# Data-Efficient Psychiatric Disorder Detection on Frequency-enhanced Brain Networks via Self-supervised Learning

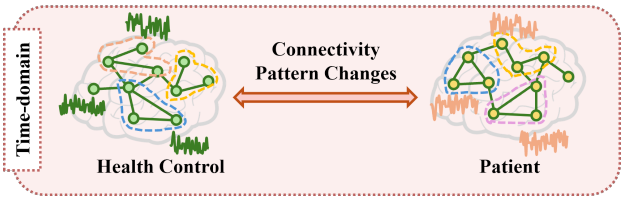
***Abstract*—Psychiatric disorders involve complex neural activity changes, with functional magnetic resonance imaging (fMRI) data serving as key diagnostic evidence. However, data scarcity and the diverse nature of fMRI information pose significant challenges. While graph-based self-supervised learning (SSL) methods have shown promise in brain network analysis, they primarily focus on time-domain representations, often overlooking the rich information embedded in the frequency domain. This omission limits the models' ability to fully exploit the available data, resulting in suboptimal data efficiency. To overcome these limitations, we propose Frequency-Enhanced Network (FENet for abbreviation), a novel SSL framework specially designed for fMRI data that integrates time-domain and frequency-domain information to improve psychiatric disorder detection in small-sample datasets. FENet constructs multi-view brain networks based on the inherent properties of fMRI data, explicitly incorporating frequency information into representation learning. Additionally, it employs domain-specific encoders to capture temporal-spectral characteristics, including an efficient frequency-domain encoder that highlights disease-relevant frequency features. Finally, FENet introduces a domain consistency-guided learning objective, balancing utilizing diverse information and generating frequency-enhanced brain graph representations. Experiments on two real-world medical datasets demonstrate that FENet outperforms state-of-the-art methods by $24.6\%$ and $17.7\%$ in detection accuracy while maintaining strong performance in minimal data conditions. Furthermore, we analyze the correlation between various frequency-domain features and psychiatric disorders, emphasizing the critical role of high-frequency information in disorder detection. Our code is available at https://anonymous.4open.science/r/FENet-1B81.**

***Index Terms*—Psychiatric disorder detection, data scarcity, brain networks, self-supervised learning, frequency-enhanced brain graph representation.**

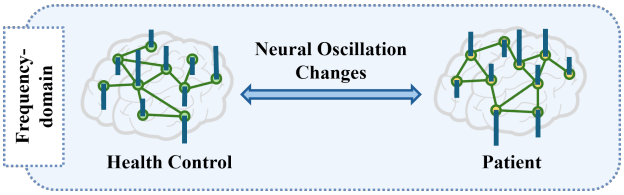
## I. Introduction

PSYCHIATRIC disorders, including depression, autism spectrum disorder (ASD), and attention deficit hyperactivity disorder (ADHD), are complex mental health conditions characterized by distinct neurological and behavioral patterns on brain networks [1]–[3]. Functional magnetic resonance imaging (fMRI) measures brain activity and connectivity by detecting blood-oxygen-level-dependent (BOLD) signal fluctuations [4], [5], offering valuable insights into neural function and dysfunction. Consequently, fMRI has been widely used to investigate neurodevelopmental and brain disorders' functional abnormalities, facilitating diagnosis and biomarker discovery [6], [7].

However, effectively analyzing fMRI data for psychiatric disorder detection remains challenging due to data scarcity and the complexity of its diverse information [8]–[10]. A major hurdle is the difficulty in collecting large, high-quality datasets, as psychiatric diagnosis often requires long-term clinical observation and expert evaluation [11], [12]. Additionally,




(a) Time-domain brain networks for the health control and patient




(b) Frequency-domain brain networks for the health control and patient

Fig. 1: Illustration of fMRI information diversity in psychiatric disorder detection. Brain network analysis from the health control and patient fMRI data reveals that disease states correlate with connection pattern changes in the time domain and neural oscillation changes in the frequency domain.

high neuroimaging costs and ethical constraints further limit data availability, resulting in small, imbalanced datasets with inconsistent preprocessing, making it difficult to train robust diagnostic models. Beyond data limitations, fMRI inherently contains both time-domain (temporal) and frequency-domain (spectral) information. However, integrating these complementary domains is challenging due to their distinct representations and the complexity of brain signals. Specifically, time-domain features capture BOLD signal fluctuations, modeling the functional connectivity of brain networks. As shown in Figure 1a, changes in connectivity patterns reflect transitions from healthy to diseased states [13]. In contrast, frequency-domain features capture neural oscillations, revealing intrinsic brain rhythms and disease-specific frequency changes [14], [15]. For example, in ASD, low-frequency correlations decline with age, while high-frequency signals highlight disorder traits [16], [17]. Figure 1b shows differences in spectral coefficients between healthy and patient groups. This diversity complicates the diagnostic process, as existing models often struggle to capture the full spectrum of information. Thus, time-frequency joint modeling is crucial for accurate psychiatric disorder detection as integrating frequency information provides complementary insights and improves data efficiency, particularly in small-sample settings.

Recent advancements have increasingly focused on graph learning-based approaches, with graph neural networks

(GNNs) [18]–[20] emerging as a powerful tool for fMRI data analysis. Given their ability to model non-Euclidean structures, GNNs are particularly well-suited for brain network construction and extracting complex representations that reflect brain function [21]–[25]. For example, BrainGNN [26] utilizes ROI-aware GNN to enable biomarker discovery, while BrainGB [27] sets a benchmark summarizing the brain network construction pipelines and modularizing the process of GNN for brain network analysis. However, these models heavily rely on large labeled datasets, often scarce in psychiatric disorder research. To mitigate data scarcity, recent studies have explored self-supervised learning (SSL) on graphs [28], [29] to enable representation learning from unlabeled data. Some SSL methods select positive and negative sample pairs in augmented brain networks, which encourages better representation learning at the node level [30]–[33]. Other approaches leverage the temporal characteristics of fMRI data by constructing augmented views across different time windows aiming to capture the inherent temporal dependencies in the representations [34], [35].

Despite their promising performance, existing SSL methods encounter several limitations when applied to fMRI data, primarily due to their neglect of frequency information, leading to suboptimal data efficiency. First, most SSL approaches rely heavily on time-domain data augmentation strategies, such as edge or attribute perturbation [30]–[33]. However, these operations often disrupt the structural integrity of brain networks [34], potentially distorting their inherent time-frequency properties and leading to inauthentic representations. To address this, we construct biologically consistent multi-view representations of fMRI data across both time and frequency domains. Second, existing SSL methods typically adopt time-domain encoders, which face inherent limitations in modeling the complex brain functions and intrinsic cyclic patterns that are better captured in the frequency domain [36], [37]. As a result, these methods often fail to exploit frequency-related neural activity, overlooking critical features that are particularly valuable in limited-sample scenarios. To remedy this, we propose carefully designed pretext tasks to incorporate frequency-domain analysis and facilitate effective time-frequency joint modeling. Third, most traditional SSL objectives are designed to learn representations within a single time domain, lacking the capability to align cross-domain features. This makes it difficult to effectively integrate complementary information from both time and frequency domains, resulting in incomplete utilization of the available data. To address this, we refine the SSL objective to support cross-domain representation alignment, enabling the model to learn joint time-frequency embeddings that are more comprehensive and discriminative.

In this paper, we propose FENet, a novel SSL framework that integrates time- and frequency-domain analysis to capture diverse information from fMRI data, enhancing data efficiency and improving diagnostic performance in limited-sample scenarios. Specifically, FENet focuses on biologically consistent modeling of multi-view brain networks in time and frequency domains, offering complementary perspectives on fMRI data. Meanwhile, FENet employs dedicated domain-specific encoders: a time-domain GNN (TGNN) and a frequency-domain GNN (FGNN)), which collaboratively perform time-frequency joint modeling, fully leveraging the complementary information to enhance data efficiency. Notably, the FGNN incorporates adaptive graph filtering to selectively capture informative frequency components, while the integration of Fourier Graph Operator (FGO) layers [14] reduces computational complexity to a log-linear scale, enabling efficient and scalable frequency-domain analysis in complex brain networks. Furthermore, FENet employs a learning objective based on canonical correlation analysis (CCA) [38], guided by the domain consistency concept to capture domain-invariant features and optimize both time- and frequency-domain representations. This enables FENet to generate frequency-enhanced brain network representations for effective psychiatric disorder detection. The contributions of this work can be summarized as follows:

- We propose FENet, a novel non-contrastive SSL framework for data-efficient psychiatric disorder detection, which is the first to integrate time- and frequency-domain analysis for brain network representation learning.
- We introduce a novel data augmentation strategy that constructs brain networks in both the time and frequency domains, ensuring biological consistency with real brain networks while effectively modeling the diverse information embedded in fMRI data.
- We design a frequency-domain information encoder, which enables the selective learning of the frequency components for the psychiatric disorders with the proposed graph filter and FGO layers.
- We conduct extensive experiments on two real-world medical datasets, demonstrating that FENet achieves state-of-the-art performance in limited-sample settings. Additionally, frequency domain analysis reveals a significant correlation between high-frequency components and psychiatric disorders.

The structure of the paper is as follows: Section II reviews related work. Then, the preliminary definitions and notations are introduced in Section III. Section IV presents the details of the proposed FENet method. Section V outlines the experimental setup, followed by the results and their discussion in Section VI. Lastly, Section VII concludes the paper.

## II. Related Work

In this section, we introduce the most related works from three perspectives: graph representation learning for brain signals, SSL on graphs, and frequency-based time series analysis.

### A. Graph Representation Learning for fMRI Data

Graph representations of brain signals are widely recognized for their effectiveness in capturing the underlying dynamics of brain functions using GNNs [39], [40]. In these models, nodes typically correspond to regions of interest (ROIs), while edges that quantify functional connectivity between different ROIs are generally defined by the correlations between their respective brain signals. Building on this concept, BrainGNN [26] introduces a novel ROI-aware GNN with an ROI-selection pooling layer highlighting the salient brain ROIs. Recently,

BrainGB [27] further established a benchmark that formulates the implementation of GNN models for brain network analysis, systematically demonstrating how different GNN mechanisms influence brain network modeling and performance. Despite the promising results achieved by these approaches in capturing the features of brain networks, they are still limited by the scarcity of labeled data. To address this challenge, SSL-based methods have emerged as an effective solution, which involves pre-training with limited data to learn effective representations, followed by fine-tuning to improve generalization and accuracy. For instance, A-GCL [30] and SF-GCL [41] investigate contrastive learning (CL) for brain signals by introducing random edge manipulation to generate augmented data views. Besides, BraGCL [31] improves model interpretability by perturbing unimportant nodes and edges. Further, GATE [34] enhances the pair constructions for SSL. UCGL [35] extends this approach by integrating a spatio-temporal graph convolution technique. While these SSL-based methods have improved the representation learning of brain signals, they focus exclusively on time-domain features. Consequently, they fail to comprehensively capture information in the frequency domain and their correlation with psychiatric disorders. To bridge this gap, our approach explores the frequency-domain characteristics of brain signals and integrates them with time-domain details, enabling a more comprehensive representation.

### B. Self-supervised Learning on Graphs

Existing SSL-based methods [42]–[45] typically construct pairs from the input through augmentation and learn the representations by maximizing agreement between positive (similar) pairs while minimizing the similarity between negative (dissimilar) pairs. The effectiveness of SSL heavily depends on negative pair setting and the learning objective functions. Data augmentation techniques are widely used in negative sample constructions. For example, GraphCL [46] employs edge perturbation, node feature masking, or subgraph sampling, while GCA [47] integrates structural and attribute information for adaptive augmentation. Additionally, some methods leverage graph spectral features, such as SpCo [48] and SFA [49] to guide the edge perturbation process, which emphasizes high-frequency information related to graph structures. However, these random perturbations can inevitably disrupt intrinsic biological structures in brain data, leading to implausible outcomes. In contrast, approaches like GATE [34] and UCGL [35] create augmented views based on the time-series properties, focusing on learning temporal dependencies instead of applying random perturbations. However, they still overlook a detailed examination of frequency information. To address this issue, we construct graphs that preserve the biological features while also incorporating the temporal-spectral characteristics of fMRI data for better multiple graph view construction, both in the time and frequency domains.

In terms of the learning objective functions, they are generally categorized into contrast-based and similarity-based approaches. Contrastive-based methods, such as GraphCL [46] and GCA [47], require careful design of positive and negative sample pairs. However, constructing negative pairs is generally challenging, especially in data-limited conditions [50], which impacts their effectiveness. On the other hand, the similarity-based methods do not rely on selecting negative pairs. Examples include CCA-SSG [38], which simplifies SSL via CCA that aims at maximizing the correlation between two augmented views [51]. Similarly, GATE [34] employs CCA for brain signal analysis to capture the correlation between representations of adjacent augmented views in the window of time. However, no prior work has explored using CCA to align time- and frequency-domain representations. We are the first to introduce the concept of domain consistency to guide CCA in maximizing the correlation of representation from different domains.

### C. Frequency-Based Time Series Analysis

Brain signals are inherently time series data, with frequency information capturing oscillatory patterns and functional connectivity that play critical roles in many diseases [14], [15]. Consequently, incorporating frequency information is essential in representation learning. Existing methods have made significant progress in this area. Techniques such as Fourier Transform decompose time series into frequency components, revealing patterns and periodicities at various scales. These methods have been widely applied in time series analysis, including FEDformer [52], Fredformer [53], FreTS [54]. A few studies have incorporated frequency information in graph learning, such as StemGNN [55] and FourierGNN [56]. However, these methods focus solely on the frequency domain and overlook the time domain, which also contains complementary information of fMRI data. However, these approaches focus solely on frequency-domain features while neglecting the complementary information present in the time domain. This limitation restricts their applicability, particularly in scenarios involving limited labeled fMRI datasets, where fully leveraging both domains is crucial for effective representation learning.

BTSF [57] is the first to integrate time-frequency affinity into the SSL paradigm, enhancing the discriminative power of frequency feature fusion. Furthermore, Zhang et al. [36] introduce a pre-training model to uncover general properties connecting time and frequency information and leverage SSL to capture these properties. Despite these advancements, there is a lack of studies examining the time-frequency integration in brain network modeling. Integrating frequency information into the traditional SSL approaches for brain networks remains challenging due to a lack of prior knowledge on biologically relevant frequency features and the high computational cost of extracting frequency-domain features from graphs. To overcome these issues, FENet introduces carefully designed pretext tasks that examine the contribution of different frequency components to disease detection, facilitating in-depth frequency analysis and enhancing representation learning for brain networks. Additionally, it incorporates efficient FGO layers, reducing the computational complexity to log-linear complexity.

TABLE I: Notations and explanations related to FENet framework. The four blocks (from top to bottom) show the notation of variables about brain networks, graph Fourier transformer, SSL, and FENet's hyper-parameters, respectively.

| Notation | Description |
|---|---|
| $\mathcal{G}_{\mathrm{T}} = (\mathcal{V}, \mathcal{E}, \mathbf{X}_{\mathrm{T}}, \mathbf{A})$ | Time-domain brain networks |
| $\mathcal{G}_{\mathrm{F}} = (\mathcal{V}, \mathcal{E}, \mathbf{X}_{\mathrm{F}}, \mathbf{A})$ | Frequency-domain brain networks |
| $\mathcal{V}$ | The node set of the brain network |
| $\mathcal{E}$ | The edge set of the brain network |
| $\mathbf{A} \in \mathbb{R}^{N \times N}$ | Adjacency matrix of the brain network |
| $\mathbf{X}_{\mathrm{T}} \in \mathbb{R}^{N \times D}$ | Time-domain node feature matrix with $D$-dimensional BOLD signals, $\mathbf{X}_{\mathrm{T}} = [\mathbf{x}_{\mathrm{T}}^1, \ldots, \mathbf{x}_{\mathrm{T}}^N], \mathbf{x}_{\mathrm{T}} \in \mathbb{R}^D$ |
| $\mathbf{X}_{\mathrm{F}} \in \mathbb{R}^{N \times D}$ | Frequency-domain node feature matrix with $D$-dimensional spectral coefficient, $\mathbf{X}_{\mathrm{F}} = [\mathbf{x}_{\mathrm{F}}^1, \ldots, \mathbf{x}_{\mathrm{F}}^N], \mathbf{x}_{\mathrm{F}} \in \mathbb{R}^D$ |
| $N$ | Number of nodes (ROIs) in the brain network |
| $\mathbf{D} \in \mathbb{R}^{N \times N}$ | Diagonal matrix of the graph |
| $\mathbf{L} \in \mathbb{R}^{N \times N}$ | Laplacian matrix of the graph |
| $\mathbf{I} \in \mathbb{R}^{N \times N}$ | Identity matrix |
| $\mathbf{U} \in \mathbb{R}^{N \times N}$ | Orthonormal matrix of eigenvectors |
| $\mathbf{\Lambda} \in \mathbb{R}^{N \times N}$ | Diagonal matrix with eigenvalues, $\mathbf{\Lambda} = \mathrm{diag}(\lambda)$ |
| $\mathbf{S} \in \mathbb{R}^{K \times K}$ | Fourier graph operator |
| $\mathbf{H} \in \mathbb{R}^{N \times N}$ | Graph filter matrix |
| $\mathbf{W} \in \mathbb{R}^{N \times N}$ | Learnable parameter matrix |
| $\mathbf{Z}_{\mathrm{T}}, \mathbf{Z}_{\mathrm{F}}, \mathbf{Z}_{\mathrm{TF}} \in \mathbb{R}^{N \times D}$ | Time-domain, frequency-domain and fused brain graph representations |
| $K_{\mathrm{L}}, K_{\mathrm{H}}$ | Parameters for frequency threshold setting |
| $\gamma, \beta$ | Trade-off coefficient for learning objective |

## III. Preliminaries

In this section, we provide preliminaries of this paper, including spectral graph theory, the concept of domain consistency, and the criterion of CCA. Table I summarizes the notations that are frequently used in this paper, where bold lowercase letters (e.g., $\mathbf{x}$), bold uppercase letters (e.g., $\mathbf{X}$), and calligraphic fonts (e.g., $\mathcal{V}$) to denote vectors, matrices, and sets, respectively.

### A. The Spectral Graph Theory

In time series data, frequency components are characterized by periodic patterns in the signal, capturing changes or rates of oscillation over time. Similarly, in graph-structured data, the frequency component is defined by variations in the signal among connected nodes, reflecting how the signal changes between the neighboring nodes. Low-frequency information in graph-structured data suggests that signals among neighboring nodes are similar or aligned, while high-frequency information indicates greater variability or differences in the signal between connected nodes. Graph Fourier Transform (GFT) is commonly adopted to represent frequency information in graph data [58].

For a given graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathbf{A})$, where $\mathcal{V}$ and $\mathcal{E}$ represent the set of nodes and edges, $\mathbf{X} \in \mathbb{R}^{N \times D}$ denotes the graph signals, $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix encoding the connection weights between node pairs, and $N$ is the number of nodes, the GFT aims to map the graph signal into the spectral domain. This is achieved through eigenvector-based feature decomposition of the graph Laplacian matrix, defined as:

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \tag{1}$$

where $\mathbf{D} \in \mathbb{R}^{N \times N}$ is a diagonal matrix encoding the degree of each node. The Laplacian matrix $\mathbf{L} \in \mathbb{R}^{N \times N}$ is real, symmetric, and positive semi-definite, allowing for eigendecomposition:

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top, \tag{2}$$

where $\mathbf{U} \in \mathbb{R}^{N \times N}$ is the orthonormal matrix of eigenvectors and $\mathbf{\Lambda} = \mathrm{diag}(\lambda)$ is the diagonal matrix of eigenvalues. The eigenvalues of the graph Laplacian, indexed as $0 < \lambda_1 \leq \cdots \leq \lambda_N$, are referred to as the graph frequency components. Their associated eigenvectors constitute the Fourier modes. These components provide a foundation for analyzing the graph signals in the spectral domain. Thus, for the given graph signal $\mathbf{X}$, using the GFT, one can transform $\mathbf{X}$ from the spatial domain to the spectral (frequency) domain as follows:

$$\mathbf{X}_{\mathrm{F}} = \mathbf{U}^\top \mathbf{X}. \tag{3}$$

As $\mathbf{U}$ is the orthonormal matrix, the inverse graph Fourier transform (IGFT) is given by:

$$\mathbf{X} = \mathbf{U}\,\mathbf{X}_{\mathrm{F}}. \tag{4}$$

To preserve certain information of graph data in the frequency domain, the process involves filtering, which is achieved by first transforming the signal using the GFT to the frequency domain, conducting the filtering, and then converting it back to the time domain as follows:

$$\widetilde{\mathbf{X}} = \mathbf{U}\,\mathbf{H}\,\mathbf{U}^\top \mathbf{X}. \tag{5}$$

Here, the $\mathbf{H} \in \mathbb{R}^{N \times N}$ is a diagonal filtering matrix that offers flexibility in analyzing specific frequency components and $\widetilde{\mathbf{X}}$ is the filtered feature matrix. For example, larger eigenvalues are associated with higher variance eigenvectors, which can be interpreted as higher graph frequencies [16].Consequently, the frequency information of graph signals can be selectively retained by applying filters $\mathbf{H}$.

### B. Domain Consistency Concept

The concept of domain consistency between time and frequency information is first introduced by Zhang et al. [36] in the context of SSL for time series modeling. It aims to align the time- and frequency-domain representations by emphasizing the domain invariant feature learning in latent space. For a given time series data $\mathbf{X} = [\mathbf{x}^1, \ldots, \mathbf{x}^N]$, it is assumed that the latent representations of the time domain $\mathbf{z}_{\mathrm{T}}^i$ and frequency domain $\mathbf{z}_{\mathrm{F}}^i$ derived from the same original time series $\mathbf{x}^i, i \in [1, N]$ should closely align, while representations from different original time series $\mathbf{x}^j, j \in [1, N]$ should be pushed farther apart. This concept guides the learning objective to learn the domain invariant features effectively, thereby achieving the integration of information from both domains effectively.

### C. Canonical Correlation Analysis

CCA [51] is a statistical technique used to analyze the correlations between two sets of variables. It aims to find a linear transformation that maximizes the correlation between the transformed variables. It has been widely applied in multi-view fusion and fMRI analysis [34], [38], [59], to enhance the understanding of relationships between different views of a graph. Specifically, assuming $\mathcal{G}_a$ and $\mathcal{G}_b$ are two views of the graph, it optimizes the correlation:

$$\begin{aligned} &\max_{\theta_1,\theta_2} \mathcal{L}_{CCA} := \mathbb{E}_{\mathcal{G}_a,\mathcal{G}_b}\left[f_{\theta_1}^{\top}(\mathcal{G}_a) f_{\theta_2}(\mathcal{G}_b)\right], \\ &\text{s.t. Cov}\left(f_i(\mathcal{G}_j), f_i(\mathcal{G}_j)\right) = \mathbf{I}, (i=\theta_1,\theta_2; j=a,b), \end{aligned} \tag{6}$$

where $f_{\theta_1}(\cdot)$ and $f_{\theta_2}(\cdot)$ are two feedforward neural networks normally sharing parameters, $\text{Cov}(\cdot,\cdot)$ represents the covariance matrix and $\mathbf{I}$ is an identity matrix. The learned feedforward neural networks provide mappings that align the representations of nodes in the two views. Since CCA does not rely on the non-trivial construction of negative sample pairs, it offers greater flexibility for graph-level representation learning in SSL settings while also being scalable to large and complex graphs. However, traditional CCA is computationally expensive. To address this issue, a recent study proposed soft CCA [60], which relaxes the hard decorrelation constraints by applying the Lagrangian theorem within convex optimization. The formulation of soft CCA is defined as follows:

$$\begin{aligned} &\min_{\theta_1,\theta_2} \mathcal{L}_{DIST}\left(f_{\theta_1}(\mathcal{G}_a), f_{\theta_2}(\mathcal{G}_b)\right) + \\ &\beta\left(\mathcal{L}_{SDL}\left(f_{\theta_1}(\mathcal{G}_a)\right) + \mathcal{L}_{SDL}\left(f_{\theta_2}(\mathcal{G}_b)\right)\right), \end{aligned} \tag{7}$$

where $\mathcal{L}_{DIST}(\cdot,\cdot)$ is used to measure the correlation between two view representations, $\mathcal{L}_{SDL}(\cdot)$, referred to as random decorrelation loss, computes the $L_1$ distance between each representation and the identity matrix functioning as the regularization, and the $\beta$ is a trade-off coefficient. The decorrelation term can be viewed as a soft constraint, as the loss is minimized rather than strictly enforced to zero. By jointly optimizing the decorrelation loss $\mathcal{L}_{SDL}(\cdot)$ alongside other objectives, such as minimizing the distance between views in the embedding space, the model can explore a broader set of globally optimal solutions. Therefore, soft CCA can effectively enhance the correlation between two graph views after transformation while also enforcing regularization to reduce computational costs and avoid overfitting in SSL settings.

## IV. The Proposed Method

The overall framework of the proposed FENet model is illustrated in Figure 2. The core idea can be summarized as follows:(i) Constructing biologically meaningful brain networks from fMRI data in both the time and frequency domains; (ii) Employing domain-specific encoders to fully exploit the time-frequency characteristics in multi-view brain networks; (iii) Leveraging a domain consistency objective to integrate time- and frequency-domain representations.

### A. Problem Definition

Let $\{\mathbf{X}^1, \mathbf{X}^2 \ldots, \mathbf{X}^M\}$ represent $M$ samples of fMRI data and $\{y^1, y^2, \ldots, y^M\}$, $y^m \in \{0,1\}$ represent the corresponding labels. Each of $\mathbf{X}^m = [\mathbf{x}^{1,m}, \ldots, \mathbf{x}^{N,m}] \in \mathbb{R}^{N\times D}$ represents the BOLD signals for $N$ ROIs in brain networks and $D$ is the dimension of each signal. Given $\mathbf{X}$, the main objective of this study is to find the representation of $\mathbf{Z}_{\text{TF}}$, which effectively integrates information from both the time and frequency domains. Specifically, $\mathbf{Z}_{\text{TF}}$ is obtained by training a mapping function $f_{\theta_{\text{T}},\theta_{\text{F}}}$, where $\theta_{\text{T}}$ and $\theta_{\text{F}}$ denote the parameters of the time-domain and frequency-domain encoders, respectively. These encoders transform the input fMRI data into meaningful representations for downstream tasks. To this end, we formulate the task as a graph classification problem, aiming to optimize the mapping function such that $f_{\theta_{\text{T}},\theta_{\text{F}}}(\mathbf{X}) \to y$.

### B. Multi-view Brain Network Construction

Unlike traditional augmentation methods that may disrupt the underlying brain network structure, our approach leverages the temporal-spectral characteristics of brain signals to construct distinct yet semantically consistent multi-view brain networks in both the time and frequency domains. This strategy provides complementary views of the fMRI data while preserving biologically meaningful structures throughout the augmentation process.

*1) Time-domain Brain Graph:* To ensure biological relevance, both brain network views are constructed based on functional connectivity, a widely adopted modeling approach in clinical research [5]. For the time-domain brain network $\mathcal{G}_{\text{T}} = (\mathcal{V}, \mathcal{E}, \mathbf{X}_{\text{T}}, \mathbf{A})$ , the $N$ ROIs are treated as the node set $\mathcal{V}$with each node's feature represented by its corresponding BOLD signals, forming the feature matrix $\mathbf{X}_{\text{T}}$. The functional connectivity defines the correlations between pairs of ROIs, computed using the Pearson correlation coefficient(PCC) [61] as follows:

$$a^{ij} = \frac{\sum_{d=1}^{D}(x^{i,d}-\bar{x}^i)(x^{j,d}-\bar{x}^j)}{\sqrt{\sum_{t=1}^{D}(x^{i,d}-\bar{x}^i)^2}\sqrt{\sum_{d=1}^{D}(x^{j,d}-\bar{x}^j)^2}}, \tag{8}$$

where $x^{i,d}$ and $x^{j,d}$ are the BOLD signals of node $i$ and $j$ at $d$-th dimension, $\bar{x}^i$ and $\bar{x}^j$ denote the mean values of BOLD signals over time. The resulting connection strengths $a^{ij}$ representing the connection strength between node $i$ and node $j$ are organized into the adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$, reflecting the pairwise relationships across all ROIs. Finally, the edge set $\mathcal{E}$ retains the top $20\%$ strongest connections to form the final time-domain brain network structure.

*2) Frequency-domain Brain Graph:* Unlike the traditional Fourier Transform performs on time series data, our approach leverages the GFT on graph structure data to generate frequency-domain brain networks. This enables the extraction of spectral representations and the identification of intrinsic neural oscillations that align with brain connectivity patterns. Specifically, given a time-domain brain graph $\mathcal{G}_{\text{T}} = (\mathcal{V}, \mathcal{E}, \mathbf{X}_{\text{T}}, \mathbf{A})$, we transform the brain signals $\mathbf{X}_{\text{T}}$ from the time domain to the spectral domain $\mathbf{X}_{\text{F}}$ using the GFT operation $F_G(\cdot)$: $\mathbf{X}_{\text{F}} = F_G(\mathbf{X}) = \mathbf{U}^{\top}\mathbf{X}_{\text{T}}$. This formulation allows

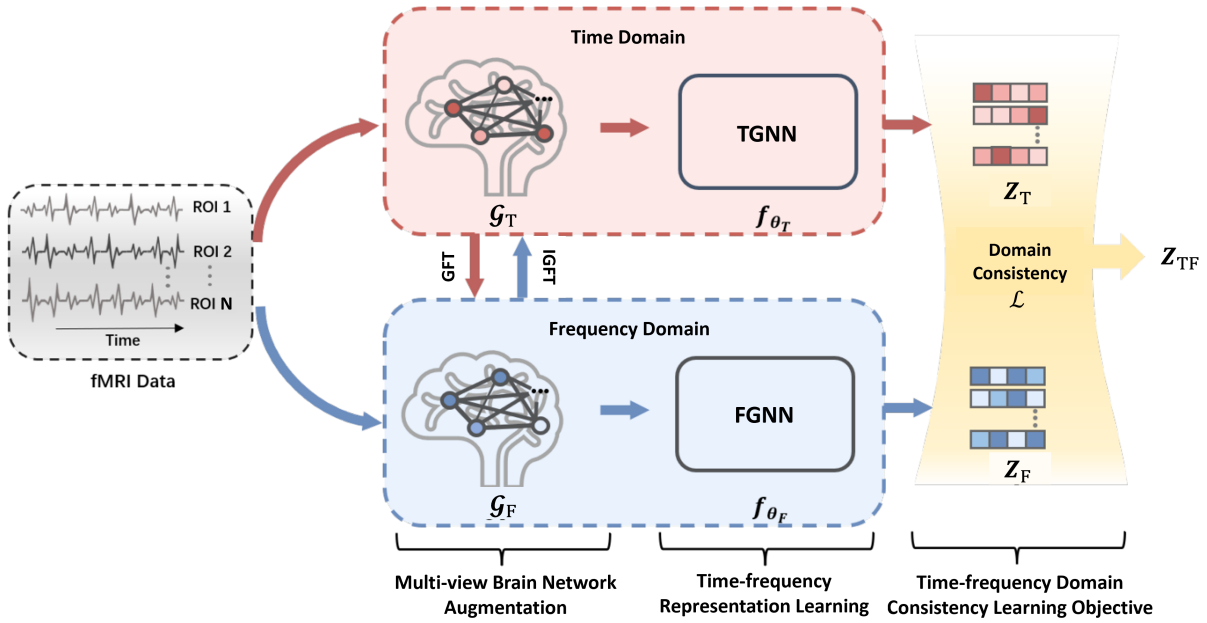



Fig. 2: Overview of FENet model. The FENet model consists of three modules: multi-view graph construction, time-frequency representation learning, and learning objective.

us to explicitly introduce frequency information in the model and analyze brain activity patterns from a spectral perspective as $\mathcal{G}_{\mathrm{F}} = (\mathcal{V}, \mathcal{E}, \mathbf{X}_{\mathrm{F}}, \mathbf{A})$, where $\mathcal{V}, \mathcal{E}$, and $\mathbf{A}$ are defined the same as its time-domain brain graph. Here, $\mathbf{X}_{\mathrm{F}} \in \mathbb{R}^{N \times D}$ is the transformed feature matrix in the graph spectral domain, where each row of $\mathbf{X}_{\mathrm{F}}$ represents a spectral coefficient corresponding to an eigenvalue $\lambda_i$.

### C. Time-frequency Representation Learning

Diverging from conventional SSL models that focus solely on encoding two graph views within the time domain, our approach employs distinct graph encoders for multi-view brain networks. Specifically, we define the TGNN encoder $f_{\theta_{\mathrm{T}}}(\cdot)$ for the time domain and FGNN encoder $f_{\theta_{\mathrm{F}}}(\cdot)$ for the frequency domain. This tailored design facilitates more effective encoding of the complementary information inherent in fMRI data.

*1) Time-domain Graph Neural Network Encoder:* Graph convolutional networks (GCNs) [18], [62] are employed as the TGNN encoder due to their effectiveness in handling non-Euclidean and irregular data structures, which are characteristic of fMRI data. Specifically, we leverage GCNs to model the interdependencies between BOLD signals across ROIs, enabling time-domain feature extraction by aggregating information from neighboring ROIs. More specifically, we obtain the node features $\mathbf{X}_{\mathrm{T}}^{(l+1)}$ at the $(l+1)$-th layer from the preceding layer as:

$$\mathbf{X}_{\mathrm{T}}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1/2}\mathbf{X}_{\mathrm{T}}^{(l)}\mathbf{W}^{(l)}\right), \tag{9}$$

where $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ is the adjacency matrix of the graph with added self-loops, $\tilde{\mathbf{D}}$ is the diagonal degree matrix of $\tilde{\mathbf{A}}$, $\mathbf{W} \in \mathbb{R}^{N \times N}$ is the learnable parameter matrix, $\sigma(\cdot)$ denotes the activation function, and $\mathbf{X}_{\mathrm{T}}^{(l)}$ denotes the node features in $l$-th hidden layer, where it's noteworthy that the node feature in the 0-th layer hidden layer is the input itself. Finally, we consider the node features learned in the last layer of graph convolution as the time-domain representation, which denotes as $\mathbf{Z}_{\mathrm{T}} = f_{\theta_{\mathrm{T}}}(\mathcal{G}_{\mathrm{T}}), \mathbf{Z}_{\mathrm{T}} \in \mathbb{R}^{N \times D}$.

*2) Frequency-domain Graph Neural Network Encoder:* To efficiently capture disease-related frequency information with acceptable computational complexity, we introduce a novel frequency-domain encoder, FGNN, as shown in Figure 3. This module consists of two primary components: a graph filter and FGO layers. The graph filter captures the crucial frequency components by integrating biological-inspired prior knowledge. Meanwhile, the FGO network introduces log-linear multiplication to replace standard linear multiplication, improving model scalability.

**Graph Filter.** Spectral graph theory provides a powerful framework for analyzing the structural properties of brain networks, where the spectral components naturally correspond to frequency characteristics and offer insights into both global and local connectivity patterns. Specifically, the GFT decomposes BOLD signals into different graph frequency components. Low-frequency components, associated with smaller Laplacian eigenvalues, capture globally coherent brain activity, reflecting stable functional connectivity across the brain network. In contrast, high-frequency components, corresponding to larger eigenvalues, emphasize localized variations and abrupt changes, often linked to abnormal or disrupted brain regions—patterns commonly observed in psychiatric disorders [14]. This spectral decomposition enables in-depth frequency-based analysis of brain networks, facilitating both the examination of frequency-specific information and the extraction of informative features for downstream tasks.

In this study, we leverage the eigenvalues $\mathbf{\Lambda}$ of the brain net-

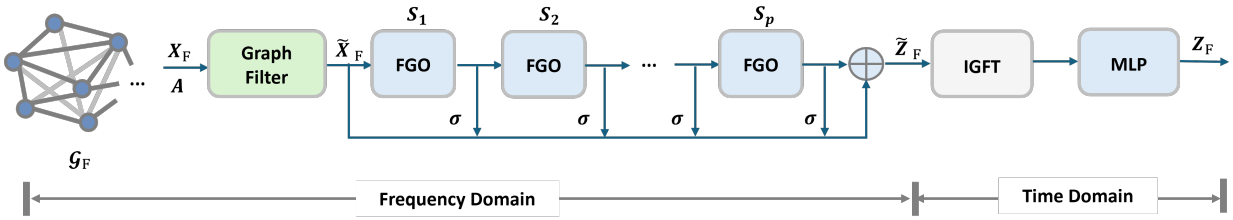



Fig. 3: The framework of the FGNN module consists of a graph filter and serval FGO network layers.

works to decompose brain frequency components into distinct bands: low-frequency components within $\lambda_{\min}$ and $\lambda_{\mathrm{L}}$, high-frequency components within $\lambda_{\mathrm{H}}$ and $\lambda_{\max}$, and the remaining range considered as middle-frequency components. To examine how different frequency components relate to cognitive states or mental disorders, we design corresponding graph filters. This enhances the learning at each specific frequency range, capturing more biologically relevant information and improving model explanation. Specifically, a low-pass graph filter $\mathbf{H}_{\mathrm{L}} = \mathrm{diag}(h_{\mathrm{L}}), \mathbf{H}_{\mathrm{L}} \in \mathbb{R}^{N\times N}$ is defined, where $h_{\mathrm{L}}$ takes the value of 1 for eigenvalues ($\lambda_{\min} < \lambda_i < \lambda_{\mathrm{L}}$) and 0 otherwise. Similarly, a band-pass filter $\mathbf{H}_{\mathrm{M}} \in \mathbb{R}^{N\times N}$ and a high-pass filter $\mathbf{H}_{\mathrm{H}} \in \mathbb{R}^{N\times N}$ are constructed based on their respective frequency bands. After filtering we obtain the filtered signals as $\mathbf{X}'_{\mathrm{F,\,H}} = \mathbf{H}_{\mathrm{H}}\mathbf{X}_{\mathrm{F}}$, $\mathbf{X}'_{\mathrm{F,\,M}} = \mathbf{H}_{\mathrm{M}}\mathbf{X}_{\mathrm{F}}$, and $\mathbf{X}'_{\mathrm{F,\,L}} = \mathbf{H}_{\mathrm{L}}\mathbf{X}_{\mathrm{F}}$ corresponding to the high-, middle-, and low-frequency components of the brain network, respectively. We then selectively integrate the frequency components that capture more salient information to construct the filtered feature matrix $\tilde{\mathbf{X}}_{\mathrm{F}} \in \mathbb{R}^{N\times K}$, where $K$ is the dimension of frequency components, which serves as input for frequency-domain representation learning.

**Fourier Graph Operator.** Encoding frequency-domain features on graphs often incurs high computational costs due to complex eigendecomposition and graph convolution operations. To alleviate this burden, FGNN incorporates an efficient learnable layer, the FGO, inspired by FourierGNN [56], significantly reducing computational overhead while effectively capturing frequency information. Specifically, FGNN extracts frequency information by using $N$-invariant FGOs, denoted as $\mathbf{S} \in \mathbb{R}^{N\times K\times K}$, enabling straightforward matrix multiplication within the frequency-domain brain networks, as shown in Figure 3. Here, $\mathbf{S}$ is a tailored Green's kernel $\kappa$ [56], a translation-invariant kernel where $\kappa[i,j] = \kappa[i-j]$ represents its invariance property when sliding from node $i$ to node $j$. This operation transforms the $([N]\times[N])$ matrix into a $\mathbb{R}^{K\times K}$ matrix-valued projection. The FGO parameterizes graph convolutions analogous to the shared-weight convolution kernels in traditional convolutional neural networks (CNNs). The operation is defined as: $F_G^{-1}(\mathbf{X}_{\mathrm{F}}\mathbf{S}_{0:p})$, which equals to $\mathbf{A}_{p:0}\mathbf{X}_{\mathrm{T}}\mathbf{W}_{0:p}$ in time domain. Here, $F_G^{-1}(\cdot)$ is the IGFT, $\mathbf{A}_p \in \mathbb{R}^{N\times N}$ represents as the $p$-th sparsity pattern of $\mathbf{A}$, $\mathbf{W}_p \in \mathbb{R}^{K\times K}$ is the $p$-th weight matrix. This convolutional kernel operates over the input features through matrix multiplications of size $([K]\times[K])$ in the graph's Fourier space. By stacking multiple FGO layers, the model can efficiently capture informative features while maintaining log-linear computational complexity. Thus, FGNN performs a recursive multiplication between the filtered frequency-domain brain graph embeddings $\tilde{\mathbf{X}}_{\mathrm{F}}$ and FGO layers $\mathbf{S}_{0:p}$ to obtain the representation $\tilde{\mathbf{Z}}_{\mathrm{F}} \in \mathbb{R}^{N\times K}$ in Fourier space. This process can be defined as:

$$\tilde{\mathbf{Z}}_{\mathrm{F}} = \sum_{p=0}^{P} \sigma\left(\tilde{\mathbf{X}}_{\mathrm{F}}\mathbf{S}_{0:p} + b_p\right), \mathbf{S}_{0:p} = \prod_{j=0}^{p} \mathbf{S}_j, \tag{10}$$

where $p$ is the number of FGO layers, $\mathbf{S}_j \in \mathbb{R}^{K\times K}$ is the $p$-th step FGO, $b_p \in \mathbb{R}^K$ are the complex-valued biases parameters.

Our model aims to minimize the distance between the time- and frequency-domain representations in the latent time-frequency space, thereby learning domain-invariant features. This objective is further discussed in Section IV-D. To achieve this, we transform the learned frequency-domain representations $\tilde{\mathbf{Z}}_{\mathrm{F}}$ in Fourier space back into the time-space, facilitating the alignment of representations from both domains and contributing to more effective model training. This conversion is achieved by applying IGFT operation $F_G^{-1}(\cdot)$ as follows: $F_G^{-1}(\tilde{\mathbf{X}}_{\mathrm{F}}) = \mathbf{U}\tilde{\mathbf{X}}_{\mathrm{F}}$. The resulting representation is then passed through a multilayer perceptron (MLP) to project it into the time-frequency latent space, ensuring alignment with the representation learned in the time domain. Finally, we denote the representation learned from the frequency domain as: $\mathbf{Z}_{\mathrm{F}} = f_{\theta_{\mathrm{F}}}(\mathcal{G}_{\mathrm{F}}), \mathbf{Z}_{\mathrm{F}} \in \mathbb{R}^{N\times D}$.

*3) Computation Complexity Analysis:* This section highlights the improved efficiency of encoding frequency-domain features in SSL settings due to the use of our FGNN. Assume the frequency-domain graphs consist of $N$ nodes and $|\mathcal{E}|$ edges, and let each node have $D$ dimension features. Traditional methods apply GCN that performs matrix multiplications for aggregating node features. This leads to a computational complexity of $O(2N|\mathcal{E}|^2+2ND^2)$. In contrast, our proposed FGNN replaces conventional GCN layers with an efficient feature encoding process involving graph filtering and recursive multiplications utilizing the efficient FGO layers. We suppose $K$ frequency components represent the filtered node features in the Fourier space, where $K \ll N$ and the FGO operator has size $K\times K$. We apply a third-order summation for the FGOs, yielding a total computational complexity for the frequency-domain encoder of $O(NKlogN+3NK^2)$, including GFT and IGFT operations. Overall, the log-linear complexity significantly reduces computational overhead while preserving essential frequency-domain information.

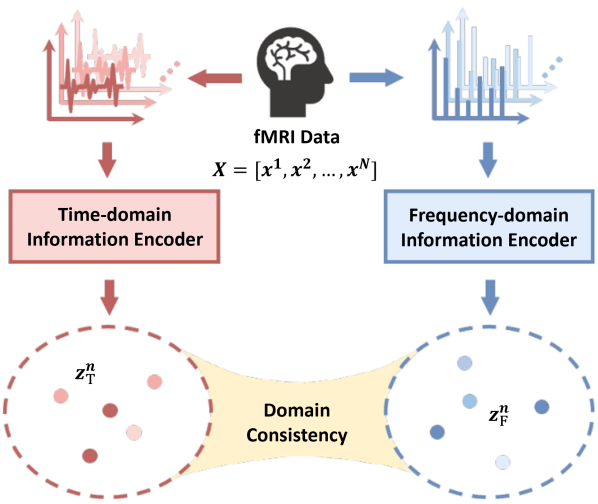



Fig. 4: Illustration of the learning objective. For each BOLD signal $\mathbf{x}^n$, its time-domain representation $\mathbf{z}_{\mathrm{T}}^n$ and the frequency-domain representation $\mathbf{z}_{\mathrm{F}}^n$ are obtained via their respective domain-specific encoders. The learning objective is to ensure that these two representations remain close in the latent time-frequency space, promoting consistency between the two domains.

### D. Time-Frequency Domain Consistency Learning Objective

Our learning objective adopts a non-contrastive SSL approach for graph-level representation learning without relying on contrastive pairs. Guided by the concept of time-frequency domain consistency (as illustrated in Figure 4), the objective enforces proximity between the time- and frequency-domain representations of the same BOLD signal in the latent space. This encourages the model to capture domain-invariant features, facilitating the comprehensive integration of diverse information from fMRI data. Specifically, we extend the application of CCA to the SSL framework. The correlation term aims to maximize the alignment between the two graph views by minimizing the distance between their representations in the latent time-frequency space, thereby extracting domain-invariant features. Additionally, we introduce a regularization term (decorrelation term) for each domain to prevent model collapse. This term encourages the learned representations to retain domain-specific semantic information by maintaining diversity across different feature dimensions. The overall process, taking the time-domain and frequency-domain representations as input, can be formulated as:

$$\mathcal{L} = \|\mathbf{Z}_{\mathrm{T}} - \mathbf{Z}_{\mathrm{F}}\|_{\mathrm{F}}^2 + \gamma \left\|\mathbf{Z}_{\mathrm{T}}^\top \mathbf{Z}_{\mathrm{T}} - \mathbf{I}\right\|_{\mathrm{F}}^2 + \beta \left\|\mathbf{Z}_{\mathrm{F}}^\top \mathbf{Z}_{\mathrm{F}} - \mathbf{I}\right\|_{\mathrm{F}}^2, \quad (11)$$

where $\gamma$ and $\beta$ are trade-off coefficients. The first term minimizes the disparity between the two graph views, enhancing their alignment. The second and third terms serve as regularization losses, preserving the semantic features of each domain. By minimizing the loss function, the model learns optimized representations from both domains. The final fused representation $\mathbf{Z}_{\mathrm{TF}}$ is then obtained by averaging these optimized time- and frequency-domain representations, effectively integrating information from both domains.

The overall training process of FENet is outlined in Algorithm 1. Initially, we apply the encoders for each domain graph separately to learn enhanced brain signal representations by optimizing Equation (11). After convergence during the pretraining, a subset of data is used for model fine-tuning, further refining the representation learning process.

**Algorithm 1** Overview of FENet Training

**Input:** The fMRI datasets for $M$ samples and their labels $\{\mathbf{X}^i, y^i\}_{i=1}^M$; Training epoch $E$.
**Output:** Trained encoders $f_{\theta_{\mathrm{T}}}(\cdot)$ and $f_{\theta_{\mathrm{F}}}(\cdot)$.
**Initialize:** The model parameters $\theta_{\mathrm{F}}$ and $\theta_{\mathrm{F}}$.
*Pretraining:*
**for** $i = 1, 2, \ldots, M$ **do**
  **Compute** multi-view brain graphs $\mathcal{G}_{\mathrm{T}}^i$ and $\mathcal{G}_{\mathrm{F}}^i$ in time and frequency domain.
  **for** $t = 1, 2, \ldots, E$ **do**
    **Compute** $\mathbf{Z}_{\mathrm{T}}^i = f_{\theta_{\mathrm{T}}}(\mathcal{G}_{\mathrm{T}}^i)$▷ Obtain time-domain representations
    **Compute** $\mathbf{Z}_{\mathrm{F}}^i = f_{\theta_{\mathrm{F}}}(\mathcal{G}_{\mathrm{F}}^i)$▷ Obtain frequency-domain representations
    **Optimize** model parameters $\theta_{\mathrm{F}}$ and $\theta_{\mathrm{F}}$ with loss function (see equation 11).
  **end for**
  **return** Brain graph representations $\mathbf{Z}_{\mathrm{TF}}^i$
**end for**
*Fine-tunning:*
**for** $t = 1, 2, \ldots, E$ **do**
  **for** Each batch in $\{\mathbf{Z}_{\mathrm{TF}}^i, y^i\}_{i=1}^M$ **do**
    **Compute** $\hat{y}^i = \phi(\mathbf{Z}_{\mathrm{TF}}^i)$▷ Predicted labels with classifier
    **Optimize** $\phi(\cdot)$ with $\mathcal{L}_{class}(\hat{y}^i, y^i)$.
  **end for**
**end for**

## V. Experiments

### A. Datasets

We conducted our experiments on two real-world medical datasets: the autism brain imaging data exchange (ABIDE) [1] and the attention deficit hyperactivity disorder (ADHD-200) [2] datasets.

**ABIDE Dataset.** This database aggregates data from 17 different acquisition sites, offering resting-state fMRI scans from 1,112 participants, including individuals diagnosed with autism spectrum disorder (ASD) and healthy controls (HC). In addition, it contains phenotypic information on executive functioning, language, visuospatial ability, motor functioning, and emotional status, assessed via self-report. For this study, we meticulously curated the data and retained a subset of 743 participants, consisting of 364 individuals with ASD (aged 7-24) and 379 HC (aged 6-22). This selection provides a relatively balanced distribution across the two groups.

[1] https://fcon_1000.projects.nitrc.org/indi/abide/
[2] https://fcon_1000.projects.nitrc.org/indi/adhd200/

**ADHD-200 Dataset.** Collected from eight independent imaging sites, the ADHD-200 dataset was designed to facilitate early diagnosis of ADHD and inform clinical treatment decisions. It includes detailed phenotypic data such as diagnostic status, dimensional ADHD symptom measures, age, sex, intelligence quotient (IQ), and lifetime medication status. For our experiments, we employed a sample of 459 participants, comprising 229 typically developing individuals and 230 children and adolescents with ADHD (aged 7-21).

### *B. Data Preprocessing*

In this study, we obtained fMRI data from both the ABIDE and ADHD datasets. We then processed these fMRI signals using the graph theoretical network analysis (GRETNA) toolbox[3], which operates in conjunction with the SPM12 software[4]. The preprocessing steps included slice timing correction, head motion correction, spatial normalization, and Gaussian smoothing. Next, the automated anatomical labeling (AAL) atlas was used as the reference space to divide the brain into 116 ROIs, from which BOLD time series were subsequently extracted.

### *C. Baselines*

For baseline methods used in the comparison, we included the GNN-based approaches, BrainGNN [26] and BrainGB [27], which employs supervised learning strategies. Additionally, we considered three contrastive SSL methods, GCA [47], SpCo [48], and MA-GCL [63], along with two similarity-based SSL methods, GATE [34] and CCA-SSG [38]. Open-source codes for all comparison methods are adopted, and the grid search technology is performed to determine their optimal hyperparameters. To ensure a fair comparison, we apply the same dynamic window strategy and utilize a grid search technique to identify the best practices for each comparison method. Details for each comparison method are outlined below.

- BrainGNN [5] [26] designs a novel ROI-aware GCN model combined with ROI-selection pooling layers to improve the interpretability of analyzing fMRI data. This serves as the baseline graph-based method for fMRI data.
- BrainGB [6] [27] comprehensively summarizes the current methods for constructing brain networks and proposes a generic GNN architecture to standardize the process of brain network analysis.
- GCA [7] [47] introduces a novel adaptive enhancement method for contrastive SSL. This approach perturbs the graph's topology and semantics as priors, highlighting important structural connections while preserving crucial semantic information at the node level.
- SpCo [8] [48] combines graph spectral information with SSL, addressing imbalances in spectral learning. The method introduces a spectral-contrastive learning module that effectively enhances existing SSL methods.
- MA-GCL [9] [63] designs different encoders to enhance view diversity, addressing the issue of semantic label changes caused by graph perturbations. It employs three techniques to reduce high-frequency noise and enhance view diversity safely.
- CCA-SSG [10] [38] proposes an efficient similarity-based SSL framework by utilizing the CCA as the learning objective. This allows the structure of this model to be concise and achieves significant results compared with other SSL methods.
- GATE [11] [34] presents a similarity-based SSL method, introducing the concept of a population graph and utilizing an enhanced CCA method to improve the model's sensitivity to spurious signals.

For supervised learning methods such as BrainGNN and BrainGB, we train the model following their original configurations. Among the aforementioned comparison methods, GCA, SpCo, CCA-SSG, and GATE represent state-of-the-art SSL models, and we adhere to their original literature and official code. For SSL methods, such as contrast-based models (i.e., GCA, SpCo, and MA-GCL), positive and negative sample pairs of nodes are constructed by corrupting the input graph using a corruption function. For similarity-based methods (i.e., CCA-SSG and GATE), we perturb the graph using a random masking function to generate two views.

### *D. Implementation Details*

We implement FENet in PyTorch and conduct all experiments on an NVIDIA Tesla P100 GPU with 16GB of memory. Model parameters are optimized using the AdamW optimizer [64] with a learning rate of $1e^{-5}$. The time-domain information encoder is configured as a standard two-layer GCN model to mitigate the over-smoothing issue in GCN. Based on empirical evaluations, the frequency-domain information encoder comprises three FGO layers for effective feature extraction in the frequency-domain graph. To effectively set the frequency thresholds for low-pass, band-pass, and high-pass filters in graph signal processing, we optimized the $\lambda_{\mathrm{L}}$ as the lowest $20\%$ of eigenvalues, while the $\lambda_{\mathrm{H}}$ as the top $20\%$ of eigenvalues. To evaluate the performance of both encoders, we tune the hyperparameters in Equation (11) and compare the model's classification accuracy; ultimately, $\gamma$ and $\beta$ are set to $1e^{-5}$ and $1e^{-4}$. The training procedure consists of two phases: pretraining and fine-tuning. In the pretraining phase, we train for 99 iterations, whereas in the fine-tuning phase, we randomly select $20\%$ labeled samples to adjust model parameters. For a comprehensive performance assessment, we employ 5-fold cross-validation and repeat the experiments five times with different random seeds. All reported results include the average performance and the corresponding standard deviation across these runs. We use the mean and standard deviation of the following metrics to assess the classification performance:

[3]https://www.nitrc.org/projects/gretna/.
[4]https://www.fil.ion.ucl.ac.uk/spm/software/spm12/.
[5]https://github.com/xxlya/BrainGNN_Pytorch
[6]https://github.com/HennyJie/BrainGB
[7]https://github.com/CRIPAC-DIG/GCA
[8]https://github.com/liun-online/SpCo
[9]https://github.com/GXM1141/MA-GCL
[10]https://github.com/hengruizhang98/CCA-SSG
[11]https://github.com/LarryUESTC/GATE

TABLE II: Performance (%) on the ABIDE and ADHD datasets. SL represents supervised learning, and SSL represents self-supervised learning. The best results are bold and the second are underlined.

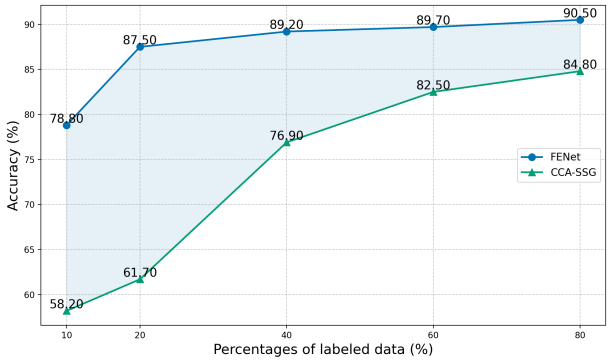
| | Methods | ABIDE | | | | ADHD | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | ACC | AUC | Recall | F1-score | ACC | AUC | Recall | F1-score |
| SL | BrainGNN | $58.4 \pm 2.8$ | $50.8 \pm 4.1$ | $55.6 \pm 4.7$ | $55.6 \pm 9.5$ | $55.8 \pm 1.2$ | $58.0 \pm 4.9$ | $46.3 \pm 4.2$ | $55.9 \pm 4.2$ |
| | BrainGB | $61.1 \pm 3.9$ | $62.3 \pm 3.6$ | $65.4 \pm 5.2$ | $63.6 \pm 10.2$ | $56.2 \pm 7.1$ | $59.6 \pm 5.2$ | $56.2 \pm 4.6$ | $\underline{61.6 \pm 6.3}$ |
| SSL | MA-GCL | $52.8 \pm 5.8$ | $52.7 \pm 4.3$ | $51.1 \pm 5.1$ | $49.7 \pm 4.8$ | $50.1 \pm 5.6$ | $48.5 \pm 6.2$ | $46.5 \pm 3.7$ | $45.2 \pm 6.4$ |
| | GCA | $53.1 \pm 6.1$ | $52.9 \pm 6.1$ | $64.4 \pm 7.4$ | $61.8 \pm 5.2$ | $50.8 \pm 2.4$ | $52.4 \pm 5.1$ | $59.4 \pm 3.6$ | $53.6 \pm 4.4$ |
| | SpCo | $61.1 \pm 5.2$ | $56.0 \pm 7.8$ | $\underline{68.5 \pm 4.9}$ | $58.8 \pm 5.3$ | $55.8 \pm 4.1$ | $58.0 \pm 4.9$ | $58.5 \pm 3.7$ | $55.9 \pm 5.1$ |
| | CCA-SSG | $61.7 \pm 5.3$ | $52.6 \pm 6.2$ | $65.8 \pm 5.9$ | $62.8 \pm 4.1$ | $56.6 \pm 5.8$ | $\underline{60.3 \pm 2.6}$ | $60.3 \pm 4.4$ | $58.0 \pm 8.6$ |
| | GATE | $\underline{62.9 \pm 3.4}$ | $\underline{62.8 \pm 6.1}$ | $64.4 \pm 7.3$ | $\underline{65.0 \pm 5.2}$ | $\underline{60.1 \pm 5.2}$ | $52.6 \pm 3.8$ | $\underline{62.8 \pm 9.3}$ | $60.9 \pm 5.5$ |
| | **FENet** (ours) | $\mathbf{87.5 \pm 3.9}$ | $\mathbf{84.7 \pm 4.1}$ | $\mathbf{83.3 \pm 4.8}$ | $\mathbf{86.9 \pm 4.8}$ | $\mathbf{77.8 \pm 4.1}$ | $\mathbf{67.9 \pm 4.4}$ | $\mathbf{66.7 \pm 2.9}$ | $\mathbf{75.0 \pm 5.2}$ |

accuracy, the area under the ROC curve (AUC), recall, and F1-score.

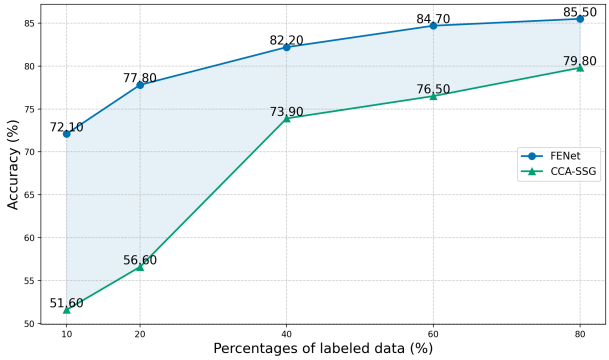
## VI. Resutlts and Discussion

### A. Performance Comparison

We evaluate the proposed model against various baseline methods, as summarized in Table II. The results indicate that our method outperforms all baselines across multiple metrics. Specifically, for autism detection, FENet shows 24.6%, 21.9%, 14.8%, and 21.9% relative improvements in terms of accuracy, AUC, recall, and F1-score. Similarly, ADHD detection demonstrates 17.7% higher accuracy, 15.3% improvement in AUC, 3.9% increase in recall, and 14.1% gain in F1-score, highlighting its effectiveness in psychiatric disorder classification. Existing SSL-based approaches outperform supervised methods overall, indicating that self-supervised training effectively learns brain graph representations under data-scarce conditions. The suboptimal performance of methods such as MA-GCL and GCA may be attributed to their reliance on manually defined contrastive learning objectives, which lack relevant prior knowledge. In contrast, similarity-based approaches offer greater flexibility but still fail to incorporate frequency information. While SpCo emphasizes high-frequency spectral information under graph perturbations, it remains encoded in the time domain, limiting its ability to capture frequency-domain semantic information. In contrast, our model integrates frequency information into a similarity-based framework and accounts for temporal, spatial, and frequency features, resulting in the best overall performance.




(a) ABIDE

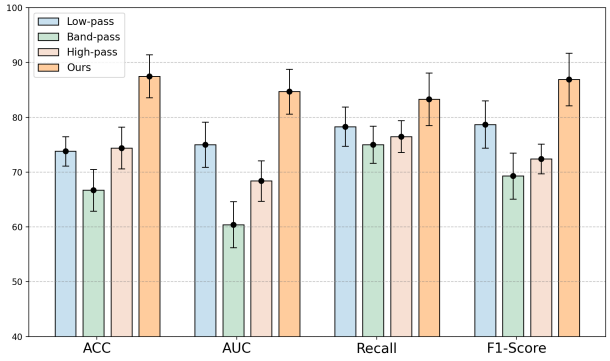



(b) ADHD

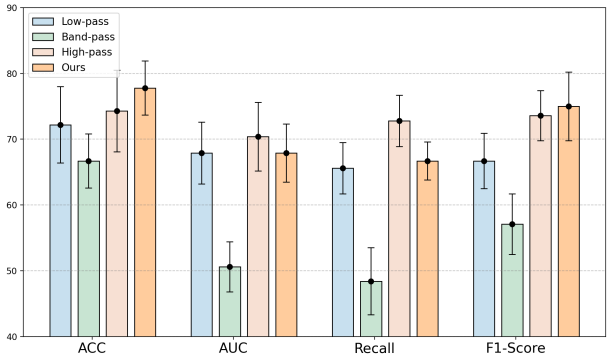
Fig. 5: Accuracy (%) of FENet and CCA-SSG at different labeled data rates for fine-tunning.

### B. Data-efficient Ability Study

We highlight that the ultimate success of SSL hinges on effectively utilizing limited samples, indicating that a well-designed SSL method can leverage additional labeled data for fine-tuning downstream tasks and further boosting performance. To validate the robustness of our model under limited-sample conditions, we keep the unlabeled training data fixed and vary the proportion of labeled data used for fine-tuning. We adopt CCA-SSG as a baseline for comparison, which shares a CCA-based learning objective and performs as the benchmark model for similarity-based SSL. As shown in Figure 5, our method consistently outperforms the baseline (CCA-SSG) across all percentages of labeled data in both datasets. Specifically, when the amount of fine-tuning data is reduced to 20% of the full dataset, the performance gap becomes most pronounced, that FENet surpasses CCA-SSG by 25.8% and 21.2% on the ABIDE and ADHD datasets, respectively, consistent with the results reported in Table II. Moreover, even with only 10% of labeled data, our model maintains strong performance, achieving 78.8% accuracy on ABIDE and 72.1% on ADHD. These findings demonstrate the data efficiency of our method, showing that it can deliver reliable results even under limited supervision.

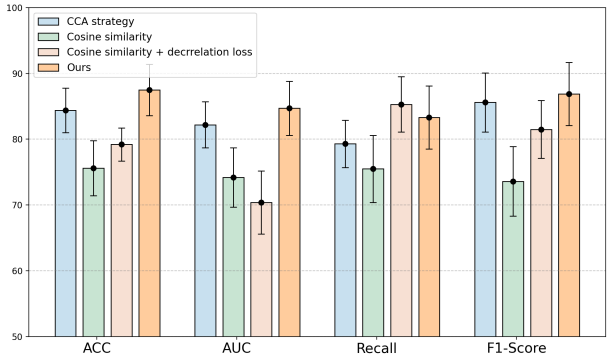



(a) ABIDE

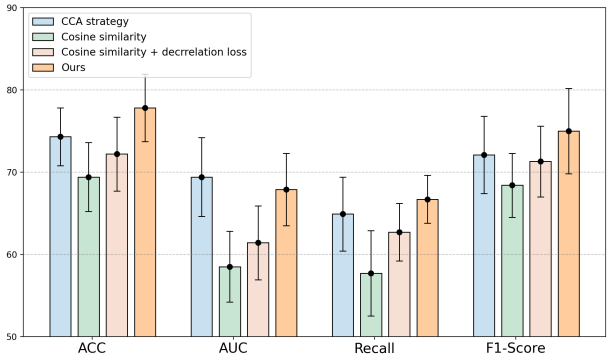



(b) ADHD

Fig. 6: Performance (%) of the model at different filter settings. The vertical lines (error bars) on the graph represent the variance of the data.




(a) ABIDE



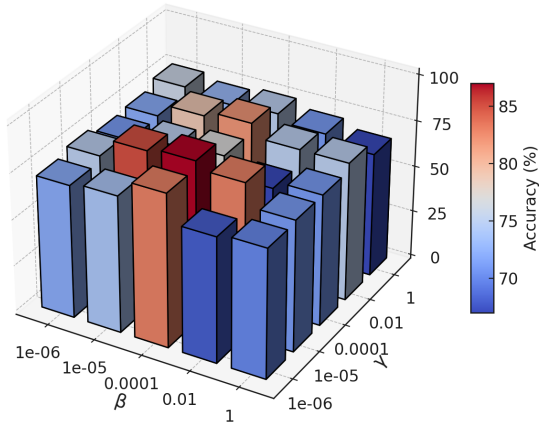

(b) ADHD

Fig. 7: Performance (%) of the model with different learning objective settings. The vertical lines (error bars) on the graph represent the variance of the data.

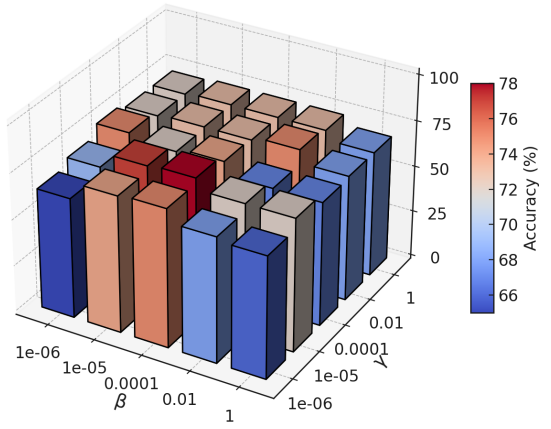
### C. Ablation Study

*1) Effectiveness of Graph Filter:* We incorporate frequency information to investigate its contribution to brain disorder detection, particularly emphasizing the role of high-frequency components. Additionally, we evaluate the impact of different graph filter settings. As shown in Figure 6, the high-pass filter attains the highest AUC and recall on the ADHD dataset, suggesting that high-frequency information, particularly related to exposure tasks, is closely associated with ADHD—where faster neural oscillations are indicative of cognitive processing deficits [15]. In contrast, the band-pass filter capturing middle-frequency components has a limited impact on classification, especially for ADHD. Moreover, detection based on low-frequency components significantly outperforms state-of-the-art baselines that ignore the frequency information, achieving 73.8% and 72.2% accuracy on the ABIDE and ADHD datasets, respectively. This highlights the importance of global brain structural changes in enhancing disease detection. These findings are consistent with existing medical studies [14], [15], confirming that both low- and high-frequency information is crucial for psychiatric disorder detection. Therefore, our method applies both low-pass and high-pass filters to extract low- and high-frequency components of brain signals while filtering out mid-frequency components to enable efficient and effective frequency-domain analysis, which achieves superior performance across both datasets.

*2) Effectiveness of Learning Objective:* To validate the effectiveness of our learning objective, namely, the improved CCA strategy guided by domain consistency for integrating time-frequency information, we conducted comparative experiments using several variants of our method with alternative objective functions. In particular, we designed three variants: i) CCA strategy: a variant using CCA but with identical trade-off coefficients applied to representations from both domains; ii) Cosine Similarity: replacing our CCA-based objective with a cosine similarity strategy following [65]; iii) Cosine Similarity + decorrelation loss: an improved version of the cosine similarity approach, enhanced by incorporating an additional decorrelation term to examine its contribution. Experimental results shown in Figure 7 demonstrate that the CCA strategy consistently outperforms the baseline cosine similarity loss, indicating its effectiveness in multi-view fusion. However, its performance is slightly inferior to our proposed method. We attribute this to the differing levels of discriminative information in the time and frequency domains; applying distinct trade-off coefficients for each decorrelation term enables better balancing of representation learning across domains.Additionally, incorporating the decorrelation term noticeably improves the cosine similarity-based model, highlighting the importance of preserving domain-specific semantic information. Nonetheless, its performance remains inferior to the CCA strategy. This gap likely arises because cosine similarity lacks explicit constraints




(a) ABIDE




(b) ADHD

Fig. 8: Accuracy (%) of FENet with different values of $\gamma$ and $\beta$ in Equation (11).

for aligning cross-domain representations, making it prone to training instability and local optima. As a result, while it achieves competitive recall, its overall performance is limited due to insufficient global alignment of time-frequency features.

*3) Effect of Decorrelation Intensity:* To achieve an optimal trade-off between time-domain and frequency-domain information, we conduct a sensitivity analysis by adjusting the decorrelation parameters in the objective function, which controls the influence of each domain. This analysis evaluates the effectiveness of different information sources. As shown in Figure 8, the results on both datasets indicate that the model performs well within a narrow parameter range ($\gamma \leq 1e^{-5}, \beta \leq 1e^{-4}$). We attribute this to the model's adherence to the domain consistency principle, which enhances overall performance by encouraging the learning of fusion representations that capture shared attributes between time-domain and frequency-domain data. Conversely, larger parameter values lead to the learning of irrelevant embeddings rather than discriminative ones, ultimately reducing performance. Notably, the parameter for frequency-domain decorrelation is more significant than that for the time domain, suggesting that the model preserves greater invariance in frequency representations. This implies that frequency-domain signals provide richer dynamic features in the integrated brain network representation. Based on these findings, we set the parameters $\gamma$ to $1e^{-5}$, and $\beta$ to $1e^{-4}$ to achieve a balanced utilization of time-domain and frequency-domain information.

## VII. Conclusion

This paper introduces FENet, an innovative and effective SSL framework designed to learn diverse information about fMRI data under limited-sample conditions, facilitating efficient disease detection. FENet explicitly integrates frequency and time information to provide a holistic understanding of brain activities. Specifically, multi-view brain graphs in both time and frequency domains are constructed, and corresponding information encoders in two domains are employed to extract temporal-spectral features while exploring their correlations with brain function. Further, to balance the utilization of these features, FENet uses a domain consistency-guided learning objective to maximize the invariance across two views, constructing frequency-enhanced brain network representations. Experimental results on two real-world neural imaging datasets demonstrate that FENet significantly outperforms existing baselines. Moreover, our analysis highlights that frequency information provides critical insights into disease-related brain dynamics, with high-frequency components playing a particularly significant role in capturing localized disruptions and neural oscillation abnormalities associated with psychiatric disorders. In future research, we will further enhance the modeling of temporal dependencies across multiple domains. Additionally, we aim to design an adaptive representation fusion module that leverages frequency-domain information, enabling a more effective joint analysis of time-domain and frequency-domain features.

## References

[1] G. Collin, L. J. Seidman, M. S. Keshavan, W. S. Stone, Z. Qi, T. Zhang, Y. Tang, H. Li, S. A. Anteraper, M. A. Niznikiewicz *et al.*, "Functional connectome organization predicts conversion to psychosis in clinical high-risk youth from the sharp program," *Molecular Psychiatry*, vol. 25, no. 10, pp. 2431–2440, 2020.

[2] J. Du, S. Wang, R. Chen, and S. Wang, "Improving fmri-based autism severity identification via brain network distance and adaptive label distribution learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 33, pp. 162–174, 2025.

[3] M. Bayat, M. Hernandez, M. Curzon, D. Garic, P. Graziano, and A. S. Dick, "Reduced recruitment of inhibitory control regions in very young children with adhd during a modified kiddie continuous performance task: a fmri study." *Cortex*, 2025.

[4] O. J. Arthurs and S. Boniface, "How well do we understand the neural origins of the fmri bold signal?" *TRENDS in Neurosciences*, vol. 25, no. 1, pp. 27–31, 2002.

[5] Y. Du, S. Fang, X. He, and V. D. Calhoun, "A survey of brain functional network extraction methods using fmri data," *Trends in Neurosciences*, 2024.

[6] S. S. Menon and K. Krishnamurthy, "A comparison of static and dynamic functional connectivities for identifying subjects and biological sex using intrinsic individual brain connectivity," *Scientific Reports*, vol. 9, no. 1, p. 5729, 2019.

[7] M. Xia and Y. He, "Functional connectomics from a "big data" perspective," *Neuroimage*, vol. 160, pp. 152–167, 2017.

[8] N. Sartorius, "Comorbidity of mental and physical disorders: A major challenge for medicine in the 21st century," *European Psychiatry*, vol. 41, no. S1, pp. S9–S9, 2017.

[9] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo *et al.*, "Dynamic functional connectivity: promise, issues, and interpretations," *Neuroimage*, vol. 80, pp. 360–378, 2013.

[10] N. C. Dvornek, P. Ventola, K. A. Pelphrey, and J. S. Duncan, "Identifying autism from resting-state fmri using long short-term memory networks," in *Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings 8*. Springer, 2017, pp. 362–370.

[11] D. A. Regier, E. A. Kuhl, and D. J. Kupfer, "The dsm-5: Classification and criteria changes," *World Psychiatry*, vol. 12, no. 2, pp. 92–98, 2013.

[12] S. L. Hyman, S. E. Levy, S. M. Myers, D. Kuo, S. Apkon, T. Brei, L. F. Davidson, B. E. Davis, K. A. Ellerbeck, G. H. Noritz *et al.*, "Executive summary: identification, evaluation, and management of children with autism spectrum disorder," *Pediatrics*, vol. 145, no. 1, 2020.

[13] A. Zou, J. Ji, M. Lei, J. Liu, and Y. Song, "Exploring brain effective connectivity networks through spatiotemporal graph convolutional models," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[14] W. Huang, L. Goldsberry, N. F. Wymbs, S. T. Grafton, D. S. Bassett, and A. Ribeiro, "Graph frequency analysis of brain signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 7, pp. 1189–1203, 2016.

[15] S. Sasai, T. Koike, S. K. Sugawara, Y. H. Hamano, M. Sumiya, S. Okazaki, H. K. Takahashi, G. Taga, and N. Sadato, "Frequency-specific task modulation of human brain functional networks: A fast fmri study," *NeuroImage*, vol. 224, p. 117375, 2021.

[16] A. T. Jafadideh and B. M. Asl, "Rest-fmri based comparison study between autism spectrum disorder and typically control using graph frequency bands," *Computers in Biology and Medicine*, vol. 146, p. 105643, 2022.

[17] S. Karavallil Achuthan, K. L. Coburn, M. E. Beckerson, and R. K. Kana, "Amplitude of low frequency fluctuations during resting state fmri in autistic children," *Autism Research*, vol. 16, no. 1, pp. 84–98, 2023.

[18] F. Xia, K. Sun, S. Yu, A. Aziz, L. Wan, S. Pan, and H. Liu, "Graph learning: A survey," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 109–127, 2021.

[19] C. Peng, J. He, and F. Xia, "Learning on multimodal graphs: A survey," *ArXiv Preprint ArXiv:2402.05322*, 2024.

[20] X. Luo, J. Wu, J. Yang, S. Xue, A. Beheshti, Q. Z. Sheng, D. McAlpine, P. Sowman, A. Giral, and S. Y. Philip, "Graph neural networks for brain graph learning: a survey," in *33rd International Joint Conference on Artificial Intelligence, IJCAI 2024*. International Joint Conferences on Artificial Intelligence, 2024, pp. 8170–8178.

[21] M. Liu, Q. Dong, C. Wang, X. Cheng, F. G. Febrinanto, A. N. Hoshyar, and F. Xia, "Motif-induced subgraph generative learning for explainable neurological disorder detection," in *Australasian Joint Conference on Artificial Intelligence*. Springer, 2024, pp. 376–389.

[22] H. Cai, X. Sheng, G. Wu, B. Hu, Y.-M. Cheung, and J. Chen, "Brain network classification for accurate detection of alzheimer's disease via manifold harmonic discriminant analysis," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[23] M. Liu, H. Zhang, F. Shi, and D. Shen, "Hierarchical graph convolutional network built by multiscale atlases for brain disorder diagnosis using functional connectivity," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[24] X.-A. Bi, K. Chen, S. Jiang, S. Luo, W. Zhou, Z. Xing, L. Xu, Z. Liu, and T. Liu, "Community graph convolution neural network for alzheimer's disease classification and pathogenetic factors identification," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[25] X. Luo, J. Wu, J. Yang, H. Chen, Z. Li, H. Peng, and C. Zhou, "Knowledge distillation guided interpretable brain subgraph neural networks for brain disorder exploration," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 2, pp. 3559–3572, 2024.

[26] X. Li, Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L. H. Staib, P. Ventola, and J. S. Duncan, "Braingnn: Interpretable brain graph neural network for fmri analysis," *Medical Image Analysis*, vol. 74, p. 102233, 2021.

[27] H. Cui, W. Dai, Y. Zhu, X. Kan, A. A. C. Gu, J. Lukemire, L. Zhan, L. He, Y. Guo, and C. Yang, "Braingb: A benchmark for brain network analysis with graph neural networks," *IEEE Transactions on Medical Imaging*, vol. 42, no. 2, pp. 493–506, 2022.

[28] Y. Liu, M. Jin, S. Pan, C. Zhou, Y. Zheng, F. Xia, and S. Y. Philip, "Graph self-supervised learning: A survey," *IEEE transactions on knowledge and data engineering*, vol. 35, no. 6, pp. 5879–5900, 2022.

[29] P. Kumar, P. Rawat, and S. Chauhan, "Contrastive self-supervised learning: review, progress, challenges and future research directions," *International Journal of Multimedia Information Retrieval*, vol. 11, no. 4, pp. 461–488, 2022.

[30] S. Zhang, X. Chen, X. Shen, B. Ren, Z. Yu, H. Yang, X. Jiang, D. Shen, Y. Zhou, and X.-Y. Zhang, "A-gcl: Adversarial graph contrastive learning for fmri analysis to diagnose neurodevelopmental disorders," *Medical Image Analysis*, vol. 90, p. 102932, 2023.

[31] X. Luo, G. Dong, J. Wu, A. Beheshti, J. Yang, and S. Xue, "An interpretable brain graph contrastive learning framework for brain disorder analysis," in *Conference on Web Search and Data Mining (WSDM'24)*, 2024.

[32] J. Xu, Q. Bian, X. Li, A. Zhang, Y. Ke, M. Qiao, W. Zhang, W. K. J. Sim, and B. Gulyás, "Contrastive graph pooling for explainable classification of brain networks," *IEEE Transactions on Medical Imaging*, 2024.

[33] Y. Zong, Q. Zuo, M. K.-P. Ng, B. Lei, and S. Wang, "A new brain network construction paradigm for brain disorder via diffusion-based graph contrastive learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[34] L. Peng, N. Wang, J. Xu, X. Zhu, and X. Li, "Gate: graph cca for temporal self-supervised learning for label-efficient fmri analysis," *IEEE Transactions on Medical Imaging*, vol. 42, no. 2, pp. 391–402, 2022.

[35] X. Wang, Y. Chu, Q. Wang, L. Cao, L. Qiao, L. Zhang, and M. Liu, "Unsupervised contrastive graph learning for resting-state functional mri analysis and brain disorder detection," *Human Brain Mapping*, vol. 44, no. 17, pp. 5672–5692, 2023.

[36] X. Zhang, Z. Zhao, T. Tsiligkaridis, and M. Zitnik, "Self-supervised contrastive pre-training for time series via time-frequency consistency," *Advances in Neural Information Processing Systems*, vol. 35, pp. 3988–4003, 2022.

[37] Y. Liu and B. Du, "Frequency domain-oriented complex graph neural networks for graph classification," *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[38] H. Zhang, Q. Wu, J. Yan, D. Wipf, and P. S. Yu, "From canonical correlation analysis to self-supervised graph neural networks," *Advances in Neural Information Processing Systems*, vol. 34, pp. 76–89, 2021.

[39] K. Sun, C. Peng, S. Yu, Z. Han, and F. Xia, "From eeg data to brain networks: Graph learning based brain disease diagnosis," *IEEE Intelligent Systems*, pp. 1–9, 2024.

[40] Y. Chen, J. Yan, M. Jiang, T. Zhang, Z. Zhao, W. Zhao, J. Zheng, D. Yao, R. Zhang, K. M. Kendrick *et al.*, "Adversarial learning based node-edge graph attention networks for autism spectrum disorder identification," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[41] C. Peng, M. Liu, C. Meng, S. Xue, K. Keogh, and F. Xia, "Stage-aware brain graph learning for alzheimer s disease," *MedRxiv*, pp. 2024–04, 2024.

[42] J. Ren, F. Xia, I. Lee, A. Noori Hoshyar, and C. Aggarwal, "Graph learning for anomaly analytics: Algorithms, applications, and challenges," *ACM Transactions on Intelligent Systems and Technology*, vol. 14, no. 2, Feb 2023.

[43] F. Xia, L. Wang, T. Tang, X. Chen, X. Kong, G. Oatley, and I. King, "Cengcn: Centralized convolutional networks with vertex imbalance for scale-free graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 4555–4569, 2023.

[44] S. Fu, Q. Peng, Y. He, X. Wang, B. Zou, D. Xu, X.-Y. Jing, and X. You, "Multilevel contrastive graph masked autoencoders for unsupervised graph-structure learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 2, pp. 3464–3478, 2024.

[45] H. Tang, G. Ma, L. Guo, X. Fu, H. Huang, and L. Zhan, "Contrastive brain network learning via hierarchical signed graph pooling model," *IEEE transactions on neural networks and learning systems*, 2022.

[46] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," *Advances in Neural Information Processing Systems*, vol. 33, pp. 5812–5823, 2020.

[47] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, "Graph contrastive learning with adaptive augmentation," in *Proceedings of the Web Conference 2021*, 2021, pp. 2069–2080.

[48] N. Liu, X. Wang, D. Bo, C. Shi, and J. Pei, "Revisiting graph contrastive learning from the perspective of graph spectrum," *Advances in Neural Information Processing Systems*, vol. 35, pp. 2972–2983, 2022.

[49] Y. Zhang, H. Zhu, Z. Song, P. Koniusz, and I. King, "Spectral feature augmentation for graph contrastive learning and beyond," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 9, 2023, pp. 11 289–11 297.

[50] Y. Wu, T. Dang, D. Spathis, H. Jia, and C. Mascolo, "Statiocl: Contrastive learning for time series via non-stationary and temporal contrast," in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024, pp. 2575–2584.

[51] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *International Conference on Machine Learning*. PMLR, 2013, pp. 1247–1255.
[52] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting," in *International Conference on Machine Learning*. PMLR, 2022, pp. 27 268–27 286.
[53] X. Piao, Z. Chen, T. Murayama, Y. Matsubara, and Y. Sakurai, "Fredformer: Frequency debiased transformer for time series forecasting," in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 2400–2410.
[54] K. Yi, Q. Zhang, W. Fan, S. Wang, P. Wang, H. He, N. An, D. Lian, L. Cao, and Z. Niu, "Frequency-domain mlps are more effective learners in time series forecasting," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
[55] D. Cao, Y. Wang, J. Duan, C. Zhang, X. Zhu, C. Huang, Y. Tong, B. Xu, J. Bai, J. Tong *et al.*, "Spectral temporal graph neural network for multivariate time-series forecasting," *Advances in Neural Information Processing Systems*, vol. 33, pp. 17 766–17 778, 2020.
[56] K. Yi, Q. Zhang, W. Fan, H. He, L. Hu, P. Wang, N. An, L. Cao, and Z. Niu, "Fouriergnn: Rethinking multivariate time series forecasting from a pure graph perspective," *Advances in Neural Information Processing Systems*, vol. 36, 2023.
[57] L. Yang and S. Hong, "Unsupervised time-series representation learning with iterative bilinear temporal-spectral fusion," in *International Conference on Machine Learning*. PMLR, 2022, pp. 25 038–25 054.
[58] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.
[59] X. Zhuang, Z. Yang, and D. Cordes, "A technical review of canonical correlation analysis for neuroscience applications," *Human Brain Mapping*, vol. 41, no. 13, pp. 3807–3833, 2020.
[60] X. Chang, T. Xiang, and T. M. Hospedales, "Scalable and effective deep cca via soft decorrelation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1488–1497.
[61] I. Cohen, Y. Huang, J. Chen, J. Benesty, J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," *Noise Reduction in Speech Processing*, pp. 1–4, 2009.
[62] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," *Computational Social Networks*, vol. 6, no. 1, pp. 1–23, 2019.
[63] X. Gong, C. Yang, and C. Shi, "Ma-gcl: Model augmentation tricks for graph contrastive learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 4, 2023, pp. 4284–4292.
[64] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *ICLR*, 2019.
[65] S. Thakoor, C. Tallec, M. G. Azar, R. Munos, P. Veličković, and M. Valko, "Bootstrapped representation learning on graphs," in *ICLR 2021 Workshop on Geometrical and Topological Representation Learning*, 2021.