Neural Networks 133 (2021) 193–206

Contents lists available at ScienceDirect

Neural Networks

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

An EEG channel selection method for motor imagery based brain–computer interface and neurofeedback using Granger causality

Hesam Varsehi [a], S. Mohammad P. Firoozabadi [b,*]

[a] Department of Biomedical Engineering, Tarbiat Modares University, Tehran, Iran
[b] Department of Medical Physics, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Article history:
Received 4 July 2020
Received in revised form 8 October 2020
Accepted 5 November 2020
Available online 10 November 2020

Keywords:
Motor imagery (MI)
Electroencephalogram (EEG)
EEG channel selection
Granger causality
Neurofeedback
Brain–computer interface (BCI)

ABSTRACT

Motor imagery (MI) brain–computer interface (BCI) and neurofeedback (NF) with electroencephalogram (EEG) signals are commonly used for motor function improvement in healthy subjects and to restore neurological functions in stroke patients. Generally, in order to decrease noisy and redundant information in unrelated EEG channels, channel selection methods are used which provide feasible BCI and NF implementations with better performances. Our assumption is that there are causal interactions between the channels of EEG signal in MI tasks that are repeated in different trials of a BCI and NF experiment. Therefore, a novel method for EEG channel selection is proposed which is based on Granger causality (GC) analysis. Additionally, the machine-learning approach is used to cluster independent component analysis (ICA) components of the EEG signal into artifact and normal EEG clusters. After channel selection, using the common spatial pattern (CSP) and regularized CSP (RCSP), features are extracted and with the k-nearest neighbor (k-NN), support vector machine (SVM) and linear discriminant analysis (LDA) classifiers, MI tasks are classified into left and right hand MI. The goal of this study is to achieve a method resulting in lower EEG channels with higher classification performance in MI-based BCI and NF by causal constraint. The proposed method based on GC, with only eight selected channels, results in 93.03% accuracy, 92.93% sensitivity, and 93.12% specificity, with RCSP feature extractor and best classifier for each subject, after being applied on Physionet MI dataset, which is increased by 3.95%, 3.73%, and 4.13%, in comparison with correlation-based channel selection method.

1. Introduction

One of the brain–computer interface (BCI) definitions is a complete system that controls different communication devices by processing human brain signals (Ramadan & Vasilakos, 2017). In a more generalized definition, brain–machine interface (BMI) can restore the lost neurological functions caused by stroke or other nervous system injuries (Donoghue, 2002). Motor imagery (MI)-based BCI uses motion imagination instead of motion execution, to issue control commands to the BCI external devices (Feng et al., 2018). Unlike BCI, neurofeedback uses visual, auditory or other feedback to help the participant in self-regulation of neural substrates systems instead of directly commanding external devices (Ghoshuni, Firoozabadi, Khalilzadeh, & Hashemi Golpayegani, 2012; Sitaram et al., 2017). In addition to motor areas and pathways improvements in post-stroke subjects, MI-based neurofeedback training also amends motor function in healthy subjects (Donoghue, 2002; Karácsony, Hansen, Iversen, & Puthusserypady, 2019).

Among brain signals, electroencephalogram (EEG) signal is the most common signal in BCI systems as it is non-invasive, portable, low cost, and with high temporal resolution (Padfield, Zabalza, Zhao, Masero, & Ren, 2019). Although the high number of EEG channels provide more information about the background's neural activity, it increases the redundancy due to noise and creates high-dimensional data (Handiru & Prasad, 2016). Besides the above-mentioned reasons, the feasibility of BCI implementation, BCI cost reduction, and improvement of the BCI performance are the main reasons to reduce the number of EEG signal channels. The noted procedure is called channel selection (Arvaneh, Guan, Ang, & Quek, 2011; Kim et al., 2016; Shan, Xu, Zhu, & He, 2015).

The channel selection algorithms for event related potentials are discussed in Feess, Krell, and Metzen (2013). In the field of motor imagery, various methods have been proposed to select the appropriate EEG signal channels. Popular techniques for channel selection of MI EEG signals are filtering and wrapper techniques (Alotaiby, El-Samie, Alshebeili, & Ahmad, 2015). Filtering techniques are based on EEG signal statistics which mostly

* Corresponding author.
E-mail address: pourmir@modares.ac.ir (S. Mohammad P. Firoozabadi).

https://doi.org/10.1016/j.neunet.2020.11.002
0893-6080/

use common spatial pattern (CSP) coefficients and mostly based on Fisher criterion (Yang, Kyrgyzov, Wiart, & Bloch, 2013) or mutual information (Ang, Chin, Zhang, & Guan, 2012). Wrapper techniques include methods such as Fisher's discriminant analysis and particle swarm optimization algorithm (Alotaiby et al., 2015). Another technique for channel selection uses class SVM classifier. In this way, in a recursive process, the length of the feature vector is reduced by removing unrelated channels (Lal et al., 2004). There are a number of other channel selection methods, such as artificial neural network genetic algorithm (Yang et al., 2012) and sparse CSP (Arvaneh et al., 2011), that are computationally time consuming. Although the easiest method for channel selection is based on neurophysiologic knowledge (Jin et al., 2019). Briefly, these methods for channel selection propose low performances. Also, most of these methods are not based on neurophysiology. Recently Jin et al. used correlation-based channel selection for the MI EEG signal (Jin et al., 2019).

Neural connectivity in the brain can be analyzed at three levels: anatomical, functional, and effective connectivity between neural units. According to the level of study, neural units can be considered as single neurons, neuronal populations, or brain areas (Friston, 1994). Anatomical or structural connectivity can provide a description of the brain network which refers to the static anatomical structure of the brain (Greenblatt, Pflieger, & Ossadtchi, 2012; Haufe, 2012). "Functional connectivity is defined as statistical dependencies among remote neurophysiological events"(Friston, 2011). Linear or nonlinear covariation between the recorded EEG signals from multiple channels is derived from functional connectivity. Correlation between data of two channels is one of the linear measures of functional connectivity (Cohen, 2014). J. Friston in Friston (2011) proposed that "effective connectivity refers explicitly to the influence that one neural system exerts over another, either at a synaptic or population level". Also, effective connectivity relies on the activity of the causal interactions between neural units within the nervous system as an interconnected network (Cohen, 2014; Friston, 2011). Numerous connectivity measures have been developed and applied in neuro-scientific studies which Granger Causality (GC) is one of them. GC is the method for causal influence detection of time series like EEG signals. It is the most popular method for analyzing effective connectivity between brain regions (Hu et al., 2015).

In addition to the applied functional connectivity perspective in Jin et al. (2019), for channel selection, effective connectivity may be used which is considered in this manuscript. The idea of this work is fundamentally based on the greater significance of some special EEG channels while measuring the causality of these channels onto others, during the MI action. These special channels may then be repeated throughout the experiment and during all trials.

Motor imagery EEG signal processing includes preprocessing, feature extraction, and classification. Nevertheless, there is another important part that is not usually considered in other research, which is the channel selection part (Aggarwal & Chugh, 2019; Shan et al., 2015). Aggarwal and Chugh (2019) divided the feature extraction part into two spectral and spatial domains. In the spectral domain, different methods have been proposed such as wavelet and Fourier transform (Park, Took, & Mandic, 2013; Tolić & Jović, 2013), autoregressive model (Park et al., 2013), and phase information (Loboda, Margineanu, Rotariu, & Lazar, 2014). Spatial domain feature extraction contains different models of CSP methods that are mostly used in MI-based BCI (Aggarwal & Chugh, 2019; Handiru & Prasad, 2016; Kim et al., 2016; Park et al., 2013). The small number of training samples reduces the performance of CSP, which regularized CSP (RCSP) has been introduced as an improved method to solve this problem (Friedman, 1989). In the classification step, features of EEG signals are classified into MI tasks (such as left and right MI) with support vector machines (SVM) (Handiru & Prasad, 2016; Park et al., 2013), neural networks (Dose, Møller, Iversen, & Puthusserypady, 2018; Karácsony et al., 2019; Tolić & Jović, 2013), random forest (Kim et al., 2016), and linear discriminant analysis (LDA) methods (Aggarwal & Chugh, 2019).

The rest of this paper is as follows. Dataset description and proposed methods are described in Section 2. The results are shown in Section 3. Results discussion and comparison with other works are presented in Section 4. Finally, a summary of the study is given in Section 5.

2. Materials and methods

2.1. Main dataset

The used dataset in this study is from the BCI2000 instrumentation system (Schalk, McFarland, Hinterberger, Birbaumer, & Wolpaw, 2004) which is available in Physionet (Goldberger et al., 2000). It consists of 64-channel EEG, recorded from 109 subjects. Subjects were asked to perform different motor movement/imagery tasks while EEG signals were recorded from 64 electrodes along the surface of the scalp. Each subject performed 14 experimental runs:

1. 1-minute baseline run, while eyes are open.
2. 1-minute baseline run, while eyes are close.
3. Three 2-minutes runs of each of the four Tasks 1, 2, 3, and 4.

In this work, only Task 2 events are used. A target is displayed on either the left or right side of the monitor for each event of Task 2. The subject is asked to imagine opening and closing the corresponding hand until the target gets disappeared. Then, the subject relaxes.

In our study, we have used data from 105 subjects out of the 109 subjects available in the dataset. All the selected subjects have at least 42 trials in their MI tasks (21 trials for each left and right hand MI). In other words data of 4 subjects (S088, S092, S100, and S104) were excluded because of the number of trials or damaged data. Subjects' data are provided in EDF+ format and contain 64 raw EEG signals as per the international 10–10 system, sampled at 160 Hz.

2.2. Preprocessing

In the preliminary analysis, the acquired EEG data were refined before the channel selection part and features extraction for classification. Fig. 1 exhibits preprocessing procedure.

2.2.1. Filtering

In the filtering part, two consecutive low-pass and high-pass filters are applied to raw EEG data. Finally, EEG data is filtered from 0.5 to 50 Hz (the frequency of power line noise was 60 Hz).

2.2.2. Segmentation

The filtered continuous EEG data is segmented by extracting data segments that are related to each event type. The MI events consist of three time intervals for both left and right hand MI. It begins with two seconds for rest which is followed by 4.1 s of MI and ends with two seconds for rest again. Fig. 2 shows the time interval of one trial for events of Task 2.

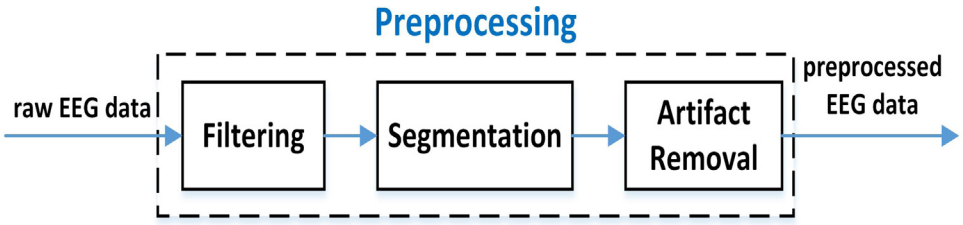

Fig. 1. preprocessing procedure.

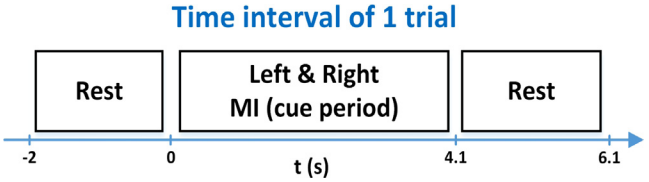

Fig. 2. The time interval of one trial for Task 2 events.

2.2.3. Artifact removal

The recorded EEG data usually contains noise and artifacts that are arisen due to various factors. Some common artifacts are eye blinks, ocular movements, muscle artifacts like chewing, and facial muscle movements (Winkler, Haufe, & Tangermann, 2011). In one perspective, artifact removal of the EEG signal can be separated into cancellation or correction of the artifacts by two methods: filtering or extracting source components of the signal (Islam, Rastegarnia, & Yang, 2016). In the BCI criterion, two major physiological artifacts are EOG (with higher power than normal EEG) and EMG (containing higher frequencies than normal EEG) (Mahajan & Morshed, 2014; Winkler et al., 2011). From another point of view, artifact removal methods can be separated into 3 categories: hand-optimized, semi-automatic, and fully-automatic (Winkler et al., 2011). An alternative solution to detect artifacts in the EEG signal is based on blind source separation (BSS) techniques that separate EEG signals into brain source components. After detecting artifact components, undesired components are omitted and remained ones are used to reconstruct EEG signals (Fatourechi, Bashashati, Ward, & Birch, 2007). Let E be the recorded multichannel EEG signal, which is assumed to be a linear mixture of the sources, S, with white noise, N, as follows (Islam et al., 2016):

$$E = AS + N \tag{1}$$

which A is the linear mixture matrix. The aim is to estimate matrix A, denoted by W (un-mixing matrix $W = A^{-1}$) in an iterative procedure to estimate source signals as follows (Islam et al., 2016):

$$\hat{S} = WE \tag{2}$$

The most commonly used method in BSS is independent component analysis (ICA) (Fatourechi et al., 2007). The desired EEG signal and the artifact sources are independent (Islam et al., 2016). As a result, filtering EEG signals (removing EMG signal) and ICA method (for removing eyeblink and other similar artifacts) will be used to correct artifacts in the proposed work. Two methods for correcting artifacts of EEG signal with ICA are proposed. Both of the methods use the runICA algorithm of EEGLAB.

2.2.4. Method 1

First ICA is applied to EEG data for each trial which is followed by removing two components of ICA with the highest powers. So in the following subsections, this method is called MAXAR. Finally, EEG data are reconstructed from remained components.

Algorithm 1 KMAR (k-means artifact removal)

1: Load EEG signals' data of dimension $(N_s \times N_{ch} \times N_{tr})$. Where N_s: is the number of samples, N_{ch}: is the number of all channels, N_{tr}: is the number of trials.
2: **for** $i = 1 : N_{tr}$ **do**
3: Normalize data using Z-score normalization method
4: **end for**
5: Arrange EEG trials' data below each other to form a matrix of dimension $(N_t \times N_{ch})$ that $N_t = N_s \times N_{tr}$
6: Transpose the matrix of the previous step to form the E matrix of dimension $(N_{ch} \times N_t)$
7: Calculate ICA weights' matrix, W, using the EEGLAB runICA algorithm.
8: Form the estimation of EEG signal sources' matrix $\hat{S} = W \times E$
9: Extract features: calculate variance, skewness, and IQR (interquartile range) for each row of $\hat{S}$ matrix.
10: Apply the k-means clustering method to extracted features for 2 clusters (artifacts and normal EEG data). That cluster is known as an artifact, which includes the component with maximum variance.
11: Remove artifact cluster's components from $\hat{S}$ matrix to form $\hat{S}_{clean}$ matrix
12: Reconstruct EEG data using W and $\hat{S}_{clean}$ matrices as $E_{clean} = W^{-1} \times \hat{S}_{clean}$

2.2.5. Method 2

Nowadays the methods for artifact removal of EEG signal with the ICA, adopt the idea of machine learning by supervised or unsupervised classifiers. The most common unsupervised learning methods in this criterion is k-means clustering (Islam et al., 2016). Therefore ICA is applied to EEG data for each trial. After that, the ICA component with the highest power is selected as cluster 1. Cluster 1 is supposed to contain artifacts. So as eyeblinks have higher power than normal EEG signals, they are labeled with cluster 1 (Mahajan & Morshed, 2014; Winkler et al., 2011). It is followed by detecting other artifacts. Then, some features are calculated, like variance, skewness, and IQR (interquartile range) of the EEG signal. In the third part, ICA components are clustered into two clusters, artifact, and normal EEG. This part uses k-means machine learning to remove artifact components of ICA. Finally, EEG data are reconstructed from remained components. Because of using k-means machine learning, this method is called KMAR. The procedures of this method are shown in algorithm 1.

2.3. Channel selection

The aim of this section is to select some EEG channels that have much more influence on the process of MI in the brain. For this purpose, we consider two aspects, functional and effective connectivity. The interactions of spatially separated brain areas can be observed in these two aspects (Friston, 2011). Functional connectivity reflects statistically related activation of brain areas,

while effective connectivity explains the causality of these observed dependencies. Therefore they can provide crucial knowledge about the task-related direction of information flow (Cadotte, DeMarse, He, & Ding, 2008).

2.3.1. Functional connectivity

Cichocki et al. in Jin et al. (2019) proposed that EEG channels related to the same MI task should contain common information. In order to measure the similarity between any two channels, the correlation is used. First of all the Z-score normalization method is applied to EEG data. After that correlation coefficients are computed with Pearson's correlation analysis which is defined as:

$$\rho(C1, C2) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{C1_i - \overline{C1}}{\sigma_{C1}}\right)\left(\frac{C2_i - \overline{C2}}{\sigma_{C2}}\right) \tag{3}$$

where $C1$ and $C2$ are EEG data of two arbitrary channels, n is the number of data samples, $\overline{C1}$ and $\overline{C2}$, and σ_{C1} and σ_{C2} are the means and standard deviation of the two channels' data respectively. The algorithm for channel selection using the correlation-based method (CCS: Correlation-based Channel Selection) is discussed in Jin et al. (2019).

2.3.2. Effective connectivity

The basic idea of effective connectivity is that the cause overrides its effect over time. Therefore, when the activity of a nervous unit (cause) leads to the activation of the next unit (effect), there is a significant sequence between their activities (Cohen, 2014). "A time series (random process) X is said to Granger-cause Y if X provides predictive information about (future) values of Y over and above what may be predicted from past values of Y (and, optionally, from past values of other observed time series Z1, Z2,…)" (Greenblatt et al., 2012). In the area of linear Granger causality (GC) and bivariate case, let $x_1(t)$ and $x_2(t)$, to be zero-mean signals with $t = 1, 2, 3, \ldots, M$. Then the P-order univariate autoregressive (AR) model of $x_1(t)$ and $x_2(t)$ is as below:

$$x_1(t) = \sum_{\tau=1}^{P} a_{1\tau}\, x_1(t-\tau) + e_1(t) \tag{4}$$

$$x_2(t) = \sum_{\tau=1}^{P} a_{2\tau}\, x_2(t-\tau) + e_2(t) \tag{5}$$

where $e_1(t)$ and $e_2(t)$ are the estimation error of $x_1(t)$ and $x_2(t)$ respectively. They depend only on each signal's past and $a_{1\tau}$ and $a_{2\tau}$ are the AR coefficients (Gourévitch, Le Bouquin-Jeannès, & Faucon, 2006). The P-order bivariate ARX (autoregressive with external input) model is considered by Granger as (Granger, 1969):

$$x_1(t) = \sum_{\tau=1}^{P} a_\tau\, x_1(t-\tau) + \sum_{\tau=1}^{P} b_\tau\, x_2(t-\tau) + e_1(t) \tag{6}$$

$$x_2(t) = \sum_{\tau=1}^{P} c_\tau\, x_1(t-\tau) + \sum_{\tau=1}^{P} d_\tau\, x_2(t-\tau) + e_2(t) \tag{7}$$

where a_τ, b_τ, c_τ, and d_τ are the ARX coefficients and $e_1(t)$ and $e_2(t)$ are the prediction errors.

MVAR models consist of univariate and multivariate terms. Univariate terms model individual time series, whereas multivariate terms describe how time series depends on past values of other time series. Vector autoregressive (VAR) models are related to the concept of GC (Cohen, 2014). Assuming that time series are stationary, a P-order MVAR model can be written as shown below (Brockwell, Davis, & Fienberg, 1991):

$$\mathbf{x}(t) = \sum_{\tau=1}^{P} \mathbf{A}_\tau \mathbf{x}(t-\tau) + \mathbf{e}(t) \tag{8}$$

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{bmatrix} = \sum_{\tau=1}^{P} \mathbf{A}_\tau \begin{bmatrix} x_1(t-\tau) \\ x_2(t-\tau) \\ \vdots \\ x_N(t-\tau) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \\ \vdots \\ e_N(t) \end{bmatrix} \tag{9}$$

$$\mathbf{A}_\tau = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1N} \\ a_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{N1} & \cdots\cdots & & a_{NN} \end{bmatrix} \tag{10}$$

where $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_N(t)]^T$ represents an N-channel EEG data (for one subject) which $[.]^T$ denotes the transpose of the matrix, $x_k(t)$ is the kth channel data in time t, $\mathbf{A}_\tau$ is the $N \times N$ matrix of autoregressive coefficients, $\mathbf{e}(t)$ is vector of residuals or prediction errors and $e_k(t)$ is the estimation of kth channel's error. A P-order MVAR model for $\mathbf{x}_{(j)}(t)$ (that is equal to $\mathbf{x}(t)$ whose row j (jth channel data) has been removed) is as:

$$\mathbf{x}_{(j)}(t) = \sum_{\tau=1}^{P} \mathbf{B}_\tau \mathbf{x}_{(j)}(t-\tau) + \boldsymbol{\eta}(t) \tag{11}$$

$$\mathbf{B}_\tau = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1(N-1)} \\ b_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ b_{(N-1)1} & \cdots & & b_{(N-1)(N-1)} \end{bmatrix} \tag{12}$$

where $\mathbf{x}_{(j)}(t) = [x_1(t), \ldots, x_{j-1}(t), x_{j+1}(t), \ldots, x_N(t)]^T$ represents an $(N-1)$-channel EEG data (for one subject), $\mathbf{B}_\tau$ is the $(N-1) \times (N-1)$ matrix of autoregressive coefficients and $\boldsymbol{\eta}(t) = [\eta_1(t), \ldots, \eta_{j-1}(t), \eta_{j+1}(t), \ldots, \eta_N(t)]^T$ is vector of residuals or prediction error. Then the multivariate Granger causality ($MVGC$) from channel i to channel j ($MVGC_{j\to i}$) is defined as:

$$MVGC_{j\to i} = ln\frac{\frac{1}{M}\sum_{t=1}^{M}\eta_i^2(t)}{\frac{1}{M}\sum_{t=1}^{M}e_i^2(t)} \tag{13}$$

where M is the number of EEG data observations (Krumin & Shoham, 2010). The value of $e_i^2(t)$ is less than η_i^2. Since $e_i^2(t)$ is related to the case where the estimation error is obtained by observing all the channels' data, which more data is available for better estimation, while η_i^2 is related to the case where the data of jth channel has been removed. If the removed channel data has no effect on ith channel estimation, these two error values are equal and then $MVGC$ value will be zero. In other words, when the value of $MVGC$ from jth channel to ith channel ($MVGC_{j\to i}$) is zero, the data of these channels have no effect on estimating each other and it can be said that they are not cause of each other, in terms of Granger causality criterion.

Most practical systems for determining model order use Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). In proposed work we used the BIC method to determine MVAR model order (Gurka & Edwards, 2007) as follows:

$$BIC = log(E) + \frac{P\ log(M)}{M} \tag{14}$$

where P is the order MVAR model, E is the estimation error and M is the number of EEG data observations. In this work, according to similar articles, MVAR model order (P) range was from 3 to 25.

Algorithm 2 Proposed Adaptive MVGC Thresholding (AMVGCT)

1: Load EEG signals' data of dimension ($N_s \times N_{ch}$), where N_s: is the number of samples, N_{ch}: is the number of all channels.
2: Add a random noise with a normal distribution (with zero mean and variance equal to mean variances of other channels) as a new channel. New data dimension will be ($N_s \times N_{ch+1}$) which is named $\mathbf{x}(t)$ matrix.
3: Calculate $\mathbf{e}(t)$ (prediction error vector with N_{ch+1} rows), refer to Eq. (8)
4: **for** $j = 1 : N_{ch+1}$ **do**
5: Remove j^{th} channel data from $\mathbf{x}(t)$ forming $\mathbf{x}_{(j)}(t)$
6: Calculate $\boldsymbol{\eta}(t)$ (prediction error vector with N_{ch} rows) for estimation of $\mathbf{x}_{(j)}(t)$
7: Calculate $MVGC_{j\to i}$ (i: is all channels number and $i \neq j$, refer to Eq. (13)), where this vector is the j^{th} row of $MVGC_{noise}$ matrix
8: **end for**
9: Form $MVGC_{noise}$ matrix (with ($N_{ch+1} \times N_{ch+1}$) dimension), from N_{ch+1} rows of the previous step matrix
10: Subtract the values in the last column of $MVGC_{noise}$ matrix from the other columns and zero the negative values.
11: Remove the last row and column of $MVGC_{noise}$ matrix, to form $MVGC$ matrix of dimension $N_{ch} \times N_{ch}$

2.3.3. Proposed adaptive MVGC thresholding

According to Eq. (13) while considering $E[\eta_i^2] \geq E[e_i^2]$, $MVGC_{j\to i}$ will be a positive value. To detect the significance of $MVGC$ values we used an adaptive thresholding method. In this method, we add a normal random noise with zero mean and the same power of other channels, as a new channel to EEG data, named noise channel. More details are given in the algorithm 2 (the result of this algorithm will be the MVGC matrix with the dimension of ($N_{ch} \times N_{ch}$)). By subtracting the achieved $MVGC_{noise\to i}$ values, for $i = 1, 2, 3, \ldots, M$, from other relating calculated values, the effect of noise on the final values of $MVGC$ will be reduced. Therefore it will improve the validity of Granger's causality values. Values of $MVGC$ that are less than the $MVGC$ values of noise channel are set to zero. This means that if there is spurious causal connectivity between two channels due to noise, it will not be considered as Granger causality between them further.

2.3.4. Proposed channel selection method

The idea of this work is fundamentally based on the greater significance of some special EEG channels while measuring the causality of these channels onto others, during the MI action. These special channels may then be repeated throughout the experiment and during all trials. However, these channels may differ for the EEG signal of each subject (Arvaneh, Guan, Ang, & Quek, 2010). Of course, it is noteworthy that the perspective of this work is subject-specific too. This means that the outcoming selected channels are specific for each subject.

In our procedure, first of all, the preprocessed data for each subject is divided into train and test parts for each of the two classes. In a way that 70 percent of the whole data to be in the train part. Secondly in each MI event or trial, only two seconds of data are selected (from 0.5 s to 2.5 s after displaying target) (Kumar, Sharma, & Tsunoda, 2019). After that, these data are given to the channel selection step.

In Granger causality channel selection (GCCS) step, $MVGC_{j\to i}$ is calculated between any two arbitrary channels of i and j, independently for each subject. Then for each EEG data trial, the $MVGC$ matrix is calculated. The dimension of the $MVGC$ matrix is $N_{ch} \times N_{ch}$, which N_{ch} is the number of EEG recorded channels.

Algorithm 3 Proposed Granger Causality Channel Selection (GCCS)

1: Load EEG signals' data (train data) of dimension ($N_s \times N_{ch} \times N_{tr}$), where N_s: is the number of samples, N_{ch}: is the number of all channels, N_{tr}: is the number of trials.
2: **for** $i = 1 : N_{tr}$ **do**
3: Perform algorithm 2 to calculate $MVGC$ matrix (with ($N_{ch} \times N_{ch}$) dimension) for i^{th} trial of EEG channels' data
4: Calculate the mean of each column of $MVGC$ matrix
5: Select N_c channels with maximum mean values (in i^{th} trial)
6: **end for**
7: Find N_c (number of selected channels) channels appearing more number of times in all N_{tr} trials from the previous step

In each trial, we find channels that are the cause rather than the effect of other channels, using $MVGC$ matrices. During MI tasks, these obtained channels are the ones with significant causal connectivity with others. It is expected that the involved neural connections for each MI task, are possibly initiated from the previously obtained channels.

The ith row of calculated $MVGC$ matrices, represent GC values from channel i to other channels. The average of GC values at ith row of $MVGC$ matrices in each trial, is interpreted as the causality amount of ith channel to other channels. After sorting these mean GC values in descending order, the first N_c channels are selected in each trial and this process is repeated for all of the trials. In the end, we introduce N_c channels with higher repetition as selected channels in all trials. Discussed procedures are expressed in more detail in algorithm 3.

2.4. Data extraction from selected channels

According to the previous section, the preprocessed data for each subject is divided into two parts for each of 2 classes, 70 percent as train data and 30 percent as test data. After applying algorithm 3. on the train data, N_c channels are selected for each subject. These channels are used in all next coming steps, both for the test and the train data.

2.5. Bandpass filtering

In this part, the selected channels for each subject are filtered using a 3rd-order butterworth filter between 8 and 30 Hz.

2.6. Common spatial pattern (CSP)

For the first time, Common Spatial Pattern (CSP) for classifying MI EEG signals was proposed by Ramoser, Muller-Gerking, and Pfurtscheller (2000). By using a linear transform, CSP projects EEG multi-channel data into a spatial subspace with a lower dimension. This is done by a projection matrix which includes weights for EEG channels that maximize the differences in variance between two classes of MI EEG signals (Müller-Gerking, Pfurtscheller, & Flyvbjerg, 1999). Let $\mathbf{C}_L$ and $\mathbf{C}_R$ be the normalized covariance matrices of left and right hand MI of EEG signals that are averaged over trials. Then the composite covariance matrix will be $\mathbf{C}_c = \mathbf{C}_L + \mathbf{C}_R$. Suppose $\mathbf{C}_c = \mathbf{U}\boldsymbol{\lambda}\mathbf{U}^T$, where $\boldsymbol{\lambda}$ and $\mathbf{U}$ are eigenvalue and eigenvector matrices of $\mathbf{C}_c$ respectively. $\mathbf{U}$ is a ($N_c \times N_c$) matrix of normalized eigenvectors and $\boldsymbol{\lambda}$ is a diagonal ($N_c \times N_c$) matrix of eigenvalues. Then the whitening transformation will be:

$$\mathbf{P} = \boldsymbol{\lambda}^{-1/2}\mathbf{U}^T \tag{15}$$

that equalizes the variances in the eigenvectors space. Then $\mathbf{C}_L$ and $\mathbf{C}_R$ can be whitened by using $\mathbf{P}$, as follows:

$$\hat{\mathbf{C}}_L = \mathbf{P}\mathbf{C}_L\mathbf{P}^T \tag{16}$$

$$\hat{\mathbf{C}}_R = \mathbf{P}\mathbf{C}_R\mathbf{P}^T \tag{17}$$

So we have $\hat{\mathbf{C}}_L + \hat{\mathbf{C}}_R = \mathbf{I}$. It means that any orthonormal matrix, $\mathbf{V}$, satisfy:

$$\mathbf{V}^T(\hat{\mathbf{C}}_L + \hat{\mathbf{C}}_R)\mathbf{V} = \mathbf{I} \tag{18}$$

With this orthonormal matrix $\mathbf{V}$, we can decompose $\hat{\mathbf{C}}_L$ and $\hat{\mathbf{C}}_R$ as:

$$\hat{\mathbf{C}}_L = \mathbf{V}\mathbf{\Lambda}_L\mathbf{V}^T \tag{19}$$

$$\hat{\mathbf{C}}_R = \mathbf{V}\mathbf{\Lambda}_R\mathbf{V}^T \tag{20}$$

that $\mathbf{\Lambda}_L + \mathbf{\Lambda}_R = \mathbf{I}$. Finally, the projection matrix (CSP weight matrix) of whitened EEG signals of two classes on $\mathbf{V}$, is as follows:

$$\mathbf{W} = \mathbf{P}^T\mathbf{V} \tag{21}$$

which will result in optimal features in the sense of least squares (Müller-Gerking et al., 1999). $\mathbf{W}$ will be a $(N_c \times N_c)$ matrix. The vector of features is calculated as follows:

$$\mathbf{f} = log(var(\mathbf{W}\mathbf{x}(t))) \tag{22}$$

where $\mathbf{x}(t)$ is EEG data and is a $(N_c \times N_s)$ matrix that N_s is the number of EEG data observations.

2.7. Regularized common spatial pattern (RCSP)

CSP classification performance for the EEG signal may be low for a small number of training samples. To solve this problem, a new method has been introduced called regularized CSP (RCSP) (Friedman, 1989; Lu, Eng, Guan, Plataniotis, & Venetsanopoulos, 2010). Consider two L and R classes of MI EEG signals ($cl = \{L, R\}$). Then for each class, the regularized average spatial covariance matrices, with regularization parameters, δ and ϵ ($0 \leq \delta, \epsilon \leq 1$) are as follows:

$$\mathbf{\Phi}_{cl}(\delta, \epsilon) = (1-\epsilon)\mathbf{\Psi}_{cl}(\delta) + \frac{\epsilon}{N_{ch}} tr[\mathbf{\Psi}_{cl}(\delta)] \; . \; \mathbf{I} \tag{23}$$

where $tr[.]$ is the trace of a matrix, I is the identity matrix (with $(N_{ch} \times N_{ch})$ dimension) and $\mathbf{\Psi}_{cl}(\delta)$ is formed from the combination of covariance matrices of all subjects ($\delta \in \{0, 0.001, 0.01, 0.1, 0.2\}$ and $\epsilon \in \{0, 0.01, 0.1, 0.2, 0.4, 0.6\}$). Cichocki et al. in Jin et al. (2019), proposed a new RCSP method that uses only one subject's data, then the $\mathbf{\Psi}_{cl}(\delta)$ matrix will be as follows:

$$\mathbf{\Psi}_{cl}(\delta) = \frac{(1-\delta)\sum_{t=1}^{N_{tr}} \mathbf{C}_{(cl,t)} + \delta \sum_{t=1}^{N_{tr}} \hat{\mathbf{C}}_{(cl,t)}}{N_{tr}} \tag{24}$$

where $\mathbf{C}_{(cl,t)}$ and $\hat{\mathbf{C}}_{(cl,t)}$ are the normalized covariance matrix and pairwise covariance matrix of one subject's EEG data respectively. $\mathbf{C}_{(cl,t)}$ and $\hat{\mathbf{C}}_{(cl,t)}$ are for cl^{th} MI class and for tth trial, and can be obtained using:

$$\mathbf{C}_{(cl,t)} = \frac{\mathbf{S}_t\mathbf{S}_t^T}{tr[\mathbf{S}_t\mathbf{S}_t^T]} \tag{25}$$

$$\hat{\mathbf{C}}_{(cl,t)} = cov(\mathbf{S}_t^T) \tag{26}$$

which T denotes the transpose of a matrix, $\mathbf{S}_t$ is the tth EEG trial data for one subject and $cov(.)$ is the covariance matrix function that calculates pairwise covariance for each EEG channel data. Now in order to maximize the differences in variance between two classes of MI EEG signals, a spatial filter $\mathbf{\Omega}$ is identified in the Rayleigh form as:

$$\mathbf{\Omega} = \arg\min_{\Omega} \frac{\mathbf{\Omega}^T\mathbf{\Psi}_L\mathbf{\Omega}}{\mathbf{\Omega}^T\mathbf{\Psi}_R\mathbf{\Omega}} \tag{27}$$

The Rayleigh form problem can be transformed into the generalized eigenvalue problem, with the Lagrange multiplier method, as follows:

$$\mathbf{\Psi}_L\mathbf{\Omega} = \lambda\mathbf{\Psi}_R\mathbf{\Omega} \tag{28}$$

By solving this problem, for each pair of (δ, ϵ), same as Eq. (22), a feature matrix $\mathbf{G}$ will be obtained. Finally, Fisher's discriminant analysis (FDA) is used to maximize scatter between different classes in one hand, and on the other hand, it is used to minimize within-class scatter (Lu et al., 2010). FDA on matrix $\mathbf{G}$ results in:

$$\mathbf{z}_{FDA} = \arg\max_{\mathbf{z}} \frac{\mathbf{z}^T\mathbf{\Theta}_b\mathbf{z}}{\mathbf{z}^T\mathbf{\Theta}_w\mathbf{z}} \tag{29}$$

where $\mathbf{\Theta}_b$ and $\mathbf{\Theta}_w$ are between and within-class scatter matrices of $\mathbf{G}$, respectively. Using a generalized eigenvalue problem, $\mathbf{z}_{FDA}$ will be obtained and then FDA feature vector will be calculated as follows (for each pair of (δ, ϵ)):

$$\mathbf{F} = \mathbf{z}_{FDA}^T\mathbf{G} \tag{30}$$

2.8. Proposed framework

In the first part, MI EEG raw data for each subject (from all 105 subjects) are preprocessed. The preprocessed data for each subject is divided into two parts for each of 2 classes; 70 percent as train data and 30 percent as test data. The GCCS and CCS approaches for channel selection are applied on the train data and N_c channels are selected. Then this N_c channels' data are extracted from the train and test data. These extracted data from selected channels are filtered by a bandpass filter from 8 to 30 Hz. In the next step, two spatial domain filters, CSP, and RCSP are applied to filtered train data to obtain weight matrices for each method will be obtained. Then these weight matrices are applied to train and test data to create feature matrices for CSP and RCSP methods. Only in the RCSP method, after obtaining the feature matrix, the Fisher discriminant analysis (FDA) is applied to it to reduce the dimensions of the feature matrix. Finally using these feature matrices, classifiers are trained and classification performances with 10-fold cross-validation will be obtained. Fig. 3 shows a block diagram of our proposed framework.

2.9. Other datasets

In order to evaluate the proposed work in this article, in addition to the dataset of Section 2.1, we also used two widely used datasets in this field. It is noteworthy that in this article only the MI data of left and right hands are used.

2.9.1. BCI competition III-dataset IIIa

This dataset is Dataset-IIIa from BCI Competition III (Blankertz et al., 2006). This dataset is recorded over 60 channels of EEG signal from 3 subjects (k3, k6, l1). The data of this motor imagery EEG signal consists of 4 classes (left hand, right hand, foot, tongue) and each class has 90, 60 and 60 trials for subjects k3, k6 and l1 respectively. The EEG was sampled with 250 Hz and was filtered between 1 and 50 Hz. Subjects perform MI tasks according to a cue and timeline of MI paradigm is shown in Fig. 4(a).

2.9.2. BCI competition IV-dataset 1

This dataset is Dataset-1 from BCI Competition IV (Tangermann et al., 2012). This dataset is recorded over 59 channels of EEG signal from 7 subjects. We only use the data of 2 subjects (b and g), because the data of 3 subjects are artificially generated and the 2 other subjects have not left and right hand class data. The data of this motor imagery EEG signal consists of 3 classes (left hand, right hand, foot). The EEG was downsampled to 100 Hz. Subjects perform MI tasks according to a cue and timeline of MI paradigm is shown in Fig. 4(b).

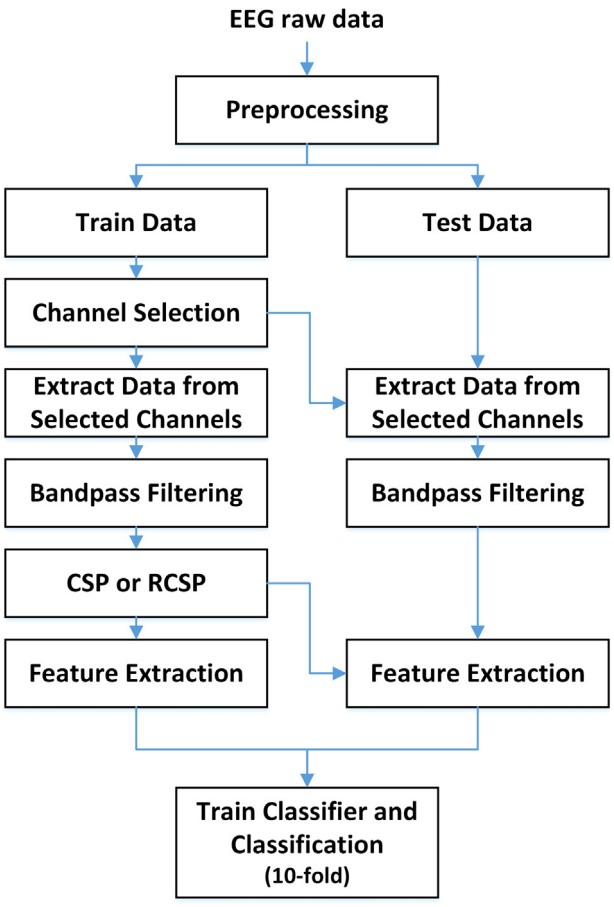

Fig. 3. Block diagram of our proposed framework.

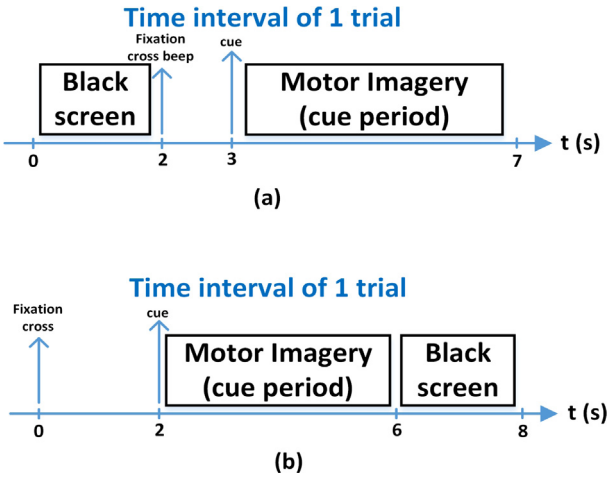

Fig. 4. Timeline of MI paradigm (for one trial).

3. Results

In this section, the results of the described methods in Section 2 for Physionet dataset are reported. The results of other 2 datasets are reported in Section 3.3.6.

3.1. Feature extraction

In this study, we used a 64-channel MI EEG recordings data from 105 subjects which is available in Physionet (Goldberger et al., 2000). For each subject's MI data, 42 trials are considered (21 trials for each left and right hand MI data). By using two channel selection approaches, CCS and GCCS, N_c channels of EEG data are selected from train data for each subject. The data of these N_c channels are selected and filtered (with a 3-rd order butterworth bandpass filter, from 8 to 30 Hz) for the next steps, for both the test and the train data.

3.1.1. CSP features

After applying the CSP method on filtered train data, CSP weight matrices for each subject are obtained. Then feature vector of each subject's trial (for train and test data) is extracted from

$$\mathbf{f}_{CSP} = log(var(\mathbf{W}_{CSP}\mathbf{x}_c(t))) \tag{31}$$

where $\mathbf{W}_{CSP}$ (CSP weight matrix) is a $(N_c \times N_c)$ matrix, $\mathbf{x}_c(t)$ is one trial EEG data of each subject and is a $(N_c \times N_s)$ matrix (N_s is the number of EEG data samples in 2-seconds of MI). $\mathbf{f}_{CSP}$ is a feature vector with $(N_c \times 1)$ dimension. Finally by using all N_{tr} trials, feature matrix is achieved with dimension of $(N_c \times N_{tr})$ for each subject.

3.1.2. RCSP features

At this stage, similar to the previous step, after applying the RCSP method to train part's filtered data, RCSP weight matrices for each subject are obtained. Then feature vector of each subject's trial (for train and test data) is extracted from

$$\mathbf{f}_{RCSP} = log(var(\mathbf{W}_{RCSP}\mathbf{x}_c(t))) \tag{32}$$

where $\mathbf{W}_{RCSP}$ (RCSP weight matrix) is a $(N_c \times N_c)$ matrix and $\mathbf{f}_{RCSP}$ is the feature vector, with $(N_c \times 1)$ dimension. Regularization parameters, δ and ϵ are selected from $\delta \in \{0, 0.001, 0.01\}$ and $\epsilon \in \{0, 0.01, 0.1\}$ which result in nine (δ, ϵ) pairs (N_p, defined as the number of (δ, ϵ) pairs). To prevent overfitting, N_p value is limited to 9. This feature vector is calculated for each pair of (δ, ϵ) and for all trials, which result in nine RCSP feature matrices all with the dimension of $(N_c \times N_{tr})$. In the following for each pair of regularization parameters, after applying FDA, as a dimensionality reduction technique, a feature vector of $(N_{tr} \times 1)$ dimension will be obtained. Finally using all N_p pairs of regularization parameters we have a $(N_p \times N_{tr})$ feature matrix for each subject. It should be mentioned that in order to reduce the dimension of RCSP feature matrices, after each RCSP an FDA is applied on the results too. Otherwise, it is noted.

3.2. Classifiers

Classification of EEG-based BCI tasks is performed commonly with multiple classifiers such as linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), kernel Fisher discriminant (KFD), support vector machine (SVM), multilayer perceptron (MLP), learning vector quantization (LVQ) neural network, k-nearest neighbor (k-NN), and decision tree (DT) (Wang et al., 2009). The mostly used classifiers for MI based on BCI are k-NN, SVM, and LDA (Aldea, Fira, & Lazăr, 2014; Chen, Wu, Lin, & Chen, 2014; Kayikcioglu & Aydemir, 2010; Ma et al., 2016; Tyagi & Nehra, 2016). In this study, we used LDA, SVM, and k-NN classifiers to classify MI tasks into the left and right hand. As mentioned earlier, 70 percent of data (train data) are selected for extracting features using CSP and RCSP method weights and then train and test data are used in the classification step with 10-fold cross-validation.

3.3. Classification performance

As mentioned earlier, we used two algorithms to remove EEG signal artifacts, two approaches for channel selection, and two methods for feature extraction. In the next subsections, the obtained results in different conditions are presented.

Table 1
Comparison of classification performances of two artifact removal methods, MAXAR (Section 2.2.4) and KMAR (algorithm 1), by LDA classifier and feature extraction by CSP method, with 5 random subjects. Accuracy, Sensitivity, and Specificity are in %.

Subject	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
	GCCS with MAXAR, Nc = 10			GCCS with KMAR, Nc = 10		
1	75.56	78.26	72.73	100.00	100.00	100.00
9	55.56	58.33	52.38	73.33	75.00	71.43
13	80.00	78.26	81.82	100.00	100.00	100.00
23	60.00	59.09	60.87	82.22	81.82	82.61
37	60.00	65.22	54.55	68.89	78.26	59.09
mean (all 105 subjects)	71.64	71.40	71.70	**89.41**	**89.17**	**89.63**

3.3.1. Artifact removal methods

As previously noted, the desired EEG signals and the artifact sources are independent and the most commonly used method in BSS is ICA. As a result, the ICA (removing eyeblink and other similar artifacts) will be used to correct artifacts in this work with two methods, MAXAR and KMAR. After applying ICA on EEG data of each trial, in the first method, MAXAR, two ICA components with the highest powers are removed. In the second method, KMAR, the ICA components are clustered into two artifact and normal EEG clusters. By using k-means machine learning ICA components in artifact cluster are removed. Finally, EEG data is reconstructed from the remained components. Classification performance of these two artifact removal methods reported in Table 1. In the following, due to better performance, the KMAR method is used to remove artifacts. Otherwise, it is noted.

3.3.2. Different classifiers

Supervised classification of MI tasks is performed with LDA, SVM, and k-NN classifiers, with multiple kernels. So MI tasks of each subject are classified into two left and right hand MI classes, with 8 classifiers:

1- C k-NN: k-NN classifier with cosine distance and 10 number of neighbors with equal distance weight.

2- fine k-NN: k-NN classifier with euclidean distance, one neighbor, and equal distance weight.

3- W k-NN: k-NN classifier with euclidean distance and 10 number of neighbors with squared inverse distance weight.

4- LDA: Linear discriminant analysis.

5- P3 SVM: SVM classifier with polynomial kernel function of order 3.

6- L SVM: Linear SVM classifier.

7- P2 SVM: SVM classifier with polynomial kernel function of order 2.

8- RBF SVM: SVM classifier with RBF kernel function.

The process is such that all classifiers are used for each subject, and in the end, the classifier with the highest accuracy is selected as the final and the best classifier for that subject. So this process is performed by considering, GCCS and CCS as channel selection approaches and CSP and RCSP methods for feature extraction. The selected classifiers distribution diagrams for all subjects, with CSP and RCSP methods, are shown in Figs. 5 and 6, respectively. Classification performances of GCCS and CCS with CSP and RCSP methods, in terms of different classifiers, are reported in Tables 2 and 3 respectively.

3.3.3. Channel selection approaches

In this study, two approaches are used to select some EEG channels that have much more effect on the process of MI in the brain, functional and effective connectivity. The functional connectivity approach uses statistical measures (correlation coefficients between EEG channels' data), while effective connectivity benefits causality measures (Granger causality values between EEG channels' data).

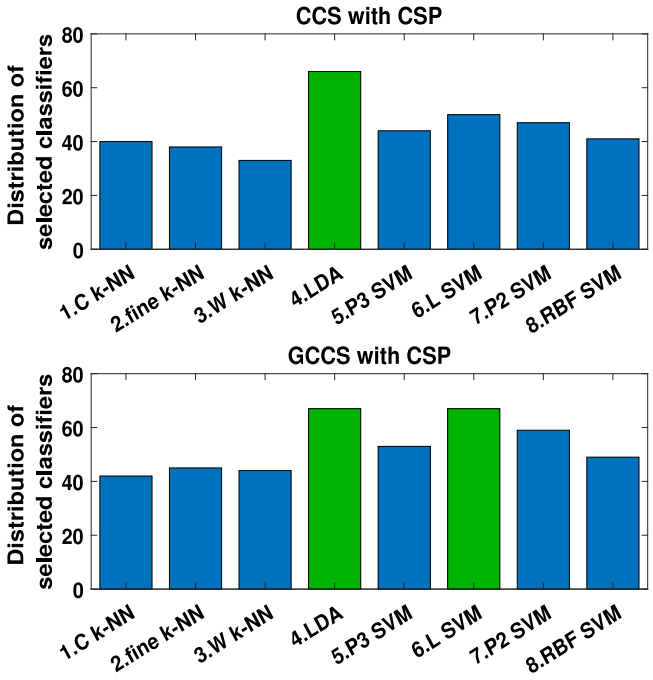

Fig. 5. Distribution of selected classifiers for all subjects for GCCS (algorithm 3) and CCS (Jin et al., 2019), with $N_c = 8$ and CSP method for feature extraction. The vertical axis indicates the number of times each classifier has been selected as the best classifier for all 105 subjects, in terms of classification accuracy. The green color bars show classifiers with maximum abundance as the best classifier.

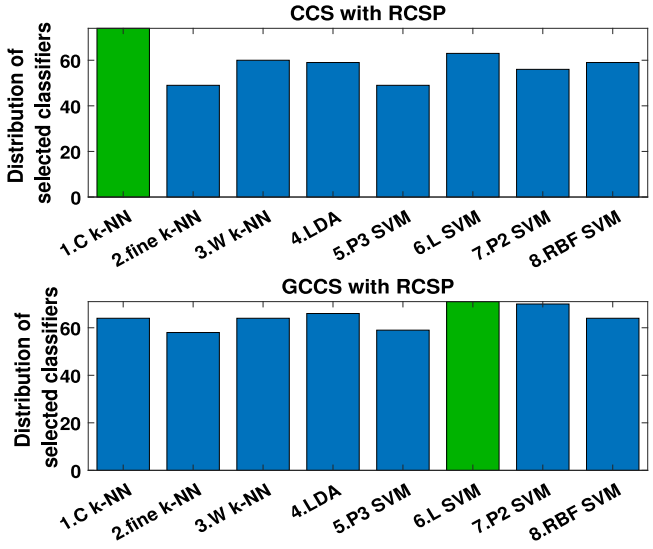

Fig. 6. Distribution of selected classifiers for all subjects for GCCS (algorithm 3) and CCS (Jin et al., 2019), with $N_c = 8$ and RCSP method for feature extraction. The vertical axis indicates the number of times each classifier has been selected as the best classifier for all 105 subjects, in terms of classification accuracy. The green color bars show classifiers with maximum abundance as the best classifier.

The correlation-based channel selection (CCS) algorithm is adopted from Jin et al. (2019). This algorithm is similar to algorithm 3, except that the correlation coefficients must be calculated instead of the GC values.

GCCS is proposed in this work with the novel criterion for channel selection. In GCCS channels with higher causalities are selected. In other words, the channel is ultimately chosen, which

Table 2
Comparison of classification performances in terms of different classifiers, for GCCS (algorithm 3) and CCS (Jin et al., 2019), with $N_c = 8$, CSP for extracting features. Accuracy, Sensitivity and Specificity are in %. Performance values are mean values of all 105 subjects' performance results.

Classifiers	CCS			GCCS		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
1-C k-NN	78.25	77.75	78.76	**84.34**	**84.65**	**83.93**
2-fine k-NN	77.04	77.15	76.85	**83.80**	**84.34**	**83.16**
3-W k-NN	78.47	79.18	77.64	**84.60**	**85.71**	**83.45**
4-LDA	81.64	81.95	81.30	**87.06**	**87.32**	**86.76**
5-P3 SVM	78.90	78.47	79.25	**85.46**	**86.44**	**84.46**
6-L SVM	80.58	81.45	79.65	**87.25**	**87.66**	**86.78**
7-P2 SVM	79.65	80.55	78.69	**86.71**	**87.07**	**86.31**
8-RBF SVM	78.36	79.51	76.94	**85.37**	**85.67**	**84.88**

Table 3
Comparison of classification performances in terms of different classifiers, for GCCS (algorithm 3) and CCS (Jin et al., 2019), with $N_c = 8$, RCSP for extracting features. Accuracy, Sensitivity and Specificity are in %. Performance values are mean values of all 105 subjects' performance results.

Classifiers	CCS			GCCS		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
1-C k-NN	87.27	87.66	86.90	**90.96**	**91.23**	**90.69**
2-fine k-NN	84.60	84.40	84.76	**89.30**	**89.13**	**89.46**
3-W k-NN	86.24	86.78	85.63	**90.72**	**90.63**	**90.77**
4-LDA	86.41	86.90	85.84	**91.05**	**91.71**	**90.40**
5-P3 SVM	84.25	85.28	83.14	**89.92**	**91.16**	**88.61**
6-L SVM	87.10	87.64	86.52	**91.22**	**91.66**	**90.75**
7-P2 SVM	85.93	86.52	85.31	**90.85**	**91.68**	**89.98**
8-RBF SVM	85.50	86.57	84.28	**90.23**	**90.77**	**89.62**

Table 4
Classification performance comparison of GCCS (algorithm 3) and CCS (Jin et al., 2019), with CSP for extracting features and by using the best classifier. N_c is the number of selected channels. Accuracy, Sensitivity, and Specificity are in %. Performance values are mean values of all 105 subjects' performance results.

Nc	CCS with CSP			GCCS with CSP		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
3	82.91	83.10	82.67	**87.30**	**88.05**	**86.46**
5	83.52	84.75	82.17	**88.31**	**89.06**	**87.55**
10	85.80	85.90	85.65	**89.41**	**89.17**	**89.63**
15	86.26	86.37	86.07	**89.69**	**89.83**	**89.55**
20	87.44	88.22	86.61	**90.59**	**90.50**	**90.70**

Table 5
Classification performance comparison of GCCS (algorithm 3) and CCS (Jin et al., 2019), with RCSP for extracting features and by using the best classifier. N_c is the number of selected channels. Accuracy, Sensitivity, and Specificity are in %. Performance values are mean values of all 105 subjects' performance results.

Nc	CCS with RCSP			GCCS with RCSP		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
3	84.31	84.91	83.64	**87.75**	**87.51**	**87.88**
5	86.56	86.89	86.17	**90.05**	**90.55**	**89.52**
10	91.56	91.54	91.61	**94.50**	**95.03**	**93.96**
15	94.91	94.77	95.04	**96.63**	**96.86**	**96.39**
20	96.53	96.71	96.34	**98.27**	**98.20**	**98.35**

is more possible to be the cause of other channels. Classification performances of GCCS and CCS with different classifiers for CSP and RCSP methods are shown in Tables 4 and 5 respectively. Classification is a subject-specific process that means all classification and feature extraction steps should be performed for each subject independently.

3.3.4. Classification accuracy vs. the number of selected channels

Comparison of GCCS and CCS classification accuracies, with CSP and RCSP methods, for all 105 subjects according to the number of selected channels (from 3 to 20) is shown in Fig. 7. Also comparison of GCCS and CCS classification accuracies, with RCSP method, for all 105 subjects according to the number of selected channels, from 3 to 64 (with no channel selection) is shown in Fig. 8.

Table 6
Classification performances of BCI competition III-Dataset IIIa (3 subjects) and BCI competition IV-Dataset 1 (2 subjects), by channel selection approaches GCCS (algorithm 3) and CCS (Jin et al., 2019), with RCSP for extracting features and by using the linear SVM classifier. N_c is the number of selected channels. Accuracy (Acc.), Sensitivity (Sen.), and Specificity (Spec.) are in %.

Dataset	Sub	CCS with RCSP, Nc = 13			GCCS with RCSP, Nc = 13		
		Acc.	Sen.	Spec.	Acc.	Sen.	Spec.
BCI III	k3	100.00	100.00	100.00	**100.00**	**100.00**	**100.00**
Dataset	k6	86.67	86.67	86.67	**88.33**	**86.67**	**90.00**
IIIa	l1	**94.33**	**92.33**	**94.67**	90.33	88.67	92.00
BCI IV	b	100.00	100.00	100.00	**100.00**	**100.00**	**100.00**
Dataset 1	g	100.00	100.00	100.00	**100.00**	**100.00**	**100.00**

3.3.5. Map of selected channels

Channel selection approaches GCCS and CCS, select the most appropriate channels with different perspectives. In the next steps, the data of these channels are used to extract the features with CSP and RCSP methods. Since this study is subject-specific, a set of channels is selected for each subject. The most frequently selected channels for all subjects are shown in Fig. 9.

3.3.6. Results of other datasets

The block diagram of the work done on these datasets is as shown in Fig. 3, except that no filter is applied and no artifact is removed during the preprocessing step and the time window of the selected signal is from 3 to 6 s and 2.5 to 4.5 s for BCI competition III-Dataset IIIa and BCI competition IV-Dataset 1 respectively. Classification performances of BCI competition III-Dataset IIIa (3 subjects) and BCI competition IV-Dataset 1 (2 subjects) datasets are reported in Table 6.

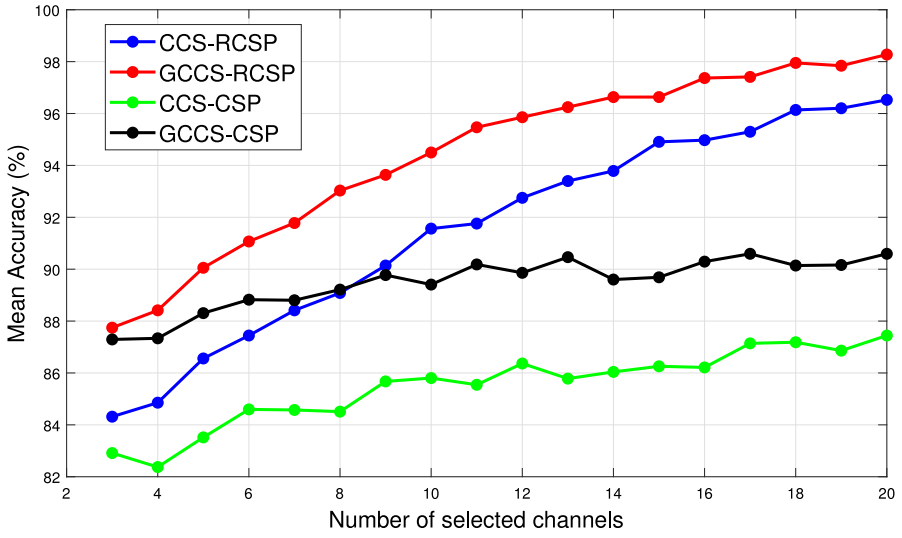

Fig. 7. Classification accuracy vs. the number of selected channels for GCCS (algorithm 3) and CCS (Jin et al., 2019), with CSP and RCSP methods for extracting features and by using the best classifier. Accuracies are mean values from all 105 subjects' accuracies.

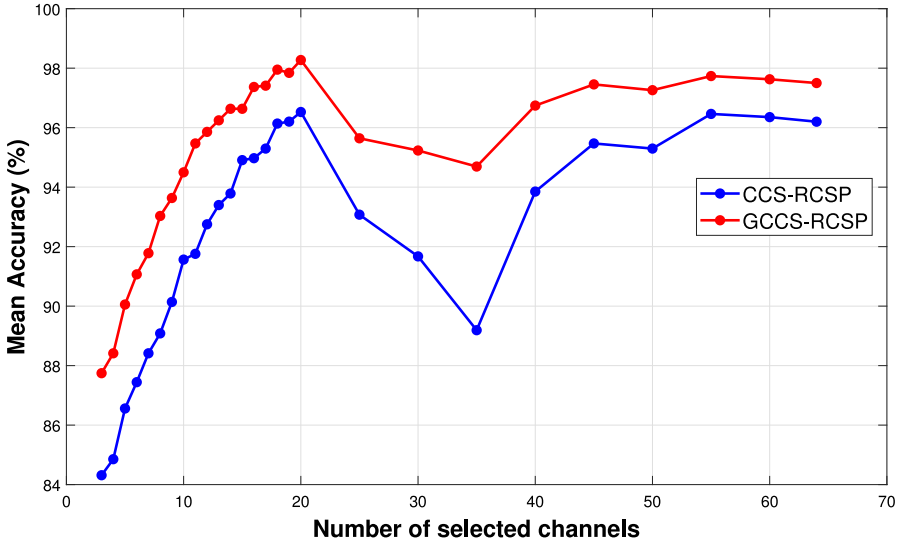

Fig. 8. Classification accuracy vs the number of selected channels for GCCS (algorithm 3) and CCS (Jin et al., 2019), with RCSP method for extracting features and by using the best classifier. Accuracies are mean values from all 105 subjects' accuracies.

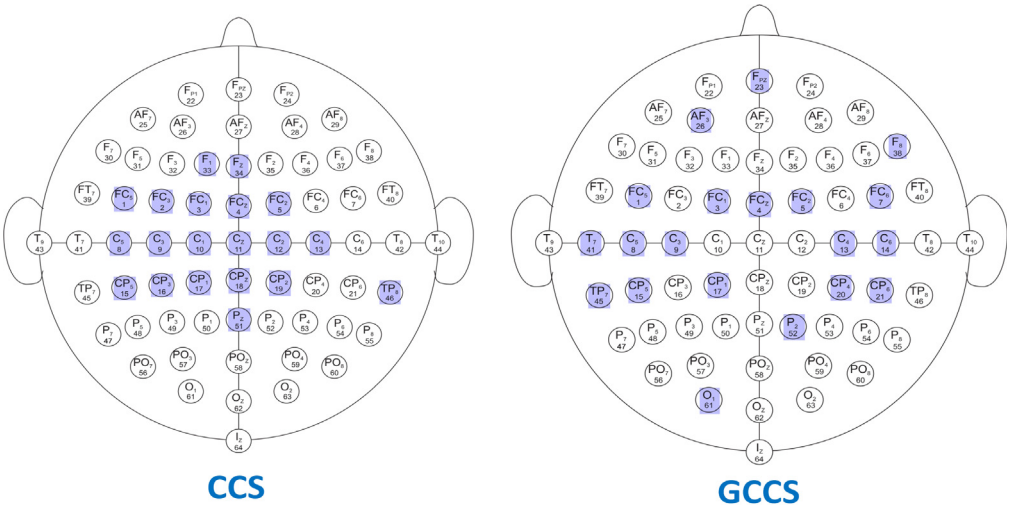

Fig. 9. Map of the 20 frequently selected channels among all 105 subjects for two GCCS and CCS approaches, (in the EEG international 10–10 system).

4. Discussion

4.1. Different classifiers

Results of CSP and RCSP methods are discussed in more detail in the following subsections.

4.1.1. CSP method

According to Table 2, in the CSP method, LDA and linear SVM (6-L SVM) classifiers propose better performances for two CCS and GCCS approaches in terms of classification accuracy respectively. As shown in Fig. 5, LDA classifier has been selected as the best classifier for each subject, in both approaches. Of course, by checking all the values of N_c from 3 to 20, two linear SVM (6-L SVM) and LDA classifiers performed better than other classifiers.

4.1.2. RCSP method

In the RCSP method, according to the results of Table 3, for CCS and GCCS, the best classifiers are k-NN (1-C k-NN) and linear SVM (6-L SVM) in terms of classification accuracy, respectively. As shown in Fig. 6, k-NN (1-C k-NN) and linear SVM (6-L SVM) classifiers are nominated more than others as the best classifiers for all subjects, for CCS and GCCS approaches respectively. Similar to the CSP method, linear SVM (6-L SVM) and LDA classifiers show better performances for N_c from 3 to 20.

Finally, it is inferred from Tables 2 and 3, that the results of GCCS are better than CCS in terms of classification performance. Additionally, Figs. 5 and 6 confirm that due to different channel selection for each subject which results in different feature spaces, it is more expected that various classifiers to be selected as the best classifier, for each subject.

4.2. Channel selection approaches

4.2.1. Classification performance comparison

Classification performances of GCCS and CCS, with CSP and RCSP methods, are shown in Tables 4 and 5 respectively. It is shown that the accuracy of GCCS is 4.39, 4.79, 3.61, 3.43, and 3.15 percent better than CCS by using CSP for N_c values 3, 5, 10, 15, and 20. Additionally by using RCSP, results of GCCS are 3.44, 3.49, 2.94, 1.72, and 1.74 percent better than CCS for N_c values 3, 5, 10, 15, and 20. So it is inferred that by increasing the value of N_c, increasing the rate of classification accuracy decreases. Obviously GCCS, with an effective connectivity perspective, performs better than CCS, with a functional connectivity perspective, in terms of classification performance (accuracy, sensitivity, and specificity).

The improvement in the accuracy of classification from CSP to RCSP method is 1.4, 3.04, 5.76, 8.65, and 9.09 percent and 0.45, 1.74, 5.09, 6.94, and 7.68 percent, for N_c values 3, 5, 10, 15, and 20, in CCS and GCCS approaches, respectively. In this case, by increasing the value of N_c, the rate of increment in classification accuracy increases too. Results show that the RCSP method outperforms the CSP method.

4.2.2. Computational cost comparison

To compare the computational cost of these two approaches, the results of a sample (subject S001) have been used. The channel selection computational time for the subject S001 in 2 MI trials, as an average (running the channel selection algorithm several times), for the CCS and GCCS approaches was 0.012 and 1.984 s, respectively. As can be seen, CCS approach for channel selection is faster than GCCS approach (proposed work). In BCI and neurofeedback experiments, in the initial session or sessions, called training sessions (Enriquez-Geppert, Huster, & Herrmann, 2017), the considered parameters are adjusted offline according to the subject's recorded signals. So the processing speed for the channel selection stage is not a concern. To perform these calculations, MATLAB software (R2018b) and hardware with the following specifications were used: CPU Core i7-7700 @3.6 GHz with 16 GB RAM.

Table 7
Comparison of 4 methods ('channel selection' with 'feature extraction') by using the best classifier for each subject, in terms of classification accuracy and significance. N_c is the number of selected channels. Accuracies are reported in the form of mean±std%, from all 105 subjects' performance results.

Methods	CCS with CSP	CCS with RCSP	GCCS with CSP	GCCS with RCSP
Nc = 3	82.91 ± 15.84	84.31 ± 14.77	87.3 ± 13.05	87.75 ± 12.70
P-value	p < 0.001	p < 0.001	p < 0.01	–
Nc = 5	83.52 ± 15.58	86.56 ± 13.68	88.31 ± 12.53	90.05 ± 11.05
P-value	p < 0.001	p < 0.001	p < 0.001	–
Nc = 8	84.51 ± 15.20	89.08 ± 11.49	89.21 ± 12.27	93.03 ± 8.45
P-value	p < 0.001	p < 0.001	p < 0.001	–
Nc = 10	85.8 ± 14.47	91.56 ± 9.45	89.41 ± 11.74	94.5 ± 6.98
P-value	p < 0.001	p < 0.001	p < 0.001	–
Nc = 15	86.26 ± 14.42	94.91 ± 6.55	89.69 ± 11.44	96.63 ± 4.95
P-value	p < 0.001	p < 0.001	p < 0.001	–
Nc = 20	87.44 ± 13.38	96.53 ± 4.54	90.59 ± 11.42	98.27 ± 2.82
P-value	p < 0.001	p < 0.001	p < 0.001	–

4.2.3. Classification accuracy and significance

As the final conclusion, the GCCS approach for channel selection with the RCSP method for feature extraction performs better than other methods in terms of classification accuracy. The mean value and standard deviation of classification accuracies, for previously mentioned methods, along with p-values (with Wilcoxon signed-rank test), are shown in Table 7. It shows that the standard deviation of accuracies decreases by increasing N_c values. Also, the standard deviation values of classification accuracy in the GCCS approach are less than values in the CCS approach. GCCS with the RCSP method, propose significantly lower values of the standard deviation of accuracies than other methods.

4.3. Classification accuracy vs. the number of selected channels

According to Fig. 7, it is observed that by increasing the number of selected channels (N_c), the classification accuracy increases. Of course, this is not unexpected, because increasing the number of selected channels increases the data of the problem. Also, in all values of N_c, from 3 to 20, the performance of the GCCS approach is better than the CCS approach, while using the same feature extraction method.

It is noteworthy that because of overfitting occurrence for high values of N_c in the CSP method, in this study N_c is limited to 20. Overfitting occurs as in the CSP method, the number of features for each trial equals N_c, while in the RCSP method it equals N_p. In this study $N_p = 9$ (refer to Section 3.1.2).

According to Fig. 7, by increasing the values of N_c, the increasing rate of the RCSP method in classification accuracy is higher than the CSP method.

Comparing the accuracy of classification in the GCCS approach, with the RCSP method, for N_c values 8 and 20, the difference is around 4%. Due to the compromise between the accuracy of the classification and the shorter EEG setup time and the comfort of the subject, it is possible to ignore the 4% increment in accuracy. So the suggested value for N_c is 8.

According to Fig. 8, as the values of N_c increases, the overall trend of classification accuracy is first increased and then decreased. This may be due to the following reasons: at low values of N_c, the number of selected channels is low or insufficient;

Table 8
Comparison of classification performances of proposed work and other works, for classification of MI left and right hand tasks, on Physionet data (Goldberger et al., 2000) (105 subjects included). N_c is the number of selected channels. The type of training depends on whether the classification was performed on all subjects or each subject. Accuracies (Acc) are in %.

Reference	Author	Nc	Max Acc	Methods	Training
Park et al. (2013)	Park et al. (2013)	58	72.37	CCSP SVM	Global
Tolić and Jović (2013)	Tolic Jovic (2013)	3	68.21	Wavelet transform Feed-forward NN	Subject specific
Loboda et al. (2014)	Loboda et al. (2014)	9	71.55	Phase information	Global
Handiru and Prasad (2016)	Handiru Prasad (2016)	16	63.62	CSP SVM	Global
Kim et al. (2016)	Kim et al. (2016)	14	80.05	CCSP Random forest	Subject specific
Dose et al. (2018)	Dose et.al. (2018)	14	82.66	Convolutional neural network	Subject specific
		9	81.82		
		3	79.2		
		14	76.66		Global
		9	75.85		
		3	73.20		
Karácsony et al. (2019)	Karacsony et al. (2019)	64	80.45	Convolutional neural network	Global
		16	72.81		
Proposed work		3	**87.75**	GCCS with RCSP best classifier	Subject specific
		5	**90.10**		
		10	**94.50**		
		15	**96.63**		
		20	**98.27**		
		3	**84.38**	GCCS with RCSP linear SVM	
		5	**87.77**		
		10	**92.99**		
		15	**95.66**		
		20	**97.52**		
		3	**84.92**	GCCS with CSP linear SVM	
		5	**85.83**		
		10	**87.08**		
		15	**87.21**		
		20	**88.46**		
		9	**81.26**	GCCS with RCSP KNN	Global
		14	**83.63**		

At high values of the N_c, increasing the number of redundant channels reduces the values of classification accuracy. It is also observed that for both channel selection approaches the highest accuracy values are obtained for $N_c = 20$.

4.4. Artifact removal methods

The used data in this study (Goldberger et al., 2000) is completely raw and without any preprocessing. So preprocessing is one of the performed tasks. Two methods MAXAR (Section 2.2.4) and KMAR (Section 2.2.5) are used in the artifact removal step. According to Table 1, much better performances of KMAR method are achieved (about 19% improvement in mean classification accuracy). It seems that one of the important reasons for this improvement in performance is finding the ICA components related to artifact in the KMAR method. In the MAXAR method, only the ICA components with the highest values of variance are removed, which may have been related to the normal EEG signal in one hand and on the other hand, too many artifact components may have remained.

4.5. Map of selected channels

There is an important point in Fig. 9 that selected channels from both GCCS and CCS approaches are approximately located in the motor areas of the brain (primary motor cortex and premotor area). Only in the CCS approach, the channels are concentrated in one area, while in the GCCS approach there is some scattering in the location of the channels.

In the CCS approach, due to volume conduction effects in the EEG signal, the correlation value of the close channels may be high, and this is a reason for the proximity of the selected channels to each other. But in the GCCS approach, due to lower effects of volume conduction on Granger causality (Cohen, 2014) and also the directional connectivity between two EEG signal channels, further channels may be selected. In this study, these further selected channels are supposed to be effective in MI tasks too.

4.6. Comparison with other studies

The convenience of the subject in BCI and neurofeedback applications is one of the most important issues. Besides resolving this issue, EEG channel selection will improve the BCI and neurofeedback performance, by removing unrelated or noisy channels (Arvaneh et al., 2011). Our main goal in this research is to reduce the number of recorded EEG signal channels in order to find effective channels in classifying MI tasks (left and right hand MI). It is achieved with a novel causality perspective by considering effective connectivity.

In recent years several methods are proposed for EEG channel selection with different perspectives. Studies on the Physionet dataset (Goldberger et al., 2000), in this section, have been compared with our proposed work, in terms of classification performance and the number of selected channels. Table 8 shows the comparison results of our proposed work with other researches. It shows that proposed methods, with GCCS approach for channel selection and CSP and RCSP methods for feature extraction, outperform the other methods, in subject-specific training conditions.

In order to fully compare the results with the results of other studies, the global training condition was implemented on the data, which the results can be seen in Table 8. The proposed block diagram for this condition is shown in Fig. 3, but it should be

noted that in this condition the data of all subjects are given to the input at once. In the data division, the data of 80% of the subjects were considered as training data and the rest as test data. According to Table 8 in the global training condition for $N_c = 9$, the maximum classification accuracies of Loboda et al. (2014) and Dose et al. (2018) and the proposed work are 71.55%, 75.85% and 81.26% respectively. Also the maximum classification accuracies of Handiru Prasad (Handiru & Prasad, 2016) (for $N_c = 16$), Dose et al. (2018) (for $N_c = 14$), Karácsony et al. (2019) (for $N_c = 16$) and the proposed work (for $N_c = 14$) are 63.62%, 76.66%, 72.81% and 83.63% respectively. These results, in the global training condition, show the superiority of the proposed channel selection method over other works in terms of classification accuracy.

According to Section 2.3.4, the effect of causality can be observed in the proposed method of channel selection. Because in this case, we observe the process of the MI event from casual point of view, and our assumption is that the channels that have had the most impact on this process, show themselves with a higher causality. In other words, it can be assumed that the channel that was the cause of most of the other channels during MI, had the greatest impact on the occurrence of this event and therefore is most likely selected as the final and effective channel. Finally, it can be said that because the effective channels have been selected with a more appropriate and relevant criterion, better results have been obtained in the classification.

4.7. Other datasets

Classification performances of two BCI competition datasets were shown in Table 6. As can be seen, the proposed approach for channel selection has worked very well in these two datasets.

4.8. Future works

In this study, a channel selection approach, based on effective connectivity, is proposed to select some EEG channels that have much more influence on the process of MI in the brain. Cichocki et al. in Jin et al. (2019), presented a channel selection approach, based on functional connectivity which uses statistical measures. Whereas our proposed method benefits from causality measures.

GCCS method for channel selection in MI EEG signal may be used in later BCI and neurofeedback applications. It can be a good starting point for finding causal interactions and the sequence of activity of brain neurons, during motor imagery tasks.

Johnson et al. showed that subject's performance in motor imagery does not depend on the subject's remaining motor ability (Johnson, 2000; Johnson, Sprehn, & Saykin, 2002). In another study, Sharma, Baron, and Rowe (2009) showed that active motor areas of the brain during motor imagery and motor execution are largely similar. MI neurofeedback experiments can be designed for people with motor impairments, such as stroke patients to restore their motor functions. Because of more knowledge from motor imagery activities in the brain, and by understanding their neural interactions besides the sequence of activity of brain neurons which can be achieved from Granger causality perspective. In this regard, the time interval of MI tasks in the EEG signal can be divided into smaller time windows instead of 2 s after displaying target. After that, the changes in the Granger causality values during MI activities in the brain can be examined.

The calculations of Granger causality in this study are all in the time domain, which can also be in frequency domain too in future works. Also, the considered frequency range can be in the form of mu (8–13 Hz) and beta (15–30 Hz) rhythm modulations in motor imagery, instead of the total range of 8 to 30 Hz.

5. Conclusion

The main idea of this work is based on the assumption of causal interactions between EEG channels in motor imagery (MI) tasks. These causal interactions are stronger in some channels in comparison with others. According to this, a novel method for motor imagery EEG signal channel selection is proposed based on Granger causality analysis, called GCCS, for brain–computer interface (BCI) and neurofeedback (NF) applications. In the preprocessing part, EEG signal artifacts are removed using machine learning with new features. Also, a new adaptive multivariate Granger causality thresholding algorithm is presented. Features are extracted using common spatial pattern (CSP) and regularized CSP (RCSP). Finally, with different classifiers, classification performances for MI tasks, left and right hand, are obtained. The GCCS proposes a method for channel selection, showing better performance than the correlation-based channel selection method. Furthermore, in the feature extraction criterion, RCSP performed much better than CSP, in terms of classification performances.

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.

References

Aggarwal, S., & Chugh, N. (2019). Signal processing techniques for motor imagery brain computer interface: A review. *Array*, *1*, Article 100003.

Aldea, R., Fira, M., & Lazăr, A. (2014). Classifications of motor imagery tasks using k-nearest neighbors. In *12th symposium on neural network applications in electrical engineering (NEUREL)* (pp. 115–120). IEEE.

Alotaiby, T., El-Samie, F. E. A., Alshebeili, S. A., & Ahmad, I. (2015). A review of channel selection algorithms for eeg signal processing. *EURASIP Journal on Advances in Signal Processing*, *2015*(1), 66.

Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2012). Mutual information-based selection of optimal spatial–temporal patterns for single-trial eeg-based bcis. *Pattern Recognition*, *45*(6), 2137–2144.

Arvaneh, M., Guan, C., Ang, K. K., & Quek, H. C. (2010). EEG channel selection using decision tree in brain-computer interface. In *Proceedings of the second APSIPA annual summit and conference* (pp. 225–230).

Arvaneh, M., Guan, C., Ang, K. K., & Quek, C. (2011). Optimizing the channel selection and classification accuracy in eeg-based bci. *IEEE Transactions on Biomedical Engineering*, *58*(6), 1865–1873.

Blankertz, B., Muller, K.-R., Krusienski, D. J., Schalk, G., Wolpaw, J. R., Schlogl, A., et al. (2006). The bci competition iii: Validating alternative approaches to actual bci problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *14*(2), 153–159.

Brockwell, P. J., Davis, R. A., & Fienberg, S. E. (1991). *Time series: theory and methods: theory and methods*. Springer Science & Business Media.

Cadotte, A. J., DeMarse, T. B., He, P., & Ding, M. (2008). Causal measures of structure and plasticity in simulated and living neural networks. *PloS One*, *3*(10).

Chen, C.-Y., Wu, C.-W., Lin, C.-T., & Chen, S.-A. (2014). A novel classification method for motor imagery based on brain-computer interface. In *2014 international joint conference on neural networks (IJCNN)* (pp. 4099–4102). IEEE.

Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT press.

Donoghue, J. P. (2002). Connecting cortex to machines: recent advances in brain interfaces. *Nature Neuroscience*, *5*(11), 1085–1088.

Dose, H., Møller, J. S., Iversen, H. K., & Puthusserypady, S. (2018). An end-to-end deep learning approach to mi-eeg signal classification for bcis. *Expert Systems with Applications*, *114*, 532–542.

Enriquez-Geppert, S., Huster, R. J., & Herrmann, C. S. (2017). Eeg-neurofeedback as a tool to modulate cognition and behavior: a review tutorial. *Frontiers in Human Neuroscience*, *11*, 51.

Fatourechi, M., Bashashati, A., Ward, R. K., & Birch, G. E. (2007). Emg and eog artifacts in brain computer interface systems: A survey. *Clinical Neurophysiology*, *118*(3), 480–494.

Feess, D., Krell, M. M., & Metzen, J. H. (2013). Comparison of sensor selection mechanisms for an erp-based brain-computer interface. *PLoS One*, *8*(7), Article e67543.

Feng, J., Yin, E., Jin, J., Saab, R., Daly, I., Wang, X., et al. (2018). Towards correlation-based time window selection method for motor imagery bcis. *Neural Networks*, *102*, 87–95.

Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, *84*(405), 165–175.

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping*, *2*(1–2), 56–78.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connectivity*, *1*(1), 13–36.

Ghoshuni, M., Firoozabadi, M., Khalilzadeh, M. A., & Hashemi Golpayegani, M. R. (2012). The effect of sensorimotor rhythm enhancing neurofeedback on power of adjacent frequency bands. *Biomedical Engineering: Applications, Basis and Communications*, *24*(04), 307–312.

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, *101*(23), e215–e220.

Gourévitch, B., Le Bouquin-Jeannès, R., & Faucon, G. (2006). Linear and nonlinear causality between signals: methods, examples and neurophysiological applications. *Biological Cybernetics*, *95*(4), 349–369.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424–438.

Greenblatt, R. E., Pflieger, M., & Ossadtchi, A. (2012). Connectivity measures applied to human brain electrophysiological data. *Journal of Neuroscience Methods*, *207*(1), 1–16.

Gurka, M. J., & Edwards, L. J. (2007). 8 mixed models. In *Handbook of statistics, Vol. 27* (pp. 253–280). Elsevier.

Handiru, V. S., & Prasad, V. A. (2016). Optimized bi-objective eeg channel selection and cross-subject generalization with brain–computer interfaces. *IEEE Transactions on Human-Machine Systems*, *46*(6), 777–786.

Haufe, S. (2012). Towards eeg source connectivity analysis.

Hu, S., Wang, H., Zhang, J., Kong, W., Cao, Y., & Kozma, R. (2015). Comparison analysis: Granger causality and new causality and their applications to motor imagery. *IEEE Transactions on Neural Networks and Learning Systems*, *27*(7), 1429–1444.

Islam, M. K., Rastegarnia, A., & Yang, Z. (2016). Methods for artifact detection and removal from scalp eeg: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*, *46*(4–5), 287–305.

Jin, J., Miao, Y., Daly, I., Zuo, C., Hu, D., & Cichocki, A. (2019). Correlation-based channel selection and regularized feature optimization for mi-based bci. *Neural Networks*, *118*, 262–270.

Johnson, S. H. (2000). Imagining the impossible: intact motor representations in hemiplegics. *Neuroreport*, *11*(4), 729–732.

Johnson, S. H., Sprehn, G., & Saykin, A. J. (2002). Intact motor imagery in chronic upper limb hemiplegics: evidence for activity-independent action representations. *Journal of Cognitive Neuroscience*, *14*(6), 841–852.

Karácsony, T., Hansen, J. P., Iversen, H. K., & Puthusserypady, S. (2019). Brain computer interface for neuro-rehabilitation with deep learning classification and virtual reality feedback. In *Proceedings of the 10th augmented human international conference 2019* (pp. 1–8).

Kayikcioglu, T., & Aydemir, O. (2010). A polynomial fitting and k-nn based approach for improving classification of motor imagery bci data. *Pattern Recognition Letters*, *31*(11), 1207–1215.

Kim, Y., Ryu, J., Kim, K. K., Took, C. C., Mandic, D. P., & Park, C. (2016). Motor imagery classification using mu and beta rhythms of eeg with strong uncorrelating transform based complex common spatial patterns. *Computational Intelligence and Neuroscience*, *2016*.

Krumin, M., & Shoham, S. (2010). Multivariate autoregressive modeling and granger causality analysis of multiple spike trains. *Computational Intelligence and Neuroscience*, *2010*.

Kumar, S., Sharma, A., & Tsunoda, T. (2019). Brain wave classification using long short-term memory network based optical predictor. *Scientific Reports*, *9*(1), 1–13.

Lal, T. N., Schroder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N., et al. (2004). Support vector channel selection in bci. *IEEE Transactions on Biomedical Engineering*, *51*(6), 1003–1010.

Loboda, A., Margineanu, A., Rotariu, G., & Lazar, A. M. (2014). Discrimination of eeg-based motor imagery tasks by means of a simple phase information method. *International Journal of Advanced Research in Artificial Intelligence*, *3*(10).

Lu, H., Eng, H.-L., Guan, C., Plataniotis, K. N., & Venetsanopoulos, A. N. (2010). Regularized common spatial pattern with aggregation for eeg classification in small-sample setting. *IEEE Transactions on Biomedical Engineering*, *57*(12), 2936–2946.

Ma, Y., Ding, X., She, Q., Luo, Z., Potter, T., & Zhang, Y. (2016). Classification of motor imagery eeg signals with support vector machines and particle swarm optimization. *Computational and Mathematical Methods in Medicine*, *2016*.

Mahajan, R., & Morshed, B. I. (2014). Unsupervised eye blink artifact denoising of eeg data with modified multiscale sample entropy, kurtosis, and wavelet-ica. *IEEE Journal of Biomedical and Health Informatics*, *19*(1), 158–165.

Müller-Gerking, J., Pfurtscheller, G., & Flyvbjerg, H. (1999). Designing optimal spatial filters for single-trial eeg classification in a movement task. *Clinical Neurophysiology*, *110*(5), 787–798.

Padfield, N., Zabalza, J., Zhao, H., Masero, V., & Ren, J. (2019). Eeg-based brain-computer interfaces using motor-imagery: Techniques and challenges. *Sensors*, *19*(6), 1423.

Park, C., Took, C. C., & Mandic, D. P. (2013). Augmented complex common spatial patterns for classification of noncircular eeg from motor imagery tasks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *22*(1), 1–10.

Ramadan, R. A., & Vasilakos, A. V. (2017). Brain computer interface: control signals review. *Neurocomputing*, *223*, 26–44.

Ramoser, H., Muller-Gerking, J., & Pfurtscheller, G. (2000). Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, *8*(4), 441–446.

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., & Wolpaw, J. R. (2004). Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on Biomedical Engineering*, *51*(6), 1034–1043.

Shan, H., Xu, H., Zhu, S., & He, B. (2015). A novel channel selection method for optimal classification in different motor imagery bci paradigms. *Biomedical Engineering Online*, *14*(1), 93.

Sharma, N., Baron, J.-C., & Rowe, J. B. (2009). Motor imagery after stroke: relating outcome to motor network connectivity. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *66*(5), 604–616.

Sitaram, R., Ros, T., Stoeckel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., et al. (2017). Closed-loop brain training: the science of neurofeedback. *Nature Reviews Neuroscience*, *18*(2), 86–100.

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the bci competition iv. *Frontiers in Neuroscience*, *6*, 55.

Tolić, M., & Jović, F. (2013). Classification of wavelet transformed eeg signals with neural network for imagined mental and motor tasks. *Kinesiology: International Journal of Fundamental and Applied Kinesiology*, *45*(1), 130–138.

Tyagi, A., & Nehra, V. (2016). Classification of motor imagery eeg signals using svm, k-nn and ann. *CSI Transactions on ICT*, *4*(2–4), 135–139.

Wang, B., Wong, C. M., Wan, F., Mak, P. U., Mak, P. I., & Vai, M. I. (2009). Comparison of different classification methods for eeg-based brain computer interfaces: a case study. In *2009 international conference on information and automation* (pp. 1416–1421). IEEE.

Winkler, I., Haufe, S., & Tangermann, M. (2011). Automatic classification of artifactual ica-components for artifact removal in eeg signals. *Behavioral and Brain Functions*, *7*(1), 30.

Yang, Y., Kyrgyzov, O., Wiart, J., & Bloch, I. (2013). Subject-specific channel selection for classification of motor imagery electroencephalographic data. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 1277–1280). IEEE.

Yang, J., Singh, H., Hines, E. L., Schlaghecken, F., Iliescu, D. D., Leeson, M. S., et al. (2012). Channel selection and classification of electroencephalogram signals: an artificial neural network and genetic algorithm-based approach. *Artificial Intelligence in Medicine*, *55*(2), 117–126.