Inverse Diffusion-Based Elimination of Volume Conduction Effect for Functional Connectivity

Yuzeng XU[1], Sho OTSUKA[2,3], Seiji NAKAGAWA[2,3]

***1.** Graduate School of Science and Engineering, Chiba University, Japan. **2.** Center for Frontier Medical Engineering, Chiba University, Japan, **3.** Graduate School of Engineering, Chiba University, Japan*

***E-mail:** yuzeng-xu@chiba-u.jp*

Abstract: Electroencephalogram (EEG)-based functional connectivity is valuable for cognitive recognition, but often distorted by volume conduction effects (VCE). We propose an elimination by modeling VCE as a diffusion process and applying inverse filtering to suppress spurious connections. The improved classification performance demonstrates its effectiveness in enhancing functional connectivity representations.

Keywords: Volume Conduction Effects, Functional Connectivity

Introduction

Electroencephalography (EEG)-based cognitive recognition plays a key role in advancing brain-machine interfaces due to its ability to non-invasively capture brain activity. However, EEG signals are indirect and often contaminated by the volume conduction effect (VCE), where electrical signals spread through conductive tissues, causing spatial blurring and pseudo-connections of functional connectivity.

Functional connectivity captures statistical dependencies between brain regions, forming functional networks useful for cognitive recognition. Yet, VCE can distort these networks by introducing spurious associations between unrelated regions. To address this, we model VCE as a diffusion-like process and apply an inverse filtering approach to reduce its impact, aiming to recover cleaner functional networks and enhance recognition performance in actual cognitive classification tasks.

Methods

Modeling of Diffusion & Inverse-Diffusion

Assuming the VCE behaves as a diffusion-like process and can be modeled by a diffusion filtering: $M' = G \cdot M \cdot G^T$ (1), here G is a spatially-informed diffusion kernel, such as a Gaussian function derived from true inter-electrode distances. This formulation preserves spatial structure [1] and offers a principled framework for modeling VCE-induced distortions in functional networks.

To incorporate physical distance and simulate signal decay, the diffusion kernel is defined as Gaussian distribution:

$$G_{i,j} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{{d_{i,j}}^2}{2\sigma^2}\right) \tag{2}$$

where $d_{i,j}$ denotes the Euclidean distance between electrode pairs and preserved in the distance matrix.

Assuming that VCE distort the genuine functional network through a diffusion-like process, the observed functional network and its recovered counterpart-based on the inverse diffusion assumption-can be modeled as follows:

$$FN_{Observed} = G \cdot FN_{Genuine} \cdot G^T \tag{3}$$

$$FN_{Recovered} = G^{-1} \cdot FN_{Observed} \cdot {G^{-1}}^T \tag{4}$$

However, due to the instability and potential ill-conditioning of directly inverting the Gaussian kernel G, a pseudo-inverse PI is introduced as a substitute for G^{-1}:

$$PI = (G^T \cdot G + \lambda_{reg} I)^{-1} \cdot G^T \approx G^{-1} \tag{5}$$

Finally, for reinforcing the data structure:

$$FN_{Recover} = PI \cdot FN_{Observed} \cdot PI^T + FN_{Observed} \tag{6}$$

Feature Engineering

EEG signals were filtered into five bands, with alpha, beta, and gamma selected for feature extraction. Signals were segmented into 1-second windows, and functional connectivity networks were computed for each using Pearson correlation coefficient (PCC). These networks across time formed the input features for CNN-based classification.

Subnetworks Extraction

For comprehensively evaluating the recovered networks, subnetworks extraction was executed. The extraction is experiment-dependent, for each experiment, node importance was ranked by average strength across all time windows:

$$s_j = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{C} FN_{i,j}^{(n)} \tag{7}$$

where $FN_{i,j}^{(n)}$ is the functional connectivity between channels i, j in the n-th window, s_j is the average node strength of channel j across all windows.

Top-k nodes were selected based on sorted s_j, and corresponding $k \times k$ subnetworks SFN were extracted:

$$SFN = FN^{(n)}[rank_{1:k}, rank_{1:k}] \tag{8}$$

where k is determined by the selection rate, $SFN \in \mathbb{R}^{k \times k}$ is the reduced subnetwork for window n.

Dataset and Classification Settings

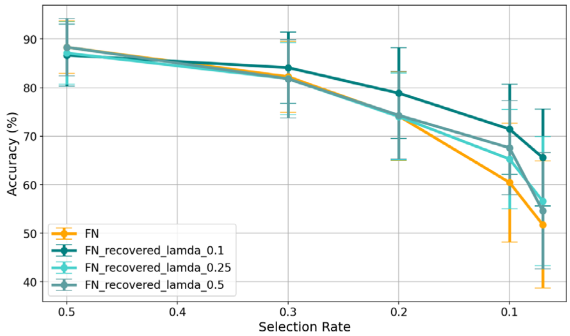
We evaluated the method on the SEED dataset [2], which includes 62-channel EEG recordings from 15 subjects across 3 emotion categories (negative, neutral, positive) induced by movie clips. Each subject completed three sessions of 15 clips.

A lightweight 10-layer multiscale CNN was used, featuring three parallel branches with 3×3, 5×5, and 7×7 convolutions, followed by batch normalization, ReLU, max pooling, and concatenation. The output was passed through three fully connected layers (64×3, 128, 3 neurons), ending with softmax for 3-class prediction.

A subject-dependent 5-fold cross-validation was applied, with training and testing performed individually on each subject's data, ensuring no data leakage across subjects.

Results

The evaluation results of classification experiments are shown in Figure 1-2. To assess the robustness of the proposed method, evaluations were performed under four different values of the regularization parameter λ. Additionally, classification tests were carried out on subnetworks extracted at various selection rates: 0.5, 0.3, 0.2, 0.1, and 0.07. For a full functional connectivity matrix of size 62×62, these correspond to subnetworks of size 31×31, 18×18, 12×12, 6×6, and 4×4, respectively.



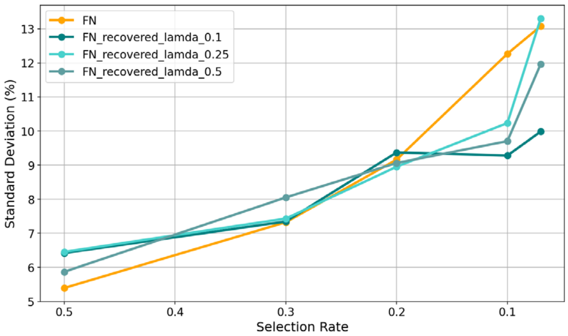


Figure 1 (left): Line plot comparing classification accuracy across different selection rates using the original functional networks FN and the recovered networks under varying regularization parameters (λ=0.1,0.25,0.5).

Figure 2 (right): Standard deviation of classification accuracy across different selection rates for the original functional networks and recovered networks with varying regularization parameters.

Discussion

Figure 1-2 demonstrate the effectiveness of the proposed inverse diffusion-based recovery method in EEG emotion classification. Recovered functional networks consistently outperformed the original networks, especially at low selection rates (<0.3). This indicates that the recovery method effectively reduces spurious correlations-likely due to volume conduction-thus preserving meaningful features. At extreme sparsity (0.1, 0.07), original networks performance drops sharply, while recovered networks remain stable, showing its robustness under dimensionality reduction. Recovered networks also yield lower standard deviation across validation folds, particularly with λ = 0.1, indicating improved model stability.

These results confirm that inverse diffusion recovery enhances both accuracy and consistency of EEG-based classification. It enables the use of smaller, denoised subnetworks without sacrificing performance, making it suitable for efficient and compact affective computing systems.

Conclusions

This study introduced a framework of recovery of functional network by mitigating spurious connections, such as those from volume conduction, using inverse diffusion filtering. The recovered networks served as input to a CNN classifier and were evaluated across varying selection rates and regularization levels.

The results obtained showed that recovered networks consistently outperformed original ones in both accuracy and stability, particularly under aggressive subnetwork reduction. The use of average node strength enabled efficient and interpretable subnetwork selection.

Overall, the method offers a principled approach to denoising functional connectivity and a robust strategy for compact, effective classification—suitable for real-time affective computing. Future work will target subject-independent models, adaptive subnetwork strategies, and multimodal integration.

References

[1] O. Scherzer and J. Weickert, "Relations between regularization and diffusion filtering," J. Math. Imaging Vis., vol. 12, pp. 43–63, Feb 2000.

[2] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," IEEE Trans. Auton. Ment. Dev., vol. 7, no. 3, pp. 162–175, May 2015.