Neuroscience Informatics 2 (2022) 100091

Contents lists available at ScienceDirect

Neuroscience Informatics

www.elsevier.com/locate/neuri

Original article

Neural decoding of imagined speech from EEG signals using the fusion of graph signal processing and graph learning techniques

Aref Einizade *, Mohsen Mozafari, Shayan Jalilpour, Sara Bagheri, Sepideh Hajipour Sardouie

Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 23 May 2022
Received in revised form 19 June 2022
Accepted 23 June 2022

Keywords:
Brain-Computer Interface (BCI)
Imagined Speech (IS)
Electroencephalography (EEG)
Machine Learning (ML)
Graph Signal Processing (GSP)
Graph Learning (GL)
Brain connectivity

ABSTRACT

Imagined Speech (IS) is the imagination of speech without using the tongue or muscles. In recent studies, IS tasks are increasingly investigated for the Brain-Computer Interface (BCI) applications. Electroencephalography (EEG) signals, which record brain activity, can be used to analyze BCI-based tasks utilizing Machine Learning (ML) methods. The current paper considers decoding IS brain waves using the fusion of classical signal processing, Graph Signal Processing (GSP), and Graph Learning (GL) based features. The proposed fusion method, named GraphIS (short for a Graph-based Imagined Speech BCI decoder), is applied to the four-class classification (three classes of the imagined words, in addition to the `rest` state) on EEG recordings of fifteen subjects. Results show that GSP and GL-based features can highly improve the performance of classification outcomes compared to using only classical signal processing features and over the state-of-the-art Common Spatial Pattern (CSP) feature extractor by considering the spatial information of the signals as well as interactions between channels in regions of interest. The proposed GraphIS method leads to a mean accuracy of 50.10% in the studied four-class IS classification task, compared to using only one feature set with an accuracy of 47.86% and 46.10%, and also the state-of-the-art CSP with an accuracy of 47.10%. Additionally, using an EEG connectivity map of the electrode signals obtained from GL methods, we also found a strong connection in the right frontal region as well as in the left frontal regions during IS, which had not been focused on in the previous IS papers.

© 2022 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

A Brain-Computer Interface (BCI) device can connect brain activity to external devices and use the brain information to help people with disorders. Paralyzed or locked-in syndrome (from now on, disabled) patients need to communicate with the world. A BCI device can use the patient's brain signals and pave the way for this person's control over surrounding devices or connect the people. For over three decades, scientists have used brain signals to develop BCI algorithms, and devices [1]. One of the branches of BCI applications goes toward speech imagination or silent speech, which has been considered recently [2]. Imagined Speech (IS) is defined as the imagination of speech with no use of the tongue or its muscles [3]. In [4,5], it is mentioned that Broca's and Wernicke's brain regions participate in IS. Therefore, Electroencephalography (EEG) signals of channels close to these regions can be used to achieve the IS prediction [6]. Other methods of recording brain activity, such as Magnetoencephalography (MEG) [7], functional Near-Infrared Spectroscopy (fNIRS) [8], and Electrocorticography (ECoG) [9], can also be used in silent speech prediction. However, EEG recording is a better option for BCI devices because of being cheaper, non-invasive, and has a high temporal resolution.

A thorough review of articles related to IS, not only in EEG but also in other methods of brain activity recording such as ECoG, is contained in [10]. IS research can be categorized into three subgroups containing imagined words [11], imagined vowels [12], and imagined syllables [13]. The objective of this paper is the classification of EEG signals in the task of imagined words. In the following paragraph, some articles related to IS tasks are summarized.

Singh et al. [14] studied binary IS tasks, including long vs. short words and IS vs. `rest` state EEG signals. To this end, they transformed covariance matrices from raw EEG signals using the Tangent Space method (applied in Riemannian geometry) to create a projected matrix. The channels located in Broca's and Wernicke's areas were selected for this process. Then they flattened this

* Corresponding author.
E-mail addresses: einizade.aref@ee.sharif.edu (A. Einizade), Mozaffary.moh@gmail.com (M. Mozafari), shayanjalilpour@gmail.com (S. Jalilpour), sara.bagheri96@gmail.com (S. Bagheri), hajipour@sharif.edu (S. Hajipour Sardouie).

https://doi.org/10.1016/j.neuri.2022.100091
2772-5286/© 2022 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

new matrix to reach a feature vector for each trial. The number of features was reduced using Principal Component Analysis (PCA). The artificial neural network was used as a base classifier for the bootstrap bagging scenario (10-fold cross-validation for each subject). This team achieved a two-class accuracy of 85%. Hashim et al. [15] used the Mel Frequency Cepstral Coefficient (MFCC) and k-Nearest Neighbor (k-NN) to classify two imagined words (*yes* and *no*). In this study, electrodes are reduced from 14 to 6 to have near motor cortex electrodes (Broca's and Wernicke's areas). The results showed that this method achieved an accuracy of 58%. Saha et al. [16] used raw signal and phonological features as inputs of a Convolutional Neural Network (CNN) and Temporal CNN (TCNN). The features of the last layer of CNN were used in the gradient boosting classifier. This team achieved an accuracy of 83% for six two-class classifications of 4 different words, including "pat," "pot," "knew," and "gnaw." Authors in [17] reduced the artifacts of recorded EEG signals using Independent Component Analysis (ICA). The reconstructed signal was filtered using a subspace-based Wiener filter to remove remnant noise and artifacts from the previous step. The classification was performed using a k-NN classifier based on the Euclidean distance between Auto-Regressive (AR) coefficients of noise-free signals. This method led to an accuracy of 61% in the two-class classification of imagined syllables /ba/ and /ku/. In another study, Nguyen et al. [18] used the Riemannian manifold and Relevance Vector Machine (RVM) to classify three classes of imagined words. They achieved an accuracy of 70%. García et al. [19] used Discrete Wavelet Transform (DWT) coefficients with the frequency range below 32 Hz as features for the five-class classification of imagined Spanish words using EEG signals. The mentioned features were extracted from F7, FC5, T7, and P7 channels. In the classification steps, they utilized Support Vector Machine (SVM), Random Forest (RF), and Naive Bayes (NB) and chose a classifier with the highest accuracy for the bagging ensemble. Results showed an accuracy of 41%. In [20], the problem of the five-class and six-class imagined words were investigated. In this work, Salinas et al. formed a three-way tensor for each epoch by concatenating each channel's Continuous Wavelet Transform (CWT). This team applied Parallel Factor Analysis (PARAFAC) to decompose created tensor into the sum of source tensors. K-means was applied as a clustering method to data corresponding to epochs of each subject and class. Histogram of this clustered data was considered as features. For the classification of signal epochs, linear SVM was applied to each subject. This team achieved an accuracy of 61%. Cooney et al. [21] used Convolutional Neural Networks (CNNs) to implement the transfer learning approach in a five-class classification of vowels. They applied two approaches of fine-tuning the first or last CNN layers and reached better performance than a traditional within-subject CNN classifier. Qureshi et al. [22] utilized sigmoid activation function-based linear extreme learning machines to classify five-class word imagination. The features were extracted using connectivity methods such as covariance and cross-correlation. The results showed a classification rate of 40%. Torres et al. [23] classified five classes of imagined words with an accuracy of 70% using the RF classifier. This team used a fuzzy inference system to find the best channels for their aim and found seven channels through the proposed method. Al-Saleh et al. [24] extracted spectral and temporal features using raw and Common Spatial Filter (CSP) filtered signals. This team utilized the mentioned features to classify eleven classes of imagined words. The best result of their method was 62% accuracy, which was achieved by the SVM classifier. Using CSP in IS analyses with a spatial viewpoint makes it a desirable choice in many other works [25,26].

Previous research [22] has rarely used the characteristics of the distance between electrodes and their distribution in three-dimensional space or the connectivity between signals from different channels to investigate the IS problem. In recent years, Graph Signal Processing (GSP) has been recognized as a powerful tool for analyzing signals that live on an irregular grid (generally a graph) [27]. Brain imaging data (such as EEG, MEG, ECoG, etc.) is also considered a potential option for using GSP techniques due to the nature of Region of Interests (ROIs) distribution on the structural graph of regions (the ones with structural connectivity), and interactions between different regions (i.e., functional connectivity ones) [28–33].

To utilize the structural connectivity of the EEG electrodes, in this paper, a popular technique of GSP, namely Graph Fourier Transform (GFT), is used. Similar approaches like the mentioned one are addressed in [32,34] and [35] to reduce the dimensions and computational complexity for the analyses of different brain imaging modalities.

This fresh study employs Graph Learning (GL) methods [36] are also used to obtain and analyze functional connectivity. There are two major drawbacks to using conventional methods (such as correlation, coherence, Phase Lag Index (PLI), etc.) to calculate functional brain connectivity [33]: 1) the limitations of the available training data can negatively affect the analysis obtained from these methods, and 2) because of the volume conductive effect of signals such as EEG and MEG, as well as the presence of various types of noise and artifacts in nearby electrodes, sometimes these methods yield to erroneous statistical analyzes that may lead to inaccurate conclusions. Because of the aforementioned deficiencies, GL methods were used in this paper, in which the assumption that the signals on the graph are smooth is fundamental. If the signal values associated with the two end vertices of graph edges with large weights tend to be similar, the corresponding signal on this graph is considered smooth [37,38]. Relative robustness against noise is another advantage of these methods. This paper uses no particular noise and artifact removal method prior to applying the GL methods. Also, despite the limited available training data, the ability to appropriately learn graphs using these methods is well illustrated in this paper.[1]

The rest of this paper is organized as follows. Section 2 details the used dataset, and the corresponding IS task. Section 3 then introduces an overview of the fundamental concepts underlying GSP methods, feature extraction process, and classification scheme. Section 4 brings the results of the proposed algorithm, named GraphIS (short for a Graph-based Imagined Speech BCI decoder), to the present IS dataset, discusses the results in detail, and analyses the GL method outputs for obtaining EEG brain connectivity. Section 5 presents the proposed method's pros and cons and future research direction. Finally, Section 6 concludes the discussions and analysis of the different results.

2. Dataset description

This paper has employed a dataset corresponding to the iBCIC2020 competition,[2] in which the EEG signals were recorded from fifteen subjects (five females). The average age was 31, and all subjects were healthy and right-handed. Brain signals were acquired using a biosignal amplifier for non-intrusive measurements called the g.HIamp (from G.Tech Company) device[3] with electrode locations depicted in Fig. 1. This headset has 64 channels following the international 10-10 system, and its sampling rate is 2400 Hz. The signal receiving system was grounded to the forehead, and one of the channels was connected to the right ear as a reference. Also, band-pass filters were applied in the range of 1–130

[1] The preliminary work on ADHD detection has been published in [39].
[2] https://nbml.ir/FA/scientific-tournament/102640.
[3] https://www.gtec.at/product/ghiamp/.

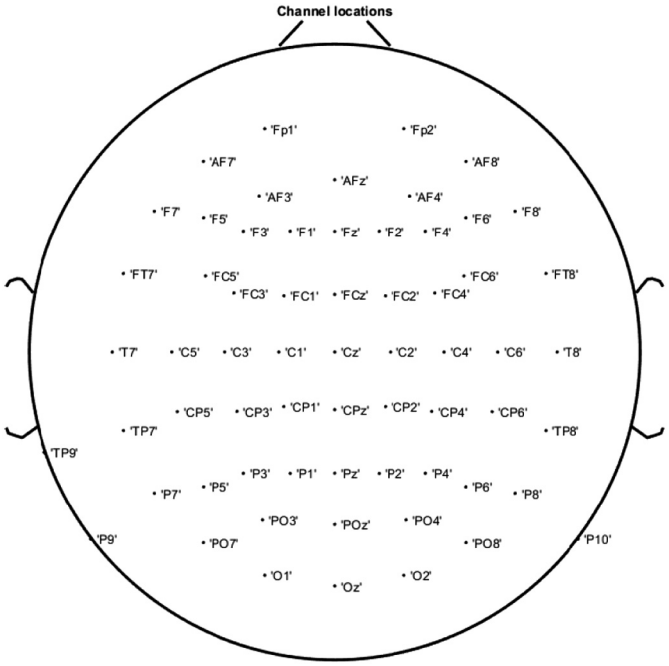


Fig. 1. Locations of the EEG electrodes according to their names.

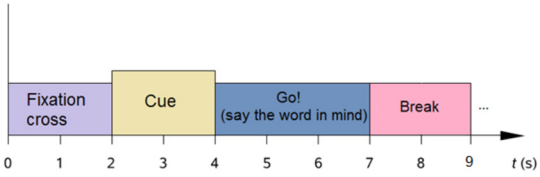


Fig. 2. Time scheme paradigm followed during data acquisition. First, the subjects are asked not to think about anything for 2 seconds (Fixation cross). Then, the imagination of the three words (`rock`, `paper`, and `scissors`) or the `rest` state lasts three seconds after the Cue [40].

Hz, the signals were re-referenced by the average of the signals of all channels, and line noise was reduced using a 50 Hz notch filter.

The experiment is conducted in a quiet environment. A subject is seated on a chair facing a screen while a visual cue of the word is displayed. Subjects are asked to think of the word when they see a hand posed as rock, paper, or scissors (as in the hand game of *rock-paper-scissor*).

The stimulus protocol is shown in Fig. 2. First, the "+" symbol appears in the center of the screen for 2 seconds. During this time, the subject is asked not to think about anything and be ready to observe the sign. In the next 2 seconds, the desired symbol appears; three types of symbols are used in this protocol (hand signs of rock, paper, or scissors). After 2 seconds, the symbol disappears, and "Go" appears for 0.5 seconds, and the person has 3-second to think of the corresponding word. The data recorded from each subject consists of 100 three-second trials, with the number of trials per class (for imagining `rock`, `paper`, `scissors`, or `rest` state) being 25. The goal is to classify EEG trials of each subject into their correct classes of "rest" and three imagined speech words "rock," "paper," and "scissors."

3. Methodology

3.1. Background

A graph $\mathcal{G}$ is known as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathbf{W}\}$, where $\mathcal{V}$ is the set of the M vertices (EEG electrodes) of the graph, $\mathcal{E}$ is the set of the edges of the graph, and $\mathbf{W} \in \mathbb{R}^{M\times M}$ is a non-negative symmetrical matrix known as the adjacency matrix of the graph. The (i, j)-th element of the matrix $\mathbf{W}$, i.e., $\mathbf{W}_{ij}$, represents the weight of the connection of the vertices i and j of the graph $\mathcal{G}$, which can be kernelized using the Euclidean distance between vertices i and j. A graph can be considered a generalization of the distribution of the Cartesian vertices. Based on this terminology, a graph signal is a signal function of graph vertices that is not necessarily defined in the Cartesian coordinates. In other words, a graph signal can be represented by a function $f : \mathcal{V} \to \mathbb{R}^M$ that assigns a scalar value to each vertex. Therefore, this graph signal can also be considered a generalization of the signal used by classical signal processing, defined on the regular coordinates. Also, the Laplacian matrix of a graph with M vertices is defined as follows:

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \tag{1}$$

where $\mathbf{D}$ is a diagonal matrix, called degree matrix, whose diagonal elements are obtained from the equation $\mathbf{D}_{ii} = \sum_{j=1}^{M} \mathbf{W}_{ij}$. Since $\mathbf{L}$ is a real, positive semi-definite matrix, it has a complete set of orthonormal eigenvectors and non-negative eigenvalues. The eigenvectors of $\mathbf{L}$ are $\mathbf{v}_i \in \mathbb{R}^M, i = 0, 1, \ldots, M-1$ and its eigenvalues are $0 = \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{M-1}$. Graph Fourier Transform (GFT) vectors are defined as $\{\mathbf{v}_i\}_{i=0}^{M-1}$. The set of eigenvalues is also denoted as the GFT frequencies. The GFT matrix can be written as follows:

$$\mathbf{V} = [\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_{M-1}]. \tag{2}$$

Graph Learning (GL) [36,37] is the case where the adjacency matrix is learned directly from the data. One of the advantages of GL is that (almost) no particular assumption is required to define the similarities between vertices and use them to minimize the desired cost function [36].

3.2. Feature extraction

In this study, there are three ways to extract features. Note that the selection of the relevant features using the Fisher criterion [41] in which the lower number of the selected features is specified with no significant changes in consecutive values of the sorted Fisher scores, that is $\frac{\sigma_{k+1}-\sigma_k}{\sigma_k} < 0.01$, where σ_k denotes the Fisher score of the k-th feature.

1) Classical signal processing features. Feature Group 1 ($\mathbf{G}_1$): At this stage, first, a wide range of classical signal processing was extracted from all EEG channels. These features were the db2 wavelet coefficients (in six decomposition levels), their statistics (min, max, mean and standard deviation), Shannon entropy, band power, and relative band power (the percentage of a band power relative to the total power) [42]. Common EEG signals' frequency bands (Delta, Theta, Alpha, Beta, and Gamma) of EEG signals of each channel were extracted for each channel. Likewise, Higuchi [43], peak frequency, line length [44], and flatness of each frequency band (Wiener entropy) [45] were extracted. Then, the important features were identified by calculating the Fisher criterion on the features extracted from the training data. Depending on the channels from which these features were extracted, it was possible to select frequent (and, therefore, dominant) channels to reduce the problem dimensionality to the least possible depending on the number of channels and information present. As a result, 16 dominant channels out of 63 were selected, which are: T8, Fp1, Fp2, F5, F7, F8, P7, P9, Fc1, Fc3, FT7, FT8, AF4, PO8, O1, and Cp6.

2) Structural graph-based features. Feature Group 2 ($\mathbf{G}_2$): The spatial distribution of the signal of channels, or equivalently, the structural connectivity, is ignored in the extracted features of the previous step. In this step, the weighted and undirected graphs are first formed using the spatial information of the channels. In this

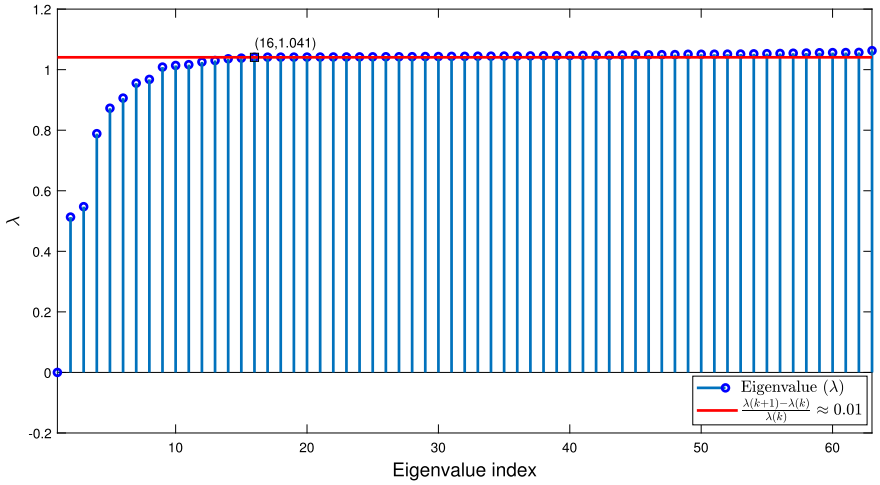


Fig. 3. Eigenvalues of the Laplacian matrix **L** obtained from the EVD of the geometric graph of the EEG channels. The lower number of eigenvalues with no significant change in consecutive values is set as the optimal r, that is $\frac{\lambda_{k+1}-\lambda_k}{\lambda_k} < 0.01$, where k denotes the eigenvalue index.

manner, firstly, an adjacency matrix is constructed using spatial information of the EEG channels. The Laplacian matrix is then obtained, and eigenvalue decomposition is performed on this matrix (graph spectral decomposition). Finally, dominant eigenvalues are used to reduce the graph dimensionality and complexity as far as graph signals go (the dimension of the EEG channel space). After this process, only the smooth components of the geometric signals are preserved, and others are suppressed, potentially assumed to be the noisy part of the EEG channel space.

In detail, first, the elements of the graph adjacency matrix **W** are formed using the structural information of the EEG channels as follows:

$$\mathbf{W}_{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{d}_i - \mathbf{d}_j\|_2^2}{2\sigma^2}\right) & i \neq j \\ 0 & i = j \end{cases}, \tag{3}$$

where $\mathbf{d}_i$ is the vector of the Cartesian 3D coordinates of the i-th EEG electrode and σ is a parameter (an exponential decay rate) for the correct matching of the formation of this matrix to achieve a better result. Due to the cross-validation accuracy analysis, this work sets $\sigma = 0.4$. To reduce the number of EEG channels, spectral decomposition of this structural graph, i.e., **W**, is needed.

Then, the projection matrix $\mathbf{V}_r$ is defined using the eigenvectors of **L** (eq. (2)) as follows:

$$\mathbf{V}_r = [\mathbf{v}_0, \mathbf{v}_1, ..., \mathbf{v}_{r-1}]; \quad r < M. \tag{4}$$

$\mathbf{V}_r$ is the projection matrix used to reduce the number of EEG channels. To define the dimension of the projected space is denoted by r. Fig. 3 plots the eigenvalues of **L**. The EEG channel space dimension r is set to the lowest number of eigenvalues with no significant change in consecutive values ($r = 16$ was selected), that is $\frac{\lambda_{k+1}-\lambda_k}{\lambda_k} < 0.01$, where k denotes the eigenvalue index. Therefore, the number of really used EEG channels became 16. Note that these selected 16 electrodes using the graph eigenvalues are actually obtained by the linear combination of the previous 63 electrodes. But, the selected 16 electrodes using $\mathbf{G}_1$ are selected among real EEG electrodes.

$\mathbf{X}_{M\times T}$ is an EEG trial data matrix, where M and T represent the number of EEG channels and the number of time samples, respectively. So, the number of EEG channels is reduced using $\mathbf{V}_r$ as follows:

$$\tilde{\mathbf{X}}_{r\times T} = \mathbf{V}_r^T \mathbf{X}; \quad r < M. \tag{5}$$

Then, the set of features described in $\mathbf{G}_1$ is extracted from these converted signals $\tilde{\mathbf{X}}$, forming the second feature vector for the corresponding trial.

3) Functional Connectivity Graph-based Features. Feature Group 3 ($\mathbf{G}_3$):

Based on the GL methods, the third and fourth feature vectors are extracted as follows:

Feature Group 3, subset 1 ($\mathbf{G}_{31}$): This part and the next part of the feature-extraction process are based on GL methods [37,38]. The goal is to learn the adjacency matrix of a graph with M vertices (i.e., the functional connectivity matrix **W** between the EEG channels) in a way the matrix $\mathbf{X}_{M\times J}$ can be thought of as a set of J vector signals, where the j-th column of **X** belongs to any $j = 1, ..., J$. This j-th vector is a graph signal as smooth as possible on this graph. The smoothness assumes the j-th column of **X**, i.e., a vector (graph signal) $\mathbf{x}$ on $\mathcal{G}$ can be measured in terms of a quadratic form **L**, as follows [37,38]:

$$\mathbf{x}^T\mathbf{L}\mathbf{x} = \frac{1}{2}\sum_{i\sim j}^{M} \mathbf{W}_{ij}\,(\mathbf{x}(i) - \mathbf{x}(j))^2, \tag{6}$$

where $\mathbf{W}_{ij}$ represents the weight of the edge connecting two vertices i and j. $\mathbf{x}(i)$ and $\mathbf{x}(j)$ are the signal values associated with these two vertices. On the other hand, $\sum_{j=1}^{J} \mathbf{x}_j^T\mathbf{L}\mathbf{x}_j = trace\,(\mathbf{X}^T\mathbf{L}\mathbf{X})$, where $\mathbf{x}_j$ is the j-th column of **X**. Therefore, the smaller the term $trace(\mathbf{X}^T\mathbf{L}\mathbf{X})$ is, the smoother the $\mathbf{X}_{M\times J}$ is on the graph $\mathcal{G}$ with **L**. In [38], this goal and the ability to have learnable Gaussian weights are led to the following cost function, optimizable over the weight matrix **W**, not **L**:

$$\min_{\mathbf{W}\in\mathcal{W}_m} \|\mathbf{W}\circ\mathbf{Z}\|_{1,1} - \alpha\mathbf{1}^T\log(\mathbf{W}\mathbf{1}) + \frac{\beta}{2}\|\mathbf{W}\|_F^2, \tag{7}$$

where $\|\mathbf{A}\|_{1,1} = \sum_{i,j}|\mathbf{A}_{ij}|$ is the element-wise norm-1 of **A**, $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j}\mathbf{A}_{ij}^2}$ denotes the Frobenius norm of **A**, the vector $\mathbf{1}$ is an all-one vector and $\circ$ denotes the Hadamard (element-wise) product between two matrices **A** and **B** with equal size as $(\mathbf{A}\circ\mathbf{B})_{ij} = \mathbf{A}_{ij}\mathbf{B}_{ij}$. The set $\mathcal{W}_m$ is defined as $\mathcal{W}_m = \{\mathbf{W}\in\mathbb{R}_+^{M\times M} : \mathbf{W} = \mathbf{W}^T, diag(\mathbf{W}) = \mathbf{0}\}$ and the matrix $\mathbf{Z}\in\mathbb{R}^{M\times M}$ is a distance matrix formed by the following relation:

$$\mathbf{Z}_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|^2, \tag{8}$$

where $\mathbf{r}_i$ is the i-th row of **X**. As abovementioned, one of the critical assumptions of this algorithm is the smoothness of the graph

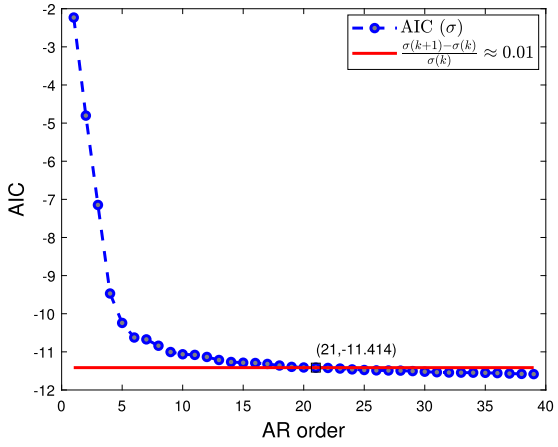


Fig. 4. The averaged AIC values over the AR models of all trials and all subjects. The candidate for the optimum model order is determined based on the insignificant changes, i.e., $\frac{\sigma_{k+1}-\sigma_k}{\sigma_k} < 0.01$, where σ_k denotes the kth AIC value.

signal on the graph. On the other hand, in a matrix containing EEG signals, the number of time samples (T) is very high, and in general, these graph signals cannot be assumed to be smooth on this graph. So, one needs an embedding algorithm that can reduce the time samples (T) and smooth the new signals on the graph. In this paper, the coefficients of the Auto-Regressive (AR) model applied to the signals of the EEG channels are used. On the other hand, if the raw EEG channel signals are used to learn the interactions between graph vertices (EEG channels) and two specific channel signals are slightly delayed versions of each other, the connectivity strength is obtained at a low value. Therefore, to consider both the phase and amplitude of graph signals, the AR coefficients have been used. These justifications have also been mentioned in [33], in which the GL has been applied to Electrocorticogram (ECoG) signals. The AR model with order p can be written as:

$$\mathbf{y}(n) = \sum_{i=1}^{p} a_i\, \mathbf{y}(n-i) + \mathbf{e}(n), \tag{9}$$

where p is the model order, and $\mathbf{y}(n)$ is the signal to be modeled. The a_i values are the AR coefficients, and $\mathbf{e}(n)$ is the Gaussian white noise. Here, one of the most important parameters determining the quality of problem-solving is the order of this model (p). Calculating the Akaike Information Criterion (AIC) [46] helps with this issue as follows for the signals of different channels and different EEG trials for all subjects:

$$AIC(p) = T\log(\epsilon_p) + 2p, \tag{10}$$

where T is the number of time samples of signals, ϵ_p is the modeling error variance, and p is the order of the AR model. Fig. 4 shows this calculated criterion for $p = 2, \ldots, 40$ and plots the mean of AIC values for all signals (all channels, trials, and subjects). The lowest p that AIC changes are no longer significant, i.e., $\frac{\sigma_{k+1}-\sigma_k}{\sigma_k} < 0.01$, where σ_k denotes the k-th AIC value, is selected as the optimal $p(p_{opt})$. By analyzing Fig. 4, one finds $p_{opt} = 21$ is selected. Using p_{opt}, the new representation of the EEG signals for each trial can be obtained ($\mathbf{X}^{AR} \in \mathbb{R}^{M \times p_{opt}}$).

The $\mathbf{Z}$ matrix is then calculated from $\mathbf{X}^{AR}$ for each trial, and the connectivity matrix ($\mathbf{W}$) is obtained by solving the problem (7). $\mathbf{W}$'s strictly upper triangle part is vectorized and considered the third feature vector for each trial. This process leads to a vector consisting of $M(M-1)/2$ elements (denoted by feature group $\mathbf{G}_{31}$).

Feature Group 3, subset 2 ($\mathbf{G}_{32}$): Dong et al. [37] proposed the following model for learning a graph:

$$\min_{\mathbf{L}} \; trace(\mathbf{X}^T\mathbf{L}\mathbf{X}) + \alpha\|\mathbf{L}\|_F^2 \quad \text{s.t.: } \mathbf{L} \in \mathcal{L} \tag{11}$$

where parameters $s > 0$, and $\alpha > 0$ control the scale and the density of the solution, respectively, and

$$\mathcal{L} = \{\mathbf{L}_{ii} \geq 0,\ \mathbf{L}_{ij} = \mathbf{L}_{ji} \leq 0,\ \mathbf{L}\mathbf{1} = \mathbf{0},\ trace(\mathbf{L}) = s\}. \tag{12}$$

Kalofolias [38] showed that problem (11) is convertible to

$$\min_{\mathbf{W} \in \mathcal{W}_m} \|\mathbf{W} \circ \mathbf{Z}\|_{1,1} + \alpha\|\mathbf{W}\mathbf{1}\|_F^2 + \alpha\|\mathbf{W}\|_F^2 \quad \text{s.t.: } \|\mathbf{W}\|_{1,1} = s\,. \tag{13}$$

The $\mathbf{Z}$ matrix is calculated similarly to the previous section, and finally, the connectivity matrix $\mathbf{W}$ is obtained by solving the problem (13). Like $\mathbf{G}_{31}$, the strictly upper triangular part of $\mathbf{W}$ is selected as the corresponding trial feature (denoted by feature group $\mathbf{G}_{32}$).

Details of obtaining $\mathbf{W}$ from the previous two methods and the role of control parameters (α, β, and s) are given in [38].

For more comparison with the classic classification scenarios in the next section, the subject-dependent and subject-independent scenarios use data from the specific subject and all other subjects (except the specific subject), respectively.

3.3. Classification

The classification scheme is given in Fig. 5. In the classification step, after extraction of the $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{G}_{31}$, and $\mathbf{G}_{32}$ features, the first two sets ($\mathbf{G}_1$ and $\mathbf{G}_2$) of features together form the first parent feature set (PS1), and the second two sets ($\mathbf{G}_{31}$ and $\mathbf{G}_{32}$) form the second parent feature set (PS2). Two consecutive classification steps are executed per the proposed GraphIS algorithm for predicting IS. First, a binary classification model predicts the `rest` vs. `non-rest` besides three classes of imagined labels. If the predicted label becomes `rest`, it will be accepted. Otherwise, the features go through the second classifier that predicts labels corresponding to the three classes of imagined words. This method is performed on each subject to predict final labels for four classes via a 10-fold cross-validation procedure. It should be noted that for the first classification stage, only the data of the same subject is used (nine out of ten folds are used as training data), while in the second stage, the data of all the other subjects is used. The classification process for the i-th subject in detail has the following parts.

`rest` vs. `non-rest` classification: Firstly, the trials of the i-th subject are shuffled and divided into ten equal parts. One part is considered the final test data ($PS1_{Test}$ and $PS2_{Test}$ in Fig. 5). The other parts become the training data ($PS1_{Train}$ and $PS2_{Train}$ in Fig. 5). Normalization is performed on $PS1_{Train}$ and $PS2_{Train}$ and, then, based on the parameters obtained from train data, is applied on $PS1_{Test}$ and $PS2_{Test}$. The reduction of the features using the Fisher criteria obtains feature-reduced versions of these data (FR-$PS1_{Train}$, FR-$PS2_{Train}$, FR-$PS1_{Test}$, and FR-$PS2_{Test}$ in Fig. 5). The number of PS1 features is reduced to 200, and the number of PS2 features is reduced to 600. These numbers are chosen based on the cross-validation of $PS1_{Train}$ and $PS2_{Train}$ separately. The Dimensionality Reduction step is performed on the training data (i.e., $PS1_{Train}$ and $PS2_{Train}$). The optimal hyperparameters have been found and applied to the unseen test data based on the accuracy obtained from cross-validation on the $PS1_{Train}$ and $PS2_{Train}$. Then, two binary SVM models with Radial Basis Function (RBF) kernels (binary SVM_1 and binary SVM_2 in Fig. 5) are trained on FR-$PS1_{Train}$ and FR-$PS2_{Train}$. Then the prediction of the two-class labels (`rest` vs. `non-rest`) of FR-$PS1_{Test}$ and FR-$PS2_{Test}$ is made using these

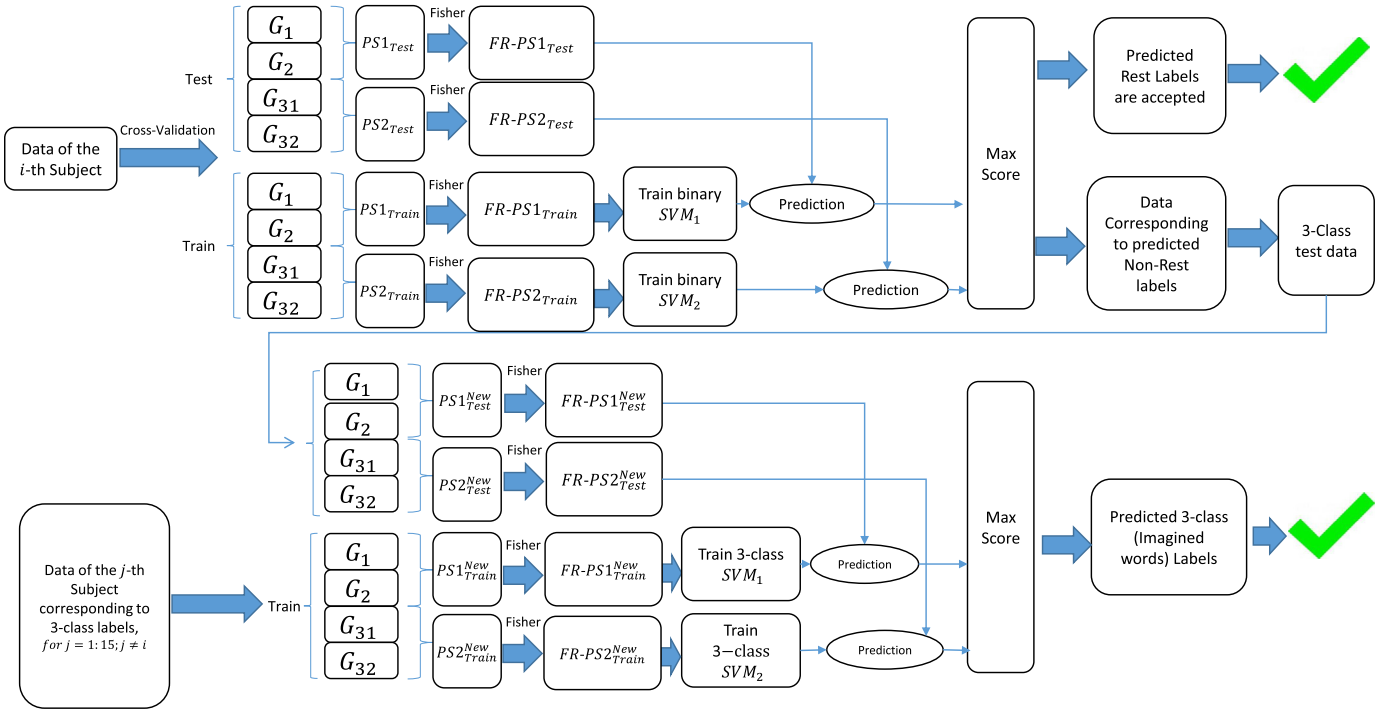


Fig. 5. Classification scheme of the proposed GraphIS method. Two-class (`rest` vs. `non-rest`) classification is performed. Then, predicted `non-rest` samples are classified using three-class data of the other subjects as the training data. In both classification steps, final labels are obtained using the maximum score of the SVM classifiers trained on PS1 and PS2 feature parents.

two SVMs. Finally, by comparing the SVM class scores (the distance between the samples from the separator boundary obtained by the SVM algorithm), the two-class labels of the test data for the current fold of the i-th subject are predicted.

Three-class classification: The predicted `non-rest` samples of $PS1_{Test}$ and $PS2_{Test}$ form new test data ($PS1^{New}_{Test}$ and $PS2^{New}_{Test}$ in Fig. 5). Also, PS1 and PS2 are extracted from other subjects' three imagined word trials. After normalizing these data, they are considered new training data ($PS1^{New}_{Train}$ and $PS2^{New}_{Train}$ in Fig. 5). Then, similar to the previous section, the reduction of the features using the Fisher criteria is done on $PS1^{New}_{Train}$, and $PS2^{New}_{Test}$, and the optimal reduction number is obtained from these training data. Then, the number of features of the $PS1^{New}_{Train}$, $PS2^{New}_{Train}$, $PS1^{New}_{Test}$ and $PS2^{New}_{Test}$ are reduced and leads to FR-$PS1^{New}_{Train}$, FR-$PS2^{New}_{Train}$, FR-$PS1^{New}_{Test}$ and FR-$PS2^{New}_{Test}$, respectively, (as illustrated in Fig. 5). The number of PS1 features is reduced to 200, and the number of PS2 features is reduced to 1450. These numbers are determined by cross-validation on $PS1^{New}_{Train}$ and $PS2^{New}_{Train}$. Then two three-class SVM models with RBF kernels (3-class SVM_1 and 3-class SVM_2 in Fig. 5) are trained on FR-$PS1^{New}_{Train}$ and FR-$PS2^{New}_{Train}$ for the three-class classification task (`non-rest` samples). Finally, by comparing the SVM classification scores of the FR-$PS1^{New}_{Test}$ and FR-$PS2^{New}_{Test}$, the three-class labels of the test samples for the current fold are predicted.

This process is performed on every ten folds of the i-th subject data, and the four-class classification accuracy is computed for each subject.

4. Results

The four-class accuracy and its average and standard deviation over all subjects corresponding to the proposed fusion GraphIS model and using only one of the PS1 or PS2 feature sets are shown in Fig. 6 and Table 1, respectively. The results of exploiting the CSP [25,26] as the feature extractor instead of extracting $\mathbf{G}_1$, $\mathbf{G}_2$, and $\mathbf{G}_3$ in Fig. 5 also appear in Fig. 6 to compare the results with the state-of-the-art. Based on these provided results, it can be admitted that the two sets of features, PS1, and PS2, are complementary. Both boost the performance of the proposed method for predicting imagined words. Fig. 6 shows that the fusion of these two sets of features (PS1 and PS2) has significantly improved the accuracy for at least 10 out of 15 subjects, i.e., p-value $< \alpha$ for the difference between the performance of the fusion method and only exploiting PS1 and PS2 (Table 2), where $\alpha = 0.05$ is the significance level. It should be noted that in many cases, the set of active electrodes for each of the three imagined word classes is very similar. Therefore, the feature of classical signal processing methods cannot clearly distinguish the classes. Moreover, the performance improvement of the GraphIS method over the state-of-the-art CSP is illustrated based on these provided results (Fig. 6).

Table 1
Mean and standard deviation of the accuracy of four-class classification using PS1, PS2, a fusion of these feature sets on all subjects, as well as exploiting the CSP feature extractor.

Feature set	Mean accuracy (%)	Standard deviation (%)
PS1	47.86	0.38
PS2	46.10	0.53
CSP	47.10	0.24
PS1+PS2	50.10	0.56

The p-values of the difference between the performance of the proposed GraphIS approach and the one using only PS1 or PS2 feature sets across all subjects and their ten folds obtained from ANOVA with Bonferroni's multiple-comparison statistical test under the null hypothesis helped investigate the statistical significance of the superiority of the fusion model. The difference in the averaged performance when using only the PS1/PS2 feature set against the fusion method is insignificant and appears in Table 2. These p-values have been corrected using a positive False Discovery Rate (pFDR). From this table and considering significance level $\alpha = 0.05$, it can be seen that, on average, the performance

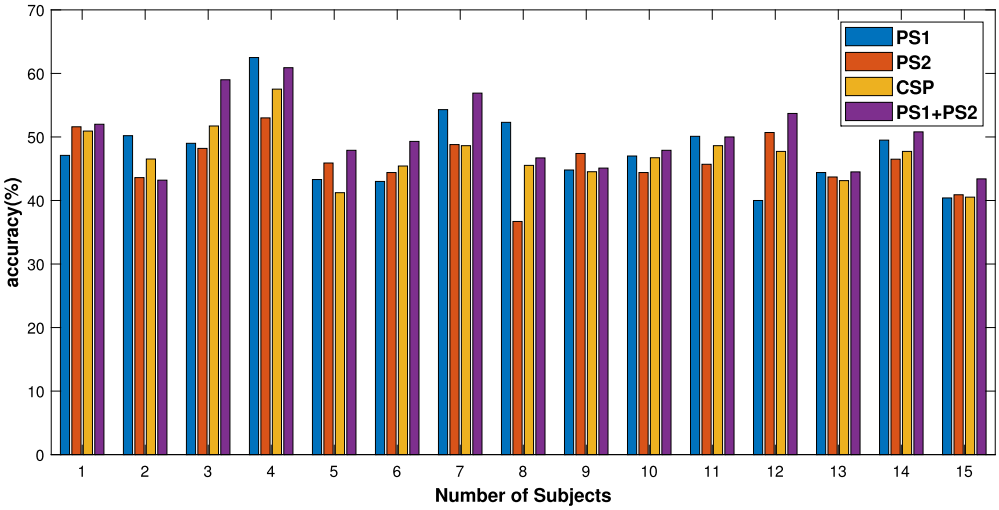


Fig. 6. Accuracy of four-class classification using PS1, PS2, and fusion of these feature sets. It can be seen that the fusion method could significantly improve the classification accuracy for 10 out of 15 subjects, i.e., p-value $< \alpha$ for the difference between the performance of the fusion method and only exploiting PS1 and PS2 (Table 2), where $\alpha = 0.05$ is the significance level.

Table 2
p-values of the differences of the accuracies between the proposed GraphIS fusion method and using only PS1 or PS2 obtained from ANOVA with Bonferroni's multiple-comparison test.

Sub	PS1 (ANOVA)	PS2 (ANOVA)
#1	$\mathbf{2.00 \times 10^{-3}}$	6.56×10^{-1}
#2	$\mathbf{2.00 \times 10^{-3}}$	6.17×10^{-1}
#3	$\mathbf{2.00 \times 10^{-3}}$	$\mathbf{2.00 \times 10^{-3}}$
#4	$\mathbf{1.95 \times 10^{-2}}$	$\mathbf{2.00 \times 10^{-3}}$
#5	$\mathbf{2.00 \times 10^{-3}}$	$\mathbf{1.56 \times 10^{-2}}$
#6	$\mathbf{2.00 \times 10^{-3}}$	$\mathbf{5.90 \times 10^{-3}}$
#7	$\mathbf{3.90 \times 10^{-3}}$	$\mathbf{2.00 \times 10^{-3}}$
#8	$\mathbf{2.00 \times 10^{-3}}$	$\mathbf{2.00 \times 10^{-3}}$
#9	7.19×10^{-1}	$\mathbf{1.17 \times 10^{-2}}$
#10	3.13×10^{-1}	$\mathbf{3.90 \times 10^{-2}}$
#11	8.75×10^{-1}	$\mathbf{2.00 \times 10^{-3}}$
#12	$\mathbf{2.00 \times 10^{-3}}$	$\mathbf{3.90 \times 10^{-3}}$
#13	1.00×10^{0}	5.82×10^{-1}
#14	7.81×10^{-2}	$\mathbf{2.00 \times 10^{-3}}$
#15	$\mathbf{1.17 \times 10^{-2}}$	$\mathbf{3.52 \times 10^{-2}}$

Table 3
Comparison of subject-based accuracy (%) results between the proposed GraphIS method and subject-independent scenario.

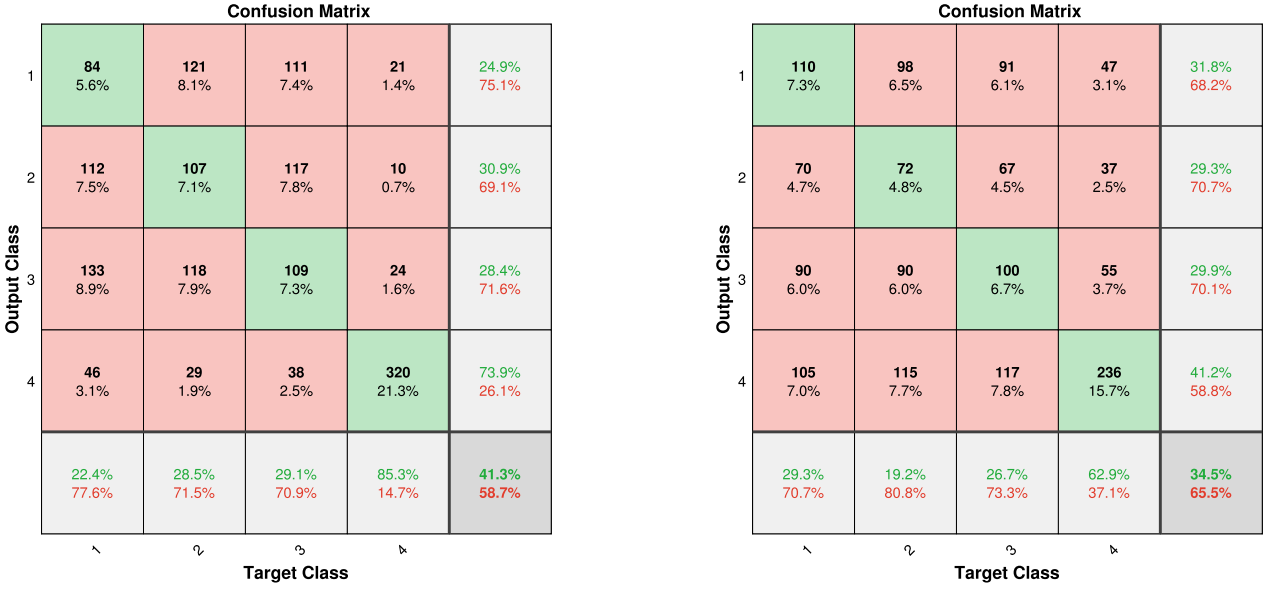
Sub	The proposed GraphIS method	subject-independent scenario
#1	**48**	30
#2	**42**	38
#3	**58**	26
#4	**62**	31
#5	**48**	28
#6	**52**	32
#7	**56**	40
#8	**46**	29
#9	**41**	35
#10	**48**	37
#11	**48**	37
#12	**57**	32
#13	**46**	34
#14	**51**	36
#15	**44**	43

improvement is quite statistically significant, i.e., p-value $< \alpha$, in almost 12 out of 15 subjects.

For further analysis and comparison of the proposed GraphIS approach to the common classification scenarios, the performance results of the subject-independent and subject-dependent scenarios are provided in Fig. 7. Besides, the subject-based classification results corresponding to the proposed GraphIS method and one-step subject-independent scenario are placed in Table 3. The one-step four-class classification leads to lower performance (41.30% and 34.5% accuracy in the subject-independent and subject-dependent scenarios, respectively) compared to the current approach (50.10% accuracy). This improved performance can probably be explained as follows: The `rest` state is a more specialized state across subjects and, therefore, it is more helpful to process the `rest` data in a subject-dependent manner, while the IS task is a more global stimulus-based task and maybe it is the reason that the processing of IS is more discriminative and helpful in a subject-independent scenario.

The performance details of the `rest` vs. `non-rest` and three-class classification steps of the proposed GraphIS scheme are provided in Fig. 8 to analyze the major bottleneck of the IS classification problem, which is the imagined word (three-class) classification step. Based on these provided results. The difficulty of this three-class classification problem can also be admitted in Fig. 11, in which the learned connectivity maps of the imagined words `rock`, `paper` and `scissors` from GL methods (7) and (13) are not very discriminative. The details of Fig. 11 and connectivity maps are provided in the following paragraphs. Moreover, the confusion matrix of exploiting the state-of-the-art CSP approach as a feature extractor is illustrated in Fig. 8, which shows poorer performance than the proposed GraphIS fusion approach. Recall that, in the first step of the GraphIS method, the two-class (`rest` vs. `non-rest`) classification is performed. Then, the `non-rest` classified samples are classified as one of the IS words. However, because these `non-rest` classified samples can have a true `rest` label (due to the possible miss-classifications in the first step) in the second step, the `rest` label can also exist in the obtained confusion matrix, as depicted in Fig. 8 (b). Moreover, it is worth noting that the three-class performance should be calculated corresponding to only three-class labels, as depicted in the red rectangle in Fig. 8 (b), and is, on average, calculated as $\frac{164+157+126}{164+157+126+118+123+102+117+93+92} \approx 41\%$.

Besides, the subject-independent step (the second step) of the proposed GraphIS approach (Fig. 8 (b)) is associated with the only imagined words (a three-class classification); however, the separate subject-independent scenario learns from four-class classification samples. Therefore, it seems that adding the `rest` class samples leads to lower performance and a higher miss-classification



(a) subject-dependent one-step four-class classification

(b) subject-independent one-step four-class classification

Fig. 7. Confusion matrices obtained from (a) subject-dependent and (b) subject-independent one-step four-class classification scenarios.

rate of the classifier in Fig. 7 (b). In summary, the mentioned four-class classification problem is different from the three-class one, even in a similar subject-independent scenario to analyze.

The one-tailed t-test on the obtained Fisher values of the extracted features of different EEG channels in a pair-wise scenario was applied. This trick aimed at determining the most significant EEG channels using the first feature set ($\mathbf{G}_1$) after calculating the Fisher criterion on ($\mathbf{G}_1$) for all trials and subjects. After obtaining p-values between these pair-wise fisher vectors, the average of these p-values is considered a significant metric for each EEG channel. Finally, for each channel, a 15-length significance vector is obtained over all fifteen subjects, and Fig. 9 shows the box plots of these significance vectors for all EEG channels. Based on these results, it can be seen that the T8 and Fp2 are the most statistically significant EEG channels for IS task. Table 4 shows the ranking of the EEG channels based on p-values provided in Fig. 9. Besides, the most repetitive features of each channel chosen in the final reduction step using the Fisher criterion are listed in this table. Table 4 shows that the predominant features were the wavelet coefficients and their statistical features in Alpha and Beta bands.

For analysis of the learned brain connectivity from GL methods (7) and (13), a statistical significance-based approach is designed and applied to determine thresholds (to binarize the learned graphs) that result in the most statistically significant connectivity illustrations. First, the pairwise p-values between sparsified normalized connectivity graphs are calculated by thresholding in the span of $[0.1, 0.15.0.2, ..., 0.95]$ provided in Fig. 10 (top row: GL method (7), bottom row: GL method (13)). Then, the optimal thresholds corresponding to either GL method (7) or (13) are determined by the maximum intersected threshold across pairwise p-value plots in Fig. 10 leading to p-value< 0.05. This way, the optimal thresholds are set at 0.7 and 0.6, corresponding to GL methods (7) and (13), respectively. The averages of the binarized learned graphs (adjacency matrices) based on these determined thresholds, over all class-specific samples and subjects, and for a particular class and GL method are illustrated in Fig. 11, showing the discriminative difference between `rest` vs. `non-rest` but similar patterns corresponding to imagined words `rock`, `paper`, and `scissors`. These findings are consistent with the classification performance obtained in Fig. 8 (i.e., high and low classification accuracy in `rest` vs. `non-rest` and three-class classification steps, respectively). From these brain graphs, it can be seen that there are considerable differences between the `rest` state and three other classes of the imagined words. In the `rest` state, most parts of the brain are connected, and there is no significant activity in a specific brain region. This sentence relies on the learned connectivity strengths, and those brain regions sparsified after thresholding were considered non-significant activities. However, in three classes of the imagined words, there are three strong connections (large edge values) that one of them (Broca's and Wernicke's areas) has been frequently mentioned in the IS literature [4,5]. This connection is placed on the top left hemisphere of the brain. Because the whole subjects were right-handed and their dominant hemisphere was on the left side, one expected to be seen. Surprisingly, a group of interactions can be seen on the right side of the brain, which is weaker than the connections on the left side. To the best knowledge, no article related to IS classification has strongly mentioned this phenomenon.

The question of why one has found a group of connections in the right hemisphere of the right-handed subjects is answered in [47]. This book mentions that "The fact that varying amount of language function may remain after dominant hemispherectomy in adults with glioma also suggest a definite though the limited capacity of the adult non-dominant hemisphere for language production." The connections on the right hemisphere are clearly showing this fact. In [48], it is also mentioned that some language-capable regions exist in the brain's right hemisphere. In another study, Belin et al. [49] experimented on two Broca's aphasia and five global aphasia patients. The patients were asked to repeat intoned words. They wanted to find the left hemisphere's activity in this task. Some findings of their study are usually ignored because of this study's focus on changes in the left hemisphere. Belin and his team found that the blood flow of the right hemisphere was highly altered during the mentioned task. This area consists of the right temporal lobe and right central operculum. Both of these areas are in the regions found in this study's right hemi-

Fig. 8. Confusion matrices obtained from a) Step1: `rest` vs. `non-rest`, b) Step2: Three-class classification, and c) Step1 + Step2, corresponding to the different steps of the proposed GraphIS approach to illustrate the bottleneck of the IS problem, and d) CSP.

sphere (Fig. 11). Therefore, electrodes from this part of the brain can also be efficiently used in imagined word classification. The third region of the brain shown in Fig. 11 (a) is placed in the occipital area. In [50], it is mentioned that sustained attention to visual stimuli can lead to connections in the occipital area of the brain. Because the stimulus used in this experiment was visual, there is a group of interactions in the occipital region. Fig. 11 (b) also has identical neurological signatures for depicted connections, except for interaction on frontal electrodes and `rest` state. The frontal connections maybe are related to eye artifacts. In the `rest` state, the second GL algorithm shows the visual stimuli effect on the occipital region, which is the only thing affecting the subject in `rest` epochs. Also, the non-occipital active electrodes in these graphs for each learning method are shown in Table 5. In this table, the electrodes that can be the source of further research have been listed to investigate the problem of IS problem with visual stimulation.

5. Discussion and future work

In this section, we briefly express the pros and cons of the proposed GraphIS method besides suggesting improvements. As well as the superior accuracy performance and robustness over the state-of-the-art CSP approach, the interpretability of the learned graphs on the brain areas is another considerable advantage. On the other hand, learning only *one* graph for a specific class instead of learning from each class sample can significantly lower the computational cost and is considered the future research direction. Besides, we seek to design graph-based classifiers to benefit from the graph structures of the data at hand instead of using classic SVM ones.

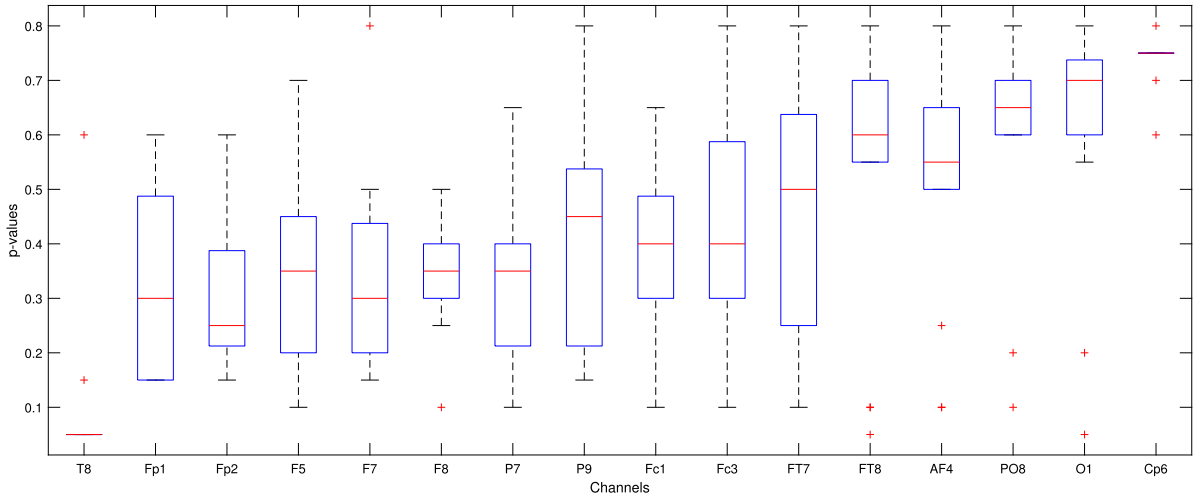


Fig. 9. *p*-values obtained for the mean pair-wise difference of the fisher vectors of the EEG channels over all fifteen subjects.

Table 4
The significance of the different EEG channels and repetition of their features in PS1 obtained from all trials and subjects.

Channel Name	Number of Features in final selected features using Fisher	Dominant features
T8	10	wavelet coefficients of the Delta and Alpha bands, max of wavelet coefficients of Beta band, min of wavelet coefficients of the Delta and Alpha bands, line length of the wavelet coefficients for all frequency bands
Fp2	9	Alpha band power, standard deviation of wavelet coefficients of the Delta and Alpha bands, max of wavelet coefficients of the Delta band, min of wavelet coefficients of the Alpha and Beta, line length wavelet coefficients for all frequency bands
F7	7	standard deviation of wavelet coefficients of the Delta band, max of wavelet coefficients of the Delta and Alpha bands, min of wavelet coefficients of the Alpha and Beta bands, line length of the wavelet coefficients for all frequency bands
Fp1	6	Alpha band power, standard deviation of wavelet coefficients of the Delta band, max of wavelet coefficients of the Delta band, min of wavelet coefficients of the Alpha band, power of wavelet coefficients for all frequency bands
F5	6	Theta peak frequency, mean of wavelet coefficients of the Theta band
F8	5	Theta higuchi, mean of wavelet coefficients of the Beta band, min of wavelet coefficients of the Alpha band
P7	3	Beta entropy, power of wavelet coefficients for all frequency bands
Fc1	2	Beta relative power, line length of the raw signal
Fc3	2	Wavelet coefficients
P9	2	Gamma highuchi, mean of wavelet coefficients of the Beta band
FT7	2	mean of wavelet coefficients of the Theta band
AF4	2	Alpha flatness, Alpha entropy
FT8	2	Wavelet coefficients
PO8	1	Wavelet coefficients
O1	1	standard deviation of wavelet coefficients of the Theta band
CP6	1	Wavelet coefficients

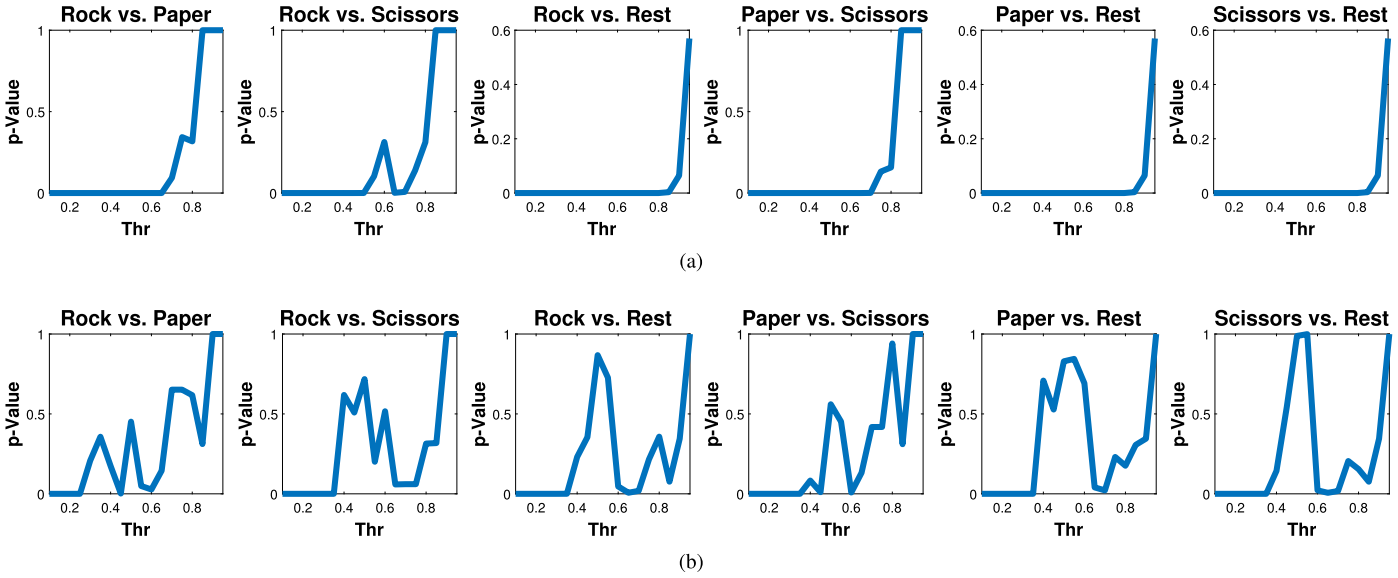


Fig. 10. The pairwise p-values between sparsified normalized connectivity graphs by thresholding in the span of $[0.1, 0.15.0.2, ..., 0.95]$ for GL methods a) (7) and b) (13).

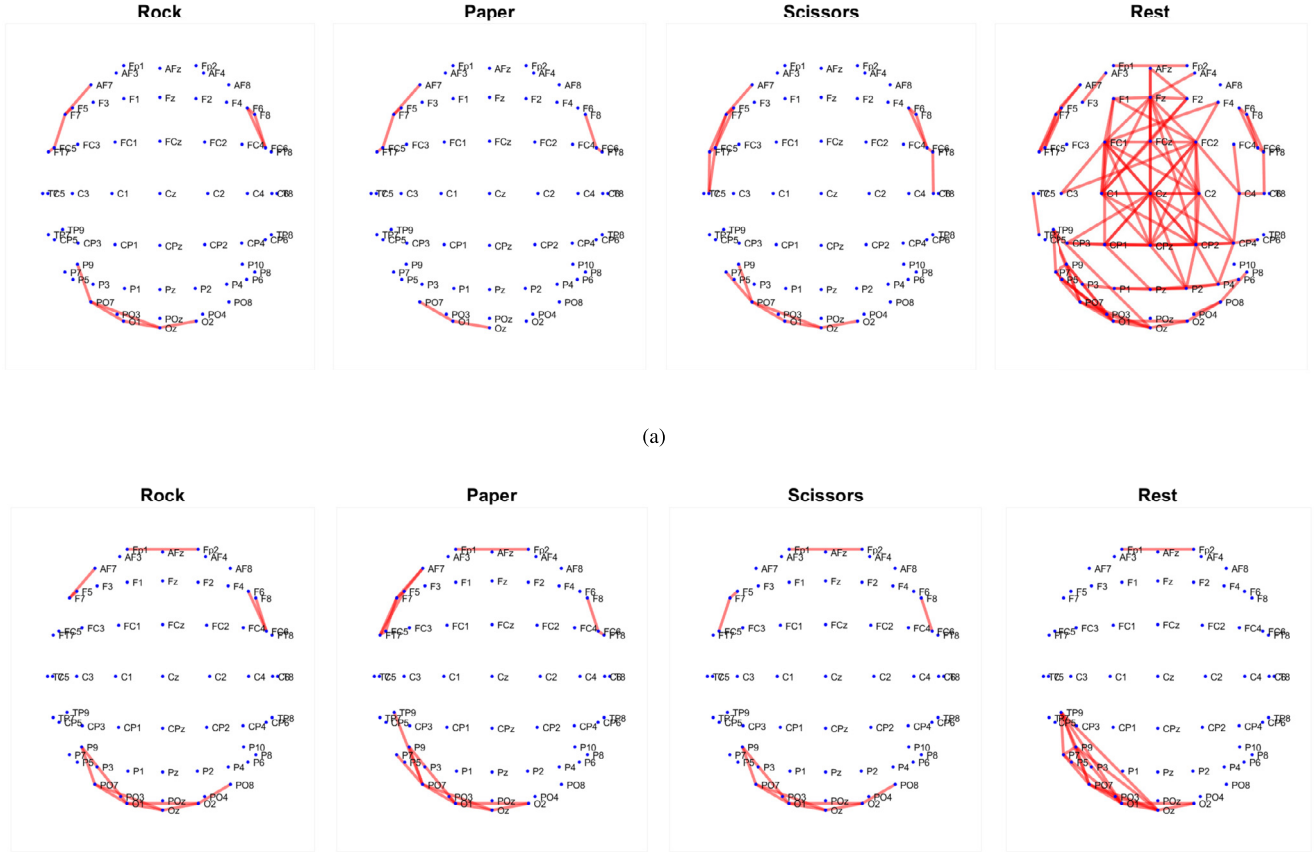


Fig. 11. Connectivity obtained using the feature subsets a) $\mathbf{G}_{31}$ and b) $\mathbf{G}_{32}$. Discrimination of `non-rest` vs. `rest` is well illustrated in this figure. The left and right hemispheres participate in the task of IS. Therefore, in addition to the frequently mentioned left brain regions (Broca's and Wernicke's areas), considerable interactions and connections are observed in the right brain regions, which had not been focused on the IS literature well. Also, the two GL methods (i.e. (7) and (13)) have discriminated against the `rest` state and IS tasks with acceptable accuracy.

Table 5
Non-occipital active electrodes for classes 1, 2, and 3, corresponding to specific GL methods.

Method	Rock	Paper	Scissors
G_{31}	F7, F8, FC5, FC6, P7, AF7, F5, F6, FT7, FT8, C6, P9	F7, F8, FC5, FC6, P7, AF7, F5, F6, FT7, FT8, C6, P9	F7, F8, FC5, FC6, Cz, P7, AF7, F5, F6, FT7, FT8, C5, C6, CPz, P9
G_{32}	Fp1, Fp2, F7, F8, FC5, FC6, P7, AF7, F5, F6, FT7, FT8, P9, TP9	Fp1, Fp2, F7, F8, FC5, FC6, P7, AF7, F5, F6, FT7, FT8, C6, P9, TP9	Fp1, Fp2, F7, F8, FC5, FC6, P7, AF7, F5, F6, FT7, FT8, C6, P9, TP9

From another point of view and providing that the datasets have a high number of class samples, designing and exploiting Graph Neural Network (GNN) can also benefit from the high flexibility of the neural networks.

6. Conclusion

This paper addressed the Imagined Speech task for application in BCI systems via Artificial Intelligence. EEG signals were recorded from the scalps of fifteen subjects during the imagination of the words `rock`, `paper`, or `scissors` or during subjects' `rest` state. The main goal of this work was to classify these classes. This work extracted classical signal processing features from the EEG channel signals. It also used graph signal processing and graph learning techniques to benefit from structural and functional connectivity information of EEG recordings. This project provided results that showed that the combination (fusion) of these feature sets and classical signal processing ones could significantly boost classification accuracy. Precisely, the proposed GraphIS method leads to a mean accuracy of 50.10% in the studied four-class IS classification task, compared to using only one feature set with an accuracy of 47.86% and 46.10%, and also the state-of-the-art CSP with an accuracy of 47.10%. Also, connected brain regions have been located using graph learning algorithms, which, based on neuroscientific studies, are relevant to the Imagined Speech task. These regions, placed in the left and right brain areas, and their roles during IS tasks are compatible with neurological information studied before.

Human and animal rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans as well as in accordance with the EU Directive 2010/63/EU for animal experiments.

Funding

This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Declaration of competing interest

The authors declare that they have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper.

Acknowledgements

Authors would like to acknowledge National Brain Mapping Laboratory (NBML), Tehran, Iran, for providing data acquisition (analysis) service for this research work.

References

[1] Lawrence Ashley Farwell, Emanuel 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.
[2] Bruce Denby, Tanja Schultz, Kiyoshi Honda, Thomas Hueber, Jim M. Gilbert, Jonathan S. Brumberg, Silent speech interfaces, Speech Commun. 52 (4) (2010) 270–287.
[3] Stephanie Martin, Peter Brunner, Iñaki Iturrate, José del, R. Millán, Gerwin Schalk, Robert T. Knight, Brian N. Pasley, Word pair classification during imagined speech using direct brain recordings, Sci. Rep. 6 (1) (2016) 1–12.
[4] Robert J. Zatorre, Alan C. Evans, Ernst Meyer, Albert Gjedde, Lateralization of phonetic and pitch discrimination in speech processing, Science 256 (5058) (1992) 846–849.
[5] Jean-François Demonet, Francois Chollet, Stuart Ramsay, Dominique Cardebat, Jean-Luc Nespoulous, Richard Wise, André Rascol, Richard Frackowiak, The anatomy of phonological and semantic processing in normal subjects, Brain 115 (6) (1992) 1753–1768.
[6] Netiwit Kaongoen, Jaehoon Choi, Sungho Jo, Speech-imagery-based brain–computer interface system using ear-EEG, J. Neural Eng. 18 (1) (2021) 016023.
[7] Lingxi Lu, Jingwei Sheng, Zhaowei Liu, Jia-Hong Gao, Neural representations of imagined speech revealed by frequency-tagged magnetoencephalography responses, NeuroImage 229 (2021) 117724.
[8] Alborz Rezazadeh Sereshkeh, Rozhin Yousefi, Andrew T. Wong, Tom Chau, Online classification of imagined speech using functional near-infrared spectroscopy signals, J. Neural Eng. 16 (1) (2018) 016005.
[9] Xiaomei Pei, Dennis L. Barbour, Eric C. Leuthardt, Gerwin Schalk, Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans, J. Neural Eng. 8 (4) (2011) 046028.
[10] Areej H. Al-Anbary, Salih Al-Qaraawi, A survey of EEG signals prepossessing and classification for imagined speech application, Int. J. Innov. Eng. Sci. Res. 4 (3) (2020) 1–9.
[11] Li Wang, Xiong Zhang, Xuefei Zhong, Yu Zhang, Analysis and classification of speech imagery EEG for BCI, Biomed. Signal Process. Control 8 (6) (2013) 901–908.
[12] L.C. Sarmiento, P. Lorenzana, C.J. Cortes, W.J. Arcos, J.A. Bacca, A. Tovar, Brain computer interface (BCI) with EEG signals for automatic vowel recognition based on articulation mode, in: 5th ISSNIP-IEEE Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC), IEEE, 2014, pp. 1–4.
[13] Michael D'Zmura, Siyi Deng, Tom Lappas, Samuel Thorpe, Ramesh Srinivasan, Toward EEG sensing of imagined speech, in: International Conference on Human-Computer Interaction, Springer, 2009, pp. 40–48.
[14] Abhiram Singh, Ashwin Gumaste, Decoding imagined speech and computer control using brain waves, J. Neurosci. Methods 358 (2021) 109196.
[15] Noramiza Hashim, Aziah Ali, Wan-Noorshahida Mohd-Isa, Word-based classification of imagined speech using eeg, in: International Conference on Computational Science and Technology, Springer, 2017, pp. 195–204.
[16] Pramit Saha, Muhammad Abdul-Mageed, Sidney Fels, Speak your mind! towards imagined speech recognition with hierarchical deep learning, preprint, arXiv:1904.05746, 2019.
[17] Katharine Brigham, B.V.K. Vijaya Kumar, Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy, in: 2010 4th International Conference on Bioinformatics and Biomedical Engineering, IEEE, 2010, pp. 1–4.
[18] Chuong H. Nguyen, George K. Karavas, Panagiotis Artemiadis, Inferring imagined speech using EEG signals: a new approach using Riemannian manifold features, J. Neural Eng. 15 (1) (2017) 016002.
[19] Alejandro Antonio Torres García, Carlos A. Reyes García, Luis Villaseñor Pineda, Toward a silent speech interface based on unspoken speech, in: Biosignals, 2012, pp. 370–373.
[20] Jesús S. García-Salinas, Luis Villaseñor-Pineda, Carlos Alberto Reyes-García, Alejandro Torres-García, Tensor decomposition for imagined speech discrimination in EEG, in: Mexican International Conference on Artificial Intelligence, Springer, 2018, pp. 239–249.
[21] Ciaran Cooney, Folli Raffaella, Damien Coyle, Optimizing input layers improves CNN generalization and transfer learning for imagined speech decoding from EEG, in: IEEE International Conference on Systems, Man, and Cybernetics, 2019: Industry 4.0, 2019.
[22] Muhammad Naveed Iqbal Qureshi, Beomjun Min, Hyeong-jun Park, Dongrae Cho, Woosu Choi, Boreom Lee, Multiclass classification of word imagination speech with hybrid connectivity features, IEEE Trans. Biomed. Eng. 65 (10) (2017) 2168–2177.

[23] Alejandro A. Torres-García, Carlos A. Reyes-García, Luis Villaseñor-Pineda, Gregorio García-Aguilar, Implementing a fuzzy inference system in a multi-objective EEG channel selection model for imagined speech classification, Expert Syst. Appl. 59 (2016) 1–12.
[24] Mashael AlSaleh, Roger Moore, Heidi Christensen, Mahnaz Arvaneh, Discriminating between imagined speech and non-speech tasks using EEG, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 1952–1955.
[25] Charles S. DaSalla, Hiroyuki Kambara, Makoto Sato, Yasuharu Koike, Single-trial classification of vowel speech imagery using common spatial patterns, Neural Netw. 22 (9) (2009) 1334–1339.
[26] Jerrin Thomas Panachakel, Ramakrishnan Angarai Ganesan, T.V. Ananthapadmanabha, Common spatial pattern based data augmentation technique for decoding imagined speech, in: 2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), IEEE, 2021, pp. 1–5.
[27] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, Pierre Vandergheynst, Graph signal processing: overview, challenges, and applications, Proc. IEEE 106 (5) (2018) 808–828.
[28] Keith Smith, Benjamin Ricaud, Nauman Shahid, Stephen Rhodes, John M. Starr, Augustin Ibáñez, Mario A. Parra, Javier Escudero, Pierre Vandergheynst, Locating temporal functional dynamics of visual short-term memory binding using graph modular Dirichlet energy, Sci. Rep. 7 (1) (2017) 1–12.
[29] Madiha J. Jafri, Godfrey D. Pearlson, Michael Stevens, Vince D. Calhoun, A method for functional network connectivity among spatially independent resting-state components in schizophrenia, NeuroImage 39 (4) (2008) 1666–1681.
[30] Urs Braun, Sarah F. Muldoon, Danielle S. Bassett, On human brain networks in health and disease, eLS (2015) 1–9.
[31] Cornelis J. Stam, Jaap C. Reijneveld, Graph theoretical analysis of complex networks in the brain, Nonlinear Biomed. Phys. 1 (1) (2007) 1–19.
[32] Liu Rui, Hossein Nejati, Ngai-Man Cheung, Dimensionality reduction of brain imaging data using graph signal processing, in: 2016 IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pp. 1329–1333.
[33] Siddhi Tavildar, Brian Mogen, Stavros Zanos, Stephanie C. Seeman, Steve I. Perlmutter, Eberhard Fetz, Ashkan Ashrafi, Inferring cortical connectivity from ECoG signals using graph signal processing, IEEE Access 7 (2019) 109349–109362.
[34] Toshihisa Tanaka, Takashi Uehara, Yuichi Tanaka, Dimensionality reduction of sample covariance matrices by graph Fourier transform for motor imagery brain-machine interface, in: 2016 IEEE Statistical Signal Processing Workshop (SSP), IEEE, 2016, pp. 1–5.
[35] Golnar Kalantar, Hamidreza Sadreazami, Arash Mohammadi, Amir Asif, Adaptive dimensionality reduction method using graph-based spectral decomposition for motor imagery-based brain-computer interfaces, in: 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, 2017, pp. 990–994.
[36] Xiaowen Dong, Dorina Thanou, Michael Rabbat, Pascal Frossard, Learning graphs from data: a signal representation perspective, IEEE Signal Process. Mag. 36 (3) (2019) 44–63.
[37] Xiaowen Dong, Dorina Thanou, Pascal Frossard, Pierre Vandergheynst, Learning Laplacian matrix in smooth graph signal representations, IEEE Trans. Signal Process. 64 (23) (2016) 6160–6173.
[38] Vassilis Kalofolias, How to learn a graph from smooth signals, in: Artificial Intelligence and Statistics, PMLR, 2016, pp. 920–929.
[39] Aref Einizade, Mohsen Mozafari, Mohammadreza Rezaei-Dastjerdehei, Elnaz Aghdaei, Amir Mohammad Mijani, Sepideh Hajipour Sardouie, Detecting ADHD children based on EEG signals using graph signal processing techniques, in: 2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME), IEEE, 2020, pp. 264–270.
[40] Nasser Samadzadeh Aghdam, Mohammad Hassan Moradi, Mohammad Bagher Shamsollahi, Ali Motie Nasrabadi, Seyed Kamaledin Setarehdan, Vahid Shalchyan, Farhad Faradji, Bahador Makkiabadi, The 2017 and 2018 Iranian Brain–Computer Interface Competitions, J. Med. Signal. Sens. 10 (3) (2020) 208.
[41] Richard O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley-Interscience, 2001.
[42] Pierre A. Balthazard, David A. Waldman, Robert W. Thatcher, Sean T. Hannah, Differentiating transformational and non-transformational leaders on the basis of neurological imaging, Leadersh. Q. 23 (2) (2012) 244–258.
[43] Umut Güçlü, Yağmur Güçlütürk, Chu Kiong Loo, Evaluation of fractal dimension estimation methods for feature extraction in motor imagery based brain computer interface, Proc. Comput. Sci. 3 (2011) 589–594.
[44] Rosana Esteller, Javier Echauz, T. Tcheng, Brian Litt, Benjamin Pless, Line Length: an Efficient Feature for Seizure Onset Detection, 2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 2, IEEE, 2001, pp. 1707–1710.
[45] Shlomo Dubnov, Generalization of spectral flatness measure for non-Gaussian linear processes, IEEE Signal Process. Lett. 11 (8) (2004) 698–701.
[46] Hirotugu Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (6) (1974) 716–723.
[47] H.R. Allan, Diseases of the peripheral nerves, in: H.R. Allan, A.S. Martin, P.K. Joshua, P. Sashank (Eds.), Adams and Victor's Principles of Neurology, 2019.
[48] Gottfried Schlaug, Sarah Marchina, Andrea Norton, From singing to speaking: why singing may lead to recovery of expressive language function in patients with Broca's Aphasia, Music Percept. 25 (4) (2008) 315.
[49] Pascal Belin, M. Zilbovicius, Ph Remy, C. Francois, S. Guillaume, F. Chain, G. Rancurel, Y. Samson, Recovery from nonfluent aphasia after melodic intonation therapy: a pet study, Neurology 47 (6) (1996) 1504–1511.
[50] Thomas Z. Lauritzen, Mark D'Esposito, David J. Heeger, Michael A. Silver, Top–down flow of visual spatial attention signals from parietal to occipital cortex, J. Vis. 9 (13) (2009) 18.