

# Granger Causal Inference Based on Dual Laplacian Distribution and Its Application to MI-BCI Classification

Peiyang Li<sup>1</sup>, Xiaohui Gao, Cunbo Li<sup>2</sup>, Chanlin Yi<sup>2</sup>, Weijie Huang, Yajing Si, Fali Li<sup>3</sup>,  
Zehong Cao<sup>4</sup>, *Member, IEEE*, Yin Tian<sup>5</sup>, and Peng Xu<sup>6</sup>, *Member, IEEE*

**Abstract**—Granger causality-based effective brain connectivity provides a powerful tool to probe the neural mechanism for information processing and the potential features for brain computer interfaces. However, in real applications, traditional Granger causality is prone to the influence of outliers, such as inevitable ocular artifacts, resulting in unreasonable brain linkages and the failure to decipher inherent cognition states. In this work, motivated by constructing the sparse causality brain networks under the strong physiological outlier noise conditions, we proposed a dual Laplacian Granger causality analysis (DLap-GCA) by imposing Laplacian distributions on both model parameters and residuals. In essence, the first Laplacian assumption on residuals will resist the influence of outliers in electroencephalogram (EEG) on causality inference, and the second Laplacian assumption on model parameters will sparsely characterize the intrinsic interactions among multiple brain regions. Through simulation study, we quantitatively verified its effectiveness in suppressing the influence of complex outliers, the stable capacity for model estimation, and sparse network inference. The application to motor-imagery (MI) EEG further reveals that our method can effectively capture the inherent hemispheric lateralization of MI tasks with sparse patterns even under strong noise conditions. The MI classification based on the network features derived from the proposed approach shows higher accuracy than other existing traditional approaches, which is attributed to the discriminative network structures being captured in a timely manner by DLap-GCA even under the single-trial online condition. Basically, these results consistently show its robustness to the influence of complex

outliers and the capability of characterizing representative brain networks for cognition information processing, which has the potential to offer reliable network structures for both cognitive studies and future brain-computer interface (BCI) realization.

**Index Terms**—Bayesian, electroencephalograms (EEGs), Granger causality, Laplacian distribution, motor-imagery (MI) brain-computer interface (BCI).

## I. INTRODUCTION

IN NEUROSCIENCE studies, brain networks characterize information integration and propagation patterns among multiple brain areas [1], [2], [3], [4]. Different from analyzing activation in a single encephalic region, brain network analysis focuses on transforming the information interactions of brain neurons into connected graphs [1], [2] and reveals the differences among various cognitive states by graphical indices, such as shortest paths and global efficiency [1], [2].

Theoretically, the methods for brain network estimation can be divided into two categories: undirected connections and directed connections [3], [5], [6]. Compared with undirected connections, directed connections reveal information integration patterns in the brain, providing quantitative indicators for cognition deciphering [3], [7]. Based on those neural mechanisms, researchers can further develop novel heuristic neural networks [8], [9]. For example, Li et al. [8] proposed bihemisphere domain adversarial neural networks to model the lateralized information processing patterns in the brain for emotional recognition. Wang et al. [9] defined a novel reinforcement learning algorithm to utilize instantaneous rewards for spatial credit assignment, showing powerful decoding ability for sophisticated brain-computer interface (BCI) tasks.

Among the technologies of directed network estimation, Granger causality analysis (GCA) is one of the most commonly used [3], [5]. Different from other directed network estimations, GCA focuses on measuring causal influence among sources [3], through which we can infer whether there is a driving relationship between arbitrary brain region pairs [3]. These merits of GCA directly reveal the asynchronous causality influence of neural activities [3], [5], [6]. Since GCA was proposed, many related studies have been carried out [3], [5]. Yan and He [10] investigated the directed connections of the whole brain with resting-state signals and identified several driving hubs from subsystems of the

Manuscript received 12 June 2022; revised 12 January 2023 and 13 April 2023; accepted 24 June 2023. This work was supported in part by the STI 2030-Major Projects under Grant 2022ZD0211400 and Grant 2022ZD0208500; in part by the National Natural Science Foundation of China under Grant #61901077, Grant #62171074, Grant #U19A2082, and Grant #61961160705; in part by the Natural Science Foundation of Chongqing under Grant CSTB2022NSQ-MSX1171; and in part by the Scientific and Technological Research Program of Chongqing Municipal Education Commission under Grant KJQN202200640. (Peiyang Li and Xiaohui Gao contributed equally to this work.) (Corresponding authors: Peng Xu; Yin Tian; Zehong Cao.)

Peiyang Li, Xiaohui Gao, Weijie Huang, and Yin Tian are with the School of Bioinformatics, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: tianyin@cqupt.edu.cn).

Cunbo Li, Chanlin Yi, Fali Li, and Peng Xu are with the Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation and School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China (e-mail: xupeng@uestc.edu.cn).

Yajing Si is with the School of Psychology, Xinxiang Medical University, Xinxiang 453000, China.

Zehong Cao is with STEM, University of South Australia, Adelaide, SA 5095, Australia (e-mail: zhcao@unsa.edu.au).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TNNLS.2023.3292179>.

Digital Object Identifier 10.1109/TNNLS.2023.3292179

attentional networks. Xue et al. [11] analyzed resting-state functional magnetic resonance imaging (fMRI) data by GCA and found that patients with Alzheimer's disease and mild cognitive impairment showed certain changes in the hippocampal directional functional connection area. Zhang et al. [12] found that the network connections between the hippocampus and the amygdala in heroin addicts were stronger and more effective than those in normal groups. These studies consistently demonstrated that GCA is effective in revealing causal relationships among brain regions [10], [11], [12].

Although traditional GCA has been successfully applied in various studies [3], [5], [10], [11], [12], it is prone to the influence of impulse noise due to the assumption that the model noise follows a Gaussian distribution [3], [13]. Thus, artifact removal preprocessing is often necessary before applying GCA. However, this preprocessing is mainly performed offline, while applications, such as online neural feedback and brain computer interfaces, urgently require online network analysis [14]. In essence, the model under the Gaussian noise assumption is solved by minimizing an L2-norm cost function [3], [6], [15]. However, the L2-norm has been proven to be sensitive to outlier influence, indicating that traditional GCA may not be effective in suppressing the influence of noise with a heavy-tailed distribution [3], [5], [6]. Alternatively, methods such as reweighted Bayesian-based models [6], [16], [17] that would enhance the robustness of causal inference in GCA have been proposed. However, these methods are still restricted by the influence of impulse noise, particularly in situations where prior structural information is not available before analysis [3], [13].

Compared with the Gaussian distribution, the Laplacian distribution has shown robustness to outliers [3], [5], [13], [14], [18]. Notably, the models with this kind of "outlier-robust" distribution can usually be transformed into the least L1-norm problems [5], [13], [14], and the robustness of the L1-norm space to outliers has been proven by various studies [14]. In recent years, the Laplacian distribution has been successfully applied to various fields [3], [13], [19]. Gai et al. [19] introduced the Laplacian distribution into single-gene wavelet transformation using a Bayesian framework, which can effectively restrict impulse noise in ultrasound images. Iriawan et al. [20] developed a new image segmentation method based on the Markov random field and the Laplacian distribution, which successfully suppressed noise in the image recognition process. Essentially, most of these improvements either focus on the restriction of the model residuals or impose the Laplacian distribution on model coefficients [13], [17]. Theoretically, the Laplacian distribution imposed on the model residuals can improve the model stability to resist outliers in observed data [13], [20], while the model coefficients following the Laplacian distribution might further improve the accuracy of parameter estimation [16], [17], especially when the network has sparse connections [18]. However, in network analysis, no study has ever imposed Laplacian restrictions on both model coefficients and residuals, mainly due to the challenge of simultaneously realizing these two goals by a direct method. In this work, we assumed that both the model residuals and the coefficients

follow a Laplacian distribution and designed a reasonable iteration by expectation maximization (EM) with variational Bayesian structure to obtain a reliable sparse GCA model under strong impulse noise conditions. Our source code and supplemental materials will be available online.<sup>1</sup> The main contributions of this article are summarized as follows.

(1) We develop a novel Granger causality estimation method for brain network analysis, which not only resists the influence of outliers in electroencephalograms (EEGs) but also sparsely characterizes the intrinsic interactions between multiple brain regions.

(2) We design two novel indices to quantitatively measure the contralateral patterns for cognitive states and the discriminability of network structures as features for EEG classification.

(3) The network topologies estimated by the proposed method have superior classification performance to a wide variety of energy-distribution-based features, such as standard common spatial pattern (CSP), and competitive performance with some state-of-the-art methods, as proposed in [21], [22], and [23].

(4) The network patterns characterized by our proposed method contain neural interpretable features that are consistent with the working mechanism of the motor-imagery (MI) task.

The remainder of this article is organized as follows. Section II presents the algorithmic principles of traditional GCA, its related improvements, our proposed dual Laplacian distribution-based GCA, the setting of the simulation experiments, and the setup of the real MI experiments in detail. Section III describes the results for both the simulation study and the real EEG experiment. Section IV analyzes the performance difference among different GCAs and the possible superiority of our proposed algorithm to other methods. Section V briefly presents our findings in this work.

## II. METHODS

### A. Granger Causality Analysis Based on Gaussian Distribution

GCA aims to measure the causal interactions among time series with a multivariate autoregression (MVAR) model [3], [6]. Given a  $D$ -dimensional time series  $\mathbf{X}_n = [x_1(n), x_2(n), \dots, x_D(n)]^T$  at the  $n$ th time point, its corresponding MVAR system can be written as

$$\mathbf{X}_n = \mathbf{U}_n \mathbf{G} + \mathbf{v}_n, \quad n \in [1, N] \quad (1)$$

where  $\mathbf{v}_n = [v_1(n), v_2(n), \dots, v_D(n)]^T$  represents the noise process, and  $N$  denotes the number of time points.  $\mathbf{U}_n \in \mathbb{R}^{D \times ((D \times q) \times D)}$  is the transfer function, which holds the form

$$\mathbf{U}_n = \begin{bmatrix} \mathbf{X}_{n-1}^T, \dots, \mathbf{X}_{n-q}^T & 0 & \dots & 0 \\ 0 & \mathbf{X}_{n-1}^T, \dots, \mathbf{X}_{n-q}^T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{X}_{n-1}^T, \dots, \mathbf{X}_{n-q}^T \end{bmatrix}. \quad (2)$$

<sup>1</sup><https://github.com/Gaitxh/DLap-GCA-Granger-causal-inference-based-on-dual-Laplacian-distribution>

$\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_D]^T \in R^{[(D \times q) \times q] \times 1}$  denotes the model coefficients, with  $\mathbf{g}_k \in R^{(D \times q) \times 1}$ ,  $k = 1, 2, \dots, D$ , representing the causal influences from the other time series to the  $k$ th time series. Theoretically, the matrix form of (1) can be written as

$$\mathbf{Y} = \mathbf{U}\mathbf{G} + \mathbf{V} \quad (3)$$

where

$$\begin{aligned} \mathbf{Y} &= [\mathbf{X}_{q+1}, \mathbf{X}_{q+2}, \dots, \mathbf{X}_N]^T \in R^{[(N-q) \times D] \times 1} \\ \mathbf{V} &= [v_{q+1}, v_{q+2}, \dots, v_N]^T \in R^{[(N-q) \times D] \times 1} \\ \mathbf{U} &= [\mathbf{U}_{q+1}^T, \mathbf{U}_{q+2}^T, \dots, \mathbf{U}_k^T, \dots, \mathbf{U}_{N-q}^T]^T \\ &\in R^{[(N-q) \times D] \times [(D \times q) \times D]} \\ \mathbf{G} &= [\mathbf{g}_1^T, \mathbf{g}_2^T, \dots, \mathbf{g}_D^T]^T \in R^{[(D \times q) \times D] \times 1}. \end{aligned} \quad (4)$$

Thus, for the  $k$ th time series, (3) is further rewritten as

$$Y_k = \mathbf{u}\mathbf{g}_k + E_k \quad (5)$$

with  $Y_k = [x_k(q+1), x_k(q+2), \dots, x_k(N)]^T \in R^{(N-q) \times 1}$ , and  $E_k = [v_k(q+1), v_k(q+2), \dots, v_k(N)]^T \in R^{(N-q) \times 1}$ , which is assumed to follow Gaussian distribution with zero mean and variance  $\Sigma_{v_k} = \sigma_k^2 \mathbf{I}_{(N-q) \times (N-q)}$ .  $\sigma_k^2$  is the hyperparameter of  $\Sigma_{v_k}$ , and  $\mathbf{u}$  is the transfer matrix, which holds the form of

$$\mathbf{u} = \begin{bmatrix} \mathbf{X}_1^T & \mathbf{X}_2^T & \dots & \mathbf{X}_q^T \\ \mathbf{X}_2^T & \mathbf{X}_3^T & \dots & \mathbf{X}_{q+1}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{N-q}^T & \mathbf{X}_{N-q+1}^T & \dots & \mathbf{X}_{N-1}^T \end{bmatrix} \in R^{(N-q) \times (D \times q)}. \quad (6)$$

By assuming that the prior distribution of  $\mathbf{g}_k$  follows identical independent normal distributions with zero mean and variance  $\alpha_k$  as

$$\begin{aligned} p(\mathbf{g}_k | \alpha_k) &= \prod_{m=1}^{D \times q} p(\mathbf{g}_{k,m} | \alpha_{k,m}) = (2\pi)^{-\frac{D \times q}{2}} |\alpha_k|^{-\frac{1}{2}} e^{-\frac{1}{2} \mathbf{g}_k^T \alpha_k^{-1} \mathbf{g}_k} \\ \alpha_k &= \text{diag}\{\alpha_{k,1}, \alpha_{k,2}, \dots, \alpha_{k,D \times q}\} \in R^{(D \times q) \times (D \times q)} \end{aligned} \quad (7)$$

the prior distribution of the noise model can be further represented as

$$p(Y_k | \mathbf{g}_k, \sigma_k^2) = p(E_k | \sigma_k^2) = (2\pi)^{-\frac{N-q}{2}} |\Sigma_{v_k}|^{-\frac{1}{2}} e^{-\frac{1}{2} E_k^T \Sigma_{v_k}^{-1} E_k}. \quad (8)$$

Therefore, the posteriors of  $\mathbf{g}_k$  are calculated as

$$p(\mathbf{g}_k | Y_k, \alpha_k, \sigma_k^2) = N(\mathbf{g}_k | C_{\text{GCA}} Y_k, P_{\text{GCA}}) \quad (9)$$

where

$$P_{\text{GCA}} = (\mathbf{u}^T \Sigma_{v_k}^{-1} \mathbf{u} + \alpha_k^{-1})^{-1}, C_{\text{GCA}} = P_{\text{GCA}} \mathbf{u}^T \Sigma_{v_k}^{-1}. \quad (10)$$

By integrating all of the hyperparameters into one set, i.e.,  $\theta_{\text{GCA}} = [\alpha_k, \sigma_k^2]$ , the model coefficients can be further

estimated by the following EM-based iterative scheme [13]:

$$\begin{aligned} \theta_{\text{GCA}}^{(t+1)} &= \arg \min_{\theta_{\text{GCA}}} E_{p(\mathbf{g}_k | Y_k, \theta^{(t)})} \ln[p(Y_k | \mathbf{g}_k, \sigma_k^2) p(\mathbf{g}_k | \alpha_k)] \\ &= \int p(\mathbf{g}_k | Y_k, \theta_{\text{GCA}}^{(t)}) \ln[p(Y_k | \mathbf{g}_k, \sigma_k^2) p(\mathbf{g}_k | \alpha_k)] d\mathbf{g}_k \\ &= \left(-\frac{1}{2}\right) \cdot \left\{ (N-q) \cdot \ln 2\pi + \ln |\Sigma_{v_k}| + Y_k^T (\Sigma_{v_k})^{-1} Y_k \right. \\ &\quad \left. - 2 Y_k^T (\Sigma_{v_k})^{-1} \mathbf{u} C^{(t)} Y_k + \ln |\alpha_k| \right. \\ &\quad \left. + \text{trace}\left\{ \left[ \mathbf{u}^T (\Sigma_{v_k})^{-1} \mathbf{u} + (\alpha_k)^{-1} \right] P_{\text{GCA}}^{(t)} \right\} \right. \\ &\quad \left. + Y_k^T (C_{\text{GCA}}^{(t)})^T (\mathbf{u}^T (\Sigma_{v_k})^{-1} \mathbf{u} + (\alpha_k)^{-1}) C_{\text{GCA}}^{(t)} Y_k \right\} \\ P_{\text{GCA}}^{(t)} &= \left[ \mathbf{u}^T (\Sigma_{v_k}^{(t)})^{-1} \mathbf{u} + (\alpha_k^{(t)})^{-1} \right]^{-1} \\ C_{\text{GCA}}^{(t)} &= P_{\text{GCA}}^{(t)} \mathbf{u}^T (\Sigma_{v_k}^{(t)})^{-1}. \end{aligned} \quad (11)$$

Theoretically, after the MVAR model is solved, the causal interactions between  $Y_k$  and  $Y_h$  can be measured in accordance with the following principle [3]: if  $Y_h$  improves the prediction performance of  $Y_k$  when all the other processes are also included, we can state that  $Y_h$  G-causes  $Y_k$ . Mathematically, this causal interaction can be further evaluated by the logarithmic ratio between unrestricted noise covariance and restricted noise covariance as

$$\ell_{h \rightarrow k | \Delta} = \ln \frac{\Sigma_{k,k}^*}{\Sigma_{k,k}}, \quad (\Delta = [1, \dots, D] - [h, k], \Delta \in R^{D-2}) \quad (12)$$

where  $\Sigma_{k,k}^*$  denotes the noise covariance of the  $k$ th equation from the restricted model (which omits the influence of the  $k$ th time series), and  $\Sigma_{k,k}$  is the noise covariance from the unrestricted model (considering the influence of all of the time series). Obviously, the noise estimation in both the restricted model and the unrestricted model is also very important for inferring the Granger causal interactions between two time series. We refer to this algorithm as dual Gaussian-constraints GCA (Gau-GCA) for further comparison.

## B. Granger Causality Analysis-Related Improvements

Recent findings in network neuroscience have revealed the efficiency of the brain network, indicating its sparse information propagation and integration patterns during cognitive tasks [3], [24]. Specifically, when one performs a given cognitive task, there are always some hub nodes handling information processing in the brain, guiding the researchers to reveal the mechanism of information integration in brain networks and promoting the development of novel neuromorphic computing [25]. However, the network estimated by traditional Granger causality is usually too dense to identify these crucial nodes [3], [14]. Thus, the sparse-constraint-based strategy has been proposed. Usually, these types of technologies hold the form of

$$\arg \min_{\mathbf{g}_k} f(\mathbf{g}_k) = \|\mathbf{Y}_k - \mathbf{u}\mathbf{g}_k\|_2 + \lambda \|\mathbf{g}_k\|_1 \quad (13)$$

where  $\|\cdot\|_2$  denotes the L2-norm, and  $\|\cdot\|_1$  denotes the L1-norm. In essence, (13) is equivalent to solving a sparse coding problem, which assumes that model noise  $E_k$  follows

the Gaussian distribution and model coefficients  $\mathbf{g}_k$  follow the Laplacian distribution [16] as

$$p(E_k|\sigma_k^2) = (2\pi)^{-\frac{N-q}{2}} |\Sigma_{v_k}|^{-\frac{1}{2}} e^{-\frac{1}{2} E_k^T \Sigma_{v_k}^{-1} E_k}$$

$$p(\mathbf{g}_k|\boldsymbol{\beta}_k) = \prod_{m=1}^{D \times q} \left( \frac{2}{\beta_{k,m}} \right)^{-1} e^{-\beta_{k,m} |\mathbf{g}_{k,m}|} \quad (14)$$

where  $|\cdot|$  denotes the absolute operator, and  $\beta_k$  is the scale parameter of  $\mathbf{g}_k$ , with  $\beta_{k,m} > 0$  being the  $m$ th element of  $\beta_k$ . Practically, the parameters in (14) can be solved with EM as described in the Supplementary Materials (see Appendix E). In current study, we refer to this kind of method as GCA improvement with Laplacian restriction imposed on model coefficients (Lap-GCA).

Although sparsity has been realized, another challenge still exists, i.e., the inevitable influence of outliers contaminating EEG recordings [3]. In EEG recordings, outliers caused by ocular artifacts and head movements make the model residuals no longer meet the Gaussian distribution assumption [3], [26] and result in many pseudocausal connections [3], [6]. Even the sparse constraint imposed on model coefficients cannot eliminate the influence of outliers. To address this problem, we propose restricting the model residuals in the L1-norm space [5] as

$$\operatorname{argmin}_{\mathbf{g}_k} f(\mathbf{g}_k) = \|\mathbf{Y}_k - \mathbf{u}\mathbf{g}_k\|_1 \quad (15)$$

which has a similar form to the maximum likelihood solution of the model with the assumption that the noises follow the Laplacian distribution [14]. In current study, we refer to this kind of method as GCA improvement with Laplacian restriction imposed on model noise (GLap-GCA). Although this strategy successfully improves the robustness of model solutions to outliers, it cannot allow for the sparse structure of neural networks.

Theoretically, when assuming that both the noise and coefficients obey Gaussian distribution, the model coefficients estimated in (11) were equivalent to the least square solution of (3), which is not sparse and sensitive to outliers with large amplitudes [3], [5]. By assuming that either the model noise or the coefficients obey a Laplacian distribution, the inferred network structure becomes either robust to noise influence or sparse. Thus, in the current study, we attempted to solve the model by assuming that both the model coefficients and residuals in (3) follow the Laplacian distribution to improve the robustness of network inference while preserving the sparse network structure.

### C. Granger Causality Analysis Based on Dual Laplace Distribution

Theoretically, this assumption is equivalent to solving the following optimal problem:

$$\operatorname{argmin}_{\mathbf{g}_k} f(\mathbf{g}_k) = \|\mathbf{Y}_k - \mathbf{u}\mathbf{g}_k\|_1 + \lambda \|\mathbf{g}_k\|_1. \quad (16)$$

By supposing that  $E_k$  follows the Laplacian distribution, the likelihood of observation in (5) can be represented as

$$p(Y_k|\mathbf{g}_k, \sigma_k^2) = p(E_k|\sigma_k) = \prod_{m=1}^{N-q} \frac{1}{\sqrt{2}\sigma_k} e^{-\frac{\sqrt{2}|E_k(m)|}{\sigma_k}}. \quad (17)$$

At the same time, making the same prior assumption of  $\mathbf{g}_k$  as in (14), the posteriors of  $\mathbf{g}_k$  can be calculated as

$$p(\mathbf{g}_k|Y_k, \boldsymbol{\beta}_k, \sigma_k^2) = \frac{p(Y_k|\mathbf{g}_k, \sigma_k^2)p(\mathbf{g}_k|\boldsymbol{\beta}_k)}{p(Y_k|\boldsymbol{\beta}_k, \sigma_k^2)}. \quad (18)$$

Practically, the posteriors of  $\mathbf{g}_k$  cannot be estimated through (18) directly because the joint distribution of  $p(Y_k|\boldsymbol{\beta}_k, \sigma_k^2)$  and  $p(\mathbf{g}_k|\boldsymbol{\beta}_k)$  is no longer Gaussian [3], [13]. Alternatively, we utilize variational Bayesian framework for the model solution. By introducing variational parameter  $\xi_k$  for  $p(\mathbf{g}_k|\boldsymbol{\beta}_k)$ , we have

$$p(\mathbf{g}_k|\boldsymbol{\beta}_k, \xi_k) = \prod_{m=1}^{D \times q} \frac{1}{2} \sqrt{2\pi\beta_{k,m}|\xi_{k,m}|} e^{-\frac{1}{2}\beta_{k,m}|\xi_{k,m}|}$$

$$\cdot (2\pi)^{-\frac{D \times q}{2}} \Lambda_k^{-\frac{1}{2}} e^{-\frac{1}{2} \mathbf{g}_k^T \Lambda_k^{-1} \mathbf{g}_k}$$

$$\Lambda_k = \operatorname{diag} \left\{ \frac{|\xi_{k,1}|}{\beta_{k,1}}, \frac{|\xi_{k,2}|}{\beta_{k,2}}, \dots, \frac{|\xi_{k,D \times q}|}{\beta_{k,D \times q}} \right\} \quad (19)$$

where  $\xi_{k,m}$  is the  $m$ th element of  $\xi_k$ . Based on the theorem in a previous study [16],  $-|\mathbf{g}_{k,m}|$  holds the lower band as

$$-|\mathbf{g}_{k,m}| \geq -\frac{1}{2} \left( \frac{\mathbf{g}_{k,m}^2}{|\xi_{k,m}|} + |\xi_{k,m}| \right) \quad (20)$$

through which we can further infer that  $p(\mathbf{g}_k|\boldsymbol{\beta}_k) \geq p(\mathbf{g}_k|\boldsymbol{\beta}_k, \xi_k)$ , and  $p(\mathbf{g}_k|\boldsymbol{\beta}_k, \xi_k)$  is actually the strict lower bound of  $p(\mathbf{g}_k|\boldsymbol{\beta}_k)$ . Similarly, by assuming that  $E_k$  follows a Laplacian distribution, the likelihood of  $p(E_k(m))$  in (19) can be rewritten as

$$p(E_k(m)) = \int_0^{+\infty} p(E_k(m)|\tau_k(m)) p(\tau_k(m)) d\tau_k(m)$$

$$= \int_0^{+\infty} \frac{1}{\sqrt{2\pi\tau_k(m)}} e^{-\frac{(E_k(m))^2}{2\tau_k(m)}} \cdot p(\tau_k) d\tau_k(m), \quad \tau_k(m) \geq 0 \quad (21)$$

where  $m$  denotes the  $m$ th component of  $E_k$ .  $p(\tau_k(m)|\sigma_k^2) = (1/\sigma_k^2) e^{[-\tau_k(m)/\sigma_k^2]}$  is independent of  $E_k(m)$ . When the value of  $\tau_k(m)$  is given,  $p(E_k(m))$  can be subsequently represented by a linear combination of various Gaussian distributions [3], [13]. Thus, the covariance matrix of noise  $E_k$  can be rewritten as

$$\Sigma_{v_k} = \operatorname{diag}\{\tau_k(1), \tau_k(2), \dots, \tau_k(N-q)\}. \quad (22)$$

Consequently, we can rewrite the posterior distribution of  $\mathbf{g}_k$  as

$$p(\mathbf{g}_k|Y_k, \tau_k, \sigma_k^2, \boldsymbol{\beta}_k)$$

$$= \frac{p(Y_k|\mathbf{g}_k, \tau_k) \cdot p(\mathbf{g}_k|\boldsymbol{\beta}_k) \cdot p(\tau_k|\sigma_k^2)}{\int p(Y_k|\mathbf{g}_k, \tau_k) \cdot p(\mathbf{g}_k|\boldsymbol{\beta}_k) \cdot p(\tau_k|\sigma_k^2) d\mathbf{g}_k}$$

$$\geq p(\mathbf{g}_k|Y_k, \tau_k, \sigma_k^2, \boldsymbol{\beta}_k, \xi_k)$$

$$= \frac{p(Y_k|\mathbf{g}_k, \tau_k) \cdot p(\mathbf{g}_k|\boldsymbol{\beta}_k, \xi_k) \cdot p(\tau_k|\sigma_k^2)}{\int p(Y_k|\mathbf{g}_k, \tau_k) \cdot p(\mathbf{g}_k|\boldsymbol{\beta}_k, \xi_k) \cdot p(\tau_k|\sigma_k^2) d\mathbf{g}_k}$$



$$= N(\mathbf{g}_k | (\mathbf{u}^T \Sigma_{v_k}^{-1} \mathbf{u} + \Lambda_k^{-1})^{-1} \mathbf{u}^T \Sigma_{v_k}^{-1} Y_k, (\mathbf{u}^T \Sigma_{v_k}^{-1} \mathbf{u} + \Lambda_k^{-1})^{-1}). \quad (23)$$

Through (23), it is easy to see that  $P = (\mathbf{u}^T \Sigma_{v_k}^{-1} \mathbf{u} + \Lambda_k^{-1})^{-1}$  is the covariance matrix of  $\mathbf{g}_k$  and  $CY_k$  is the expectation of  $\mathbf{g}_k$  with  $C = P\mathbf{u}^T \Sigma_{v_k}^{-1}$ . Based on the joint probability density of  $Y_k$ ,  $\mathbf{g}_k$ , and  $\theta$ , we have

$$p(Y_k, \mathbf{g}_k | \theta) \cdot p(\theta) = p(Y_k | \mathbf{g}_k, \tau_k) \cdot p(\mathbf{g}_k | \beta_k, \xi_k) \cdot p(\tau_k | \sigma_k^2). \quad (24)$$

Thus, by maximizing the following marginal likelihood function, we can estimate all of these hyperparameters as

$$\theta^{(t+1)} = \max_{\theta} E_{p(\mathbf{g}_k | Y_k, \theta^{(t)})} \ln[p(Y_k, \mathbf{g}_k | \theta) \cdot p(\theta)] \quad (25)$$

where  $(\cdot)^{(t)}$  and  $(\cdot)^{(t+1)}$  represent the values calculated at step  $t$  and step  $t + 1$ , respectively.  $\theta$  denotes the set of all of the hyperparameters with the form of  $\theta = [\beta_{k,1}, \dots, \beta_{k,D \times q}, \xi_{k,1}, \dots, \xi_{k,D \times q}, \sigma_k^2, \tau_k(1), \dots, \tau_k(N - q)]$ .

Theoretically,  $E_{p(\mathbf{g}_k | Y_k, \theta^{(t)})} \ln[p(Y_k, \mathbf{g}_k | \theta) \cdot p(\theta)]$  holds the form of

$$\begin{aligned} & E_{p(\mathbf{g}_k | Y_k, \theta^{(t)})} \ln[p(Y_k, \mathbf{g}_k | \theta) \cdot p(\theta)] \\ &= \int p(\mathbf{g}_k | Y_k, \theta^{(t)}) \ln[p(Y_k, \mathbf{g}_k | \theta) \cdot p(\theta)] d\mathbf{g}_k \\ &= \left(-\frac{1}{2}\right) \cdot \left[ \ln |\Sigma_{v_k}| + Y_k^T \Sigma_{v_k}^{-1} Y_k + 2 \sum_{n=1}^{N-q} \frac{\tau_k(n)}{\sigma_k^2} - 2 \sum_{n=1}^{N-q} \ln \frac{1}{\sigma_k^2} \right. \\ &\quad + \ln |\Lambda_k| - \sum_{m=1}^{D \times q} \ln \sqrt{2\pi \beta_{k,m} |\xi_{k,m}|} + \frac{1}{2} \sum_{m=1}^{D \times q} \beta_{k,m} |\xi_{k,m}| \\ &\quad - 2 Y_k^T \Sigma_{v_k}^{-1} \mathbf{u} C^{(t)} Y_k + \text{trace}\{(\mathbf{u}^T \Sigma_{v_k}^{-1} \mathbf{u} + \Lambda_k^{-1}) P^{(t)}\} \\ &\quad \left. + Y_k^T (C^{(t)})^T (\mathbf{u}^T \Sigma_{v_k}^{-1} \mathbf{u} + \Lambda_k^{-1}) C^{(t)} Y_k \right]. \quad (26) \end{aligned}$$

Therefore, by maximizing (26) with respect to each hyperparameter of  $\theta$ , we can obtain their corresponding iteration solution as

$$\tau_k^{(t+1)}(n) = \frac{(\sigma_k^2)^{(t)}}{4} \left( \sqrt{1 + \frac{8 \cdot \{(\Xi^{(t)}(n))^2 + S^{(t)}(n, n)\}}{(\sigma_k^2)^{(t)}}} - 1 \right) \quad (27)$$

$$(\sigma_k^2)^{(t+1)} = \frac{1}{N - q} \cdot \sum_{n=1}^{N-q} \tau_k^{(t+1)}(n) \quad (28)$$

$$(\xi_{k,m}^{(t+1)})^2 = \Psi^{(t)}(m) \quad (29)$$

$$\frac{1}{\beta_{k,m}^{(t+1)}} = \frac{1}{2} \left( \left| \xi_{k,m}^{(t+1)} \right| + \frac{\Psi^{(t)}(m)}{\left| \xi_{k,m}^{(t+1)} \right|} \right) = \left| \xi_{k,m}^{(t+1)} \right| \quad (30)$$

where

$$\begin{aligned} \Xi^{(t)} &= (Y_k - \mathbf{u} C^{(t)} Y_k), S^{(t)} = \mathbf{u} P^{(t)} \mathbf{u}^T \\ \Psi^{(t)} &= \text{diag} \left[ \left( \mathbf{u}^T (\Sigma_{v_k}^{(t)})^{-1} \mathbf{u} + \Lambda_k^{-1} \right)^{-1} + (C^{(t)} Y_k) (C^{(t)} Y_k)^T \right] \\ P^{(t)} &= \left( \mathbf{u}^T (\Sigma_{v_k}^{(t)})^{-1} \mathbf{u} + (\Lambda_k^{(t)})^{-1} \right)^{-1}, C^{(t)} = P^{(t)} \mathbf{u}^T (\Sigma_{v_k}^{(t)})^{-1}. \quad (31) \end{aligned}$$

Subsequently, the expectation of  $\mathbf{g}_k$  can be updated as

$$\mathbf{g}_k^{(t+1)} = C^{(t+1)} Y_k = P^{(t+1)} \mathbf{u}^T (\Sigma_{v_k}^{(t+1)})^{-1} Y_k \quad (32)$$

where

$$\begin{aligned} \Sigma_{v_k}^{(t+1)} &= \text{diag} \left\{ \tau_k^{(t+1)}(1), \tau_k^{(t+1)}(2), \dots, \tau_k^{(t+1)}(N - q) \right\} \\ \Lambda_k^{(t+1)} &= \text{diag} \left\{ \frac{|\xi_{k,1}^{(t+1)}|}{\beta_{k,1}^{(t+1)}}, \frac{|\xi_{k,2}^{(t+1)}|}{\beta_{k,2}^{(t+1)}}, \dots, \frac{|\xi_{k,D \times q}^{(t+1)}|}{\beta_{k,D \times q}^{(t+1)}} \right\}. \quad (33) \end{aligned}$$

Through limited iteration steps, the objective function defined in (26) will converge, and the optimal model coefficients can be further estimated as

$$\mathbf{g}_k^* = C^* Y_k \quad (34)$$

where  $C^*$  is the final output of  $C^{(t)}$ . Detailed information can be found in the Supplementary Materials (see Appendices A–D), and the parameter estimation is summarized in Algorithm 1. With these hyperparameters, we can further infer the causal interactions between multiple time series as mentioned above. We refer to this algorithm as dual Laplacian GCA (DLap-GCA) to differentiate it from other GCAs.

---

#### Algorithm 1 Model Estimation Using DLap-GCA

---

**Require:**  $\theta \in R^{(1+N-q+2 \times D \times q) \times 1}$ ,  $Y_k \in R^{(N-q) \times 1}$ ,  $\mathbf{u} \in R^{(N-q) \times (D \times q)}$ ,  $s \geq 1$ ,  
 $\theta = [\beta_{k,1}, \dots, \beta_{k,D \times q}, \xi_{k,1}, \dots, \xi_{k,D \times q}, \sigma_k^2, \tau_k(1), \dots, \tau_k(N - q)]$   
1. Initializing:  $\beta_k^0 \leftarrow \beta_k^{Ini}$ ,  $\xi_k^0 \leftarrow \xi_k^{Ini}$ ,  $(\sigma_k^2)^0 \leftarrow (\sigma_k^2)^{Ini}$ ,  $\tau_k^0 = (\sigma_k^2)^0$ ,  $\varepsilon = 1e - 004$   
2.  $t \leftarrow 0$   
3. **while**  $t \leq s$  or  $\frac{\|E_{p(\mathbf{g}_k | Y_k, \theta^{(t+1)})} \ln[\Phi(\theta)] - E_{p(\mathbf{g}_k | Y_k, \theta^{(t)})} \ln[\Phi(\theta)]\|}{\|E_{p(\mathbf{g}_k | Y_k, \theta^{(t)})} \ln[\Phi(\theta)]\|} < \varepsilon$   
 $\Phi(\theta) = p(Y_k, \mathbf{g}_k | \theta) \cdot p(\theta)$   
**do**  
4.  $\tau_k^{(t+1)}(n) = \frac{(\sigma_k^2)^{(t)}}{4} \cdot \left( \sqrt{1 + \frac{8 \cdot \{(\Xi^{(t)}(n))^2 + S^{(t)}(n, n)\}}{(\sigma_k^2)^{(t)}}} - 1 \right)$   
5.  $(\sigma_k^2)^{(t+1)} = \frac{\sum_{n=1}^{N-q} \tau_k^{(t+1)}(n)}{N - q}$   
6.  $(\xi_{k,m}^{(t+1)})^2 = \Psi^{(t)}(m)$   
7.  $\frac{1}{\beta_{k,m}^{(t+1)}} = \frac{1}{2} \left( \left| \xi_{k,m}^{(t+1)} \right| + \frac{\Psi^{(t)}(m)}{\left| \xi_{k,m}^{(t+1)} \right|} \right) = \left| \xi_{k,m}^{(t+1)} \right|$   
8.  $P^{(t+1)} = (\mathbf{u}^T \Sigma_{v_k}^{-1} \mathbf{u} + \Lambda_k^{-1})^{-1}$ ,  $C^{(t+1)} Y_k = P^{(t+1)} \mathbf{u}^T \Sigma_{v_k}^{-1} Y_k$ ,  
9.  $t \leftarrow t + 1$   
10. **end while**  
11. **return**  $\mathbf{g}_k^*$  as  $\mathbf{g}_k^* = C^t Y_k$

---

#### D. Computational Complexity

The computational burden of our proposed method mainly increases in the estimation of the covariance matrix and its inverse in Algorithm 1. Because  $\Sigma_{v_k}$  and  $\Lambda_k$  are the diagonal matrices, the complexity of calculating their inverse matrices

is  $O((N - q)^3)$  and  $O((D \times q)^3)$ , respectively. In addition, the estimation of covariance matrix  $P$  in step 8 involves the complexity of  $O((N - q)^3)$ . Similarly, the complexity for computing the expectation  $CY_k$  of  $g_k$  is  $O((D \times q)(N - q)^2)$ . Subsequently, the maximum computational complexity value of each subitem in (26) at the  $i$ th iteration is  $O(E_{p(g_k|Y_k, \theta^{(q+1)})} \ln[\Phi(\theta)]) = O((N - q)^3)$ .

### E. Experiments

In this study, we designed the experiments of both simulation and real EEG applications to evaluate the robustness of DLap-GCA and reveal its performance difference from that of Gau-GCA, GLap-GCA, and GLap-GCA. During the simulation stage, we mainly assessed the performance of DLap-GCA in resisting the influence of outliers on parameter estimation, data prediction, and network recovery. For real EEG applications, we mainly validated its effectiveness and reliability under noise conditions based on three aspects: 1) the capability of capturing cognition-related intrinsic brain networks and locating crucial brain regions for cognition processing; 2) the capability of offering discriminative brain-network topologies for EEG classification; and 3) the efficiency of online BCI systems.

1) *Simulation Study*: In this part, we mainly assessed the influence of outliers on four GCAs. One concern is the effects of outliers on the estimation of MVAR parameters that characterize the driving relationships among multiple time series, and another concern is their effects on the inference of causal connections. For each experiment, we adopted the Bayesian information criterion (BIC) to determine the order of the MVAR model for all the GCAs [3], [6]. The false discovery rate (FDR) with a threshold of  $p = 0.05$  was utilized to remove the unlikely connections for Gau-GCA, Lap-GCA, GLap-GCA, and DLap-GCA. To measure the parameter bias, we utilized the following indices:

$$\Delta A = \frac{\|A_b - A_a\|_1}{\|A_b\|_1} \quad (35)$$

where  $A_b$  is the predefined MVAR coefficient and  $A_a$  is the estimated parameter under noise conditions. To evaluate the performance of network recovery, the number of correctly recovered linkages (NCRLs) was further utilized as another index, holding the form of

$$Z_{\text{NCRL}} = \sum_{i=1}^D \sum_{j=1}^D (\tilde{B}_{ij} \odot B_{ij}^*) - D \quad (36)$$

where  $\tilde{B}$  and  $B^*$  denote the predefined network and the estimated network, respectively, and  $\odot$  indicates the equivalent operator.  $D$  is the number of nodes in a given network. In addition to NCRLs, we further analyzed the truly recovered causal relationship based on the confusion matrix  $A = a_{kl} \in R^{3 \times 3}$ , with

$$a_{kl} = \sum_{i=1}^D \sum_{j=1}^D P(B_{ij}^* \in l | \tilde{B}_{ij} \in k), \quad l = 1, 2, 3; k = 1, 2, 3 \quad (37)$$

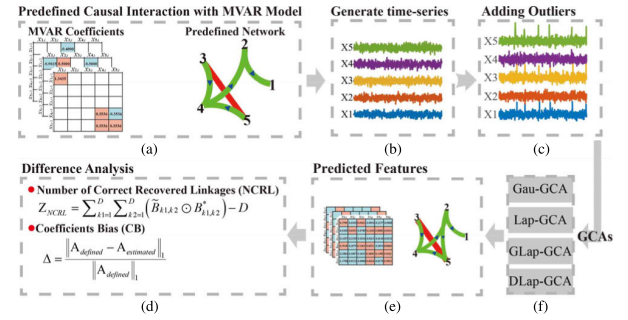


Fig. 1. Flowchart of the simulation study. (a) Predefined network and MVAR coefficients, (b) corresponding time series, (c) using simulated outliers to noise the original time series, (d) GCAs used for comparison, (e) feature estimated by GCAs, and (f) analysis of performance difference between different methods with indices.

where  $l$  and  $k$  indicate one of the connection patterns (i.e., the null connection, the bidirectional connection, and the unidirectional connection). Theoretically, the better that the network is recovered, the more predominant that the values on the diagonal of the confusion matrix are.

Subsequently, we predefined a network with five nodes that represent the causal interactions among five jointly stationary stochastic processes and can be described by the following MVAR model:

$$\begin{aligned} x_1(n) &= 0.95\sqrt{2} \cdot x_1(n-1) - 0.9025 \cdot x_1(n-2) + e_1(n) \\ x_2(n) &= 0.5 \cdot x_1(n-2) - 0.4 \cdot x_3(n-2) + e_2(n) \\ x_3(n) &= -0.4 \cdot x_1(n-3) + 0.4 \cdot x_2(n-2) + e_3(n) \\ x_4(n) &= -0.5 \cdot x_1(n-2) + 0.25\sqrt{2} \cdot x_4(n-1) \\ &\quad + 0.25\sqrt{2} \cdot x_5(n-1) + e_4(n) \\ x_5(n) &= -0.25\sqrt{2} \cdot x_4(n-1) + 0.25\sqrt{2} \cdot x_5(n-1) + e_5(n) \end{aligned} \quad (38)$$

where  $e_i(n)$  indicates the uncorrelated Gaussian noise with a zero mean and identical variance. To evaluate the aforementioned indices, we adopted the same strategy as our previous studies [3], [5], [6]. To analyze the performance differences among the aforementioned methods, we mainly considered the following aspects: 1) the influence of outlier occurrence rates; 2) the influence of outlier intensities, defined as  $c \cdot X_{\max}$ , with  $c$  being a constant within the range of  $c \in [2], [3], [4], [5], [6]$  and  $X_{\max}$  being the maximum of the  $i$ th time series; and 3) the influence caused by the proportions of the time series contaminated by outliers. Fig. 1 illustrates the pipeline of the simulation experiment. For each noise condition, we repeated the experiment 200 times. During the simulation study, both the contaminated time series and the positions of outliers were randomly selected. The paired- $t$  test was performed to reveal the significant performance differences between the proposed one and the corresponding GCA variant [3], [6].

2) *Real EEG Applications*: In real EEG analysis, the ocular artifacts introduced by eye blinking are inevitable and usually hold a large amplitude [27], which create more complex noises than outliers [3], [5], [14], [18], [26]. Thus, in this section, we mainly focused on evaluating the different performances of the aforementioned GCAs (i.e., Gau-, Lap-, GLap-, and DLap-GCAs), when applying them to estimate directed brain networks for MI EEG datasets that contain obvious ocular

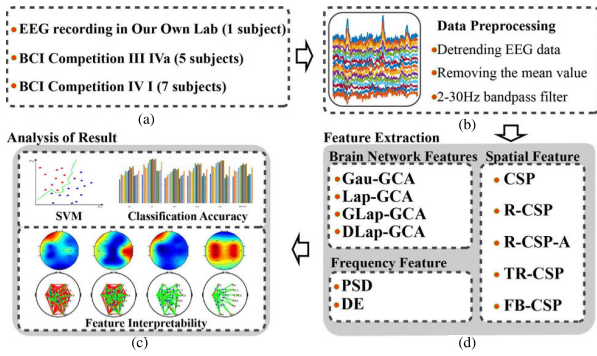


Fig. 2. Experimental pipeline of real EEG applications. (a) Dataset descriptions, (b) data preprocessing, (c) comparisons methods, and (d) analysis results.

artifacts. Fig. 2 illustrates the experimental pipeline of real EEG applications.

3) *Contralateral Network Patterns for MI EEGs*: For real EEG applications, there are no standard criteria to evaluate the performance [3], [14]. Therefore, a dataset with a clearly defined physiological network pattern must be utilized to assess the method performance [3], [5], [28]. Motivated by this requirement, we selected MI EEGs, the network pattern of which has been known to be dominant over the contralateral sensorimotor area [14], [28], [29].

a) *Data description*: In this study, the MI EEGs mentioned in our previous studies [28] were utilized to verify the robustness of DLap-GCA for brain network estimation. These data were recorded by a Symtop amplifier (Symtop Instrument, Beijing, China) of which the online sample rate was 1000 Hz, and the bandpass filter held the passband from 0.4 to 45 Hz. The 15 Ag/AgCl electrodes (i.e., F3, F4, FC3, FC4, Cz, C3, C4, C5, C6, CP3, CP4, P3, P4, O1, and O2) were placed according to the extended international 10–20 system, with AFz serving as the reference. There were 40 healthy participants (25 men, aged  $21.7 \pm 2.6$  years old, three left-handed) recruited from University of Electronic Science and Technology of China (UESTC). After the MI-BCI performance evaluation, we selected the data from one participant who had obvious event-related desynchronization (ERD) or event-related synchronization (ERS) phenomena to ensure the characterized MI network patterns in subsequent analysis [30]. Consequently, a total of 60 trials of EEG measurements were selected (half were right-handed and the others were left-handed), and the segment between 0.5 and 2.5 s after the visual cue was extracted for analysis.

b) *EEG data preprocessing*: Before inferring causal networks, the following preprocessing steps were adopted: 1) detrending EEG data and removing the means for each segment and 2) filtering each EEG segment with a 2–30-Hz bandpass filter to preserve the main frequency band of MI tasks [28].

c) *Brain network estimation of MI EEG*: For MI tasks, a physiological phenomenon is the contralateral lateralization of brain networks [3], [28]. Therefore, in the current work, we assumed that if the noise impact on EEGs could be effectively compressed, the network structures inferred by GCAs would exhibit strong contralateral lateralization. To reveal the influence of ocular artifacts on GCA and to evaluate the

robustness of DLap-GCA to artifacts, all the selected EEG segments contained obvious ocular artifacts. Subsequently, the four GCAs mentioned above were applied to these noisy EEGs to estimate directed networks. To quantitatively compare the performance differences among the four GCAs, we further designed two indices and adopted the  $t$  test to check the significant performance differences between DLap-GCA and the other methods [3].

The first index is the lateralization index, which is defined as

$$\gamma = \sum_{i=1}^7 [\text{In}_{\text{Deg}} L_i + \text{Out}_{\text{Deg}} L_i] / \sum_{j=1}^7 [\text{In}_{\text{Deg}} R_j + \text{Out}_{\text{Deg}} R_j] \quad (39)$$

where  $\text{In}_{\text{Deg}} L_i$  indicates the in-degree for node  $i$  in the left hemisphere and  $\text{In}_{\text{Deg}} R_j$  indicates the in-degree for node  $j$  in the right hemisphere. Similarly,  $\text{Out}_{\text{Deg}} L_i$  indicates the out-degree for node  $i$  in the left hemisphere, and  $\text{Out}_{\text{Deg}} R_j$  indicates the out-degree for node  $j$  in the right hemisphere. Accordingly,  $\gamma$  with value smaller than one indicates left lateralization, while values of  $\gamma$  larger than one indicate right lateralization. The closer to one that the  $\gamma$  values are, the more symmetric the network patterns are.

The second index is the ratio of network difference between different classes ( $ND_{\text{BDC}}$ ) and network difference within the same classes ( $ND_{\text{WSC}}$ ). Theoretically, the network difference is the inverse of SimiNet, which was proposed by Mheich [31]. Its detailed calculation process can be found in [31]. In the current work, we name the ratio between  $ND_{\text{BDC}}$  and  $ND_{\text{WSC}}$  as RDiffNet and adopt the following steps to estimate this index.

*Step-1*: Constructing the paired networks by bootstrapping.

*Step-2*: Eliminating the isolated nodes in the brain network.

*Step-3*: Using the steps proposed in [31], adjusting the node structure of paired networks, i.e.,  $G_1$  and  $G_2$ , and recording related cost  $NE$  (the radius was set as 1.5, according to the operation in [31]).

*Step-4*: Estimating the edge difference between  $G_1$  and  $G_2$  as

$$\text{LE} = \sum_i^D \sum_j^D |G_1(i, j) - G_2(i, j)|. \quad (40)$$

*Step-5*: Calculating the RDiffNet as

$$\text{RDiffNet} = ND_{\text{BDC}} / ND_{\text{WSC}}, \quad ND = NE + \text{LE} \quad (41)$$

through the definition of RDiffNet, we can see that the larger that the RDiffNet is, the more different that the topologies of paired networks between different classes are, and the more similar that the topologies of paired networks within the same classes are. Thus, RDiffNet actually measures the discriminability of network structures as features for MI EEG classification.

4) *EEG Classification Based on Brain Networks*: In recent years, using brain network topologies for EEG classification has gained much attention in the field of BCI [32]. Notably, their reliability mainly depends on the accurate estimation of cooperative patterns among multiple brain regions [3]. Thus, to further validate the effectiveness of DLap-GCA for EEG classification and compare its performance with other



GCAs, we designed EEG-based MI classification experiments. In addition, we compared the performance difference between EEG networks and EEG activations characterized by power spectral density (PSD) and differential entropy (DE) [32]. As spatially filtered technologies such as CSP have been proven to be successful for MI classification, we adopted CSP and its state-of-the-art improvements [21], [22], [23] for further comparisons. Specifically, the regularization CSP (R-CSP) and regularization CSP with feature aggregation (R-CSP-A) mentioned in [21], the Tikhonov regularization CSP (TR-CSP) proposed in [22], and the filter-bank CSP (FB-CSP) for decoding the directional focus of attention proposed in [23] were utilized for further comparisons. Their detailed information is provided in [21], [22], and [23].

In this study, to verify the generalization power of our proposed method, we adopted three MI EEG datasets, i.e., 1) MI EEGs mentioned in Section II-E2.1 (Dataset-1) [28]; 2) MI EEGs from BCI Competition IV dataset I (Dataset-2, sampling rate: 1000 Hz) [33]; and 3) the MI EEGs from BCI Competition III dataset IVa (Dataset-3, sampling rate: 1000 Hz) [34]. To simplify the comparison, we adopted the same 15 channels with Dataset-1 for the other datasets. Subsequently, the EEG segments between 0.5 and 2.5 s after the cue were extracted for analysis. More detailed information about the public MI EEG datasets can be found in the Supplementary Material.

During the classification experiments, all of the matrices of the brain network estimated from bandpass-filtered EEGs (BPF-EEGs) were reshaped into vectors as features for MI EEG classification. For the features of brain activity, we estimated the average PSD and DE from BPF-EEGs. For the spatially filtered EEG features, we adopted the same processing as reported in [21], [22], and [23], and the number of spatial filters was defined as 3. The support vector machine (SVM) with linear kernel was adopted as the classifier [35]. To alleviate the randomness of the classification results, a ten-fold cross-validation (CV) scheme with randomization was utilized. The ten-fold CV was repeated ten times, and the average classification accuracy was used as the evaluation index.

##### 5) EEG Brain Networks for Online BCI Applications:

In recent years, MI-BCI systems have been widely applied in various fields, such as motor rehabilitation and the development of motor skills [36]. However, in practical applications, EEG signals are often susceptible to external interferences, such as motor, electromyography, and eye blinking [3], [14], [37], making them too weak to offer reliable information. In addition, feature extraction and classification methods that are robust to these interferences have rarely been reported [3], [36]. These factors increase the setup time [38] and further affect the long-term stability of BCIs [39]. Therefore, accurately identifying the motor intention of the subjects is one of the greatest challenges for enhancing the interactive experience in an online MI-BCI system [36]. To verify the feasibility of the proposed method for possible application in online MI BCI systems, we designed an online experiment based on hand MI; its flowchart is illustrated in Fig. 3.

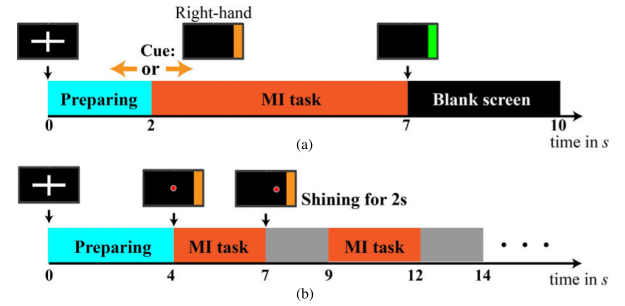


Fig. 3. Flowchart of online BCI experiments. (a) Procedure of the training stage in one trial and (b) procedure of the testing stage in one run.

The entire online experiment consisted of two stages: the training stage and online testing stage. In the training stage, there were a total of 100 trials, with half of them corresponding to the left-hand MI task. At the beginning of each trial, a white cross appeared in the middle of the screen for 2 s indicating to the subject to prepare for the MI task. Then, an orange bar appeared on the left or right side of the screen for 5 s instructing the subject to perform the left-hand or right-hand MI task. When the bar turned green, the subjects stopped the MI task and rested for 3 s. The EEG signals recorded in the training stage were utilized to estimate brain networks and train classifiers. For the online testing stage, there were six runs, among which three runs corresponded to the left-hand MI task, and the trials in the same run corresponded to the same task. For each run, the maximum number of trials was 10. At the beginning of each run, a white cross appeared in the middle of the screen for 4 s asking the subject to prepare for the related task. Then, a red ball appeared in the middle of the screen and an orange bar appeared on the left or right side of the screen to indicate to the subjects to perform corresponding MI. At the beginning of each trial, the red ball shined for 2 s, asking the subject to perform left- or right-hand MI for 3 s according to the location of the orange bar. At the end of each trial, the red ball moved one step orienting to the orange bar or its opposition according to the classification result of the recorded EEGs. Specifically, the horizontal movements of the red ball can be estimated as

$$\Delta x = m_h(t) \times v_x \quad (42)$$

where  $m_h(t)$  denotes the output of SVM at time  $t$ , with  $m_h(t) = 1$  being the left-hand MI and  $m_h(t) = -1$  being the right-hand MI, and  $v_x$  is the movement velocity in the horizontal direction. Subsequently, the horizontal position at time  $t + 1$  can be updated as

$$p_x(t + 1) = p_x(t) + \Delta x. \quad (43)$$

At the end of the last trial, if the ball arrived at the orange bar, a smiling face appeared in the middle of the screen. For both the training stage and online testing stage, the order of the left and right bars was randomized. In our current study, we recruited four subjects (age:  $22.50 \pm 1.29$  years old), and each subject completed six runs with a 5-s interval between the two online testing runs. For each subject, we recorded the correct steps moved by the red ball in every single block. The EEG signals of the online experiment were recorded by a Neuracle amplifier (Neural Wireless EEG/event-related poten-



TABLE I

CB A ND NCRL WITH INCREASING OCCURRENCE FREQUENCIES OF OUTLIERS

Methods	Frequency of outliers									
	0.2%	0.3%	0.4%	0.5%	0.6%	0.2%	0.3%	0.4%	0.5%	0.6%
	Index values of CB					Index values of NCRL				
Gau-GCA	0.87 ±0.35*	0.89 ±0.38*	0.96 ±0.41*	0.97 ±0.42*	1.02 ±0.44*	18.41 ±1.57*	18.18 ±1.56*	17.85 ±1.57*	17.67 ±1.61*	17.45 ±1.65*
Lap-GCA	0.73 ±0.37*	0.73 ±0.40*	0.82 ±0.43*	0.83 ±0.44*	0.89 ±0.46*	18.48 ±1.51*	18.21 ±1.53*	17.86 ±1.56*	17.68 ±1.61*	17.49 ±1.64*
GLap-GCA	0.71 ±0.22*	0.72 ±0.27*	0.79 ±0.29*	0.81 ±0.31*	0.86 ±0.34*	19.60 ±0.71	19.45 ±0.87	19.34 ±0.90	19.19 ±0.97*	18.97 ±1.23*
DLap-GCA	<b>0.62</b> ±0.22	<b>0.64</b> ±0.26	<b>0.72</b> ±0.29	<b>0.74</b> ±0.30	<b>0.79</b> ±0.34	<b>19.71</b> ±0.53	<b>19.60</b> ±0.62	<b>19.46</b> ±0.75	<b>19.34</b> ±0.84	<b>19.16</b> ±1.07

TABLE II

CB A ND NCRL WITH INCREASING OUTLIER STRENGTHS

Methods	Strength of outliers									
	2	3	4	5	6	2	3	4	5	6
	Index values of CB					Index values of NCRL				
Gau-GCA	0.73 ±0.26*	0.88 ±0.37*	0.93 ±0.41*	0.93 ±0.45*	1.04 ±0.46*	19.44 ±0.81*	18.44 ±1.64*	17.81 ±1.75*	17.32 ±1.77*	17.30 ±1.78*
Lap-GCA	0.59 ±0.28*	0.74 ±0.38*	0.79 ±0.43*	0.89 ±0.46*	0.92 ±0.48*	19.46 ±0.76*	18.48 ±1.59*	17.83 ±1.74*	17.35 ±1.75*	17.31 ±1.77*
GLap-GCA	0.61 ±0.16*	0.70 ±0.22*	0.74 ±0.26*	0.82 ±0.29*	0.83 ±0.31*	19.74 ±0.54	19.54 ±0.84	19.37 ±0.99	19.18 ±1.07*	18.97 ±1.22*
DLap-GCA	<b>0.52</b> ±0.16	<b>0.62</b> ±0.22	<b>0.65</b> ±0.26	<b>0.73</b> ±0.29	<b>0.74</b> ±0.30	<b>19.79</b> ±0.47	<b>19.59</b> ±0.74	<b>19.49</b> ±0.93	<b>19.33</b> ±0.97	<b>19.16</b> ±1.20

TABLE III

CB AND NCRL WITH INCREASING NUMBERS OF OUTLIER CONTAMINATED TIME SERIES

Methods	Number of influenced channels									
	1	2	3	4	5	1	2	3	4	5
	Index values of CB					Index values of NCRL				
Gau-GCA	0.70 ±0.34*	0.87 ±0.38*	1.11 ±0.36*	1.22 ±0.39*	1.38 ±0.42*	19.30 ±1.18*	18.18 ±1.71*	17.25 ±1.83*	17.12 ±1.68*	16.72 ±1.75*
Lap-GCA	0.56 ±0.36	0.73 ±0.37*	0.98 ±0.40*	1.10 ±0.40*	1.27 ±0.41*	19.33 ±1.15*	18.21 ±1.70*	17.33 ±1.79*	17.14 ±1.57*	16.79 ±1.72*
GLap-GCA	0.61 ±0.22*	0.72 ±0.25*	0.90 ±0.26*	1.00 ±0.26*	1.13 ±0.27*	19.72 ±0.59	19.45 ±0.87	19.07 ±1.00*	18.82 ±1.06*	18.33 ±1.05*
DLap-GCA	<b>0.52</b> ±0.22	<b>0.64</b> ±0.25	<b>0.82</b> ±0.25	<b>0.93</b> ±0.26	<b>1.04</b> ±0.26	<b>19.75</b> ±0.59	<b>19.60</b> ±0.72	<b>19.24</b> ±0.95	<b>19.03</b> ±0.98	<b>18.69</b> ±0.95

tial (ERP)), of which the online sampling rate was 250 Hz, and the bandpass filter held the passband from 0.5 to 70 Hz.

### III. RESULTS

#### A. Simulation Study

The results of the simulation study are integrated in Tables I–III, where the values with bold symbols indicate the best result, and “\*” indicates a significant difference between the proposed DLap-GCA and the other GCAs ( $p < 0.05$ ).

1) *Influence of Occurrence Rates*: To analyze the influence of the outlier occurrence rate, we increased the outlier occurrence rate from 0.2% to 0.6% with a step size of 0.1%. The number of contaminated time series was set to 2, and the outlier intensity was defined as three times the maximum amplitude of the corresponding time series. The coefficient bias (CB) and the NCRLs in this experiment are integrated in Table I.

The CBs in Table I indicate that the bias errors for MVAR coefficients estimated by all of the GCAs are increased when the signals are contaminated with more outliers. However, the MVAR coefficients estimated by DLap-GCA are closer to those in (38) than those estimated by the other methods. These

“better” coefficients further result in more consistent network structures with the predefined one as illustrated in Table I.

2) *Influence of Outlier Intensities*: To investigate the influence of outlier intensities, we increased the outlier intensity from 2 to 6 with a step size of 1. The number of contaminated time series was kept at 2, and the outlier occurrence rate was kept at 0.2%. The experimental results are integrated in Table II.

Similar to the results in Table I, the CBs in Table II show that the MVAR coefficients estimated by DLap-GCA are still closer to those in (38) than those estimated by the other methods, which result in more correct network structures, as illustrated in Table II.

3) *Influence Caused by the Proportions of the Time Series That Are Contaminated by Outliers*: Notably, the experiments described in Sections III-A1 and III-A2 primarily focused on the situation, in which the outliers occurred in two time series. To explore the robustness of DLap-GCA under the conditions that more time series are contaminated by outliers, we increased the number of time series contaminated by outliers from 1 to 5, with a step size of 1. The outlier occurrence rate was fixed at 0.3%, and their intensities were set as  $3 \cdot X_{\max}$ . The experimental results are listed in Table III.

Compared to the results in Tables I and II, we observed that the performance of each GCA for solving the MVAR model was more influenced. In fact, the bias of MVAR coefficients increases when more time series are contaminated by outliers. However, the bias of MVAR coefficients estimated by DLap-GCA remains closest to the predefined coefficients, which improves the accuracy of the network structures, as illustrated in Table III. Based on Tables I–III, we discovered that the networks estimated by DLap-GCA hold more consistent linkages with the predefined networks. Thus, to quantitatively evaluate whether these linkages can represent the inherent network structure predefined in (38), we further calculated the confusion matrix based on their connection directions.

Consistent with Tables I–III, the confusion matrices demonstrated in Fig. 4 further show that the DLap-GCA can almost recover the predefined network pattern under various outlier conditions, in which the values on the diagonal of the confusion matrices are larger than those of the other GCAs.

In essence, both the smallest CBs and the largest numbers of consistent linkages could indicate that DLap-GCA can better compress the outlier effect and capture the intrinsic varying trend of time series. Fig. 5 intuitively demonstrates the underlying mechanism accounting for the performance differences among the four approaches. As shown in Fig. 5, both GLap- and DLap-GCA can restrain the influence of outliers and focus on fitting the original components in the time series, while Gau- and Lap-GCA emphasize the outliers more and fail to capture the intrinsic information of the signals. When comparing GLap- and DLap-GCA, we find that some impulse components can also be observed in the predicting time series of GLap-GCA compared to that of DLap-GCA. These results indicate that DLap-GCA can better compress the outlier effect and capture meaningful information in the time series.

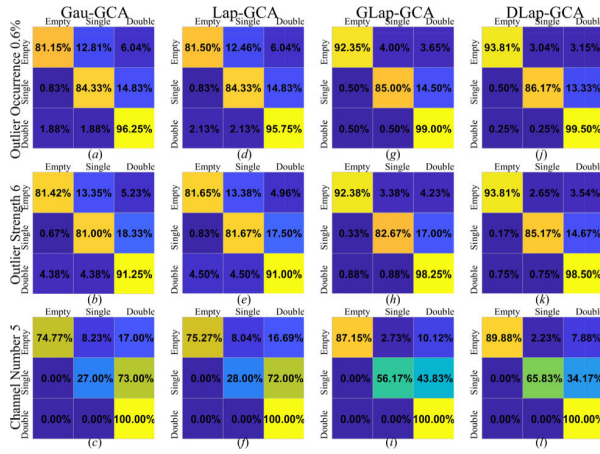


Fig. 4. Confusion matrices of different GCAs under different outlier conditions for 200 times. (a)–(c) Gau-GCA under different outlier conditions. (d)–(f) Lap-GCA under different outlier conditions. (g)–(i) GLap-GCA under different outlier conditions. (j)–(l) DLap-GCA under different outlier conditions.

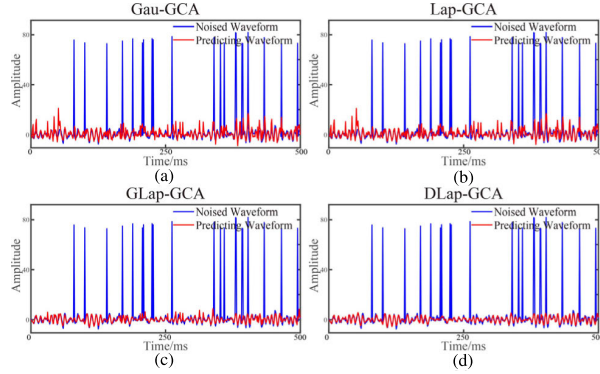


Fig. 5. Fitting performances of the four GCAs. (a) Gau-GCA. (b) Lap-GCA. (c) GLap-GCA. (d) DLap-GCA. Blue lines depict the noised time series and red lines depict the predicted time series.

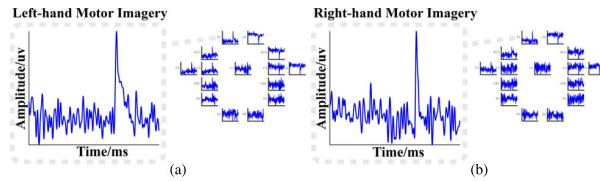


Fig. 6. EEGs of MI used for the Granger analysis. (a) Right-hand MI EEG. (b) Left-hand MI EEG.

### B. Real EEG Applications

In this section, we mainly focus on evaluating the different performances of all the aforementioned GCAs (i.e., Gau-, Lap-, GLap-, and DLap-GCAs) when applying them to estimate directed brain networks for MI EEG datasets that contain obvious ocular artifacts. Fig. 6 shows one pair of EEG segments utilized for network estimation.

1) *Contralateral Network Patterns for MI EEGs*: To visually demonstrate the network patterns estimated by different GCAs under the influence of ocular artifacts, we averaged the brain networks of all the trials [3]. In fact, due to the variability of EEGs, the brain networks estimated from EEG segments of different trials could hold some different linkages [3], [14]. When these different linkages are accumulated together, it could result in very dense connection patterns. Thus, we set the connections with weights smaller than the threshold to 0. Specifically, we calculated the lateralization

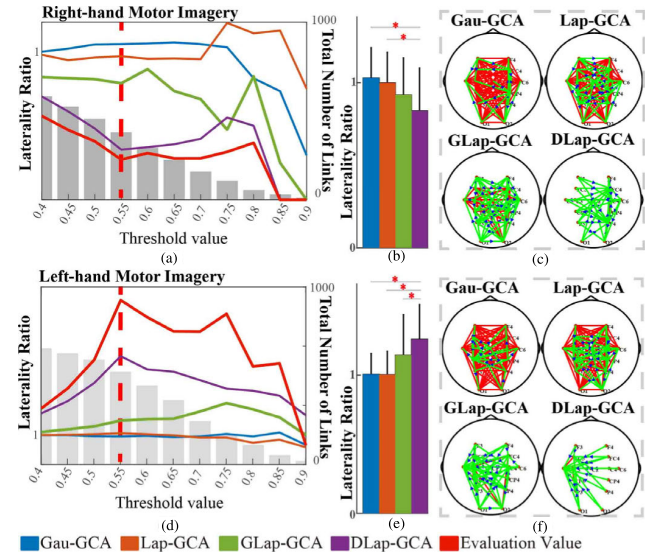


Fig. 7. Trend of the laterality ratio with increasing threshold. (a) Trend of the laterality ratio for the left-hand MI, (b) lateralization patterns of networks estimated by different GCAs with a threshold of 0.55, (c) brain networks of left-hand MI under the threshold of 0.55, (d) trend of the laterality ratio for right-hand MI, (e) lateralization patterns of networks estimated by different GCAs with a threshold of 0.55, and (f) brain networks of right-hand MI under the threshold of 0.55.

indices of these averaged networks under different occurrence frequencies and designed the following indexes to evaluate the performance of different thresholds:

$$\text{Com}L_i = \prod_{me \in Me} AL_i^{me} \quad (44)$$

where  $AL_i^{me}$  represents the average-network laterality calculated by GCAs under the  $i$ th threshold, and  $me$  is the set of methods, i.e., Gau-GCA, Lap-GCA, GLap-GCA, and DLap-GCA.

Fig. 7(a) and (d) illustrates the lateralization indices, through which we observe that the linkage threshold ranges between 0.5 and 0.7 can yield optimal lateralization indices for all the GCAs. Thus, we mainly illustrated the network topologies and  $\gamma$  values under the threshold of 0.55. In Fig. 7(b) and (e), we observed that the contralateral laterality revealed by DLap-GCA is superior to that of the other GCAs, with the smallest  $\gamma$  values for left-hand MI tasks and the largest  $\gamma$  values for right-hand MI tasks. Physiologically, MI EEGs are usually accompanied by strong ERS and ERD phenomena that could be characterized by lateralized network patterns [3], [28], with the electrodes distributed around C3 and C4 acting as the center hub for right-hand MI tasks and left-hand MI tasks [3], [14], respectively. However, as shown in Fig. 7(c) and (f), when the EEG is contaminated with ocular artifacts, the brain neural networks estimated by both Gau-GCA and Lap-GCA from noisy EEG segments lose their lateralization patterns. In contrast, both GLap-GCA and DLap-GCA preserve strong lateralization patterns in the recovered networks for right-hand MI tasks and left-hand MI tasks, respectively. Furthermore, compared with the results of GLap-GCA, DLap-GCA reveals that the center hub nodes surrounding both C3 and C4 for left-hand MI tasks and right-hand MI tasks appeared more centralized, respectively.

TABLE IV  
COMPARISON RESULTS OF RDIFNET

	Gau-GCA	Lap-GCA	GLap-GCA	DLap-GCA
RDifNet	1.22±0.29*	1.21±0.22*	1.86±0.86*	<b>2.03±0.71</b>

Consistently, through the values of RDifNet integrated in Table IV, we observed that the brain networks corresponding to different MI tasks estimated by DLap-GCA have the largest RDifNet value, which indicates that DLap-GCA can effectively characterize discriminable network structures corresponding to different cognition states, while preserving the intrinsic network structure for the same cognition state.

In addition to these two indices, we adopted three commonly used network characteristics, i.e., the local efficiency (LE) [28], the accumulation coupling strength (ACS) [40], and the local clustering coefficient (LCC) [28], to assess the contralateral patterns of brain networks estimated by different GCAs. Through these indices, we also observed that the brain networks estimated by DLap-GCA hold contralateral connection patterns, where the brain regions close to C3 play important roles in the right-hand MI task, while the brain regions close to C4 are important in the left-hand MI task. More detail information about this result could be referred in the Supplementary Materials.

#### 2) MI EEG Classification Using Brain Causal Networks:

##### a) MI classification performances of different methods:

The classification results of each dataset are shown in Table V, where \* indicates that the proposed approach has significant performance improvement ( $p < 0.05$ ). As shown in Table V, we found that the classification accuracy of the graph topologies characterized by DLap-GCA is better than those estimated by other GCAs and better than the features extracted by spatial weighted methods, such as PSD and CSP. Compared with other state-of-the-art MI EEG feature extraction methods, such as CSP improvements [21], [22], [23], our proposed method shows competitive performance.

In addition to the comparative methods utilized in classification experiments, there have been abundant novel methods for MI EEG classification [41], [42], [45], [47], [48]. Table VI mainly integrates recent improvements using brain topological features for MI EEG classification. Through these improvements, we observed that most of them focus on extracting the features of spatially filtered brain activations from EEG signals (see [7], [41], [42], [43], [44], [45]) and that some of them tried to enhance the robustness of the spatial filter by introducing outlier robust operators, such as the L1-norm (see [45]). In addition to the features of brain activity, EEG classification based on the topology of brain networks has gradually been applied. However, compared with brain activation, there have been fewer reports about using brain network topologies for MI EEG classification (see [29], [46]). These studies mainly utilized functional brain networks estimated by technologies, such as phase locking value and correlation for MI EEG classification, and none of them utilized directed brain networks. Compared with these studies, our proposed DLap-GCA offers outlier-robust network structures for EEG classification, which not only enhances the discriminative power of brain network topologies for EEG classification in outdoor environments but

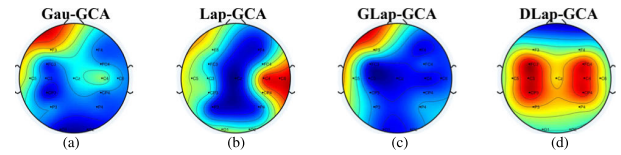


Fig. 8. Importance topologies of network nodes for classification. Node contribution of brain network estimated by (a) Gau-GCA, (b) Lap-GCA, (c) GLap-GCA, and (d) DLap-GCA.

also contains more valuable information about the interaction patterns among multiple brain regions.

##### b) Interpretable network topologies for MI classification:

In terms of EEG classification, the neural interpretability of features is also important in the field of neural science research [3], [32], [49]. Thus, we explored the neural interpretability of optimal features extracted from brain network topologies for MI EEG classification. Specifically, the method utilized in [49], which combines recursive feature elimination and SVM (SVM-RFE) for feature selection and analyzes the contribution of each feature, was utilized to estimate the importance of each brain linkage for MI EEG classification among the entire dataset. To scale the importance of each linkage into [0, 1], we further conducted min-max normalization. For each node, we calculated the frequency of its linkages being selected for classification. Fig. 8 illustrates the probability distribution of the features of each node being selected for classification, where red indicates that the linkages of the corresponding node are mostly selected. As shown in Fig. 8, we observed that the frequently selected nodes of connection patterns estimated by DLap-GCA were mainly located around the motor sensory regions that play important roles in MI tasks. This observation is consistent with the findings in various studies about MI [28], [50]. However, the optimal nodes of network topologies estimated by other GCAs are distributed in other regions, such as frontal areas, which may cause ambiguous interpretations for MI tasks [3], [5].

#### C. EEG Brain Networks for Online BCI Applications

Table VII lists the averaged online classification accuracies of all runs of four subjects, from which we observe that the lowest classification accuracy is 80.45% and that the highest classification accuracy is 90.24%. Moreover, the online classification accuracies for all subjects are larger than 80%. Based on these results, we further calculated the information transfer ratio (ITR) as an index to assess the effectiveness of DLap-GCA in online BCI applications [51]. Table VIII presents the ITRs of the four subjects, ranging from 6.05 to 13.12. According to previous studies [52], [53], [54], these results are acceptable for online BCI control, revealing that the brain-network topologies estimated by DLap-GCA may be a promising candidate for online BCI applications.

## IV. DISCUSSION

In this work, we proposed imposing a Laplacian distribution on both observation noise and model coefficients to improve the performance of GCA in both noise resistance and locating important nodes in the network. To verify the robustness of our proposed method, we conducted a simulation study that



TABLE V  
MI CLASSIFICATION ACCURACIES (%) FOR EACH SUBJECT

		PSD	DE	CSP	R-CSP	R-CSP-A	TR-CSP	FB-CSP	Gau-GCA	Lap-GCA	GLap-GCA	DLap-GCA
<i>Dataset-1</i>	<i>Our</i>	65.83 ±17.14*	82.83 ±13.50	76.00 ±19.44	77.67 ±18.38	<b>87.33</b> ±12.43	71.50 ±16.63*	78.67 ±15.73	48.00 ±20.28*	46.00 ±19.55*	68.00 ±18.90*	79.50 ±19.43
	<i>ds1a</i>	52.05 ±2.76*	63.00 ±9.77	61.30 ±11.41	58.55 ±10.28*	68.15 ±11.00	63.20 ±11.27	64.50 ±10.21	50.90 ±8.21*	45.75 ±8.60*	57.55 ±12.86*	<b>68.90</b> ±9.99
	<i>ds1b</i>	53.95 ±10.35	48.00 ±9.32*	48.50 ±7.16*	48.25 ±10.50*	51.30 ±8.75	<b>61.00</b> ±10.30	57.50 ±10.38	52.70 ±7.80	50.30 ±5.40*	45.20 ±10.27*	57.10 ±10.33
	<i>ds1c</i>	58.25 ±9.28	54.60 ±11.41	53.90 ±11.16	51.50 ±9.68	55.00 ±10.94	<b>60.25</b> ±10.33	58.65 ±10.12	45.20 ±10.94*	47.15 ±9.33*	48.70 ±8.28*	54.75 ±6.17
<i>Dataset-2</i>	<i>ds1d</i>	48.40 ±8.67*	51.55 ±8.58*	49.80 ±10.94*	46.45 ±10.78*	46.10 ±10.34*	47.00 ±12.02*	48.45 ±10.56*	47.45 ±9.99*	53.15 ±12.75*	64.45 ±10.17*	<b>72.70</b> ±7.90
	<i>ds1e</i>	50.65 ±6.50	50.00 ±9.92	47.65 ±9.44	49.35 ±11.67	48.30 ±12.29	53.20 ±10.34	44.95 ±10.60	47.85 ±10.11	<b>56.30</b> ±9.86	51.45 ±10.85	50.45 ±9.46
	<i>ds1f</i>	59.25 ±8.97	54.50 ±10.77	55.05 ±11.58	58.00 ±11.70	60.90 ±10.67	61.10 ±10.77	<b>66.50</b> ±10.74	46.95 ±8.99*	49.00 ±8.82*	49.75 ±10.38*	56.70 ±9.85
	<i>ds1g</i>	51.50 ±7.74*	<b>68.45</b> ±8.70	63.70 ±10.19	59.80 ±9.77	67.60 ±9.78	62.95 ±10.64	66.15 ±9.64	55.85 ±9.80*	49.25 ±9.06*	55.95 ±10.93	62.15 ±9.43
	<i>aa</i>	52.18 ±3.86*	62.14 ±8.07	57.79 ±9.53*	58.07 ±9.98*	<b>68.39</b> ±7.36	55.61 ±8.14*	65.39 ±9.44	54.68 ±8.57*	52.04 ±8.75*	59.64 ±9.09	66.14 ±7.71
	<i>al</i>	61.71 ±7.02*	85.18 ±6.41	69.57 ±9.26*	72.61 ±10.24*	85.46 ±6.95	59.89 ±7.48*	80.50 ±8.06*	63.14 ±8.64*	52.21 ±8.32*	79.57 ±6.89*	<b>88.36</b> ±5.91
<i>Dataset-3</i>	<i>av</i>	54.75 ±8.75*	57.57 ±7.52*	60.21 ±8.76	59.79 ±7.89*	64.79 ±9.29	61.46 ±7.92	63.18 ±7.73	51.89 ±4.55*	44.00 ±8.21*	63.18 ±8.61	<b>65.93</b> ±8.80
	<i>aw</i>	59.25 ±7.93*	72.32 ±8.13	61.61 ±9.56*	59.68 ±8.98*	<b>72.25</b> ±9.09	63.21 ±7.90*	68.14 ±8.70	59.21 ±8.05*	49.46 ±8.35*	70.25 ±7.37	70.14 ±7.10
	<i>ay</i>	63.86 ±8.24*	81.89 ±6.56	73.43 ±11.32	77.00 ±9.25	<b>83.86</b> ±6.00	59.96 ±9.73*	66.21 ±7.31*	49.50 ±3.93*	51.21 ±8.78*	63.11 ±8.79*	78.75 ±6.53
MEAN		56.28 ±8.25	64.00 ±9.13	59.89 ±10.75	59.75 ±10.70	66.11 ±9.61	60.03 ±10.27	63.75 ±9.94	51.79 ±9.22	49.68 ±9.68	59.75 ±10.26	<b>67.04</b> ±9.12

TABLE VI  
NOVEL IMPROVEMENTS OF EEG FEATURES FOR MOTOR IMAGERY CLASSIFICATION

Study	Proposed methods	Feature type	Datasets and classification accuracy	Improvements	Training free	Automatic artifacts reduction
[41]	Correlation-based channel selection RCSP(CRCSP)	Spatial filtered brain activations	1.BCI IV I (81.6%); 2.BCI III Iva (87.4%); 3.BCI III IIIa (91.9%).	Automatic channel selection	×	×
[42]	Frequency-optimized local region common spatial pattern (LRFCSPP)	Spatial filtered brain activations	1.BCI III Iva (92.93%); 2.BCI IV I (84.7%); 3.BCI IV IIb (82.12%)	Solving small sample problem	×	×
[43]	Correlation-based common spatial pattern (CCSP)	Spatial filtered brain activations	1.BCI III IIIa (83.84%), 2.BCI IV IIa (71.54%)	Solving multiclass problem	×	×
[7]	Filterbank spatial filtering + temporal-spatial convolutional neural network (FBSF-TSCNN)	brain activation along time-frequency domain	BCI IV IIa (72%)	Introducing the information of temporal information into spatial patterns	×	×
[44]	Sparse CSP +transfer CSP +multisubject CSP	Spatial filtered brain activations	1.BCI III IIIa (81.38%) 2.BCI III Iva (78.23%) 3.BCI IV IIa (79.53%)	Transfer learning, Sparse learning Cross subjects problems	×	√
[45]	L1-norm+feature difference maximization + Dempster-Shafer theory + CSP (DRL1-CSP)	Spatial filtered brain activations	1.BCI III Iva (85.1%), 2.BCI IV I (68.4%)	Suppressing outliers and discovering features with larger interclass distances	×	√
[46]	sLORETA for source location + PLV for brain network construction + graphical feature extraction	Brain network structure	EEG data recorded in their own lab (67.13%)	Using brain graphic features for EEG classification	√	×
[29]	Coherence analysis for brain network construction + graphical feature extraction	Brain network structure	EEG data recorded in their own lab (70% to 80%)	Using the brain network structures in MI EEGs	√	×

consisted of three different outlier conditions. Tables I–III consistently reveal that the MVAR coefficients estimated by DLap-GCA show the smallest bias ( $p < 0.05$ ) compared to

the other three methods. These smallest CBs subsequently produced more reliable network patterns, which held a larger NCRLs, as shown in Tables I–III. The confusion matrices



TABLE VII

ONLINE MI CLASSIFICATION ACCURACIES (%) FOR EACH SUBJECT					
	Sub-1	Sub-2	Sub-3	Sub-4	MEAN
DLap-GCA	85.45 ±8.12	82.14 ±16.75	90.24 ±12.14	80.45 ±6.43	84.57 ±10.86

TABLE VIII

ONLINE MI ITR (BIT/MIN) FOR EACH SUBJECTS					
	Sub-1	Sub-2	Sub-3	Sub-4	MEAN
DLap-GCA	8.99 ±5.65	8.78 ±6.40	13.12 ±7.83	6.05 ±2.47	9.23 ±5.59

illustrated in Fig. 4 intuitively demonstrate the topological differences of networks estimated by different GCAs, by which we observe that the causal interactions captured by GLap-GCA and DLap-GCA are more consistent with the predefined network than with the other GCAs. In fact, the reliable network structure estimated by DLap-GCA can also be attributed to its superior power of capturing the inherent varying trend in time series, as shown in Fig. 5. It is obvious that both Gau-GCA and Lap-GCA capture the components of outliers and introduce them in signal fitting, while GLap-GCA and DLap-GCA restrain these components and preserve the more expected information. Interestingly, although both GLap-GCA and DLap-GCA hold similar power in outlier resistance, the performance of DLap-GCA is superior to that of GLap-GCA. Focusing on the differences between the confusion matrices of DLap-GCA and GLap-GCA, we observed that the diagonal values in the confusion matrix of DLap-GCA appear more dominant than those of GLap-GCA. This phenomenon indicates that a more inherent topological structure can be inferred by DLap-GCA than by GLap-GCA, while the networks estimated by GLap-GCA preserve more linkages that are not consistent with the predefined connection pattern. This outcome can be attributed to the dual Laplacian structure utilized for network estimation. In essence, this structure attempts to search sparse parameters in the condition that the observation noise holds a heavy-tailed character, resulting in not only the robustness of network inference to outliers but also a sparse structure with crucial connections preserved. As illustrated in Fig. 5, the superior fitting result of DLap-GCA over GLap-GCA in return explains the better network recovery performance of DLap-GCA. This finding can also be attributed to the dual Laplacian structure utilized in our proposed method. Theoretically, by assuming that the distribution of the model coefficient follows a Laplacian distribution, we imposed a sparse restriction on the MVAR parameters to smoothen the predicted signals, which further emphasizes the inherent variation trend of time series and improves the performance of GLap-GCA for causal inference among various time series.

For real EEG applications, one crucial factor for GCA is to accurately solve the MVAR model. Consequently, it is necessary to use pre-experimental data to estimate the noise and carefully check whether it follows a Gaussian distribution [55]. Alternatively, to meet this assumption, one common method is to average many EEG segments [28], [55], which may not satisfy the timeliness for applications, such as online BCI systems [56]. Considering that the EEG segments

contaminated with eye blink artifacts usually present a heavy-tailed distribution [3], [57], we proposed DLap-GCA, which adopts an outlier-robust distribution for both model noise and coefficients. Its robustness to outliers has been quantitatively validated by simulation experiments.

To quantitatively evaluate the effectiveness of DLap-GCA for real EEG applications, we adopted MI EEGs, which are usually accompanied by obvious contralateral network patterns [5], and designed a lateralization index to measure this character. As shown in Fig. 7(c) and (f), we observed that the network patterns estimated by both Gau-GCA and Lap-GCA fail to capture this important character with similar values close to 1 for both left- and right-hand MI tasks, as shown in Fig. 7(b) and (e). This outcome can be attributed to some pseudocausal connections, violating the inherent contralateral patterns in MI EEGs [3], [14]. In fact, these pseudocausal connections can be attributed to strong synchronous oscillations of ocular artifacts [3], [5], [14]. Methods, such as Gau-GCA and Lap-GCA, cannot restrict ocular artifacts and capture their strong synchronous oscillations, resulting in symmetric network patterns. These symmetric patterns further decrease the discriminability of networks and lead to the lower RDiffNet values in Table IV. In comparison with Gau-GCA and Lap-GCA, both GLap-GCA and DLap-GCA can still characterize contralateral networks with their lateralization indices, showing the great difference between two different MI tasks. Notably, the hub nodes of MI networks inferred by DLap-GCA are concentrated more in the motor-sensory regions (C3 and C4) than those estimated by GLap-GCA. This finding further reveals the powerful capability of DLap-GCA in identifying the hub nodes of networks for cognitive processing, which could be attributed to the Laplacian distribution imposed on MVAR coefficients.

Consistent with the network patterns, the conducted MI EEG classification experiments also revealed the efficiency and stability of DLap-GCA for characterizing the identifiable network structures from EEGs referring to different MI tasks [3], [14]. In fact, the critical step for MI EEG-based BCI systems is to extract discriminative brain activities of the corresponding cognitive state. Usually, these brain activities can be described with energy distribution by methods, such as PSD [32], and their differences can be further emphasized by spatial filter methods, such as CSP, to improve the classification performance, as shown in Table V. In essence, these methods focus on characterizing the power distribution differences of local brain areas, which disregards the complex cooperation of multiple brain regions [3], [28]. Different from these methods, the brain network topologies estimated by DLap-GCA characterize the information interaction patterns in the brain [3], [14], [32] and hold superior classification performance. However, the classification performances of the other GCAs are poor, which can be explained by their lower RDiffNet values, as shown in Table IV. These results indicate that the network topologies may offer more efficient classification information if they can capture intrinsic patterns for cognition states. By analyzing the feature interpretability of network topologies in Fig. 8(d), we further observed that the frequently selected linkages estimated by DLap-GCA are

mainly located around the motor sensory regions, which play crucial roles in MI tasks [14], [28], [29]. However, the optimal nodes of network topologies estimated by other GCAs are distributed in less related regions, such as frontal areas, which could lead to a confusing explanation for MI tasks [3], [5].

Note that in addition to the methods used in our comparison experiments, many other state-of-the-art studies have focused on improving the classification performance of MI EEGs [41], [44], [45], [58]. Compared with these improvements, the purpose of DLap-GCA is to capture the inherent network patterns from EEGs under the influences of complex noises and to highlight the crucial nodes in brain networks. Essentially, the efficiency and stability of DLap-GCA can be attributed to the following merits: 1) it is robust to complex artifact noises when estimating brain networks, which guarantees the capability of characterizing intrinsic connection patterns of cognition processing [3], [5]; 2) it sparsely characterizes the network patterns, which highlights the crucial regions for the information processing of the brain and facilitates related analysis, such as source electrode identification [59]; and 3) it belongs to training free technologies [60], which will not be influenced by problems, such as small samples [21] and the setting of hyperparameters [21], [22], [23].

In addition to offline analysis, the online experimental results illustrated in Tables VII and VIII consistently validated the effectiveness of DLap-GCA, which has reached the basic requirement for real-time applications [52], [53], [54]. There have been various efforts to develop real-time BCI systems [36], [38]. Most of the efforts focus on extracting the features of EEG activation, and few have utilized the structure of network topologies. Compared with EEG activity, brain networks can capture tiny differences between two different cognition states by a variety of linkages [3], [14], [28]. These minute differences can offer new features for online applications.

By both simulation and real EEG application, we verified the efficiency of DLap-GCA in resisting the influence of notable outliers to characterize the inherent information propagation and integration patterns among multiple brain regions as well as locate the crucial brain regions for cognition processing. Thus, our current work offers a flexible and reliable alternative for related research fields.

The limitations of this study can be integrated into the following two aspects: 1) the simple and effective feature extraction technology for network topologies is scarce, which may limit the application of brain networks for EEG classification, and 2) our proposed method cannot capture dynamic interactions among time series, which may result in poor performance for time-varying brain state analysis. Therefore, our future work will use supervised learning strategies to preserve the significant linkage and introduce the time-varying prior into our current model to render it suitable for time-varying brain state analysis.

## V. CONCLUSION

In this study, a dual Laplacian distribution-based GCA was proposed to estimate reliable brain networks under strong noise conditions. Both the simulation study and real MI

EEG applications conducted in this work demonstrate the powerful capability of our proposed method for complex outlier restriction, leading to desirable network patterns from EEG time series. In addition to EEGs, heavy tail phenomena widely exist in other research, such as remote sensing, global economics, air traffic control, and electronic power grids. Thus, the proposed method might also provide an alternative in these studies for data prediction, causal inference, and hub node localization [61].

## REFERENCES

- [1] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, "Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by Granger causality," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 26, pp. 9849–9854, Jun. 2004.
- [2] A. Brovelli, J.-P. Lachaux, P. Kahane, and D. Boussaoud, "High gamma frequency oscillatory activity dissociates attention from intention in the human premotor cortex," *NeuroImage*, vol. 28, no. 1, pp. 154–164, Oct. 2005.
- [3] X. Gao et al., "A novel robust student's  $t$ -based Granger causality for EEG based brain network analysis," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104321.
- [4] C. Li et al., "Effective emotion recognition by learning discriminative graph topologies in EEG brain networks," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Feb. 2, 2023, doi: 10.1109/TNNLS.2023.3238519.
- [5] P. Li et al., "Robust Granger analysis in  $L_p$  norm space for directed EEG network analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1959–1969, Nov. 2017.
- [6] P. Li et al., "Robust brain causality network construction based on Bayesian multivariate autoregression," *Biomed. Signal Process. Control*, vol. 58, Apr. 2020, Art. no. 101864.
- [7] J. Chen, Z. Yu, Z. Gu, and Y. Li, "Deep temporal-spatial feature learning for motor imagery-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 11, pp. 2356–2366, Nov. 2020.
- [8] Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, and X. Zhou, "A bi-hemisphere domain adversarial neural network model for EEG emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 494–504, Apr. 2021.
- [9] F. Wang et al., "Quantized attention-gated kernel reinforcement learning for brain-machine interface decoding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 4, pp. 873–886, Apr. 2017.
- [10] C. Yan and Y. He, "Driving and driven architectures of directed small-world human brain functional networks," *PLoS ONE*, vol. 6, no. 8, Aug. 2011, Art. no. e23460.
- [11] J. Xue et al., "Altered directed functional connectivity of the hippocampus in mild cognitive impairment and Alzheimer's disease: A resting-state fMRI study," *Frontiers Aging Neurosci.*, vol. 11, p. 326, Dec. 2019.
- [12] Y. Zhang et al., "Granger causality reveals a dominant role of memory circuit in chronic opioid dependence," *Addiction Biol.*, vol. 22, no. 4, pp. 1068–1080, Jul. 2017.
- [13] G. Bottegal, A. Y. Aravkin, H. Hjalmarsson, and G. Pillonetto, "Robust EM kernel-based methods for linear system identification," *Automatica*, vol. 67, pp. 114–126, May 2016.
- [14] P. Li et al., "L1-norm based time-varying brain neural network and its application to dynamic analysis for motor imagery," *J. Neural Eng.*, vol. 19, no. 2, Apr. 2022, Art. no. 026019.
- [15] Y. Cheng, W. Xing, W. Pedrycz, S. Xian, and W. Liu, "NFIG-X: Non-linear fuzzy information granule series for long-term traffic flow time series forecasting," *IEEE Trans. Fuzzy Syst.*, early access, Mar. 27, 2023, doi: 10.1109/TFUZZ.2023.3261893.
- [16] M. Zhong, "A variational method for learning sparse Bayesian regression," *Neurocomputing*, vol. 69, nos. 16–18, pp. 2351–2355, Oct. 2006.
- [17] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, "Sparse Bayesian classification of EEG for brain-computer interface," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2256–2267, Nov. 2016.
- [18] K. Liu et al., "Robust Bayesian estimation of EEG-based brain causality networks," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 6, pp. 1879–1890, Dec. 2022.

- [19] S. Gai, B. Zhang, C. Yang, and L. Yu, "Speckle noise reduction in medical ultrasound image using monogenic wavelet and Laplace mixture distribution," *Digit. Signal Process.*, vol. 72, pp. 192–207, Jan. 2018.
- [20] N. Iriawan, A. A. Pravitasari, Irhamah, K. Fithriasari, S. W. Purnami, and W. Ferriastuti, "Mixture model for image segmentation using Gaussian, student's  $t$ , and Laplacian distribution with spatial dependence," in *Proc. AIP Conf.*, vol. 2194, 2019, Art. no. 020042.
- [21] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2936–2946, Dec. 2010.
- [22] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.
- [23] S. Geirnaert, T. Francart, and A. Bertrand, "Fast EEG-based decoding of the directional focus of auditory attention using common spatial patterns," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 5, pp. 1557–1568, May 2021.
- [24] L. Baldassarre, M. Pontil, and J. Mourão-Miranda, "Sparsity is better with stability: Combining accuracy and stability for model selection in brain decoding," *Frontiers Neurosci.*, vol. 11, p. 62, Feb. 2017.
- [25] M. I. Rabinovich, M. A. Zaks, and P. Varona, "Sequential dynamics of complex networks in mind: Consciousness and creativity," *Phys. Rep.*, vol. 883, pp. 1–32, Nov. 2020.
- [26] X. Gao et al., "An adaptive joint CCA-ICA method for ocular artifact removal and its application to emotion classification," *J. Neurosci. Methods*, vol. 390, Apr. 2023, Art. no. 109841.
- [27] P. Li et al., "An adaptive EOG removal method based on local density," *J. Electron. Inf. Technol.*, vol. 44, no. 2, pp. 464–476, 2022.
- [28] F. Li et al., "The dynamic brain networks of motor imagery: Time-varying causality analysis of scalp EEG," *Int. J. Neural Syst.*, vol. 29, no. 1, Feb. 2019, Art. no. 1850016.
- [29] F. Li et al., "Brain network reconfiguration during motor imagery revealed by a large-scale network analysis of scalp EEG," *Brain Topography*, vol. 32, no. 2, pp. 304–314, Mar. 2019.
- [30] J. C. Bore et al., "A long short-term memory network for sparse spatiotemporal EEG source imaging," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3787–3800, Dec. 2021.
- [31] A. Mheich, M. Hassan, M. Khalil, V. Gripon, O. Dufor, and F. Wendling, "SimiNet: A novel method for quantifying brain network similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 9, pp. 2238–2249, Sep. 2018.
- [32] P. Li et al., "EEG based emotion recognition by combining functional connectivity network and local activations," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2869–2881, Oct. 2019.
- [33] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive Berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, Aug. 2007.
- [34] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.
- [35] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.
- [36] S. He et al., "EEG- and EOG-based asynchronous hybrid BCI: A system integrating a speller, a web browser, an e-mail client, and a file explorer," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 519–530, Feb. 2020.
- [37] D. Wen et al., "Combining brain-computer interface and virtual reality for rehabilitation in neurological diseases: A narrative review," *Ann. Phys. Rehabil. Med.*, vol. 64, no. 1, Jan. 2021, Art. no. 101404.
- [38] Z. Zhou, E. Yin, Y. Liu, J. Jiang, and D. Hu, "A novel task-oriented optimal design for P300-based brain-computer interfaces," *J. Neural Eng.*, vol. 11, no. 5, Oct. 2014, Art. no. 056003.
- [39] R. Rupp, "Challenges in clinical applications of brain computer interfaces in individuals with spinal cord injury," *Frontiers Neuroengineering*, vol. 7, p. 38, Sep. 2014.
- [40] A. Gong, J. Liu, S. Chen, and Y. Fu, "Time-frequency cross mutual information analysis of the brain functional networks underlying multiclass motor imagery," *J. Motor Behav.*, vol. 50, no. 3, pp. 254–267, May 2018.
- [41] J. Jin, Y. Miao, I. Daly, C. Zuo, D. Hu, and A. Cichocki, "Correlation-based channel selection and regularized feature optimization for MI-based BCI," *Neural Netw.*, vol. 118, pp. 262–270, Oct. 2019.
- [42] Y. Park and W. Chung, "Frequency-optimized local region common spatial pattern approach for motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 7, pp. 1378–1388, Jul. 2019.
- [43] K. D. Ghanbar et al., "Correlation-based common spatial pattern (CCSP): A novel extension of CSP for classification of motor imagery signal," *PLoS ONE*, vol. 16, no. 3, pp. 1–18, 2021.
- [44] B. Wang et al., "Common spatial pattern reformulated for regularizations in brain-computer interfaces," *IEEE Trans. Cybern.*, vol. 51, no. 10, pp. 5008–5020, Oct. 2021.
- [45] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and Dempster-Shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4814–4825, Nov. 2021.
- [46] L. Gu, Z. Yu, T. Ma, H. Wang, Z. Li, and H. Fan, "EEG-based classification of lower limb motor imagery with brain network analysis," *Neuroscience*, vol. 436, pp. 93–109, Jun. 2020.
- [47] J. Wang, Z. Feng, N. Lu, L. Sun, and J. Luo, "An information fusion scheme based common spatial pattern method for classification of motor imagery tasks," *Biomed. Signal Process. Control*, vol. 46, pp. 10–17, Sep. 2018.
- [48] S. Park, D. Lee, and S. Lee, "Filter bank regularized common spatial pattern ensemble for small sample motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 2, pp. 498–505, Feb. 2018.
- [49] A. R. Hidalgo-Muñoz et al., "Application of SVM-RFE on EEG signals for detecting the most relevant scalp regions linked to affective valence processing," *Exp. Syst. Appl.*, vol. 40, no. 6, pp. 2102–2108, May 2013.
- [50] T. Zhang et al., "Structural and functional correlates of motor imagery BCI performance: Insights from the patterns of frontoparietal attention network," *NeuroImage*, vol. 134, pp. 475–485, Jul. 2016.
- [51] M. Cheng, X. Gao, S. Gao, and D. Xu, "Design and implementation of a brain-computer interface with high transfer rates," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 10, pp. 1181–1186, Oct. 2002.
- [52] X. Zhu, P. Li, C. Li, D. Yao, R. Zhang, and P. Xu, "Separated channel convolutional neural network to realize the training free motor imagery BCI systems," *Biomed. Signal Process. Control*, vol. 49, pp. 396–403, Mar. 2019.
- [53] J. Meng, T. Streitz, N. Gulachek, D. Suma, and B. He, "Three-dimensional brain-computer interface control through simultaneous overt spatial attentional and motor imagery tasks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 11, pp. 2417–2427, Nov. 2018.
- [54] A. Chowdhury, H. Raza, Y. K. Meena, A. Dutta, and G. Prasad, "Online covariate shift detection-based adaptive brain-computer interface to trigger hand exoskeleton feedback for neuro-rehabilitation," *IEEE Trans. Cognit. Develop. Syst.*, vol. 10, no. 4, pp. 1070–1080, Dec. 2018.
- [55] M. Bilal et al., "Big data in the construction industry: A review of present status, opportunities, and future trends," *Adv. Eng. Informat.*, vol. 30, no. 3, pp. 500–521, Aug. 2016.
- [56] M. Li et al., "The MindGomoku: An online P300 BCI game based on Bayesian deep learning," *Sensors*, vol. 21, no. 5, p. 1613, Feb. 2021.
- [57] S. Çınar, "Design of an automatic hybrid system for removal of eye-blink artifacts from EEG recordings," *Biomed. Signal Process. Control*, vol. 67, May 2021, Art. no. 102543.
- [58] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern.*, vol. 49, no. 9, pp. 3322–3332, Sep. 2019.
- [59] C. Wilke, L. Ding, and B. He, "Estimation of time-varying connectivity patterns through the use of an adaptive directed transfer function," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 11, pp. 2557–2564, Nov. 2008.
- [60] D. Zambon, C. Alippi, and L. Livi, "Graph random neural features for distance-preserving graph representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 10968–10977.
- [61] X. Ge, Q. Han, L. Ding, Y. Wang, and X. Zhang, "Dynamic event-triggered distributed coordination control and its applications: A survey of trends and techniques," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 50, no. 9, pp. 3112–3125, Sep. 2020.