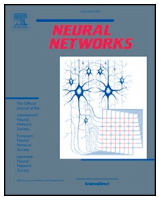
Neural Networks 196 (2026) 108405

Contents lists available at ScienceDirect

Neural Networks

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



Full Length Article

A Causal Validation augmented Temporal Convolutional Framework for Brain Effective Connectivity Networks Estimation

Aoxiang Dong [a,1], Jun Cao [a,b,1], Ptolemaios Georgios Sarrigiannis [c], Daniel Blackburn [d], Andrew Starr [a], Yifan Zhao [a,*]

[a] Faculty of Engineering and Applied Sciences, Cranfield University, Cranfield, MK43 0AL, UK
[b] College of Intelligence Science and Technology, National University of Defense Technology, Changsha, 410073, China
[c] Royal Devon and Exeter NHS Foundation Trust, Exeter, EX2 5DW, UK
[d] Department of Neurosciences, Sheffield Teaching Hospitals, NHS Foundation Trust, Royal Hallamshire Hospital, Sheffield, UK

ARTICLE INFO

Keywords:
Effective brain connectivity
Nonlinear granger causality
Alzheimer's disease
Parkinson's disease
Temporal convolutional network
Lasso

ABSTRACT

Advancements in neuroimaging have facilitated unprecedented insights into brain connectivity, making the study of brain effective connectivity networks (ECNs) essential for understanding neurological functions and diseases. Recently, neural networks (NNs) have emerged as powerful tools for ECN estimation due to their prominent universal approximation ability and less reliance on prior knowledge. However, most NN-based approaches fail to eliminate redundant temporal information and lack rigorous causal validation mechanisms. This paper introduces a novel end-to-end framework for estimating ECNs utilising Least Absolute Shrinkage and Selection Operator (Lasso) regression of Temporal Convolutional Networks (TCNs), named the Causal Validation augmented Temporal Convolutional Framework (CVTCF). In the CVTCF, a convolutional Hierarchical Group Lasso (cHGL) is proposed to detect Granger Causality (GC) inputs and eliminate redundant temporal information during GC detection. Additionally, the framework incorporates permutation importance validation based on the Wilcoxon signed-rank test to enhance the reliability of GC detection. The proposed CVTCF generally outperformed state-of-the-art methods in a controlled simulation using the chaotic Lorenz-96 model and the publicly available blood-oxygen-level-dependent (BOLD) benchmark dataset. Furthermore, the proposed CVTCF has enabled a detailed analysis of the causal interactions within the cerebral cortex, bringing to light the intricate relationships that underlie neurological functioning and impairment of neurodegenerative conditions like Alzheimer's Disease (AD) and Parkinson's Disease (PD). This study demonstrates the potential of using ECN estimation based on the CVTCF as indicators for neurodegenerative diseases and paves the way for future diagnostic and therapeutic strategies.

1. Introduction

The human brain is a complex network of interconnected regions that communicate through intricate patterns of activity (M. Cao et al., 2014). Advancements in neuroimaging technology have enabled researchers to explore brain connectivity (Morita et al., 2016). Effective brain connectivity (EBC) describes the causal effects of neural activities between various brain regions (J. Cao et al., 2021; Ji et al., 2023). The EBC can be estimated based on various neuroimaging data including functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), or electroencephalogram (EEG) (Chen et al., 2024; B. He et al., 2019; Logothetis, 2008; Sakkalis, 2011; X. Wang et al., 2024). The concept of effective connectivity networks (ECNs) refers to the aggregation of all pairs of effective brain connectivity. The exploration of ECNs is crucial for gaining insights into the intricate mechanisms underlying neurological illnesses such as Alzheimer's Disease (AD) and Parkinson's Disease (PD) (J. Cao et al., 2024; Ji et al., 2023). Therefore, developing reliable methods for estimating ECNs is highly beneficial for neuroscience research and clinical diagnosis.

The parametric approaches, which utilise mathematical models to estimate the EBC, are widely utilised due to their interpretability (F. He & Yang, 2020). The causal interactions between brain regions are

* Corresponding author.
E-mail address: yifan.zhao@cranfield.ac.uk (Y. Zhao).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.neunet.2025.108405
Received 10 March 2025; Received in revised form 10 November 2025; Accepted 28 November 2025
Available online 29 November 2025

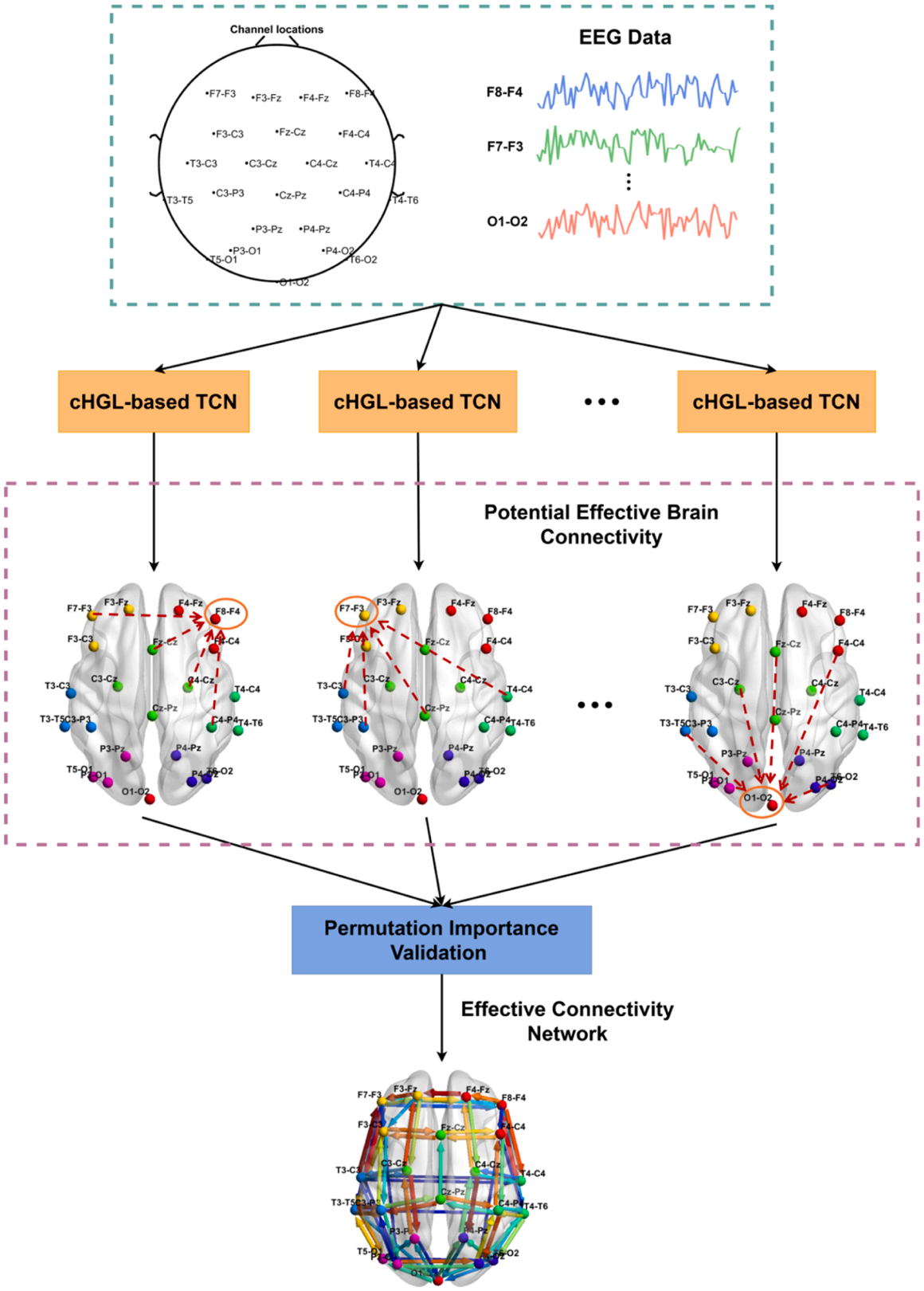


Fig. 1. The flowchart of brain effective connectivity networks (ECNs) estimation with CVTCF. The 23 channel EEG data, denoted as $Y = [F8\text{-}F4, F7\text{-}F3\ldots, O1\text{-}O2]$, are recorded from the corresponding electrodes. To detect the EBC for each element in Y, each EEG channel is sequentially selected as the target for prediction, while all time-series in Y are considered as potential causal inputs and are shifted one time step into the past. Then, 23 convolutional Hierarchical Group Lasso (cHGL)-based Temporal Convolutional Networks (TCN) are trained with the prediction task to estimate potential corresponding EBCs. Finally, the potential EBCs for each EEG channel are validated using permutation importance validation and aggregated into the final ECNs.

revealed through model structure assumptions and parameter estimations. Many conventional parametric approaches assume linear causal interactions between various brain regions, where the brain ECNs are constructed based on vector autoregressive (VAR) models (Baccalá & Sameshima, 2022). Partial directed coherence (PDC) and directed transfer function (DTF) are commonly adopted to estimate the causal interactions between various brain regions in the frequency domain based on VAR parameters (Baccalá & Sameshima, 2001; Kaminski & Blinowska, 1991). However, nonlinear dynamics often appear in brain neural systems, and linear methods normally fail to detect nonlinear causal interactions (Zhao et al., 2013). To address this problem, some nonlinear parametric methods have been developed. The Bayesian inference-based dynamic causal modelling (DCM), Nonlinear AutoRegressive with eXogenous inputs (NARX) based Granger Causality (GC) test and the error reduction ratio (ERR) based causality test have prevailed in the past few decades (Zhao et al., 2012, 2013). Nonetheless, one prominent limitation of these methods is the requirement for prior knowledge about the brain in estimating the EBC. DCM, for instance, is more suited for investigating connectivity among predefined regions of interest rather than for exploratory analysis of relatively large brain or neural networks, due to its complexity and high computational demands (F. He & Yang, 2020; Stephan et al., 2010). On the other hand, the ERR-based method risks prolonged computational times and a higher chance of overfitting when the possible maximum time delay is over-specified, exponentially increasing the number of candidate terms (Dong et al., 2023).

With the growth of computational power, neural networks (NNs) have become more popular in GC analysis for highly nonlinear and complex systems because of their prominent universal approximation ability and less reliance on prior knowledge. It is important to note that the principle of using sparsity-inducing penalties like Lasso for brain connectivity estimation has a strong foundation in classical regression techniques (Antonacci & Astolfi, 2020; Antonacci & Toppi, 2024). Our work builds upon this concept but translates it into the context of deep temporal neural networks. Furthermore, other advanced approaches have been proposed to address the limitations of Granger causality. Notably, the method in (Antonacci et al., 2021) combines an L1-penalised neural network with state-space modelling to mitigate the bias inherent in double regressions, a known challenge in GC estimation (Barnett & Seth, 2014). While this is a valuable contribution, our framework takes a different and complementary path. The utility of NNs in deciphering the enigmatic operations of the brain's signalling pathways has led to significant advancements in the field of neuroimaging. Recent literature indicates a surge in leveraging NNs for EBC estimation, capitalising on their capacity to model complex, dynamic interactions between brain regions. There are generally two types of neural network-based GC detection methods. The first type of approach evaluates whether a specific input time series contributes to output prediction by diminishing the prediction error. The multilayer perceptron (MLP), recurrent neural networks (RNNs) and their variants have been employed to capture the temporal dynamics inherent in EEG data (Montalto et al., 2015; Y. Wang et al., 2018). However, the NNs are required to be trained multiple times to evaluate the contribution of each input in this type of method, making them relatively more complicated than the second type of method (Dong et al., 2023). The second type of method is based on sparse regression of NN, where the causal inputs to the target output are detected in an end-to-end manner. In (Nauta et al., 2019), a temporal causal discovery framework (TCDF) was proposed based on temporal convolutional networks (TCN), where each input channel of the TCN is assigned a trainable attention score. Potential causal inputs are then detected by applying a threshold to these attention scores. After deriving the potential causal inputs, permutation importance validation is applied to further remove spurious results. However, causal input selection in the TCDF only relies on channel-wise attention, while the redundant temporal information is not considered in the process of GC detection (Shi et al., 2023). Furthermore, there is a lack of statistical hypothesis testing in the permutation importance validation process. Recently, the hierarchical group least absolute shrinkage and selection operator (Lasso) penalty has been incorporated into the training of MLP to both detect GC inputs and reduce the redundant input temporal information during GC detection (Tank et al., 2018). However, due to the nonconvex nature of the objective function in (Tank et al., 2018), the gradient descent-based optimisation method does not guarantee the selection of optimal GC inputs and there is a lack of a causal validation process to further remove spurious results.

Recent advances in brain connectivity research have demonstrated the value of network-based approaches across various neurological diseases. Studies have explored connectivity alterations in mild cognitive impairment (Li et al., 2022), developed cross-site harmonisation methods for autism spectrum disorder (Liu, Cui, et al., 2024) and identified network biomarkers for Alzheimer's disease and major depressive disorder (Liu, Li, et al., 2024; Liu, Yang, et al., 2024). These investigations across different neuropathologies highlight the growing importance of connectivity analysis in understanding brain disorders. Our work contributes to this expanding field by introducing a novel framework for estimating effective connectivity networks, with particular application to Alzheimer's and Parkinson's diseases. While previous studies have primarily examined functional connectivity patterns, our approach specifically targets the directed, causal influences between brain regions, offering new insights into the pathophysiological mechanisms of neurodegenerative disorders.

This study proposes an end-to-end ECN estimation framework based on Lasso regression of TCN, namely the Causal Validation augmented Temporal Convolutional Framework (CVTCF). Within the CVTCF, this study introduced the Convolutional Hierarchical Group Lasso (cHGL) in the training of the TCN to both detect GC inputs and remove the redundant temporal information in each TCN channel during GC detection. Additionally, this framework incorporates a Wilcoxon signed-rank test-based permutation importance validation following the derivation of potential EBCs, thereby ensuring that the identified connections are both accurate and causally reliable. The CVTCF is first tested with the chaotic Lorenz-96 model and a publicly available blood-oxygen-level-dependent (BOLD) dataset. Then, the CVTCF is used to investigate the complex neural interactions that characterise neurological conditions, such as AD and PD. Through the application of this framework, this study aims to identify the crucial variations in neural connectivity associated with these disorders, thereby facilitating advancements in diagnostic and therapeutic strategies.

2. Methodology

This section introduces a Causal Validation augmented Temporal Convolutional Framework (CVTCF) to estimate the effective connectivity networks (ECNs) of human subjects across three groups: Alzheimer's Disease (AD), Parkinson's Disease (PD), and healthy controls (HC). The following section begins by introducing the structure of CVTCF, followed by a description of the EEG dataset.

2.1. Causal Validation augmented Temporal Convolutional Framework

The proposed CVTCF comprises two main components: (a) effective brain connectivity (EBC) estimation with convolutional Hierarchical Group Lasso (cHGL) based Temporal Convolutional Networks (TCN) and (b) Permutation Importance Validation with Wilcoxon Signed-Rank Test. The procedure for estimating ECNs via the CVTCF is shown in Figure 1. The CVTCF commence by acquiring data through 23 EEG bipolar channels denoted by Y = [F8-F4, F7-F3, …, O1-O2]. These channels correspond to electrodes placed strategically on the scalp to capture electrical activity across various brain regions. The details about the EEG dataset used in this work are described in Section 2.2..

After the EEG data collection, the CVTCF proceeds to estimate EBC

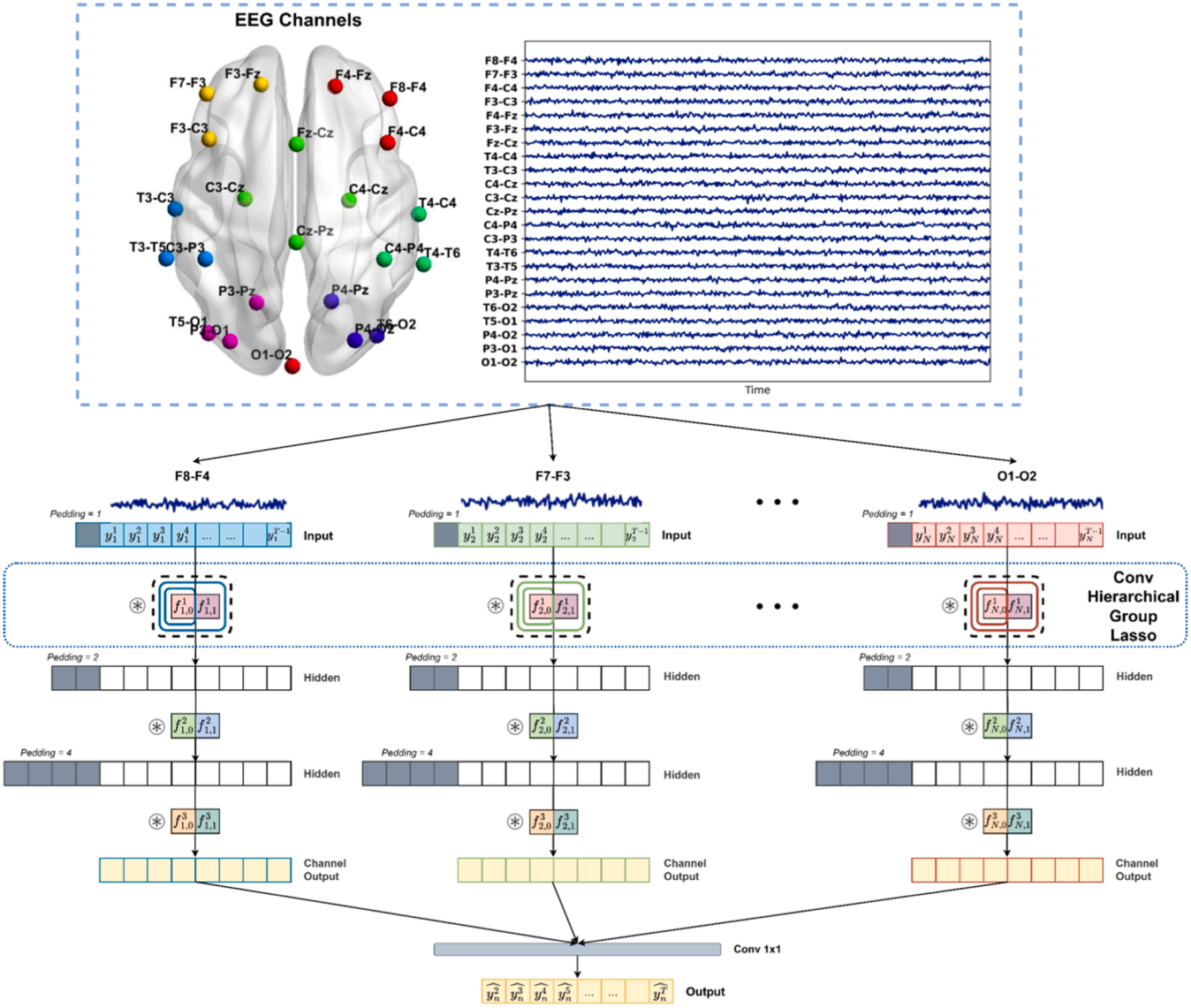


Fig. 2. An illustration of the proposed convolutional Hierarchical Group Lasso (cHGL) implemented on the TCN with a kernel size of 2. The coloured boxes within the first Conv1d layer represent the hierarchical nested groups of kernel weights. The black dash-line boxes represent the channel-wise groups.

for each of the 23 EEG bipolar channels. In the EBC estimation procedure, each EEG channel is sequentially selected as the target for prediction, while all time-series in Y are considered as potential causal inputs and are shifted one time step into the past. Then 23 separate convolutional Hierarchical Group Lasso (cHGL)-based Temporal Convolutional Networks (TCN) are adopted to predict 23 target EEG channels. The potential EBCs for each EEG channel are detected via cHGL penalty during the prediction. Further information on cHGL based TCN is provided in Section 2.1.1. Finally, the potential EBCs for each EEG channel is validated using the permutation importance validation, and the validated connections are aggregated to form the final ECNs. The details of this validation process are presented in Section 2.1.2.

2.1.1. *Effective Brain Connectivity estimation with convolutional Hierarchical Group Lasso*

Effective brain connectivity (EBC) describes the causal effects of neural activities between various brain regions (J. Cao et al., 2021). In this work, EBC is estimated via the neural network-based multivariate Granger Causality (GC) test. Granger Causality initially investigates whether the historical data of one time series, X, are helpful in forecasting the future values of another time series, Y, beyond what the past of Y can predict alone (Granger, 1969). While Granger Causality was initially investigated in a pairwise (bivariate) fashion, this approach is prone to false positives in the presence of common drivers (Antonacci et al., 2025). In (John F. Geweke, 1984; Tank et al., 2018), the GC has been extended from univariate to multivariate time series analysis. The conditional or multivariate formulation overcomes this by testing whether the historical data of one time series, X, helps forecast another, Y, given the past of all other relevant variables in the system. The multivariate definition is described as follows:

Consider a multivariate time series $\boldsymbol{Y}$ with N variables and T samples in each series, denoted as $\{\boldsymbol{Y}^t\}_{t\in[1,T]} = \left\{\left(y_1^t, y_2^t, \ldots, y_j^t, \ldots, y_N^t\right)\right\}_{t\in[1,T]}$. If Granger Causality exists between variables in $\boldsymbol{Y}$, it can be described as

$$y_n^t = g_n\left(y_1^-, \ldots, y_j^-, \ldots, y_N^-\right) + e_t^n,\ 1 \le n \le N, \tag{1}$$

where $y_j^- = \left(y_j^{t-1}, y_j^{t-2}, \ldots, y_j^{t-K_j}\right)$ denotes the historical values of the time series y_j within the maximum time delay K_j, and e_t is the additive zero mean noise. The functions $g_n(\cdot)$ specify how the past information of the time series $[y_1, \ldots, y_N]$ is mapped to the series y_n.

The null hypothesis (H_0), stating that the time series y_j is Granger non-causal for the time series y_n for all $t \in [1, T]$, is defined as

$$g_n\left(y_1^-, \ldots, y_j^-, \ldots, y_N^-\right) = g_n\left(y_1^-, \ldots, y_{j-1}^-, y_{j+1}^-, \ldots, y_N^-\right). \tag{2}$$

Detecting non-causal time series pair (y_n, y_j) in the multivariate time series $\boldsymbol{Y}$ is achieved through the application of the Lasso penalty in regression, including Group Lasso (Rosoł et al., 2022) and Hierarchical Lasso (Jenatton et al., 2010).

In the context of neural network-based GC test, $g_n(\cdot)$ are represented by neural networks. In the CVTCF, the multivariate time series $\boldsymbol{Y}$ are the 23-channel EEG recordings, denoted as $\boldsymbol{Y}$ = [F8-F4, F7-F3, …, O1-O2]. The temporal convolutional networks (TCN) are adopted as the model $g_n(\cdot)$. To identify the Granger causal inputs, we propose a novel convolutional Hierarchical Group Lasso (cHGL) in the first layer of the TCN during the prediction process. The topology of the cHGL-based TCN is illustrated in Figure 2. The cHGL-based TCN is constructed using multiple one-dimensional dilated convolutional (Conv1d) layers. When the time series y_j is fed into channel j of the cHGL-based TCN with a kernel size of k, the kernel weights in the Lth Conv1d layer can be represented as $f_j^L = \left[{f_{(j,0)}}^L, \ldots, {f_{(j,k-1)}}^L\right]$. In Figure 2, the kernel size of each Conv1d layer is set as 2.

The objective function of the sTCN with cHGL is given by

$$\widehat{\boldsymbol{f}} = \underset{f}{argmin}\left\{\frac{1}{T-1}\sum_{t=2}^{T}\left({y_n}^t - g_n\left(\boldsymbol{Y}^{t-1}\right)\right)^2 + \lambda\sum_{j=1}^{N}\| \mathcal{G}\left(\mathbf{f_j}^1\right)\|_1\right\}, \tag{3}$$

where $\cdot_1$ is the L_1 norm of the channel-wise groups. λ is the cHGL hyperparameter that controls the sparsity of the kernel weights in the first Conv1d layer. The hierarchical structured convolutional kernels $\mathcal{G}\left(\mathbf{f_j}^1\right)$ in the first layer of each channel are defined as

$$\mathcal{G}\left(\mathbf{f_j}^1\right) = \sum_{\mathbf{i=0}}^{\mathbf{k-1}} {\mathbf{f}_{(\mathbf{j},\mathbf{i})}}^1{}_{\mathbf{F}}, \tag{4}$$

where $\cdot_F$ denotes the Frobenius matrix norm and ${f_{(j,:i)}}^1 = \left[{f_{(j,0)}}^1, \ldots {f_{(j,i)}}^1\right]$.

The hierarchical structured convolutional kernel within each channel enables a more interpretable and precise temporal causal discovery. Unlike methods that select entire time series as causal inputs, our approach operates at a finer temporal resolution by identifying which specific historical time points within a channel contribute to Granger causality. This temporal refinement is particularly important in neural systems where causal interactions often occur at specific delays rather than uniformly across all past values. By pruning irrelevant time lags while preserving causally significant ones, the cHGL provides not only which brain regions interact but also when these interactions occur, offering a more detailed view of dynamic brain connectivity.

The cHGL removes the redundant temporal information within each channel-wise group to facilitate the selection of causal inputs for the TCN. To achieve this, we hierarchically allocate the kernels in the first layer of cHGL-based TCN into nested groups when implementing the Lasso penalty. Figure 1 depicts the hierarchical nested groups of kernel weights for each channel enclosed in the coloured boxes. These nested groups are then incorporated into channel-wise groups, represented as black dash-line boxes.

In the Lasso problem, the iterative soft-thresholding algorithm (ISTA) (Beck & Teboulle, 2009), a first-order proximal gradient descent method, is commonly used to optimise non-smooth objective function. To optimise the non-smooth objective function as defined in Eq. (3), we adapt the ISTA and incorporated it into the backpropagation process of hierarchical convolutional kernels.

Unlike other existing neural network-based ISTAs that are used solely for selecting causal inputs, the ISTA in this manuscript is designed to both select causal channels and induce sparsity within each TCN channel to remove redundant temporal information. To optimise the non-smooth objective function as defined in Eq. (3), the proximal gradient descent with ISTA can be derived as follows. Let the smooth part of the objective function be $F(f) = \frac{1}{T-1}\sum_{t=2}^{T}\left({y_n}^t - g_n\left(\boldsymbol{Y}^{t-1}\right)\right)^2$, then Eq. (3) can be written as:

$$\widehat{\boldsymbol{f}} = \underset{f}{argmin}\left\{F(f) + \lambda\sum_{j=1}^{N}\| \mathcal{G}\left(\mathbf{f_j}^1\right)\|_1\right\}. \tag{5}$$

At $\forall$ gradient decent step k, the Taylor Series quadratic approximation of the $F(f)$ around f^k is given by:

$$F(f) \approx F\left(f^k\right) + \nabla F\left(f^k\right)^T\left(f - f^k\right)$$
$$+\frac{1}{2}\left(f - f^k\right)^T\nabla^2 F\left(f^k\right)\left(f - f^k\right), \tag{6}$$

where $\nabla F\left(f^k\right)$ can be calculated through backpropagation. To avoid the complicated computation process of the Hessian matrix $\nabla^2 F\left(f^k\right)$, the $\nabla^2 F\left(f^k\right)$ is replaced by $\frac{1}{\eta}I$, where η is the tuneable learning rate and I is the identity matrix. As the $F\left(f^k\right)$ is a constant, which does not affect the optimisation process and can be removed. Then the objective function Eq. (5) can be written as:

$$\widehat{\boldsymbol{f}} = \underset{f}{argmin}\left\{\nabla F\left(f^k\right)^T\left(f - f^k\right) + \frac{1}{2\eta}\| f - f^k \|_2^2\right.$$
$$\left.+\lambda\sum_{j=1}^{N}\| \mathcal{G}\left(\mathbf{f_j}^1\right)\|_1\right\}. \tag{7}$$

By forming the perfect square trinomial, Eq. (7) can be written as:

$$\widehat{\boldsymbol{f}} = \underset{f}{argmin}\left\{\frac{1}{2\eta}f - f^k + \eta\nabla F\left(f^k\right)_2^2 - \frac{\eta}{2}\nabla F\left(f^k\right)_2^2\right.$$
$$\left.+\lambda\sum_{j=1}^{N}\| \mathcal{G}\left(\mathbf{f_j}^1\right)\|_1\right\}, \tag{8}$$

Where the $-\frac{\eta}{2}\nabla F\left(f^k\right)_2^2$ is also a constant term without containing the variable f, therefore it can be removed in the optimisation. Then the objective function can be written as:

$$\widehat{\boldsymbol{f}} = \underset{f}{argmin}\left\{\frac{1}{2\eta}f - \left(f^k - \eta\nabla F\left(f^k\right)\right)_2^2\right.$$
$$\left.+\lambda\sum_{j=1}^{N}\| \mathcal{G}\left(\mathbf{f_j}^1\right)\|_1\right\}, \tag{9}$$

Let the variable f take gradient descent on the smooth part of the objective function be ${Z_f}^{k+1} = f^k - \eta\nabla F\left(f^k\right)$, then the object function can be written as:

$$\widehat{\boldsymbol{f}} = \underset{f}{argmin}\left\{\frac{1}{2\eta}f - {Z_f}^{k+1}{}_2^2 + \lambda\sum_{j=1}^{N}\| \mathcal{G}\left(\mathbf{f_j}^1\right)\|_1\right\}. \tag{10}$$

When reaching any optimum, the gradient of the variable become zero. For the kernel in the first layer in the channel j, the gradient with respect to the objective function Eq. (10) can be written as:

$$\widehat{\boldsymbol{f}}'|_{{f_j}^1} = \left(\frac{1}{2\eta}f - {Z_f}^{k+1}{}_2^2 + \lambda\sum_{j=1}^{N}\| \mathcal{G}\left(\mathbf{f_j}^1\right)\|_1\right)' = 0 \tag{11}$$

If the the number of kernels in channel j equals 2. Then the channel-wise root group $\mathbf{G}\left({f_j}^1\right) = \left(\left\{{f_{j,1}}^1, {f_{j,2}}^1\right\}\right)$ contains the children group

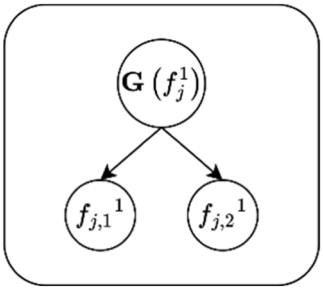


Fig. 3. The hierarchical group structure of the convolutional kernels in the first layer of channel j.

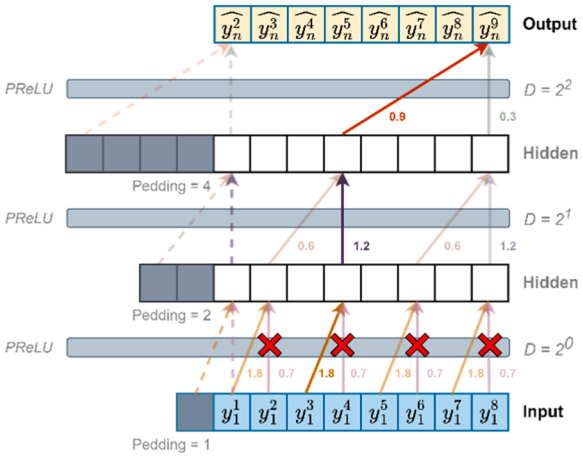


Fig. 4. The illustration of the elimination of the redundant kernels after one iteration of ISTA optimising the cHGL. The red cross represents the elimination of the kernel.

$\left(\left\{f_{j,1}{}^1\right\}\right)$ and $\left\{\left(f_{j,2}{}^1\right)\right\}$. The hierarchical structure can be illustrated in Figure 3 according to (Jenatton et al., 2010).

The children groups only exist when the root group exists (Jenatton et al., 2010). Then, for each element in the channel-wise root group $\mathbf{G}\left(f_j^1\right)$, the soft thresholding function can be derived as follows:

$$\widehat{f}'|_{f_{j,n}^1} = \left(\frac{1}{2\eta}f - Z_f{}^{k+1}{}_2^2 + \lambda \parallel \mathbf{G}\left(f_j^1\right)\parallel_F\right)' = 0 \tag{12}$$

$$\Rightarrow \frac{1}{\eta}\left(f_{j,n}^{1,k+1} - Z_{f_{j,n}^1}^{k+1}\right) + \lambda \frac{f_{j,n}{}^1}{\parallel \mathbf{G}\left(f_j^1\right)\parallel_F}\Bigg|_{Z_{f_{j,n}^1}^{k+1}} = 0 \tag{13}$$

$$\Rightarrow \frac{1}{\eta}\left(f_{j,n}^{1,k+1} - Z_{f_{j,n}^1}^{k+1}\right) + \lambda \frac{Z_{f_{j,n}^1}^{k+1}}{\parallel \mathbf{G}\left(f_j^1\right)\parallel_F} = 0 \tag{14}$$

$$\Rightarrow f_{j,n}^{1,k+1} = \left(1 - \frac{\lambda\eta}{\parallel \mathbf{G}\left(f_j^1\right)\parallel_F}\right) Z_{f_{j,n}^1}^{k+1} \tag{15}$$

This holds only when all $f_{j,n}^{1,k+1} > 0$ or in other words $\parallel \mathbf{G}\left(f_j^1\right)\parallel_F > \lambda\eta$, otherwise all $f_{j,n}^{1,k+1} = 0$. Where n is the element index in channel-wise root group $\mathbf{G}\left(f_j^1\right)$, here n = 1 or 2. $Z_{f_{j,n}^1}^{k+1}$ represents the value of the variable $f_{j,n}^1$ after taking one step of gradient descent on the smooth part of the objective function.

If $\parallel \mathbf{G}\left(f_j^1\right)\parallel_F > \lambda\eta$, then the children groups exist. For $\forall \left(\left\{f_{j,n}{}^1\right\}\right) \supset \mathbf{G}\left(f_j^1\right)$, the soft thresholding function can be derived as follows:

$$\widehat{f}'|_{f_{j,n}^1} = \left(\frac{1}{2\eta}f - Z_f{}^{k+1}{}_2^2 + \lambda \parallel f_{j,n}{}^1\parallel_F\right)' = 0 \tag{16}$$

Algorithm 1
Optimisation process of the cHGL with Iterative Soft-Thresholding Algorithm (ISTA).

Let L be the number of Conv1d layers in cHGL-based TCN.
Let k be the kernel size in each Conv1d layer.
Let $\left\{\boldsymbol{f}_{(j,:i)}^1\right\}$ be the kernel weights in the first layer of channel j of the sTCN from the first weight to index i, $\left\{\boldsymbol{f}_{(j,:i)}^1\right\} = \left[f_{(j,0)}^1, \ldots, f_{(j,i)}^1\right], \forall j \in [0, N], \forall i \in [1, k-1]$.
Require $\lambda > 0, \eta > 0$
  $s = 0$, initialize $\boldsymbol{f}^{(s)} = \left\{\boldsymbol{f}^{l,(s)}\right\}_{l=1}^{L}$
  while not converged **do**
    $s = s + 1$
    compute all $\nabla\mathscr{L}(\boldsymbol{f}^{(s)})$
    determine λ by line search
      for $j = 1$ to N **do**
        for $l = 1$ to L **do**
          $\boldsymbol{f}_j^{l,(s)} = \boldsymbol{f}_j^{l,(s-1)} - \eta\nabla_{f_j^l}\mathscr{L}\left(\boldsymbol{f}^{(s-1)}\right)$
        end for
          for $i = 0$ to $k-1$ **do**
            if $\boldsymbol{f}_{(j,:i)}^{1,(s)} F > \lambda\eta$ **then**
            $\boldsymbol{f}_{(j,i)}^{1,(s)} = \boldsymbol{f}_{(j,:i)}^{1,(s)} - \lambda\eta\dfrac{\boldsymbol{f}_{(j,:i)}^{1,(s)}}{\boldsymbol{f}_{(j,:i)}^{1,(s)} F}$
          else
            $\boldsymbol{f}_{(j,i)}^{1,(s)} = 0$
          end if
        end for
      end for
    end while
  return $\boldsymbol{f}^{(s)}$

$$\Rightarrow \frac{1}{\eta}\left(f_{j,n}^{1,k+1} - Z_{f_{j,n}^1}^{k+1}\right) + \lambda \frac{f_{j,n}{}^1}{\| f_{j,n}{}^1 \|_F}\Bigg|_{Z_{f_{j,n}^1}^{k+1}} = 0 \tag{17}$$

$$\Rightarrow \frac{1}{\eta}\left(f_{j,n}^{1,k+1} - Z_{f_{j,n}^1}^{k+1}\right) + \lambda \frac{Z_{f_{j,n}^1}^{k+1}}{\| f_{j,n}{}^1 \|_2} = 0 \tag{18}$$

$$\Rightarrow f_{j,n}^{1,k+1} = \left(1 - \frac{\lambda\eta}{\| f_{j,n}{}^1 \|_2}\right) Z_{f_{j,n}^1}^{k+1} \tag{19}$$

This holds only when $\| f_{j,n}{}^1 \|_F > \lambda\eta$, otherwise $f_{j,n}^{1,k+1} = 0$. Therefore, it creates the sparsity in the first layer within each existing channel. The sparsity is illustrated in Figure 4.

The elimination of the redundant kernel removes its contribution to the next iteration's channel-wise root group selection. Therefore, it helps remove redundant temporal information and improves Granger causality detection. This hierarchical sparsity induction also serves as an effective regularizer that improves generalisation performance, especially valuable when working with limited-length neuroimaging time series. By eliminating redundant temporal information, the cHGL reduces overfitting and focuses the model's capacity on genuinely predictive patterns. In contrast, in (Nauta et al., 2019), kernels in each channel-wise group are selected together, and redundant temporal information is considered during the selection process.

Then the general proximal gradient descent for the convolutional kernel i in the first layer of channel j at iteration step s can be written as:

$$\boldsymbol{f}_{(j,i)}^{1,s} = Prox_{\lambda\eta}^{(s)}\left(\boldsymbol{f}_{(j,i)}^{1,(s-1)} - \eta\nabla_{\boldsymbol{f}_{(j,i)}^1}\mathscr{L}\left(\boldsymbol{f}^{(s-1)}\right)\right), \forall \boldsymbol{f}^1 \in \mathbb{R}^{N\times k}, \tag{20}$$

where $\mathscr{L}(\boldsymbol{f}^{(s)})$ represents the unregularised loss function of the sTCN, defined as:

$$\mathscr{L}(\boldsymbol{f}^{(s)}) = \frac{1}{T-1}\sum_{t=2}^{T}\left(y_n{}^t - g_n\left(\boldsymbol{Y}^{t-1}\right)\right)^2. \tag{21}$$

In the channel-wise group j, the convolutional kernels in the first layer of sTCN, starting from the first kernel up to index i are denoted as $\left\{\boldsymbol{f}_{(j,:i)}^1\right\} = \left[f_{(j,0)}^1, \ldots f_{(j,i)}^1\right]$. The hierarchical soft-thresholding operator $\boldsymbol{HS}_{\lambda\eta}$ for $\left\{\boldsymbol{f}_{(j,:i)}^1\right\}$ is defined as follows:

$$\boldsymbol{HS}_{\lambda\eta}\left(\boldsymbol{f}_{(j,:i)}^1\right) := \begin{cases} \boldsymbol{f}_{(j,:i)}^1 - \lambda\eta\dfrac{\boldsymbol{f}_{(j,:i)}^1}{\| \boldsymbol{f}_{(j,:i)}^1 F \|}, \boldsymbol{f}_{(j,:i)}^1 F > \lambda\eta \\ 0, \| \boldsymbol{f}_{(j,:i)}^1 F \le \lambda\eta \| \end{cases}, \forall \boldsymbol{f}^1 \in \mathbb{R}^{N\times k} i \in [0, k-1] \tag{22}$$

where $\boldsymbol{HS}_{\lambda\eta}$ iteratively updates all k kernels in the first layer of channel j. During the i^{th} iteration, all kernels ranging from the first to the index i are updated. As a result, $\boldsymbol{HS}_{\lambda\eta}$ hierarchically updates all nested kernel groups as the iteration progresses. Finally, $\boldsymbol{HS}_{\lambda\eta}$ iteratively updates the first layer kernels $\boldsymbol{f}_j^1$ across all N channel-wise groups.

The full optimisation algorithm is outlined in Algorithm 1, where the gradient of the smooth part of the objective function is first computed through backpropagation. Then, cHGL-based TCN takes a smooth gradient descent step. Subsequently, the first layer of cHGL-based TCN takes a proximal gradient descent step using hierarchical ISTA. This process is repeated until the convergence criterion is met.

2.1.2. Permutation Importance Validation

Due to the complicated structure of neural networks, the objective function $\widehat{\boldsymbol{f}}$ may have multiple local minima and saddle points, which are characterized as nonconvex (Goodfellow et al., 2014; Yu et al., 2020). The gradient descent, which is a randomly initialised first-order method, cannot guarantee the smooth part of the objective function to converge to the global minima (Goodfellow et al., 2014). Therefore, the ISTA might be unable to select the optimal causal inputs, and the initial set of potential EBCs derived from the cHGL may contain both true causal links and spurious associations. To address this, we incorporate a permutation importance validation grounded in causal inference principles. The core idea is that if a variable X is a genuine cause of Y, then destroying the temporal information in X's past should significantly impair the prediction of Y. Permutation importance operationalises this by randomly shuffling the values of a potential cause, thereby breaking its causal link to the effect while preserving its marginal distribution. This ensures that any resulting change in prediction accuracy is due to the disruption of the causal relationship and not the variable's statistical properties (Breiman, 2001; van der Laan, 2006; Woodward, 2004). However, relying solely on a fixed threshold for the change in error is sensitive to random variations and lacks statistical rigour. To formally test the hypothesis that the permuted and original errors originate from different distributions, we employ the Wilcoxon signed-rank test. This non-parametric paired test assesses whether the degradation in prediction performance after permutation is systematic and statistically significant, providing a robust, distribution-free mechanism to control for false positives and validate the causal link (Wilcoxon, 1945). This two-step process—causal disruption via permutation followed by statistical hypothesis testing—ensures that the identified connections are not merely predictive but are causally reliable. Therefore, if the prediction error of the model with the permuted input is significantly different from the prediction error of the model with the original input, the input is regarded as a significant causal input.

The Wilcoxon signed-rank test is a paired differences hypothesis test, used to compare whether the locations of two populations are significantly different, without making any assumptions about the sample distributions (Wilcoxon, 1945). In (Rosoł et al., 2022), the Wilcoxon signed-rank test is used to assess the presence of causality between two variables by comparing the prediction error with and without the past value of the variable. In the CVTCF, the Wilcoxon signed-rank test based causal validation is extended to the multivariate scenario. After the derivation of the potential cause set $\boldsymbol{Y_c} = \left[y_i, \ldots, y_j\right]$, $\boldsymbol{Y_c} \subseteq \boldsymbol{Y}$ of the channel $y_n \in \boldsymbol{Y}$, each element in $\boldsymbol{Y_c}$ is validated sequentially. To enhance the reliability of the validation, $\boldsymbol{Y_c}$ is chronologically divided into 10 folds. For the creation of the permuted dataset, each $y_c \in \boldsymbol{Y_c}$ is randomly permuted within its fold, while the other elements in $\boldsymbol{Y_c}$ remain unchanged. This permutation process is repeated for all 10 folds. Subsequently, the permuted dataset and the original dataset are input into the TCN to calculate the mean square error (MSE) of prediction. The Wilcoxon signed-rank test is then applied to compare the MSEs of the datasets before and after permutation, with the null hypothesis H_0 stating that there is no significant difference between the MSEs. If the P-value exceeds 0.05, H_0 is accepted, indicating that y_c is Granger non-causal for y_n. Then all Granger non-causal elements of y_n in $\boldsymbol{Y_c}$ will be eliminated, resulting in a validated EBC for y_n. After that, all channels in $\boldsymbol{Y}$ will be sequentially validated. Finally, all validated EBCs are aggregated to form the final ECNs.

2.1.3. Hyperparameter Tuning

The performance of the CVTCF depends on several key hyperparameters. The most critical is the sparsity parameter λ in the cHGL penalty (Eq. (3)), which controls the number of potential causal connections identified. Other architectural hyperparameters include the number of Temporal Convolutional Network (TCN) layers, the number of filters per layer, the kernel size, and the learning rate. To systematically select these hyperparameters, we employed a grid search strategy combined with a stability-based selection criterion. For the BOLD dataset, the hyperparameters search space was defined as follows: $\lambda \in [0.001, 0.01, 0.1, 1]$, number of TCN layers $\in [2,3,4]$, number of filters $\in [32,64]$, kernel size $\in [2,3]$, and learning rate $\in [0.001, 0.01]$. The optimal set was determined by evaluating the stability of the resulting ECN across multiple runs with different random seeds, in addition to the

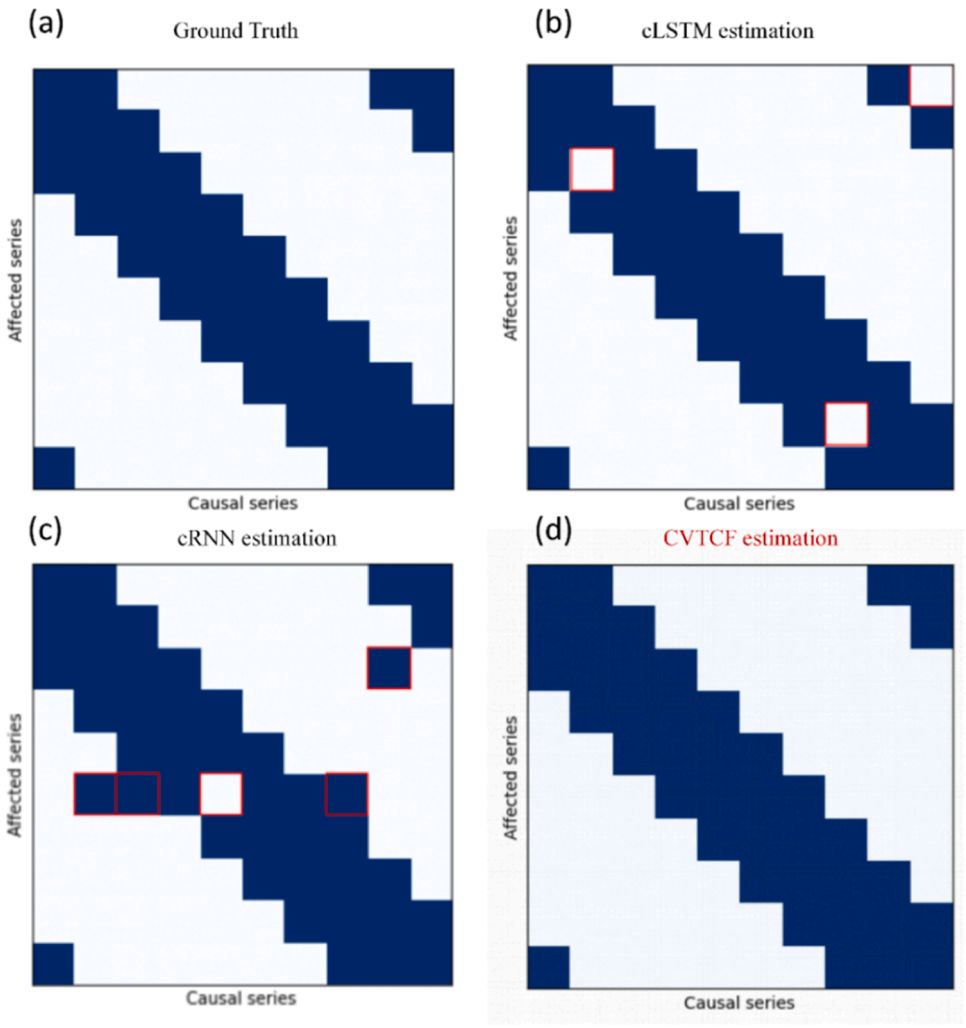


Fig. 5. The adjacency matrix representing the GC relationships on the Lorenz-96 simulation dataset (a) Ground Truth; (b) cLSTM; (c) cRNN; (d) CVTCF.

prediction error on a held-out validation set. The final configuration selected for the BOLD experiment was: λ=0.01, λ=0.01, 3 TCN layers, 32 filters per layer, a kernel size of 2, and a learning rate of 0.001.

2.2. EEG dataset

The recruitment of participants took place at the neuroscience clinic of the NHS Foundation at Sheffield Teaching Hospital. The Health Control Group (HC) was formed through various means, including educational sessions, oral communication, and social networks, with the Memory Clinic at Sheffield Teaching Hospital supplying the majority of these individuals. This clinic primarily serves geriatric patients, especially those under the age of 65. The Parkinson's Disease (PD) cohort was recruited during the Parkinson's UK Roadshow in York in July 2018, while the remaining participants were recruited from September 2014 to December 2019. A team at the University of Sheffield conducted resting-state electroencephalogram (EEG) recordings on sixty participants, aged between 48 to 70 years, from all three groups. The participants were age- and gender-matched across groups: HC (12 females and 8 males, average age 61 years, standard deviation 6.7 years), AD (8 females and 12 males, average age 60 years, standard deviation 4.4 years), and PD (10 females and 10 males, average age 60.25 years, standard deviation 4.96 years). Ethical approval for the EEG study was granted by the Leeds West Research Ethics Committee, with reference number 14/YH/1070. AD patients were diagnosed one month to two years prior to their EEG recordings, at which time they exhibited mild to moderate cognitive deficits according to their Mini-Mental State Examination. Both patients and control group members underwent brain magnetic resonance imaging (MRI) scans to rule out other causes of dementia. Health control participants required normal MRI brain scans and cognitive assessments before their EEG recordings. The final diagnosis of AD was determined using the NINCDS-ADRDA criteria (Dubois et al., 2007), based on a multidisciplinary consensus considering clinical history, neurological examination, cognitive scores, and neuroradiological findings (Blackburn et al., 2018). Additionally, all 20 PD patients received their clinical diagnosis from consultants specializing in PD, presenting normal cognitive functions but at various stages of the disease, with variations in medication and symptoms.

EEG recordings were undertaken with an XLTEK 128-channel equipment (Optima Medical LTD) with bipolar derivations and time-locked video recordings. In order to mitigate the volume conduction effects associated with the common reference electrode, a total of 23 bipolar derivations were generated, which are denoted as follows: F8-F4, F7-F3, F4-C4, F3-C3, F4-Fz, F3-Fz, Fz-Cz, T4-C4, T3-C3, C4-Cz, C3-Cz, Cz-Pz, C4-P4, C3-P3, T4-T6, T3-T5, P4-Pz, P3-Pz, T6-O2, T5-O1, P4-O2, P3-O1 and O1-O2. The sampling frequency used was 2kHz. The EEG sessions captured moments when participants had their eyes closed (EC) and alternated with times they had their eyes open (EO). Participants were in a resting state for 30 minutes, during which they were asked to relax without engaging in any specific thought processes. A neurophysiologist later reviewed these recordings at an XLTEK station, selecting three 12-second segments without any artefacts for both EC

Table 1
Performance comparison on Lorenz-96 simulation dataset.

Method	Accuracy	AUROC
cRNN	0.97±0.02	0.93 ± 0.04
cLSTM	0.95±0.05	0.89 ± 0.05
CVTCF	**0.99 ± 0.01**	**0.95 ± 0.02**

and EO conditions for each participant.

3. Results

3.1. Simulation Study using Lorenz-96 Model

To validate the CVTCF under controlled conditions with known ground truth, we employed the Lorenz-96 system, a standard benchmark for evaluating causal discovery methods in complex, nonlinear dynamical systems. The Lorenz-96 model is defined by the following differential equations:

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F \tag{23}$$

where $i = 1, \ldots, 8$ with cyclic boundary conditions, and F=8.0 to ensure chaotic behaviour. We generated synthetic time series data by numerically integrating these equations using a fourth-order Runge-Kutta method with a time step of 0.01, then sampling every 10 steps to create a multivariate time series of length 1,000.

The ground truth connectivity in this system is well-defined: each variable x_i is directly influenced by its immediate neighbours x_{i+1}, x_{i-1} and x_{i-2}, creating a known circular connectivity pattern. This provides an ideal testbed for evaluating the method's ability to recover both the existence and directionality of causal relationships in a complex,

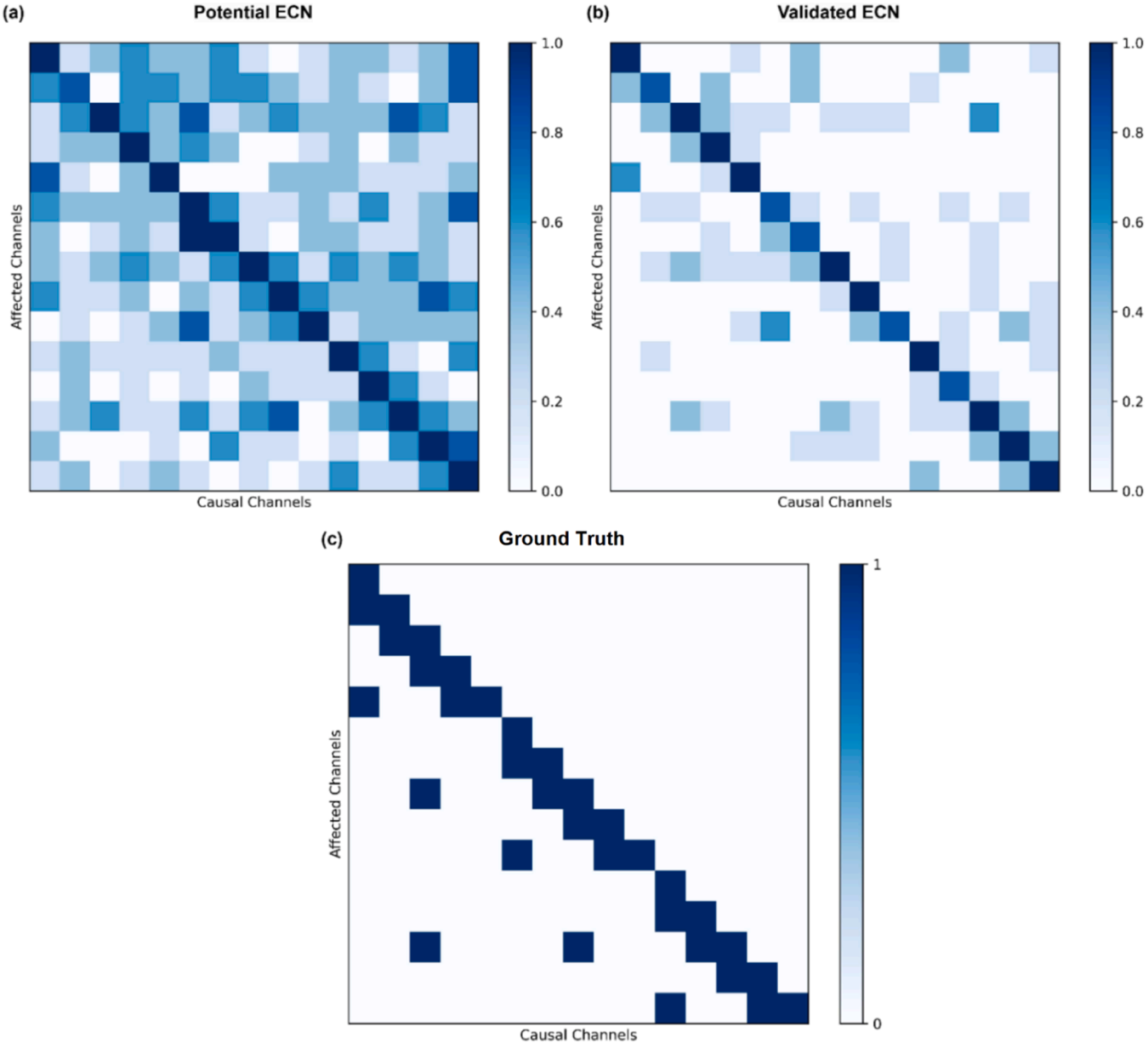


Fig. 6. The comparison between the estimated ECNs of the CVTCF and the ECN ground truth across 5 human subjects in the BOLD dataset. (a) The averaged adjacency matrix representing the estimated potential brain ECN. (b) The averaged adjacency matrix representing the estimated brain ECN after permutation importance validation. (c) The adjacency matrix representing the ECN ground truth in the BOLD dataset. The frequency of the occurrence of EBCs is shown by the colour gradation within the matrix components.

nonlinear system.

We compared the proposed CVTCF against two state-of-the-art neural network-based Granger causality methods: component-wise Recurrent Neural Network (cRNN) and component-wise Long Short-Term Memory (cLSTM). All methods were evaluated using the same dataset and experimental conditions. Performance was quantified using Accuracy and the Area Under the Receiver Operating Characteristic curve (AUROC), which measures the ability to distinguish true causal connections from non-causal ones. The ground truth GC relationships and estimations of cRNN, cLSTM and CVTCF are shown in Figure 5. Besides, Table 1 represents the performance comparison among these three approaches in terms of accuracy and AUROC. The results demonstrate that the proposed CVTCF achieved superior performance with an accuracy of 0.99 and AUROC of 0.94, outperforming both cRNN (accuracy= 0.97, AUROC = 0.93) and cLSTM (accuracy= 0.95, AUROC = 0.89). This performance advantage can be attributed to CVTCF's ability to effectively capture the temporal dependencies through its temporal convolutional architecture while simultaneously eliminating redundant temporal information via the convolutional Hierarchical Group Lasso.

The superior performance of CVTCF in this controlled simulation study validates its effectiveness in recovering known connectivity structures and demonstrates its robustness in handling complex, nonlinear dynamical systems. This provides strong evidence for the method's reliability before applying it to real-world neurophysiological data.

3.2. Bold challenge

The first experiment evaluated the brain ECN estimation performance of the proposed CVTCF using the publicly available blood-oxygen-level-dependent (BOLD) dataset (Smith et al., 2011). We compared the proposed CVTCF with three state-of-the-art approaches that can also estimate the brain ECN in an end-to-end manner, including two Lasso-based approaches: component-wise MLP (cMLP), component-wise long short-term memory (cLSTM) (Tank et al., 2018) and a TCN based temporal causal discovery framework (TCDF) (Nauta et al., 2019). The BOLD dataset was generated from a realistic simulation of functional magnetic resonance imaging (fMRI) models, where each time series represents the activity of a specific region of interest (ROI) within the human brain. The BOLD dataset comprises 15 ROIs for each human subject, where each ROI features a time series length of 200 data points. The relatively limited time series length poses a challenge for Granger Causality (GC) detection. All human subjects share the same ground truth GC relationships among 15 brain ROIs, as shown in Figure 6(c). The objective of this challenge is to detect the GC among the various ROIs based on these time series. In this experiment, the performance of ECN estimation is evaluated across the first 5 human subjects (labelled as 0 to 4). The results of the ECN estimation using the proposed CVTCF are shown in Figure 6(a)-(b). Figure 6(a) shows the averaged adjacency matrix across 5 human subjects, representing the estimated potential brain ECN using the proposed cHGL-based TCN before validation. The colour gradation within the matrix components indicates the occurrence frequency of the estimated EBCs, with darker shades representing more frequent connections. Figure 6 displays the averaged adjacency matrix across 5 human subjects, representing the brain ECN after permutation importance validation in the CVTCF. Compared to the estimated potential ECNs in Figure 6(a) and the ground truth in Figure 6(c), Figure 6(b) shows more sparse connectivity in the upper right section of the adjacency matrix. Despite the persistence of some redundant connectivity in the upper right section of the adjacency matrix, the majority of such connections have been effectively removed. Additionally, in the diagonal and lower left sections, most of the correct connectivity is preserved, although there has been incidental removal of some appropriate connections. This is probably attributable to the impact of noise affecting the GC detection and causal validation.

Table 2
Performance comparison of GC detection by estimating the brain functional connectivity with the BOLD dataset. The mean and standard deviations of AUROC were evaluated across the first 5 human subjects.

Model	AUROC
cMLP	0.80 ± 0.05
cLSTM	0.71 ± 0.04
TCDF	0.77 ± 0.03
CVTCF	0.85 ± 0.03

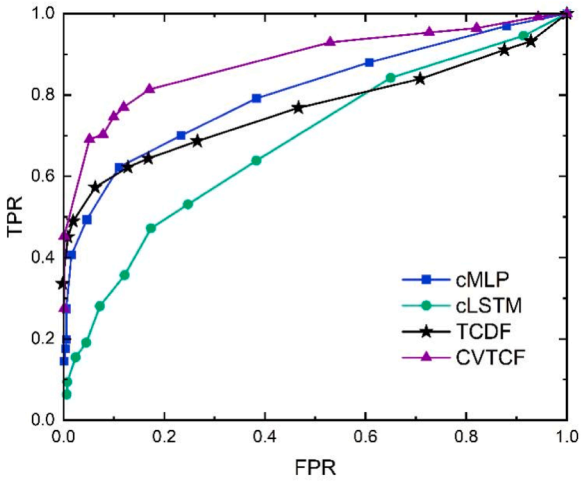


Fig. 7. The ROC curves of the brain functional connectivity estimation with the BOLD dataset. The mean TPR and FPR of each method were evaluated across the first 5 human subjects.

Generally, the application of permutation importance validation successfully mitigated the presence of redundant EBCs that the cHGL-based TCN was not able to eliminate.

In the performance comparison between the proposed CVTCF and three state-of-the-art approaches (Nauta et al., 2019; Tank et al., 2018), the mean and standard deviations of Area Under the Receiver Operating Characteristic curve (AUROC) were evaluated across 5 human subjects. The results are shown in Table 2. Additionally, the corresponding ROC curves for GC detection, depicting the mean true positive rate (TPR) and false positive rate (FPR) of each method across 5 human subjects, are illustrated in Figure 7. Table 2 reveals that the proposed CVTCF exhibits the highest mean AUROC of 0.85 with a comparatively low standard deviation of 0.03. This suggests a consistently superior performance in GC detection across the 5 human subjects. Figure 7 reinforces the superior performance of CVTCF, showcasing its ROC curve maintaining a higher TPR against FPR across all thresholds. Compared to TCDF, CVTCF achieved an approximately 5% higher mean AUROC, emphasising the benefit of incorporating temporal information into GC detection, particularly with a limited time series length. On the other hand, cMLP performed only slightly less effectively than CVTCF, potentially due to its simpler architecture with only one hidden layer in the hyper-parameter setting. This simplicity of structure may make it less prone to overfitting, even with a limited time series length. In summary, the proposed CVTCF has proven to perform well in scenarios involving a limited time series length and complex dynamics, as evidenced by both the ROC curve analysis and the AUROC scores.

3.3. ECNs of Alzheimer's Disease, Parkinson's Disease patients and Health Controls

EEG primarily measures the electrical activity of cortical neurons in

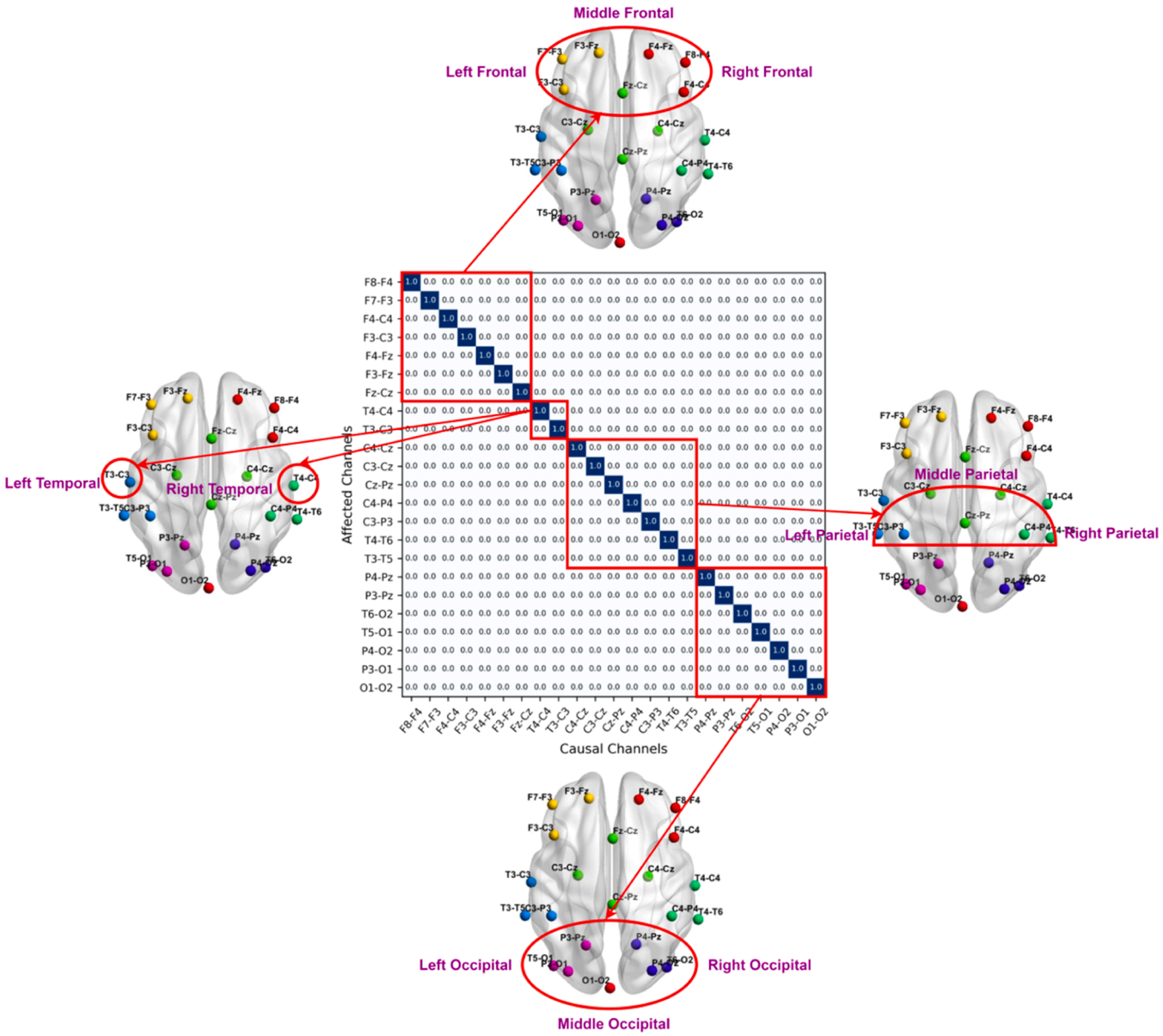


Fig. 8. An illustration depicts the locations of the cerebral cortex lobes and the corresponding channel positions in the adjacency matrix. The dark blue matrix element indicates the existence of causal relationship between two EEG channels. The elements in the red rectangle areas of the adjacency matrix indicate the causal relationships that occur within specific regions of the cerebral cortex lobes.

the cerebral cortex (Blackburn et al., 2018). In neuroanatomy, the cerebral cortex is anatomically and functionally divided into four distinct lobes, namely the frontal, temporal, parietal, and occipital lobes. The frontal lobes play a crucial role in motor functions, facilitating complex decision-making processes and interpersonal relationships fundamental to human behaviour. The temporal lobe processes sensory information into meaningful contexts, aiding in the proper storage of emotions, visual memories, and language comprehension. The parietal lobe plays a key role in processing and interpreting sensations and perceptions. And it is also crucial for melding sensory information with the visual system. The occipital lobe serves as the primary area for processing visual information (Javed et al., 2024). To discover the intricate mechanisms underlying neurological illnesses, including Alzheimer's Disease (AD) and Parkinson's Disease (PD), and to identify biomarkers for disease diagnosis, we compare the ECNs of AD, PD, and Healthy Control (HC) participants. This investigation aims to investigate the differences in EBCs within and between different lobes across these three groups.

To investigate the general differences among various groups and minimise the impact of individual variations, we first derived the ECN for each participant in the AD, PD, and HC groups using the proposed method. Then the ECNs for all participants in these three groups were independently averaged for the eyes-closed (EC) and eyes-open (EO) states respectively. Subsequently, adjacency matrices were formed based on the frequency of occurrence of the EBCs in each group. The locations of the cerebral cortex lobes and the corresponding channel positions in the adjacency matrix are illustrated in Figure 8. The dark blue square indicates the existence of a causal relationship between two EEG channels. The red rectangle areas in the adjacency matrix indicate the causal relationships that occur within specific regions of the cerebral cortex lobes. The rest of the positions in the adjacency matrix indicate the causal relationships that occur between different cortex lobes. The diagonal components of the adjacency matrix are eliminated in order to concentrate on the causal connections among different channels.

The differences in ECNs within various cortex lobes between AD, PD, and HC groups are investigated using two graph metrics: the area under curve (AUC) of the clustering coefficient (aCp) and the AUC of the global

Table 3
The comparison of graph metrics of ECNs within different lobes between AD and HC groups. Statistics were derived from a two-sample T-test with a significance level of $p < 0.05$. The significant p values are set in bold.

Eye State	Graph Metrics	p			
		Frontal	Temporal	Parietal	Occipital
EO	aCp	3.299e-1	NA	4.778e-1	6.848e-1
	aEg	1.322e-1	**3.085e-5**	5.438e-1	2.138e-1
EC	aCp	9.746e-2	NA	1.068e-1	5.924e-1
	aEg	2.203e-1	**4.043e-3**	2.993e-1	3.871e-1

network efficiency (aEg). The clustering coefficient measures the degree to which nodes in a graph tend to cluster together (Bassett & Bullmore, 2006). The aCp reflects how the network's clustering property changes over the different densities of connections and its tendency to form tightly linked groups (J. Wang et al., 2015). The aEg reflects the network's overall efficiency in facilitating communication between nodes across different densities of connections (Bassett & Bullmore, 2006). The comparison of aCp can disclose differences in brain network organisation, while the comparison of aEg can reveal differences in information processing efficiency. The graph theory network analysis (GRETNA) toolbox is employed to derive the aCp and aEg (J. Wang et al., 2015). The two-sample T-test is then used to investigate the discrepancy in graph metrics between the AD, PD, and HC groups, with a significance level of $p < 0.05$.

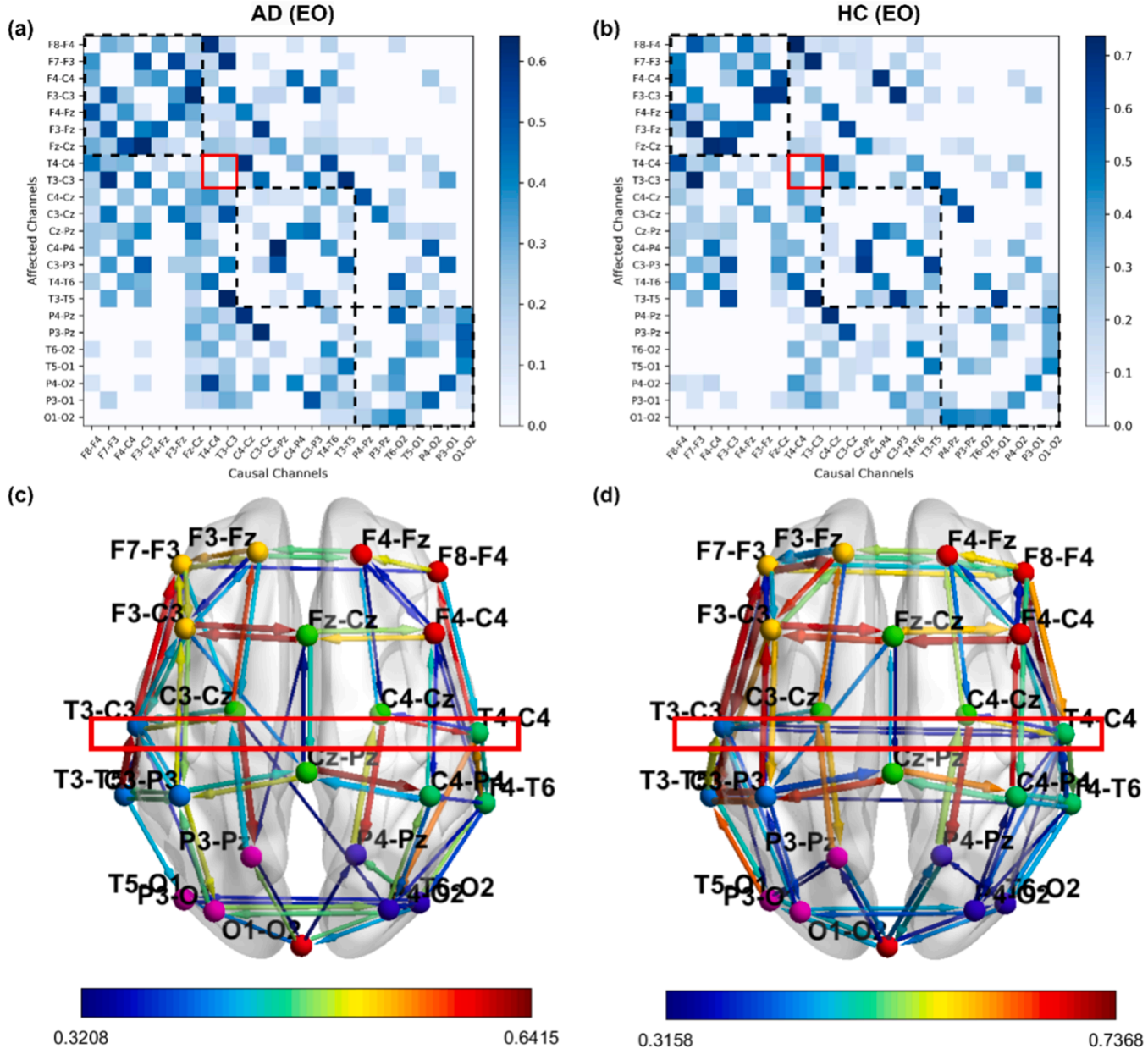


Fig. 9. The comparison of AD and HC with respect to the estimated ECNs under conditions of EO. (a) The adjacency matrix representing the averaged ECNs of the AD participants under the condition of EO. (b) The adjacency matrix representing the averaged ECNs of the HC participants under the condition of EO. The matrix element colour gradation indicates the frequency of occurrence of the EBCs within the respective group. (c) The general ECNs of AD participants under the condition of EO. (d) The general ECNs of HC participants under the condition of EO. The hue of the arrows represents the frequency of occurrence of EBCs in the respective group. Dashed line rectangles represent different cerebral cortex lobes in the adjacency matrix. Significant differences between AD and HC participants under the EO condition are highlighted by red rectangles in both the adjacency matrix and the corresponding cortex lobes.

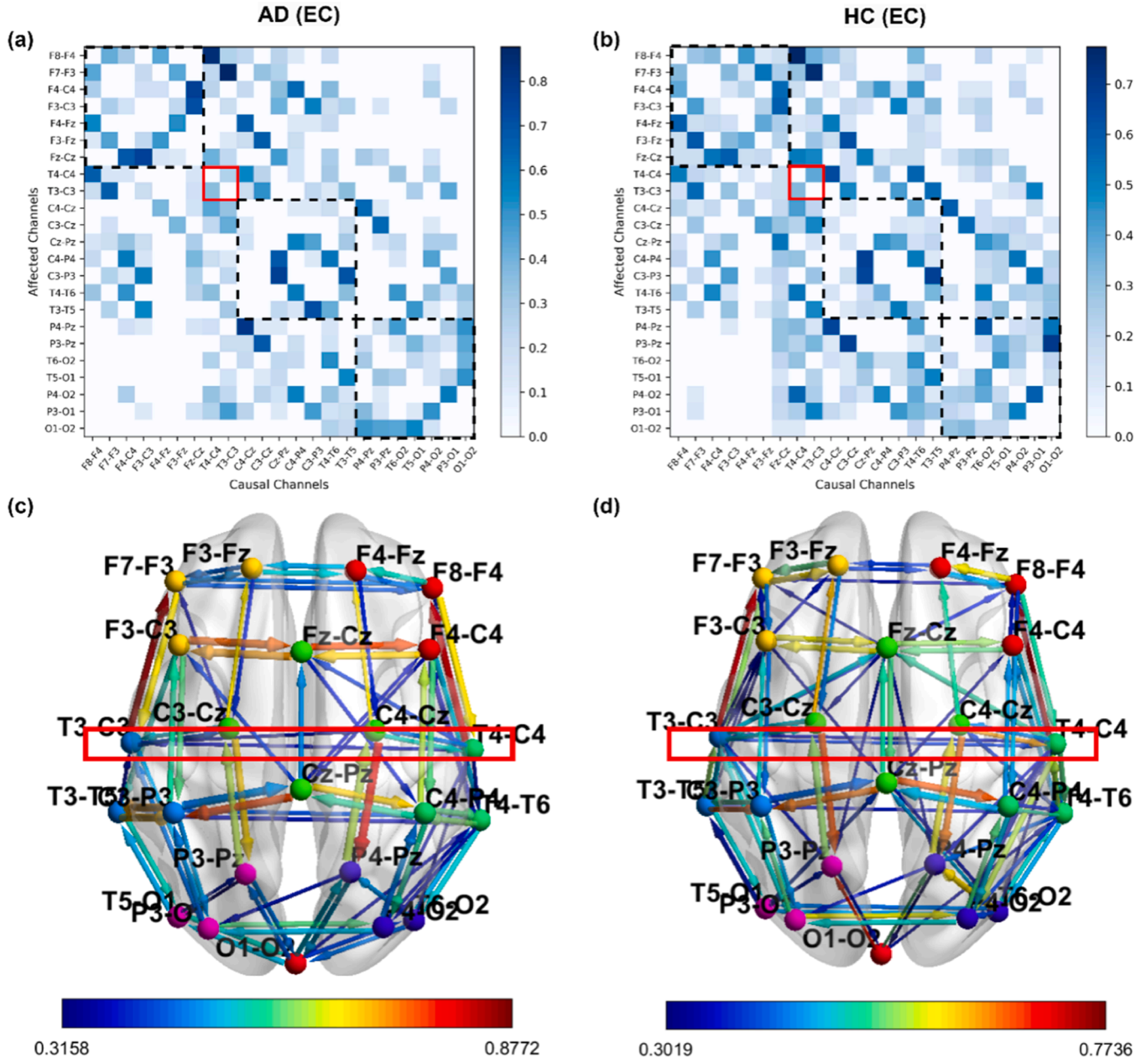


Fig. 10. The comparison of AD and HC with respect to the estimated ECNs under conditions of EC. (a) The adjacency matrix representing the averaged ECNs of the AD participants under the condition of EC. (b) The adjacency matrix representing the averaged ECNs of the HC participants under the condition of EC. The matrix element colour gradation indicates the frequency of occurrence of the EBCs within the respective group. (c) The general ECNs of AD participants under the condition of EC. (d) The general ECNs of HC participants under the condition of EC. The hue of the arrows represents the frequency of occurrence of EBCs in the respective group. Dashed line rectangles represent different cerebral cortex lobes in the adjacency matrix. Significant differences between AD and HC participants under the EC condition are highlighted by red rectangles in both the adjacency matrix and the corresponding cortex lobes.

3.3.1. Alzheimer's Disease vs. Health Controls

This section presents a comparative analysis of the estimated ECNs between Alzheimer's Disease (AD) patients and Healthy Controls (HC) during both eyes-closed (EC) and eyes-open (EO) conditions. Table 3 compares the graph metrics of ECNs in different lobes between the AD and HC groups. Since there are only two nodes in the temporal lobe, and the basis for calculating of the aCp requires at least three nodes to form a triangle, the aCp is not applicable in the temporal lobe and its p value is marked as 'NA' in Table 3. In the T-test, it has been observed the aEg in the temporal lobe of the AD and HC groups exhibit a significant difference in both EO and EC conditions. The EO condition exhibited a more pronounced difference.

Figures 9-10 illustrate the differences in ECNs between AD and HC groups. In both figures, the upper panels (a) and (b) display the adjacency matrices representing the general ECNs for the AD and HC groups respectively. The colour gradation within the matrix elements signifies the frequency of occurrence of the EBCs within each group. Dashed line rectangle areas represent different cerebral cortex lobes in the adjacency matrix. To further visualise the causal interactions within and across various lobes, the ECNs diagrams (c) and (d) are illustrated to depict these connections, with the colour intensity of the arrows reflecting the frequency of the EBCs. Only EBCs with a frequency of over 30% are plotted for better visualisation. In Figure 9, the significant differences between AD and HC groups are highlighted by red rectangles, indicating distinct connectivity patterns within Temporal lobes. In Figure 10, the significant differences between AD and HC groups are also highlighted

Table 4
The comparison of graph metrics of ECNs within different lobes between PD and HC groups. Statistics were derived from a two-sample T-test with a significance level of $p < 0.05$. The significant p values are set in bold.

Eye State	Graph Metrics	p			
		Frontal	Temporal	Parietal	Occipital
EO	aCp	**2.452e-3**	NA	7.817e-1	**1.446e-2**
	aEg	**3.997e-2**	5.663e-1	9.155e-2	2.147e-1
EC	aCp	**2.876e-2**	NA	8.983e-1	**1.239e-2**
	aEg	**6.211e-4**	1.339e-1	2.772e-1	8.757e-2

by red rectangles, indicating that distinct connectivity patterns within Temporal regions are still noticeable. Generally, especially during the EO condition, AD participants display reduced connectivity within the Temporal regions, compared to HC participants. This reduced connectivity might be related to impairments in memory processing seen in AD.

3.3.2. Parkinson's Disease vs. Health Controls

For both the EC and EO circumstances, this section compares the estimated ECNs of PD patients with those of HC. Table 4 compares the graph metrics of ECNs in different lobes between the AD and HC groups. The aCp is also not applicable in the temporal lobe and its p value is marked as 'NA' in Table 4. T-tests reveal significant differences in both aCp and aEg in the frontal lobe between the PD and HC groups under

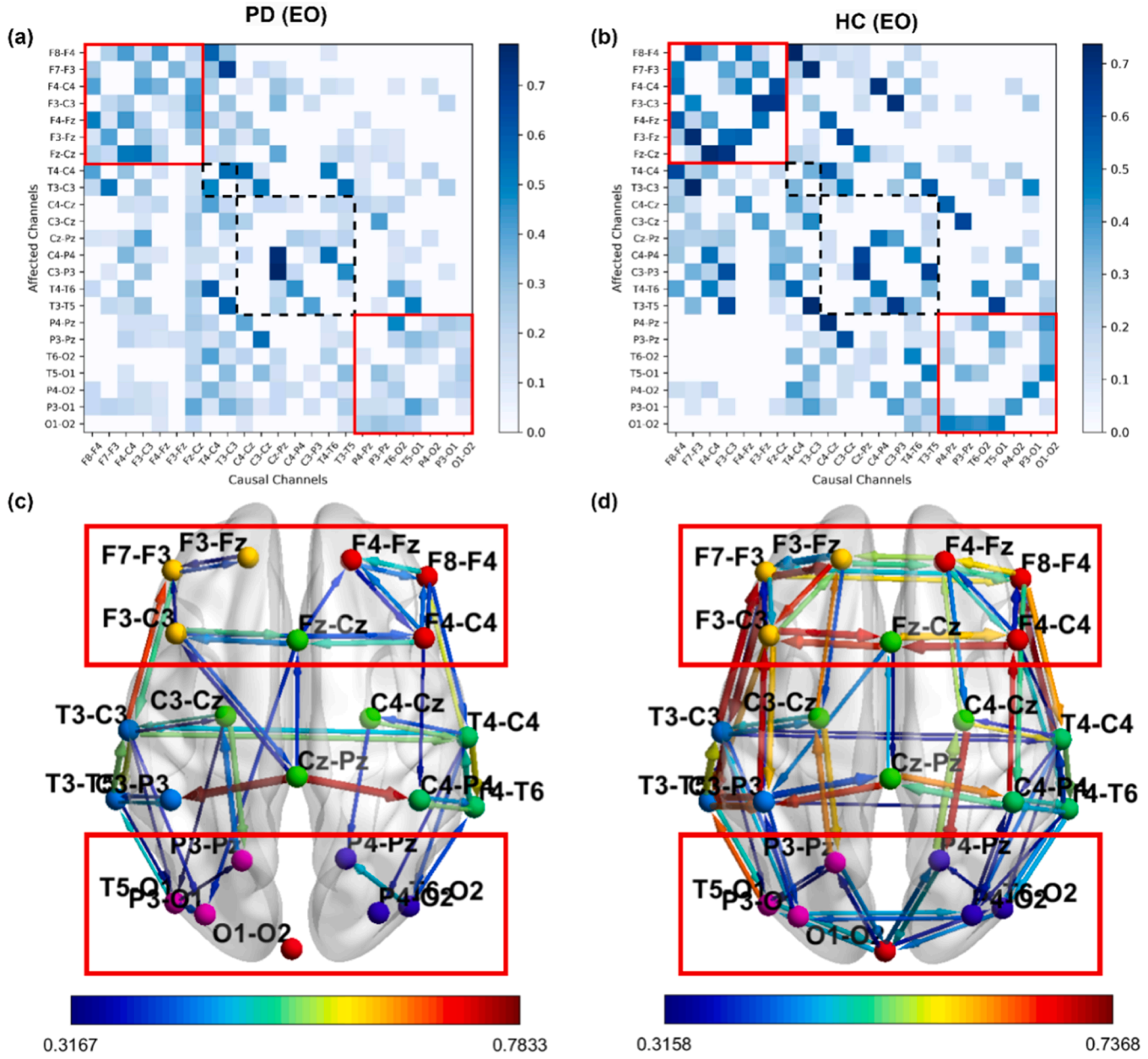


Fig. 11. The comparison of PD and HC with respect to the estimated ECNs under conditions of EO. (a) The adjacency matrix representing the averaged ECNs of the PD participants under the condition of EO. (b) The adjacency matrix representing the averaged ECNs of the HC participants under the condition of EO. The matrix element colour gradation indicates the frequency of occurrence of the EBCs within the respective group. (c) The general ECNs of PD participants under the condition of EO. (d) The general ECNs of HC participants under the condition of EO. The hue of the arrows represents the frequency of occurrence of EBCs in the respective group. Dashed line rectangles represent different cerebral cortex lobes in the adjacency matrix. Significant differences between PD and HC participants under the EO condition are highlighted by red rectangles in both the adjacency matrix and the corresponding cortex lobes.

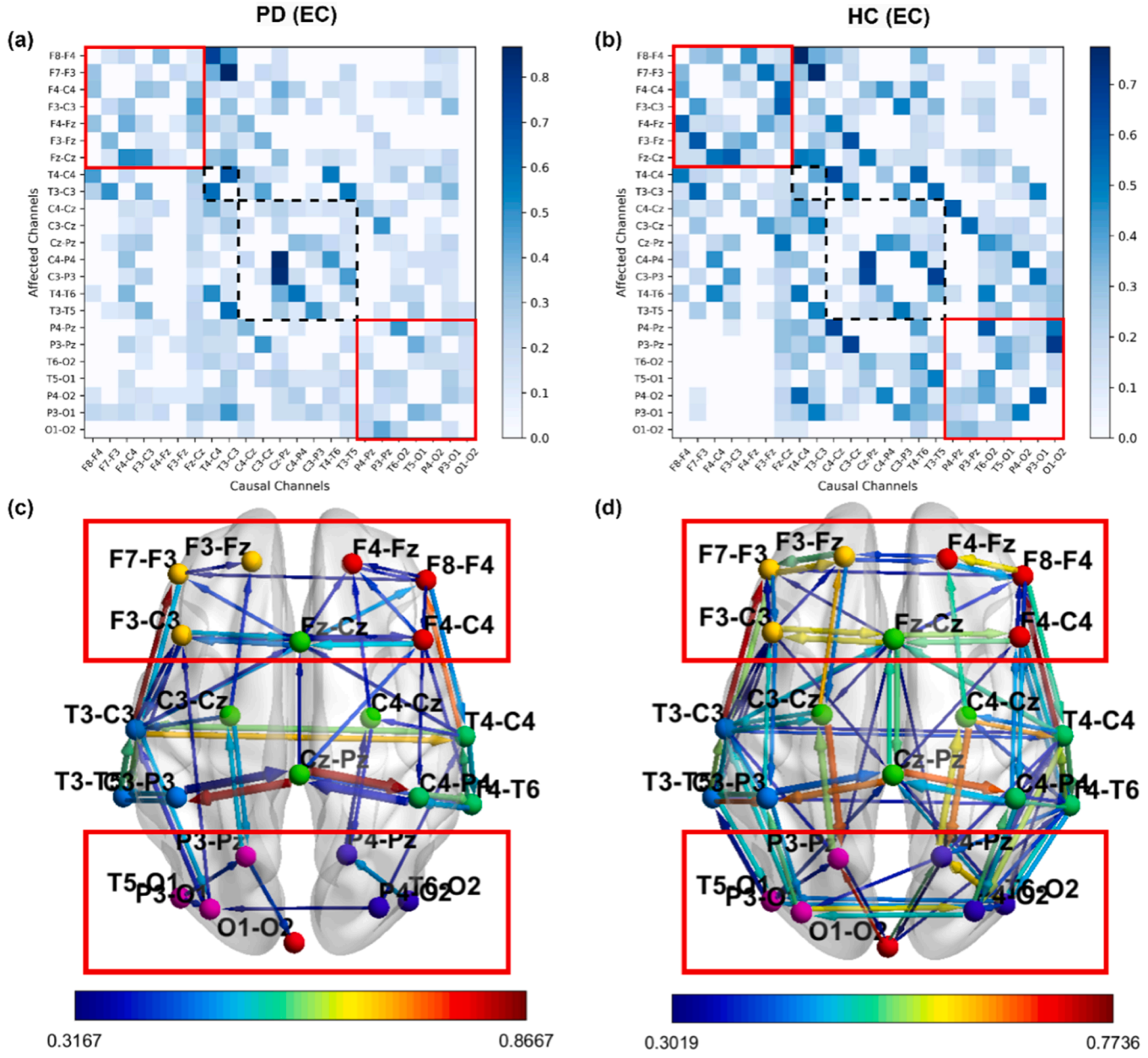


Fig. 12. The comparison of PD and HC with respect to the estimated ECNs under conditions of EC. (a) The adjacency matrix representing the averaged ECNs of the PD participants under the condition of EC. (b) The adjacency matrix representing the averaged ECNs of the HC participants under the condition of EC. The matrix element colour gradation indicates the frequency of occurrence of the EBCs within the respective group. (c) The general ECNs of PD participants under the condition of EC. (d) The general ECNs of HC participants under the condition of EC. The hue of the arrows represents the frequency of occurrence of EBCs in the respective group. Dashed line rectangles represent different cerebral cortex lobes in the adjacency matrix. Significant differences between PD and HC participants under the EC condition are highlighted by red rectangles in both the adjacency matrix and the corresponding cortex lobes.

both EO and EC conditions. This suggests significant differences in brain network organization and information processing efficiency in the frontal lobe between the two groups. Conversely, in the occipital lobe, while aCp shows significant differences, aEg does not. This indicates that although the network organisation differs in the occipital lobe under both EO and EC conditions, the information processing efficiency of the occipital lobe is not significantly affected by PD.

To further visualise the difference of ECNs between PD and HC groups, the adjacency matrices depicting the overall ECNs for the PD and HC groups are shown in Figures 11-12, respectively, in the top panels (a) and (b). The frequency of the EBCs within each group is shown by the colour gradation inside the matrix components. The rectangular dashed line sections in the adjacency matrix stand for several lobes of the cerebral cortex. The ECNs diagrams (c) and (d) show the EBCs between different lobes and how they causally interact with one another; the arrows' colour intensity corresponds to the frequency of the EBCs. Figure 11 reveals during the EO condition that PD patients exhibit alterations in connectivity compared to HCs, particularly in the Frontal and Occipital regions. The significant differences between PD and HC groups are highlighted by red rectangles. This deviation in connectivity patterns suggests a potential impact of Parkinson's Disease on the brain's communication pathways during active visual engagement. In Figure 12, under EC conditions, the differences are also significant in the Frontal and Occipital lobes. The adjacency matrix and corresponding network diagrams denote reduced interactions between these lobes in PD patients compared to HCs. As the Frontal lobes play a crucial role in

Table 5
The comparison of graph metrics of ECNs within different lobes between AD and PD groups. Statistics were derived from a two-sample T-test with a significance level of $p < 0.05$. The significant p values are set in bold.

Eye State	Graph Metrics	p			
		Frontal	Temporal	Parietal	Occipital
EO	aCp	**1.599e-3**	NA	8.853e-1	**2.312e-2**
	aEg	**8.054e-6**	**5.493e-4**	5.392e-2	**6.046e-3**
EC	aCp	4.730e-1	NA	1.272e-1	**2.603e-2**
	aEg	**5.482e-3**	**8.896e-5**	9.626e-1	**2.869e-2**

motor functions and the Occipital lobes are mainly responsible for processing visual information, the observed reduced connectivity within these two regions could be associated with the motor coordination deficits characteristic of PD. Furthermore, these findings suggest that the neural connectivity impairments in PD are not only present during active visual tasks but also during rest, highlighting the pervasive effect of the disease on brain function.

3.3.3. *Alzheimer's Disease vs. Parkinson's Disease*

An analysis is conducted in this section to compare the estimated ECNs of patients with AD and PD under both the EC and EO conditions. Table 4 compares the graph metrics of ECNs in different lobes between the AD and PD groups. The aCp is also not applicable in the Temporal lobe and its p value is marked as 'NA' in Table 5. T-tests reveal

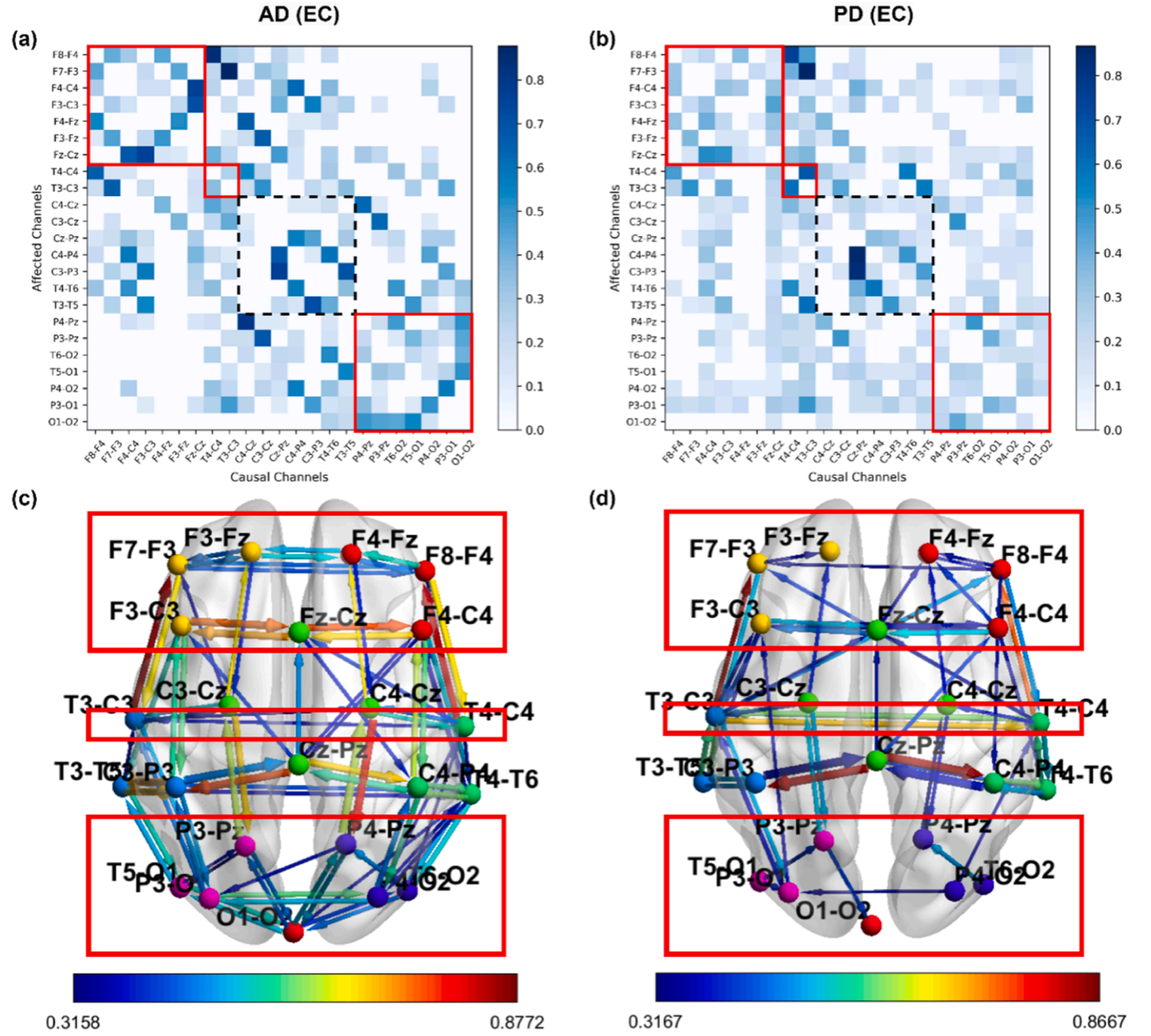


Fig. 13. The comparison of AD and PD with respect to the estimated ECNs under conditions of EC. (a) The adjacency matrix representing the averaged ECNs of the AD participants under the condition of EC. (b) The adjacency matrix representing the averaged ECNs of the PD participants under the condition of EC. The matrix element colour gradation indicates the frequency of occurrence of the EBCs within the respective group. (c) The general ECNs of AD participants under the condition of EC. (d) The general ECNs of PD participants under the condition of EC. The hue of the arrows represents the frequency of occurrence of EBCs in the respective group. Dashed line rectangles represent different cerebral cortex lobes in the adjacency matrix. Significant differences between AD and PD participants under the EC condition are highlighted by red rectangles in both the adjacency matrix and the corresponding cortex lobes.

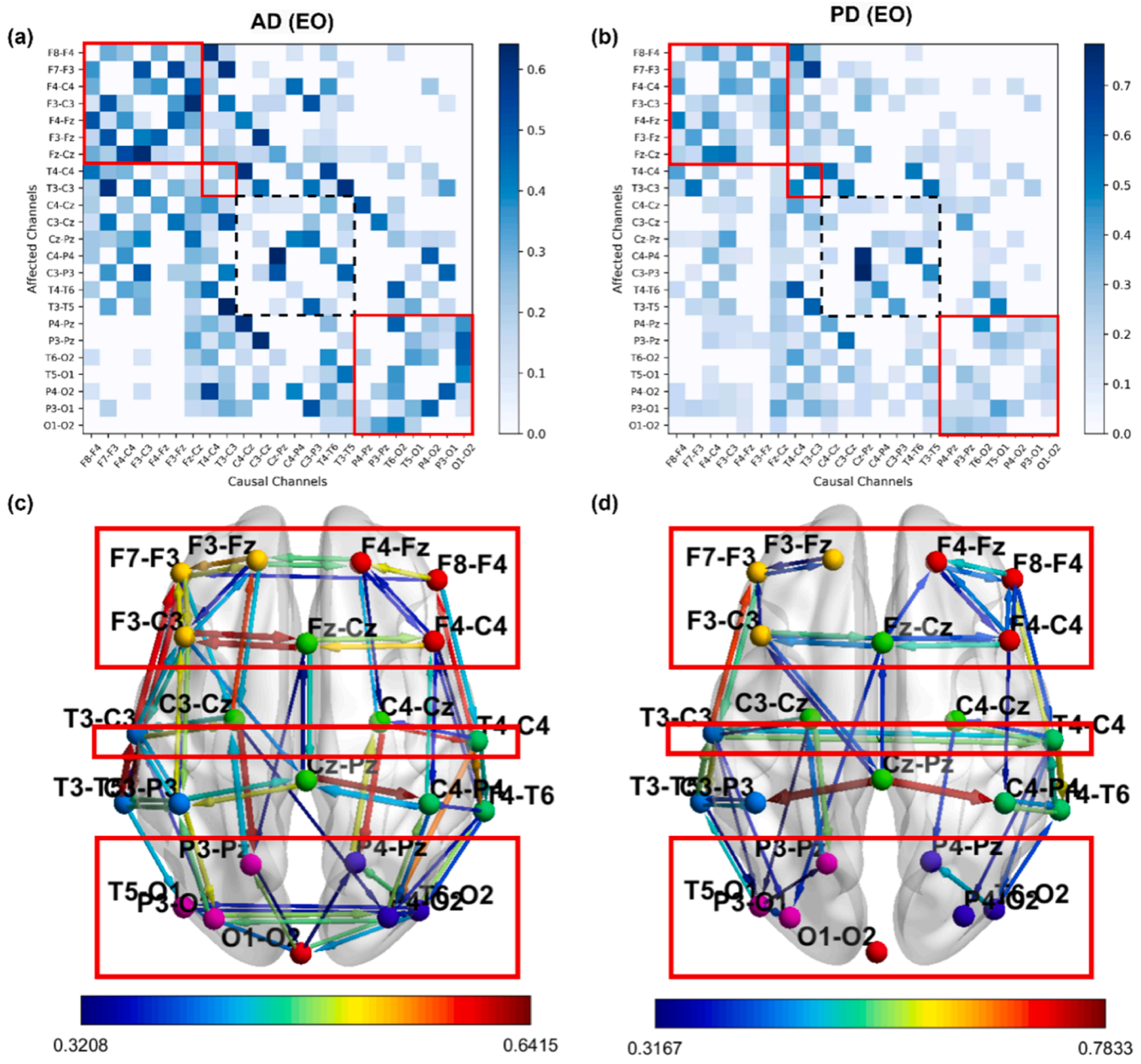


Fig. 14. The comparison of AD and PD with respect to the estimated ECNs under conditions of EO. (a) The adjacency matrix representing the averaged ECNs of the AD participants under the condition of EO. (b) The adjacency matrix representing the averaged ECNs of the PD participants under the condition of EO. The matrix element colour gradation indicates the frequency of occurrence of the EBCs within the respective group. (c) The general ECNs of AD participants under the condition of EO. (d) The general ECNs of PD participants under the condition of EO. The hue of the arrows represents the frequency of occurrence of EBCs in the respective group. Dashed line rectangles represent different cerebral cortex lobes in the adjacency matrix. Significant differences between AD and PD participants under the EO condition are highlighted by red rectangles in both the adjacency matrix and the corresponding cortex lobes.

significant differences in the Frontal, Temporal and Occipital lobes between the AD and PD groups under both EO and EC conditions. Although aCp shows significant differences in the frontal lobe only under EO conditions and not under EC conditions, aEg is significantly different under both EO and EC conditions. This indicates that while the network organization in Frontal lobe are similar between AD and PD groups under EC conditions, the information processing efficiency is still significantly different.

To better understand the distinction between ECNs in the AD and PD groups, the upper panels (a) and (b) of Figure 13 and Figure 14 display the adjacency matrices that represent the general ECNs for each group, respectively. Within the matrix components, a colour gradient indicates the incidence of EBCs within each group. The rectangular dashed line segments within the adjacency matrix correspond to distinct cerebral cortex lobes. The causal interactions between distinct regions are illustrated in Figure 13(c) and (d); the frequency of the EBCs is represented by the colour intensity of the arrows. Figure 13, representing the EO condition, illustrates that AD participants demonstrate distinct connectivity patterns from those with PD, particularly within the frontal, temporal and occipital lobes. This disparity suggests that the neural pathways involved in executive functioning, memory and visual processing may be differentially affected in AD compared to PD. For the EC condition depicted in Figure 14, the ECNs reveal a difference in connectivity patterns. The significant differences between AD and PD patients are also highlighted by red rectangles, indicating distinct connectivity patterns in the frontal, temporal and occipital regions.

These findings indicate that there may be variations in the impact on the brain, related to executive and memory functioning and visual processing between AD and PD under both EO and EC conditions. Overall, the findings suggest that AD and PD have distinct ECNs, reflecting the unique pathophysiological features of these conditions.

4. Conclusion

In this study, a novel Causal Validation augmented Temporal Convolutional Framework (CVTCF) was proposed and performed on the brain Effective Connectivity Networks (ECN) estimation. With the help of cHGL in removing redundant temporal information during GC detection and using permutation importance validation to further remove spurious results, the proposed CVTCF generally outperformed three state-of-the-art methods. This performance was first evaluated through a simulation on the Lorenz-96 system, a standard benchmark for nonlinear dynamics, where the CVTCF reliably recovered the known ground-truth connectivity with high accuracy, demonstrating its core competence in a controlled environment. The framework's effectiveness was further confirmed on the publicly available BOLD benchmark. The hierarchical structure of cHGL contributed significantly to this outcome by enhancing both interpretability and performance. It provides a more nuanced understanding of neural dynamics by revealing not only which brain regions interact causally but also the specific time delays at which these interactions occur, a refinement over methods that treat entire channels as causal units. Furthermore, the study's extensive investigation into the ECNs among individuals with Alzheimer's Disease (AD), Parkinson's Disease (PD), and healthy controls (HC) provides valuable insights into the neural connectivity patterns unique to these conditions. The proposed CVTCF has enabled a detailed analysis of the causal interactions within the cerebral cortex, bringing to light the intricate relationships that underlie neurological functioning and impairment.

The ECN comparisons between AD and HC, as well as PD and HC, have revealed significant differences in connectivity patterns. Notably, the disparity in the temporal lobe findings between AD and HC participants might be in part a reflection of the specific challenges in memory processing associated with AD. Similarly, the distinct EBCs in the frontal and occipital regions observed in PD patients may reflect the motor and sensory processing difficulties that are hallmarks of Parkinson's Disease. The direct comparison between AD and PD patients further accentuates the distinct pathophysiological features of these conditions. Our results have demonstrated unique patterns of connectivity disruption in AD as opposed to PD, suggesting differing impacts on neural pathways involved in cognitive, sensory and motor processing. One limitation of the proposed CVTCF is that EBC detection relies on the assumption of stationarity in multivariate time series, whereas EEG data are inherently non-stationary. This is particularly pertinent in the study of neurological disorders like epilepsy, where ictal activity (seizures) presents abrupt, high-amplitude changes. In such scenarios, applying the CVTCF to a long, non-stationary recording could lead to misleading inferences, as the estimated ECN would represent an average of potentially distinct connectivity states, obscuring transient but critical causal interactions. The model might fail to capture the dynamic reconfiguration of brain networks, which is often a key pathophysiological feature. Future work will aim to enhance the method's applicability by integrating it with a temporal segmentation approach. This could involve employing change-point detection algorithms or sliding window analysis to first identify quasi-stationary epochs before applying the CVTCF to each epoch separately. This would allow for the estimation of dynamic effective connectivity networks, providing a more accurate and clinically relevant characterisation of brain dynamics in non-stationary conditions.

Through extensive analysis, this study demonstrates the potential of ECN estimation using EEG recordings based on the proposed CVTCF to serve as biomarkers for neurodegenerative diseases. This research paves the way for future studies where such neural network-based approaches could be employed not only for understanding the neurological underpinnings of diseases but also for the development of diagnostic strategies.

CRediT authorship contribution statement

Aoxiang Dong: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jun Cao:** Writing – review & editing, Validation, Formal analysis. **Ptolemaios Georgios Sarrigiannis:** Writing – review & editing, Validation, Data curation. **Daniel Blackburn:** Writing – review & editing, Validation, Data curation. **Andrew Starr:** Writing – review & editing, Supervision. **Yifan Zhao:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

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

References

Antonacci, Y., & Astolfi, L. (2020). Information Transfer in Linear Multivariate Processes Assessed through Penalized Regression Techniques : Validation and Application to Physiological Networks. 1–31. https://doi.org/10.3390/e22070732.

Antonacci, Y., Barà, C., Sparacino, L., Pirovano, I., Mastropietro, A., Rizzo, G., & Faes, L. (2025). Spectral Information Dynamics of Cortical Signals Uncover the Hierarchical Organization of the Human Brain's Motor Network. *IEEE Transactions on Biomedical Engineering, 72*(5), 1655–1664. https://doi.org/10.1109/TBME.2024.3516943

Antonacci, Y., Minati, L., Faes, L., Pernice, R., Nollo, G., Toppi, J., Pietrabissa, A., & Astolfi, L. (2021). Estimation of Granger causality through Artificial Neural Networks: applications to physiological systems and chaotic electronic oscillators. *PeerJ Computer Science, 7*, e429. https://doi.org/10.7717/peerj-cs.429

Antonacci, Y., & Toppi, J. (2024). Measuring Connectivity in Linear Multivariate Processes With Penalized Regression Techniques. *IEEE Access, 12*(February), 30638–30652. https://doi.org/10.1109/ACCESS.2024.3368637

Baccalá, L. A., & Sameshima, K. (2001). Partial directed coherence: A new concept in neural structure determination. *Biological Cybernetics, 84*(6), 463–474. https://doi.org/10.1007/PL00007990/METRICS

Baccalá, L. A., & Sameshima, K. (2022). Partial Directed Coherence and the Vector Autoregressive Modelling Myth and a Caveat. *Frontiers in Network Physiology, 2*. https://doi.org/10.3389/fnetp.2022.845327

Barnett, L., & Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox : A new approach to Granger-causal inference. *Journal of Neuroscience Methods, 223*, 50–68. https://doi.org/10.1016/j.jneumeth.2013.10.018

Bassett, D. S., & Bullmore, E. (2006). Small-World Brain Networks. *The Neuroscientist, 12*(6), 512–523. https://doi.org/10.1177/1073858406293182

Beck, A., & Teboulle, M. (2009). A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences, 2*(1), 183–202. https://doi.org/10.1137/080716542

Blackburn, D. J., Sarrigiannis, P. G., De Marco, M., Zhao, Y., Venneri, A., Lawrence, S., Unwin, Z. C., Blyth, M., Angel, J., Baster, K., Wilkinson, I. D., Bell, S. M., He, F., Wei, H. L., Billings, S. A., & Farrow, T. F. D. (2018). A Pilot Study Investigating a Novel Non-Linear Measure of Eyes Open versus Eyes Closed EEG Synchronization in People with Alzheimer's Disease and Healthy Controls. *Brain Sciences, 8*(7), 134. https://doi.org/10.3390/BRAINSCI8070134, 2018, Vol. 8, Page 134.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Cao, J., Yang, L., Sarrigiannis, P. G., Blackburn, D., & Zhao, Y. (2024). Dementia classification using a graph neural network on imaging of effective brain connectivity. *Computers in Biology and Medicine, 168*, Article 107701. https://doi.org/10.1016/J.COMPBIOMED.2023.107701

Cao, J., Zhao, Y., Shan, X., Wei, H., Guo, Y., Chen, L., Erkoyuncu, J. A., & Sarrigiannis, P. G. (2021). Brain functional and effective connectivity based on electroencephalography recordings: A review. *Human Brain Mapping*. https://doi.org/10.1002/HBM.25683

Cao, M., Wang, J.-H., Dai, Z.-J., Cao, X.-Y., Jiang, L.-L., Fan, F.-M., Song, X.-W., Xia, M.-R., Shu, N., Dong, Q., Milham, M. P., Castellanos, F. X., Zuo, X.-N., & He, Y. (2014). Topological organization of the human brain functional connectome across the lifespan. *Developmental Cognitive Neuroscience, 7*, 76–93. https://doi.org/10.1016/j.dcn.2013.11.004

Chen, K., Chai, S., Xie, T., Liu, Q., & Ma, L. (2024). EEG spatial inter-channel connectivity analysis: A GCN-based dual stream approach to distinguish mental fatigue status. *Artificial Intelligence in Medicine, 157*, Article 102996. https://doi.org/10.1016/J.ARTMED.2024.102996

Dong, A., Starr, A., & Zhao, Y. (2023). Neural network-based parametric system identification: a review. *International Journal of Systems Science, 54*(13), 2676–2688. https://doi.org/10.1080/00207721.2023.2241957

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P. J., & Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *The Lancet. Neurology, 6*(8), 734–746. https://doi.org/10.1016/S1474-4422(07)70178-3

Goodfellow, I. J., Vinyals, O., & Saxe, A. M. (2014). Qualitatively characterizing neural network optimization problems. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. https://arxiv.org/abs/1412.6544v6.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica, 37*(3), 424. https://doi.org/10.2307/1912791

He, B., Astolfi, L., Valdes-Sosa, P. A., Marinazzo, D., Palva, S. O., Benar, C. G., Michel, C. M., & Koenig, T. (2019). Electrophysiological Brain Connectivity: Theory and Implementation. *IEEE Transactions on Bio-Medical Engineering, 66*(7), 2115–2137. https://doi.org/10.1109/TBME.2019.2913928

He, F., & Yang, Y. (2020). Nonlinear System Identification of Neural Systems from Neurophysiological Signals. *BioRxiv*. https://doi.org/10.1101/2020.08.09.243253, 2020.08.09.243253.

Javed, K., Reddy, V., & Lui, F. (2024). Neuroanatomy, Cerebral Cortex. In *StatPearls*. http://www.ncbi.nlm.nih.gov/pubmed/21042938.

Jenatton, R., Mairal, J., Obozinski, G., & Bach, F. (2010). Proximal Methods for Hierarchical Sparse Coding. *Journal of Machine Learning Research, 12*, 2297–2334. http://arxiv.org/abs/1009.2139.

Ji, J., Zou, A., Liu, J., Yang, C., Zhang, X., & Song, Y. (2023). A Survey on Brain Effective Connectivity Network Learning. *IEEE Transactions on Neural Networks and Learning Systems, 34*(4), 1879–1899. https://doi.org/10.1109/TNNLS.2021.3106299

Geweke, John F. (1984). Measures of Conditional Linear Dependence and Feedback between Time Series. *Journal of the American Statistical Association, 79*(388), 907–915. https://doi.org/10.1080/01621459.1984.10477110

Kaminski, M. J., & Blinowska, K. J. (1991). A new method of the description of the information flow in the brain structures. *Biological Cybernetics, 65*(3), 203–210. https://doi.org/10.1007/BF00198091/METRICS

Li, Y., Liu, J., Jiang, Y., Liu, Y., Lei, B., & Member, S. (2022). Virtual Adversarial Training-Based Deep Feature Aggregation Network From Dynamic Effective. *IEEE Transactions on Medical Imaging, 41*(1), 237–251. https://doi.org/10.1109/TMI.2021.3110829

Liu, J., Cui, W., Chen, Y., Ma, Y., Dong, Q., Cai, R., Li, Y., & Hu, B. (2024). Deep Fusion of Multi-Template Using Convolutional Networks for Brain Disease Analysis. *IEEE Transactions on Medical Imaging, 43*(2), 860–873. https://doi.org/10.1109/TMI.2023.3325261

Liu, J., Li, Z., Sheng, Y., Zhou, J., Feng, L., Cai, H., Liu, W., Tian, F., Dong, Q., & Liu, R. (2024). Adaptive Weight and Wasserstein Distance Constrained Low-Rank Sparse Representation Method for Functional Connectivity. *IEEE Transactions on Computational Social Systems, 11*(6), 7400–7410. https://doi.org/10.1109/TCSS.2024.3400029

Liu, J., Yang, W., Ma, Y., Dong, Q., Li, Y., & Hu, B. (2024). Effective hyper-connectivity network construction and learning : Application to major depressive disorder identification. 171(February). https://doi.org/10.1016/j.compbiomed.2024.108069.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature, 453*(7197), 869–878. https://doi.org/10.1038/nature06976, 2008 453:7197.

Montalto, A., Stramaglia, S., Faes, L., Tessitore, G., Prevete, R., & Marinazzo, D. (2015). Neural networks with non-uniform embedding and explicit validation phase to assess Granger causality. *Neural Networks, 71*, 159–171. https://doi.org/10.1016/J.NEUNET.2015.08.003

Morita, T., Asada, M., & Naito, E. (2016). Contribution of Neuroimaging Studies to Understanding Development of Human Cognitive Brain Functions. *Frontiers in Human Neuroscience, 10*. https://doi.org/10.3389/fnhum.2016.00464

Nauta, M., Bucur, D., & Seifert, C. (2019). Causal Discovery with Attention-Based Convolutional Neural Networks. *Machine Learning and Knowledge Extraction, 1*(1), 312–340. https://doi.org/10.3390/make1010019

Rosoł, M., Młyńczak, M., & Cybulski, G. (2022). Granger causality test with nonlinear neural-network-based methods: Python package and simulation study. *Computer Methods and Programs in Biomedicine*. , Article 106669. https://doi.org/10.1016/J.CMPB.2022.106669

Sakkalis, V. (2011). Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Computers in Biology and Medicine, 41*(12), 1110–1117. https://doi.org/10.1016/J.COMPBIOMED.2011.06.020

Shi, T., Yang, W., Qi, A., Li, P., & Qiao, J. (2023). LASSO and attention-TCN: a concurrent method for indoor particulate matter prediction. *Applied Intelligence, 53*(17), 20076–20090. https://doi.org/10.1007/S10489-023-04507-6/TABLES/5

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., Ramsey, J. D., & Woolrich, M. W. (2011). Network modelling methods for FMRI. *NeuroImage,, 54*(2), 875–891. https://doi.org/10.1016/j.neuroimage.2010.08.063

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *NeuroImage, 49*(4), 3099–3109. https://doi.org/10.1016/J.NEUROIMAGE.2009.11.015

Tank, A., Covert, I., Foti, N., Shojaie, A., & Fox, E. (2018). Neural Granger Causality. https://doi.org/10.1109/TPAMI.2021.3065601.

van der Laan, M. J. (2006). Statistical inference for variable importance. *International Journal of Biostatistics, 2*(1). https://doi.org/10.2202/1557-4679.1008/MACHINEREADABLECITATION/RIS

Wang, J., Wang, X., Xia, M., Liao, X., Evans, A., & He, Y. (2015). GRETNA: A graph theoretical network analysis toolbox for imaging connectomics. *Frontiers in Human Neuroscience, 9*(JUNE), Article 147509. https://doi.org/10.3389/FNHUM.2015.00386/BIBTEX

Wang, X., Liesaputra, V., Liu, Z., Wang, Y., & Huang, Z. (2024). An in-depth survey on Deep Learning-based Motor Imagery Electroencephalogram (EEG) classification. *Artificial Intelligence in Medicine, 147*, Article 102738. https://doi.org/10.1016/J.ARTMED.2023.102738

Wang, Y., Lin, K., Qi, Y., Lian, Q., Feng, S., Wu, Z., & Pan, G. (2018). Estimating Brain Connectivity With Varying-Length Time Lags Using a Recurrent Neural Network. *IEEE Transactions on Bio-Medical Engineering, 65*(9), 1953–1963. https://doi.org/10.1109/TBME.2018.2842769

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin, 1*(6), 80. https://doi.org/10.2307/3001968

Woodward, J. (2004). Making Things Happen: A Theory of Causal Explanation. *Making Things Happen: A Theory of Causal Explanation*, 1–418. https://doi.org/10.1093/0195155270.001.0001

Yu, X., Wu, L., Xu, C., Hu, Y., & Ma, C. (2020). A Novel Neural Network for Solving Nonsmooth Nonconvex Optimization Problems. *IEEE Transactions on Neural Networks and Learning Systems, 31*(5), 1475–1488. https://doi.org/10.1109/TNNLS.2019.2920408

Zhao, Y., Billings, S. A., Wei, H., He, F., & Sarrigiannis, P. G. (2013). A new NARX-based Granger linear and nonlinear casual influence detection method with applications to EEG data. *Journal of Neuroscience Methods, 212*(1), 79–86. https://doi.org/10.1016/j.jneumeth.2012.09.019

Zhao, Y., Billings, S. A., Wei, H., & Sarrigiannis, P. G. (2012). Tracking time-varying causality and directionality of information flow using an error reduction ratio test with applications to electroencephalography data. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 86*(5). https://doi.org/10.1103/PhysRevE.86.051919