and non-tapping, in real and imagined finger-tapping tasks. Uribe et al. [25] investigated the potential of centrality measures to discriminate between left and right-hand MI via the graph based on the difference between each pair of symmetric electrodes across hemispheres. The graph measurements including degree, betweenness, and eigenvector centrality were used to provide information regarding functional connectivity under different mental tasks. In the work of Cattai et al. [26], the author revealed brain signal amplitude/phase synchronization mechanisms during EEG-based MI vs rest tasks and detected specific brain network changes associated with MI.

Other studies suggested dynamic functional connectivity approaches in which sequences of spatial functional connectivity features were captured across shorter slices of EEG signals as an attempt to include the temporal dynamics of the interaction between different brain regions [27,28]. However, these approaches consider brain regions as nodes and the synchronizations between given-length of EEG signals from two nodes as edges, which address the spatial relationships between nodes over a period of time instead of temporal dynamics directly. The resolution of temporal dynamics for these methods depends on the EEG signal segment length, which is much lower than the temporal graph that considers timestamps as nodes and represents the temporal dynamics in the same temporal resolution as the input EEG signals.

2.2. Spectral graph theory and temporal dynamic detection

The conventional graph theoretic approaches, which consider signal sources as graph nodes and the statistical measures of association between signal slices from a pair of sources as graph edges, are valuable while investigating the overall topological patterns of a network from the spatial perspective. This strategy does not address the interdependence between the states at different timestamps which contains the temporal dynamic pattern shifting under different mental states. The spectral graph theoretic approach has been widely applied in the system design and optimization fields, such as network monitoring in real-time [29], power system vulnerability identification [30], and bearing fault detection and diagnosis [31], by building and analyzing temporal graphs that consider the timestamps as nodes and temporal synchronization between states at distinctive timestamps as edges.

Compared to the conventional node-wise graph measurement, spectral graph theory explores the properties of a graph in relation to the characteristic polynomial, eigenvalues, and eigenvectors of matrices associated with the graph. Tootooni et al. applied spectral graph theory to monitor complex dynamic processes using the Fiedler number from a sliding window as a graph measurement [32]. While spectral graph theoretic approaches show great advantages in monitoring dynamic characteristics over a process, utilizing the spectral graph theory for brain functional connectivity analysis is still rarely explored. Following the same method as [32], Fan et al. [33] proposed a spectral graph theoretic approach to detect abnormal patterns of epileptic seizures. Elisha et al. achieved seizure classification using EEG features from the spectral graph decomposition [34]. Ghaderi et al. explored the synchrony and complexity in the EEG-based brain functional network via the spectral graph theoretic approach [35]. Kirar et al. further explored combining spectral graph theory with a quantum genetic algorithm to build a BCI system for motor imagery classification [36].

In our work, we introduce the spectral graph theoretic approach into the field of BCI systems. We consider it an opportunity to combine the conventional global graph measures of temporal graphs with spectral graph theoretic features to analyze the temporal dynamic patterns of EEG data to be able to better distinguish different mental states.

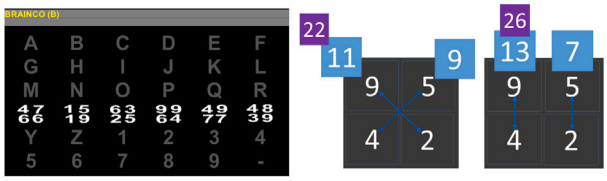

Fig. 2. The subjects are instructed to perform mental calculations using the matrix of numbers intensified over the target character. The calculation included a simple addition/subtraction either diagonally (at the first flash) or vertically (at the second flash) within the intensified matrix followed by doubling the higher result. The stimulation time was set to 300 ms, and the interstimulation interval (ISI) was set to six seconds. In total, for the VM paradigm, there were 14 target characters per run with two flashes (one row and one column flash) for each character.

3. Subjects, materials, and methods

3.1. Participants, data acquisition, and experimental protocol

Six participants with ALS were recruited for this study (age 57.0 ± 15.7 years, one female) with varying degrees of disability. The study protocol was approved by the Institutional Review Board (IRB) of the University of Rhode Island (URI) and all subjects provided informed consent or assent for the study and received financial compensation. EEG data were recorded from eight Ag/AgCl electrodes referenced to the left earlobe: Fz, Cz, P3, Pz, P4, PO7, PO8, and Oz covering all frontal, central, parietal, and occipital areas [37]. The data was recorded using a g.USBamp amplifier (g.tec Medical Tech.), digitized at 256 Hz and zero-phase bandpass filtered (1 – 30 Hz). The experimental protocol adopted in this study was based on our previously proposed Visuo-Mental (VM) paradigm for enhancing communication for ALS patients [38,39]. A diagram for the VM paradigm is shown in Fig. 2. The proposed paradigm is an adapted version of the conventional oddball paradigm [37]. Participants attended two EEG data recording sessions on separate days with the first session being used for familiarization with the experimental protocol and the second for analysis. Each session consisted of two runs. In each run, subjects were instructed to perform mental calculations when a matrix of numbers intensifies target characters in a 6 × 6 matrix of letters while ignoring the non-target ones. This is similar to the standard P300-based oddball paradigm used for spelling [40]. However, in the VM paradigm the random intensifications were performed by overlaying a 2 × 2 matrix of digits on the displayed character matrix instead of the commonly used celebrity faces. Subjects were asked to perform mental calculations including simple addition/subtraction either diagonally or vertically when target characters were intensified. Data acquisition and stimulus presentations were controlled using BCI2000 software [41].

28 total target intensifications were collected across the two runs (2 runs × 7 characters × 2 intensifications per character = 28). On the first intensification of each character, participants were instructed to add the diagonal numbers in each matrix, identify the larger sum, and multiply the result by two. On the second intensification of each character, the process was repeated using the vertical pairs of numbers instead of the diagonal pairs. If participants missed one or two of the numbers, they were instructed to replace the sum of the missed pair of numbers with a predefined number to ensure that mental arithmetic

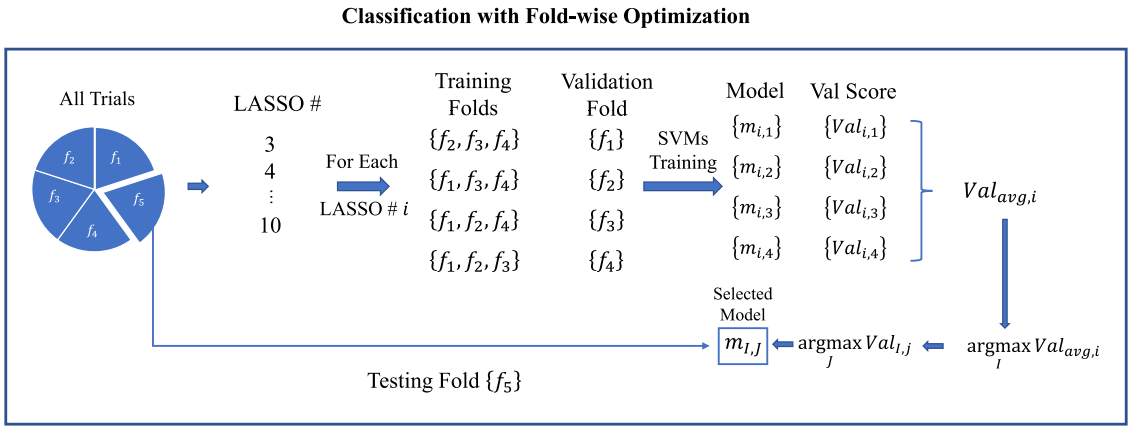

Fig. 3. The pipeline of the proposed fold-wise optimization process. The features from all trials under specific tw and f_d are segmented into five folds. In this case, f_5 is considered as the testing fold. The other four folds form up four sets of training and validation setups, each of which generates a corresponding model and validation accuracy under a given feature number for LASSO. Therefore, each LASSO feature number corresponds to an averaged validation accuracy $Val_{avg,i}$ over four-fold setups, where i refers to the corresponding LASSO feature number. We then find the LASSO feature number I that reaches the maximum averaged validation accuracy over a range of LASSO feature number, and the test fold are evaluated using the model $m_{I,J}$ corresponding to the maximum validation accuracy over four-fold setups while I features are selected. This process is repeated five times, in each of which one different fold is considered as the testing fold, and the testing accuracies are averaged.

would be attempted during the trial. For the alternative resting state, we randomly selected 28 6-s EEG recordings from the resting period between two mental calculation trials.

3.2. Method overview

An overview of the proposed graph-based feature extraction approach is shown in Fig. 1. Our subject-specific processing pipeline starts by bandpass filtering the EEG signal into 1-30 Hz. After selecting the EEG segment based on a given time window tw and downsampling rate f_d, time–frequency analysis is used to calculate the signal's instantaneous amplitude and phase sequences using complex Morlet wavelets [42]. The temporal graphs, in which timestamps are nodes and temporal state similarity are edges, are then generated from the amplitude and phase sequences. Graph-based features including eigenvalues and global graph features are then extracted from the temporal graphs. The eigenvalues are selected using a threshold to keep the most separable features. The selected eigenvalues, noted as F_E, and the global graph features, noted as F_G, are then stacked from all adjacency matrices to form the feature vector of the corresponding trial.

The classification procedure is then implemented through a fold-wise optimized pipeline for each subject. The pipeline is represented in Fig. 3. To do so, the feature vectors corresponding to all trials are split into five folds. One fold is considered as the testing fold, and the other four folds form four training and validation sets. We use the least absolute shrinkage and selection operator (LASSO) [43] method to select i features from each stacked F_E and F_G input feature vector before training models. Models and corresponding validation accuracy are saved while training with selected features under each LASSO feature selection number i from 3 to 10 to find the optimized feature selection hyperparameter. We choose the LASSO number I corresponding to the highest average validation accuracy, and then select the model $m_{I,J}$ with the highest validation accuracy from four folds. Average validation accuracy $Val_{I,J}$ over five-fold cross validation is considered as the final performance under a set of tw and f_d. Repeat the pipeline for inputs with different tw and f_d, the set of tw and f_d with maximum average $Val_{I,J}$ is considered as the optimized input hyperparameters. Then we use the corresponding $m_{I,J}$ to evaluate the testing fold, which is repeated for each fold as the testing fold. The average testing accuracy is then considered as the fold-wise optimized performance evaluation.

3.3. Time–frequency processing of EEG signals

We use the complex Morlet wavelets (CMW) [44] to extract the frequency-band–specific amplitude and phase from the multi-channel EEG signals. To calculate an overall frequency representation for each channel, the time-point-wise average of the amplitude and phase series was calculated over all the frequency bins (1 Hz wide). All channels cascaded together are then considered as the time-varying amplitude series $\mathcal{A}$ and phase series $\mathcal{P}$ for the corresponding EEG signal under 1–30 Hz. $\mathcal{A}$ and $\mathcal{P}$ share the same dimension as the corresponding EEG signal and were used for further analysis.

3.4. Graph-based dynamical multi-channel feature extraction

$$\mathcal{X}^{d\times N} = \begin{bmatrix} x_1^1 & \cdots & x_1^N \\ \vdots & \ddots & \vdots \\ x_d^1 & \cdots & x_d^N \end{bmatrix}$$

The recorded multi-channel raw or preprocessed EEG signals can be represented as above, where d is the number of sensor channels and N is the number of sample points in a given trial. Here, $\mathcal{X}$ can be amplitude $\mathcal{A}$, phase $\mathcal{P}$. To extract a series of dynamical features from the time-varying amplitude and phase series, we apply a spectral graph theoretical approach to convert $\mathcal{X}$ into an undirected graph to decode its topological information. A diagram of the spectral graph theoretical approach is shown in Fig. 4. A pseudo-code of the feature extraction algorithm is also attached in the supplementary materials section as Algorithm 1.

Since a large number of samples (N) in a given trial would increase the computational burden of feature extraction, we generate the graph from data in a window-based manner. To do so, only the first K data points, which can be considered as a window of K data points, in each trial are utilized. The segmented signal is denoted as $\mathcal{X}^{d\times K}$. First, pairwise state similarity comparison between state vector $\vec{x}_i$ at time i and state vector $\vec{x}_j$ at time j is calculated as follows:

$$w_{ij} = \Omega(\vec{x}_i, \vec{x}_j) = e^{\left(-\frac{\|\vec{x}_i - \vec{x}_j\|^2}{2\sigma^2}\right)}, \quad \forall i, j \in \{1 \dots K\} \tag{1}$$

where $\vec{x}_i$ and $\vec{x}_j$ are column i and column j of $\mathcal{X}^{d\times K}$, w is the similarity metric, and Ω is a radial basis kernel function with σ^2 as the overall statistical variation between columns of matrix $\mathcal{X}^{d\times K}$. Then, a threshold function Θ is applied to convert w_{ij} into a binary form, such that

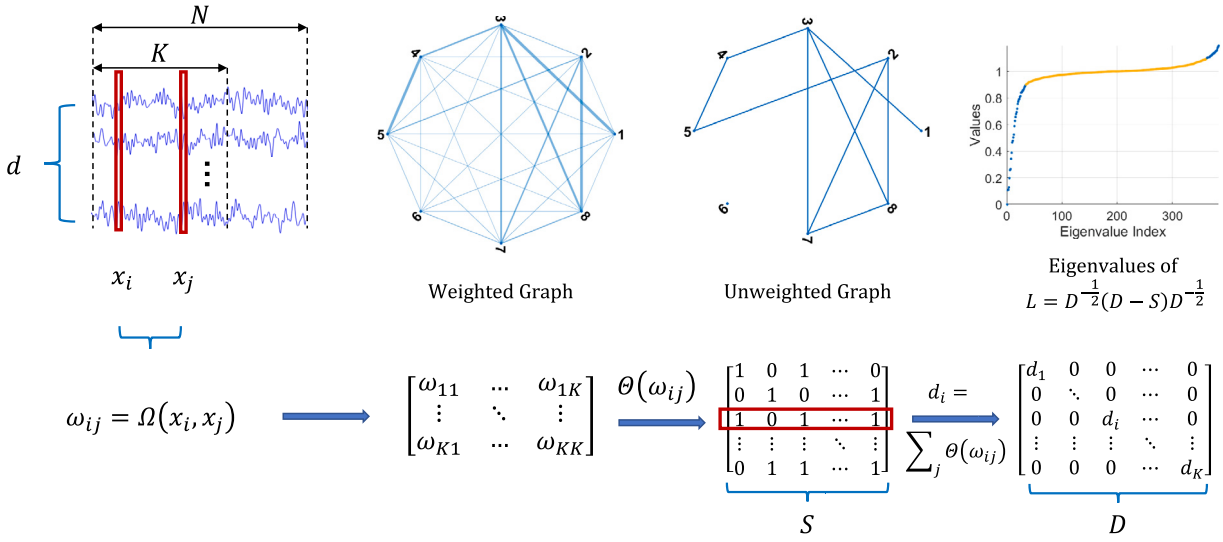

Fig. 4. The diagram of the graph-based dynamical multi-channel feature extraction process. d, N, K represent the number of EEG sensor channels, the original length of the input EEG signal, and the window size. State vectors x_i and x_j represent the states of all EEG channels at time i and j, which then are used to calculate the edge weight w_{ij} between node i and j so that the weighted graph is built up. The wider edge width represents larger edge weights w_{ij}. The weighted graph is converted into an unweighted graph via threshold function θ. The graphs above only have 8 nodes for cleaner demonstration, whereas the real graph in the proposed methods has K nodes. The normalized Laplacian L can be calculated via sparse similarity matrix S and degree matrix D. In each training fold, we cascade all K eigenvalues of L into a feature vector and then eliminate the eigenvalues between 0.9 – 1.1 in it, since the eigenvalues in that range (noted as yellow above) are too close to 1 for either EEG VM or rest states and has low separability. A detailed demonstration of using a threshold to select EEG eigenvalue features is shown in Figure S1 in the supplementary materials section.

$\Theta(w_{ij}) = 1$ if $w_{ij} \leq r$, else $\Theta(w_{ij}) = 0$ where $r = (\sum_{i=1}^{K}\sum_{j=1}^{K} w_{ij})/K^2$.

$$S^{K\times K} = \begin{bmatrix} \Theta(w_{i=1,j=1}) & \cdots & \Theta(w_{i=1,j=K}) \\ \vdots & \ddots & \vdots \\ \Theta(w_{i=K,j=1}) & \cdots & \Theta(w_{i=K,j=K}) \end{bmatrix} \tag{2}$$

The sparse similarity matrix $S^{K\times K}$ represented in (2) is an unweighted and undirected network graph $G \equiv (V, E)$. The index of rows and columns of S represent the nodes V of the graph G, and $\Theta(w_{ij}) = 1$ indicates the existence of the edge E between nodes i and j, otherwise $\Theta(w_{ij}) = 0$.

Once the graph G is generated, the topological information is extracted in the following process to measure the system dynamics. The degree d_i of a node i is calculated as the number of edges connected from i to other vertices, which is defined as: $d_i = \sum_{j=1}^{K}\Theta(w_{ij}), \forall j = \{1 \ldots K\}$, and the degree matrix is defined as: $D^{K\times K} \overset{def}{=} diag(d_1, \ldots, d_K)$.

The normalized Laplacian L of the graph G is defined as:

$$L^{K\times K} \overset{def}{=} D^{-\frac{1}{2}} \times (D - S) \times D^{-\frac{1}{2}}, \tag{3}$$

where S is the sparse similarity matrix of graph G. Thereafter, the eigenvalues λ of L are computed as $Lv = \lambda v$ where v represents eigenvectors. We cascade all K eigenvalues of L into a feature vector $f_e^{K\times 1}$ and consider it as the dynamic feature of the signal in the corresponding window of size K samples.

3.5. Global graph features

Network measures are calculated for the quantitative investigation of network properties. Seven typical network measures are included in this study. Detailed descriptions of these network measures and their interpretations are provided in several studies [45–47]

The characteristic path length $l(G)$ of a graph $G = (V, E)$ is defined as the average number of edges in the shortest paths between all nodes pairs given by

$$l(G) = \frac{1}{|V|(|V|-1)} \sum_{v\in V} \sum_{v'\in V\setminus v} spl(v, v'), \tag{4}$$

where V is the set of nodes within the graph G. It is used to measure the integration of timestamps and provides information regarding global communication. A low CPL indicates greater integration within the temporal sequence.

The global efficiency $E_{global}(G)$ is calculated as:

$$E_{global}(G) = \frac{1}{|V|(|V|-1)} \sum_{i\neq j\in V} \frac{1}{d(v_i, v_j)}, \tag{5}$$

where $d(v_i, v_j)$ is the number of edges in the shortest path between any two vertices i and j in a graph G. E_{global} is used to quantify the overall efficiency of information transfer across the whole network. A greater E_{global} value indicates a faster parallel transfer of information in a network.

The global clustering coefficient $CC(G)$ is defined as:

$$CC(G) = \frac{\#\text{ of closed triplets}}{\#\text{of all triplets}}, \tag{6}$$

where a triplet is three nodes that are connected by either two (open triplet) or three (closed) edges. It is used to measure the functional segregation of timestamps. A higher clustering coefficient corresponds to more robust and efficient local interactions.

Transitivity $T(G)$ is defined as:

$$T(G) = \frac{3\times\#\text{of triangles}}{\#\text{of all triplets}}. \tag{7}$$

Transitivity is the overall probability for the network to have adjacent nodes interconnected, thus revealing the existence of tightly connected communities.

The diameter of a graph $D(G)$ is the maximum distance between the pair of vertices. It can also be defined as the maximal distance between the pair of vertices.

The radius of graph $R(G)$ exists only if it has the diameter, which is the minimum among all the maximum distances between a node to all other nodes.

Table 1
The averaged test accuracies comparison between four kinds of features is shown in the table. Classical features represent the six spatial and temporal features mentioned in [38]. Eigenvalue features represent the thresholded eigenvalues of the amplitude and phase graph generated from the EEG signals on 1–30 Hz. The Global Graph features represent the 7 global graph features. We cascaded the thresholded eigenvalues and global graph features above to form up the Eigenvalue and Graph features. All these features are following the same inference model and optimization process on the LASSO feature selection number, the EEG downsampling factor, and the sampling window size.

Visuo-Mental (VM) task classification				
Subject index	Classical features	Eigenvalue features	Global graph features	Eigenvalue + Graph features
Sub#1	67.7	**76.8**	73.0	73.2
Sub#2	76.8	71.7	67.7	**77.1**
Sub#3	**73.3**	61.1	67.7	67.9
Sub#4	64.2	61.4	67.4	**70.2**
Sub#5	64.7	62.4	60.9	**73.5**
Sub#6	55.6	64.4	58.8	**64.5**
Avg.	67.1 ± 7.5	66.3 ± 6.5	65.9 ± 5.2	**71.1 ± 4.5**

Table 2
The validation and test accuracies of each subject for models trained with eigenvalue + global graph features under different combinations of downsampling factor and time window. tw represents the time window and f represents the downsampling factor. For each set of parameters of a given subject, The results are shown in the format of "validation accuracy | test accuracy" for each set of parameters of a given subject. The test accuracies corresponding to the validation max among all parameter combinations, both of which are noted as bold, are considered the test accuracy of the optimized model for each subject.

Effect of downsampling factor and time window on VM task classification accuracy						
Subject index	tw = 2 f = 4	tw = 4 f = 4	tw = 6 f = 4	tw = 2 f = 8	tw = 4 f = 8	tw = 6 f = 8
Sub#1	56.7 | 50.0	74.9 | 69.5	**77.7** | **73.2**	51.7 | 46.4	61.3 | 58.9	61.6 | 57.4
Sub#2	74.4 | 64.4	72.4 | 62.6	76.9 | 73.3	70.9 | 64.2	**79.6** | **77.1**	77.0 | 69.8
Sub#3	61.4 | 66.4	60.5 | 52.1	63.8 | 57.6	64.2 | 55.8	66.3 | 61.1	**68.9** | **67.9**
Sub#4	**70.8** | **70.2**	67.2 | 70.0	66.6 | 58.8	65.0 | 50.2	66.7 | 60.8	68.9 | 57.0
Sub#5	57.3 | 48.2	72.3 | 67.9	61.6 | 49.8	53.4 | 46.2	61.1 | 61.1	**73.0** | **73.5**
Sub#6	62.5 | 57.6	58.9 | 52.3	56.3 | 44.7	**72.1** | **64.4**	62.5 | 51.7	56.4 | 48.3

The density of graph $Den(G)$ is defined as:

$$Den(G) = \frac{\#\text{of existing edges}}{\#\text{of possible edges}}. \tag{8}$$

It ranges from 0 to 1. Low density indicates the graph is sparse. The global graph features above are extracted for each graph and concatenated into a feature vector $f_g^{7\times1}$

3.6. Fold-wise hyperparameter optimization

L is a symmetric positive semi-definite whose eigenvalues are non-negative and bounded between 0 and 2. Most of the eigenvalues are concentrated at 1 and only the ones far from 1 have separability as discriminative features. For each training fold, we only select the eigenvalues smaller than 0.9 or greater than 1.1 to eliminate the less-separable eigenvalues and keep the same eigenvalue indexes for the testing set. We cascade the selected eigenvalues of the graph from both amplitude series $\mathcal{A}$ and phase series $\mathcal{P}$ as F_E. The global graph feature vector f_g generated from the graph based on $\mathcal{A}$ and $\mathcal{P}$ are cascaded as F_G.

Since BCI data are usually rare and the dimension of feature vector F_E and F_G relatively high, we apply the LASSO (least absolute shrinkage and selection operator) algorithm [48] to perform feature selection after concatenating F_E and F_G together. In the meantime, to reduce the selection bias of variables such as the time window size tw, downsampling rate f_d, and the number of features N_{lasso} selected by LASSO, we propose a fold-wise optimization method to optimize the value of the variables mentioned above.

The trials of a given subject are merged into a sample pool and separated into five folds. One fold is considered as the testing fold, and the other four folds are training and validation folds. The training process is repeated 4 times, in each of which one different fold from the training and validation folds is used as the validation fold and the other three folds as training folds. The training and validation processes are performed after the LASSO feature selection under N_{lasso}. The average of four validation accuracies is considered as the classification performances of the corresponding training and validation folds. The model with the highest validation accuracy is saved for the later test. Given a set of N_{lasso}, we get a series of classification performances and models for the same training and validation folds. The N_{lasso} with the highest average validation accuracy is used to apply the corresponding selected feature indexes on the test fold, and then the model with the highest validation accuracy is used to generate the testing accuracy. Repeat the process above five times with different testing folds using different combinations of window size and downsampling rate. The validation and testing accuracies are averaged over five folds. The testing accuracy corresponding to the highest average validation accuracy is considered as the final classification performance after optimizing tw, f_d, and N_{lasso}. We employ a linear support vector machine (SVM) model as our main classifier, where its parameters including the kernel scale are automatically optimized while training.

To compare our proposed graph-based temporal dynamical feature with the traditional EEG features, we use the EEG VM features mentioned in [38] including the P300 response (the maximum peak between 250 and 400 ms post-stimulus), the later N400 (the minimum peak between 350 and 560 ms post-stimulus) and P600 (the maximum peak between 650 and 800 ms post-stimulus) components, along with spectral changes in the delta (1-3 Hz), theta (4-7 Hz), and alpha (8-12 Hz) bands.

4. Experimental results and discussion

4.1. Classification results and discussion

For each specific subject, we split the data into training, validation, and testing datasets. The training dataset and validation dataset are used to generate the optimized model, which is then evaluated with the testing dataset. For each EEG signal trial in all datasets, we perform the same trimming and downsampling preprocessing using a given time window and downsampling factor to get the input EEG data frame.

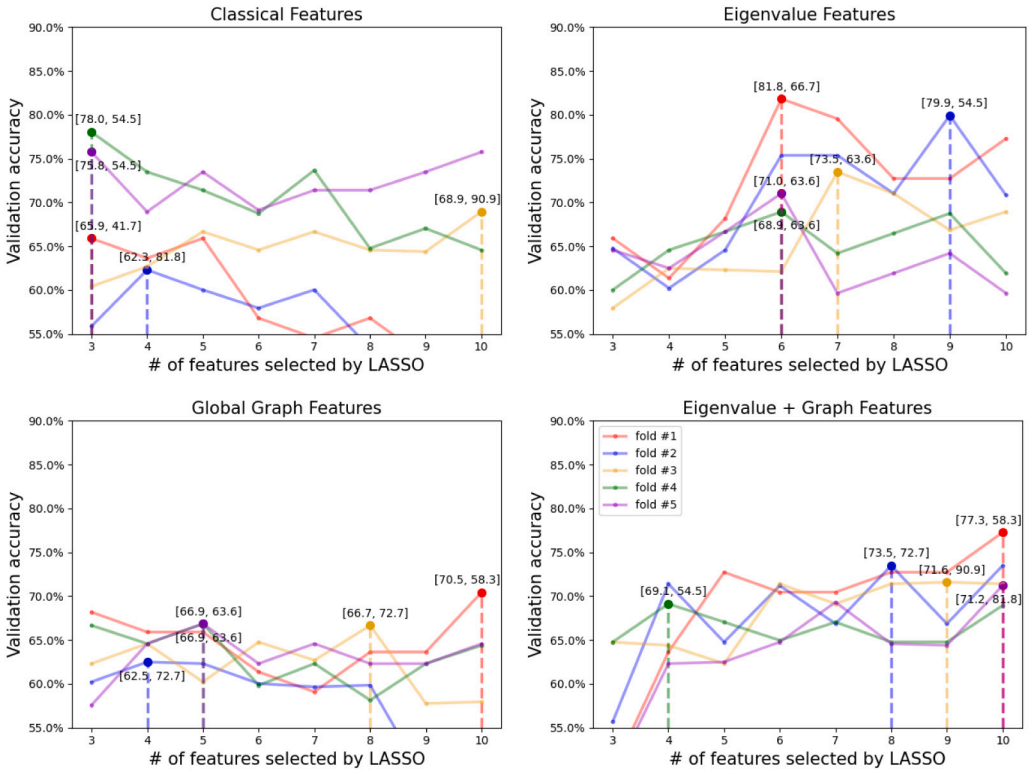

Fig. 5. The validation accuracy variation curve under different LASSO feature selection numbers for four kinds of features. The data are from Subject#5. Each case uses the corresponding best-performed downsampling factor and window size. Each line represents the validation accuracy variation of each fold within the 5-cross validation. For the LASSO selected number that maximum the validation accuracy, the corresponding validation, and test accuracy are noted in the form of [validation accuracy, test accuracy].

Table 3
The distribution of eigenvalue features and global graph features in their combination after LASSO feature selection. The results are from all five folds of subject#1. The feature distribution is shown in the form of "number of eigenvalue features: number of global graph features" for each LASSO feature selection number. The results of other subjects are attached in the supplementary materials section.

Number of eigenvalue feature versus global graph features after LASSO								
Fold #	LASSO #							
	3	4	5	6	7	8	9	10
1	2:1	3:1	4:1	3:3	6:1	7:1	6:3	8:2
2	2:1	3:1	4:1	5:1	6:1	7:1	7:2	8:2
3	2:1	2:2	2:3	3:3	4:3	5:3	6:3	7:3
4	2:1	3:1	3:2	4:2	5:2	6:2	7:2	8:2
5	1:2	3:1	4:1	5:1	5:2	5:3	7:2	8:2

The time window candidates are the first 2 s, 4 s, and 6 s, and the downsampling factor is either 4 or 8. Then, feature vectors are shrunk by LASSO into a candidate dimension. The range of LASSO feature selection number is from 3 to 10. The time window, downsampling factor, and LASSO feature selection number are considered optimizable parameters. The reported range of the parameters above is selected from enumerated experimental trials for better representation. We have tried larger ranges for the candidates of the time window, downsampling factor, and LASSO feature selection number, which resulted in little performance change yet extra training computational cost.

All the experiments were performed using Python 3.6 with scikit-learn package on an Intel i9-8950 CPU. The timing performance of finishing the fine-tuning and training processes for one subject is 50 min on average. Since we enumerated the LASSO feature selection number, time window size, and downsampling factor during fold-wise optimization, a large batch of trials in the training and validation datasets is repeatedly used to find optimized hyperparameters and generate the corresponding model. However, a testing trial only takes 0.53 s on average to be processed by our pipeline, including preprocessing, feature calculation, and final prediction. The calculation efficiency shows that our proposed pipeline has the potential to be further modified into an online BCI system with subject-specific offline training.

Table 1 shows the classification performance of four kinds of features (classical EEG VM features; eigenvalues as the spectral graph features; global graph features; and spectral graph features combined with global graph features) under the optimized parameters for each subject. All kinds of features are used to train the SVM with the same fold-wise optimization pipeline.

The results show that the test accuracy of eigenvalue features combined with global graph features reached 71.1% with 4.8% and 5.2% higher than using eigenvalue features or global graph features individually. The classical EEG VM features are then used in the same fold-wise optimization pipeline to reach the average test accuracy of 67.1% over all subjects, which is 4.0% lower than our proposed graph-based feature. Also, the standard deviation of our graph-based method is only 4.5%, 3% smaller than the one of classical features. 4 out of 6 subjects reach maximum performance while using the eigenvalue+graph features. The ablation studies for other kinds of classifiers such as Linear Discriminant Analysis (LDA) and multi-layer perceptron (MLP) are attached in the supplementary materials section. The proposed linear SVM outperformed the other classifiers.

To illustrate the influence of the time window and the downsampling factor, we show the subject-specific classification test accuracies of using eigenvalue+global graph features under different combinations of these two parameters in Table 2. As the results show, the classification performance for each subject varies under different combinations of downsampling factor and window size, which demonstrates that optimization is necessary. Also, the optimized model selected via the validation max reached the highest test accuracy compared to the rest model under non-optimal parameters, which demonstrates the optimization strategy in our proposed method is useful.

Table 3 shows the feature number distribution of eigenvalue features and global graph features in their combinations after LASSO

feature selection. Before LASSO, each trial of EEG data generates 16 graph features and 105 to 130 eigenvalue features depending on how many of them are left after the thresholding process described in Section 3.6. The features from both modalities are balanced in our method while performing feature combinations with LASSO. As the results show, more eigenvalue features are selected compared to the graph features in most cases. However, considering that the number of eigenvalue features is almost 10 times that of the graph features, this distribution is reasonable. Fig. 5 shows the validation accuracy variations of five folds across a range of LASSO feature selection numbers. In this case, we use the performance of features from Subject#5 as examples. Each feature is for the corresponding best performance time window and downsampling factors. As the results show, the average validation accuracy $Val_{avg,i}$ varies for different LASSO feature selection numbers within each fold, which would affect the choice of model selection to get the testing accuracy. The demonstration above shows the parameters are subject-dependent, which is a consensus for BCI studies [38,49–51]. Therefore, the fold-wise parameter optimization process is necessary. Our results show that the proposed eigenvalue + global graph features method performs better and is more robust than the classical spectral and temporal features method, since it reached higher accuracy while remaining with a lower standard deviation.

5. Conclusion

This study focused on extracting graph-based dynamic features from EEG signals and evaluating their performance in a fold-wise optimization framework for classifying VM-BCI tasks. We tested our proposed graph-based EEG features both alone and concatenated together using a linear SVM classifier. The proposed eigenvalue + global graph features reached greater performance than the classical spectral and temporal features while having a smaller standard deviation demonstrating the robustness of our proposed algorithm. Compared to using features separately, the combination strategy could considerably improve performance. It is shown that the influence of parameters is subject-dependent, which demonstrates our fold-wise optimization strategy is valuable and necessary. We can draw the conclusion that the proposed framework, which applies the combination of eigenvalue features and global graph features from the temporal dynamic graph with the fold-wise parameter optimization process, has potential value for improving the performance of the BCI system with higher robustness.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the National Science Foundation, USA (NSF-2005957, NSF-1913492, NSF-2006012) and the Institutional Development Award (IDeA) Network for Biomedical Research Excellence, USA (P20GM103430).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2022.106498.

References

[1] L.M. McCane, E.W. Sellers, D.J. McFarland, J.N. Mak, C.S. Carmack, D. Zeitlin, J.R. Wolpaw, T.M. Vaughan, Brain-computer interface (BCI) evaluation in people with amyotrophic lateral sclerosis, Amyotroph. Lateral Scler. Frontotemporal Degener. 15 (3–4) (2014) 207–215.
[2] S.B. Borgheai, J. McLinden, A.H. Zisk, S.I. Hosni, R.J. Deligani, M. Abtahi, K. Mankodiya, Y. Shahriari, Enhancing communication for people in late-stage ALS using an fNIRS-based BCI system, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (5) (2020) 1198–1207.
[3] D. Beukelman, S. Fager, A. Nordness, Communication support for people with ALS, Neurol. Res. Int. 2011 (2011) 714693.
[4] A. Heilinger, R. Ortner, V. La Bella, Z.R. Lugo, C. Chatelle, S. Laureys, R. Spataro, C. Guger, Performance differences using a vibro-tactile P300 BCI in LIS-patients diagnosed with stroke and ALS, Front. Neurosci. 12 (2018) 514.
[5] D. Huang, A. Ren, J. Shang, Q. Lei, Y. Zhang, Z. Yin, J. Li, K.M. Von Deneen, L. Huang, Combining partial directed coherence and graph theory to analyse effective brain networks of different mental tasks, Front. Hum. Neurosci. 10 (2016) 235.
[6] C.A. Stefano Filho, R. Attux, G. Castellano, Can graph metrics be used for EEG-BCIs based on hand motor imagery? Biomed. Signal Process. Control 40 (2018) 359–365.
[7] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world'networks, Nature 393 (6684) (1998) 440–442.
[8] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[9] H. Ayaz, F. Dehais, Neuroergonomics: The Brain at Work and in Everyday Life, Academic Press, 2018.
[10] A.F. Rabbi, K. Ivanca, A.V. Putnam, A. Musa, C.B. Thaden, R. Fazel-Rezai, Human performance evaluation based on EEG signal analysis: a prospective review, in: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2009, pp. 1879–1882.
[11] M. Rahman, W. Karwowski, M. Fafrowicz, P.A. Hancock, Neuroergonomics applications of electroencephalography in physical activities: a systematic review, Front. Hum. Neurosci. (2019) 182.
[12] M. Rubinov, O. Sporns, Complex network measures of brain connectivity: uses and interpretations, Neuroimage 52 (3) (2010) 1059–1069.
[13] D. Goldenberg, A. Galván, The use of functional and effective connectivity techniques to understand the developing brain, Dev. Cogn. Neurosci. 12 (2015) 155–164.
[14] M. Xu, F. He, T.-P. Jung, X. Gu, D. Ming, Current challenges for the practical application of electroencephalography-based brain–computer interfaces, Engineering 7 (12) (2021) 1710–1712.
[15] L.E. Ismail, W. Karwowski, A graph theory-based modeling of functional brain connectivity based on eeg: A systematic review in the context of neuroergonomics, IEEE Access 8 (2020) 155103–155135.
[16] J.-P. Lachaux, E. Rodriguez, J. Martinerie, F.J. Varela, Measuring phase synchrony in brain signals, Hum. Brain Mapp. 8 (4) (1999) 194–208.
[17] L.A. Baccalá, K. Sameshima, Partial directed coherence: a new concept in neural structure determination, Biol. Cybernet. 84 (6) (2001) 463–474.
[18] M. Vinck, R. Oostenveld, M. Van Wingerden, F. Battaglia, C.M. Pennartz, An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias, Neuroimage 55 (4) (2011) 1548–1565.
[19] F. de Vico Fallani, J. Richiardi, M. Chavez, S. Achard, Graph analysis of functional brain networks: practical issues in translational neuroscience, Philos. Trans. R. Soc. B 369 (1653) (2014) 20130521.
[20] K.J. Blinowska, Review of the methods of determination of directed connectivity from multichannel data, Med. Biol. Eng. Comput. 49 (5) (2011) 521–529.
[21] W. De Haan, Y.A. Pijnenburg, R.L. Strijers, Y. van der Made, W.M. van der Flier, P. Scheltens, C.J. Stam, Functional neural network analysis in frontotemporal dementia and alzheimer's disease using EEG and graph theory, BMC Neurosci. 10 (1) (2009) 1–12.
[22] L. Wang, X. Guo, J. Sun, Z. Jin, S. Tong, Cortical networks of hemianopia stroke patients: a graph theoretical analysis of EEG signals at resting state, in: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2012, pp. 49–52.
[23] S.C. Ponten, L. Douw, F. Bartolomei, J. Reijneveld, C. Stam, Indications for network regularization during absence seizures: weighted and unweighted graph theoretical analyses, Exp. Neurol. 217 (1) (2009) 197–204.
[24] I. Daly, S.J. Nasuto, K. Warwick, Brain computer interface control via functional connectivity dynamics, Pattern Recognit. 45 (6) (2012) 2123–2136.
[25] L.F.S. Uribe, C.A. Stefano Filho, V.A. de Oliveira, T.B. da Silva Costa, P.G. Rodrigues, D.C. Soriano, L. Boccato, G. Castellano, R. Attux, A correntropy-based classifier for motor imagery brain-computer interfaces, Biomed. Phys. Eng. Express 5 (6) (2019) 065026.
[26] T. Cattai, S. Colonnese, M.-C. Corsi, D.S. Bassett, G. Scarano, F.D.V. Fallani, Phase/amplitude synchronization of brain signals during motor imagery BCI tasks, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 1168–1177.

[27] F. Shamsi, A. Haddad, L. Najafizadeh, Early classification of motor tasks using dynamic functional connectivity graphs from EEG, J. Neural Eng. 18 (1) (2021) 016015.
[28] P.G. Rodrigues, A.K. Takahata, R. Suyama, R. Attux, G. Castellano, J.R. Sato, S.J. Nasuto, D.C. Soriano, et al., Can dynamic functional connectivity be used to distinguish between resting-state and motor imagery in EEG-BCIs? in: International Conference on Complex Networks and their Applications, Springer, 2021, pp. 688–699.
[29] R. Gera, L. Alonso, B. Crawford, J. House, J. Mendez-Bermudez, T. Knuth, R. Miller, Identifying network structure similarity using spectral graph theory, Appl. Netw. Sci. 3 (1) (2018) 1–15.
[30] R.S. Biswas, A. Pal, T. Werho, V. Vittal, A graph theoretic approach to power system vulnerability identification, IEEE Trans. Power Syst. 36 (2) (2020) 923–935.
[31] T. Wang, Z. Liu, G. Lu, J. Liu, Temporal-spatio graph based spectrum analysis for bearing fault detection and diagnosis, IEEE Trans. Ind. Electron. 68 (3) (2020) 2598–2607.
[32] M.S. Tootooni, P.K. Rao, C.-A. Chou, Z.J. Kong, A spectral graph theoretic approach for monitoring multivariate time series data from complex dynamical processes, IEEE Trans. Autom. Sci. Eng. 15 (1) (2016) 127–144.
[33] M. Fan, C.-A. Chou, Detecting abnormal pattern of epileptic seizures via temporal synchronization of EEG signals, IEEE Trans. Biomed. Eng. 66 (3) (2018) 601–608.
[34] A.E. Elisha, L. Garg, O. Falzon, G. Di Giovanni, EEG feature extraction using common spatial pattern with spectral graph decomposition, in: 2017 International Conference on Computing Networking and Informatics, ICCNI, IEEE, 2017, pp. 1–8.
[35] A.H. Ghaderi, B.R. Baltaretu, M.N. Andevari, V. Bharmauria, F. Balci, Synchrony and complexity in state-related EEG networks: an application of spectral graph theory, Neural Comput. 32 (12) (2020) 2422–2454.
[36] J.S. Kirar, R. Agrawal, A combination of spectral graph theory and quantum genetic algorithm to find relevant set of electrodes for motor imagery classification, Appl. Soft Comput. 97 (2020) 105519.
[37] D.J. Krusienski, E.W. Sellers, D.J. McFarland, T.M. Vaughan, J.R. Wolpaw, Toward enhanced P300 speller performance, J. Neurosci. Methods 167 (1) (2008) 15–21.
[38] S.B. Borgheai, J. McLinden, A.H. Zisk, S.I. Hosni, R.J. Deligani, M. Abtahi, K. Mankodiya, Y. Shahriari, Enhancing communication for people in late-stage ALS using an fNIRS-based BCI system, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (5) (2020) 1198–1207.
[39] S. Borgheai, R. Deligani, J. McLinden, A. Zisk, S. Hosni, M. Abtahi, K. Mankodiya, Y. Shahriari, Multimodal exploration of non-motor neural functions in ALS patients using simultaneous EEG-fNIRS recording, J. Neural Eng. 16 (6) (2019) 066036.
[40] L.A. Farwell, E. Donchin, Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials, Electroencephalogr. Clin. Neurophysiol. 70 (6) (1988) 510–523.
[41] G. Schalk, J. Mellinger, A Practical Guide to Brain–Computer Interfacing with BCI2000: General-Purpose Software for Brain-Computer Interface Research, Data Acquisition, Stimulus Presentation, and Brain Monitoring, Springer Science & Business Media, 2010.
[42] S. Zhu, S.I. Hosni, X. Huang, S.B. Borgheai, J. McLinden, Y. Shahriari, S. Ostadabbas, A graph-based feature extraction algorithm towards a Robust data fusion framework for brain-computer interfaces, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2021, pp. 878–881.
[43] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1) (1996) 267–288.
[44] M.X. Cohen, Analyzing Neural Time Series Data: Theory and Practice, MIT Press, 2014.
[45] E. Bullmore, O. Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems, Nat. Rev. Neurosci. 10 (3) (2009) 186–198.
[46] J. Li, J. Lim, Y. Chen, K. Wong, N. Thakor, A. Bezerianos, Y. Sun, Mid-task break improves global integration of functional connectivity in lower alpha band, Front. Hum. Neurosci. 10 (2016) 304.
[47] S. Achard, E. Bullmore, Efficiency and cost of economical brain functional networks, PLoS Comput. Biol. 3 (2) (2007) e17.
[48] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. R. Stat. Soc. Ser. B Stat. Methodol. (1996) http://dx.doi.org/10.1111/j.2517-6161.1996.tb02080.x.
[49] L.C. Schudlo, T. Chau, Dynamic topographical pattern classification of multichannel prefrontal NIRS signals: II. Online differentiation of mental arithmetic and rest, J. Neural Eng. 11 (1) (2013) 016003.
[50] S. Borgheai, M. Abtahi, K. Mankodiya, J. McLinden, Y. Shahriari, Towards a single trial fNIRS-based brain-computer interface for communication, in: 2019 9th International IEEE/EMBS Conference on Neural Engineering, NER, IEEE, 2019, pp. 1030–1033.
[51] G. Bauernfeind, R. Scherer, G. Pfurtscheller, C. Neuper, Single-trial classification of antagonistic oxyhemoglobin responses during mental arithmetic, Med. Biol. Eng. Comput. 49 (9) (2011) 979–984.