

A Deep Probabilistic Spatiotemporal Framework for Dynamic Graph Representation Learning with Application to Brain Disorder Identification

Sin-Yee Yap^{1,2}, Junn Yong Loo¹, Chee-Ming Ting¹, Fuad Noman¹, Raphaël C.-W. Phan¹, Adeel Razi³ and David L. Dowe²

¹School of Information Technology, Monash University Malaysia

²Department of Data Science and AI, Faculty of Information Technology, Monash University, Australia

³Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University
{sin.yap, loo.junnyong, ting.cheeming, fuad.noman, raphael.phan, adeel.razi, david.dowe}@monash.edu

Abstract

Recent applications of pattern recognition techniques on brain connectome classification using functional connectivity (FC) are shifting towards acknowledging the non-Euclidean topology and dynamic aspects of brain connectivity across time. In this paper, a deep spatiotemporal variational Bayes (DSVB) framework is proposed to learn time-varying topological structures in dynamic FC networks for identifying autism spectrum disorder (ASD) in human participants. The framework incorporates a spatial-aware recurrent neural network with an attention-based message passing scheme to capture rich spatiotemporal patterns across dynamic FC networks. To overcome model overfitting on limited training datasets, an adversarial training strategy is introduced to learn graph embedding models that generalize well to unseen brain networks. Evaluation on the ABIDE resting-state functional magnetic resonance imaging dataset shows that our proposed framework substantially outperforms state-of-the-art methods in identifying patients with ASD. Dynamic FC analyses with DSVB-learned embeddings reveal apparent group differences between ASD and healthy controls in brain network connectivity patterns and switching dynamics of brain states.

1 Introduction

The human brain is a complex system that consists of numerous interconnected neuronal regions. As revealed by functional magnetic resonance imaging (fMRI), spontaneous spatiotemporal fluctuations in brain activity exist even during a resting state. Functional connectivity (FC) in brain networks is typically characterized via statistical dependence (such as correlations) between blood oxygen level dependent (BOLD) fMRI signals over spatially-distinct brain regions. Alterations in resting-state FC networks have been associated with neuropsychiatric or neurodevelopmental disorders [Filippi *et al.*, 2019], for example in autism spectrum disorder (ASD) [Holiga *et al.*, 2019; Valenti *et al.*, 2020; Wang *et al.*, 2021].

Machine learning (ML) techniques have been applied to the identification and prediction of neuropsychiatric disorders in the past decade [Tanveer *et al.*, 2020; Hyde *et al.*, 2019]. Traditional ML methods such as support vector machines (SVM), artificial neural networks, and deep learning methods are very widely used ML techniques [Moridian *et al.*, 2022]. Deep learning has especially gained popularity in the studies of neuroscience as it offers an opportunity to discover and understand the underlying differences in the pattern of brain connectivity between individuals with psychiatric disorders and healthy controls using deep neural networks [RaviPrakash *et al.*, 2019; Khodatars *et al.*, 2021; Zhang *et al.*, 2020]. Conventional approaches have modelled brain networks as 2D grid-like connectivity matrices, neglecting the non-Euclidean nature of brain networks. To address this, graph-based deep learning offers flexibility in modeling individual-level pairwise brain region interactions [Yan *et al.*, 2019a], capturing group-level associations using population graphs [Jiang *et al.*, 2020a], or combining both strategies [Li *et al.*, 2022; Zhou and Zhang, 2021]. Yet, studies using graph neural networks (GNNs) in these contexts often focus on group-level network topology with phenotype data or supervised subject-level embedding using pre-computed population graphs for classifying static brain networks.

Dynamic spatiotemporal patterns in resting-state brain functional networks have been reported in several studies [Lee and Kim, 2019; Aedo-Jury *et al.*, 2020]. They emphasize the consideration of not only the stationary brain network topology but also the dynamic temporal structure. Studies show that GNNs have potential in studying brain networks. However, the application of GNNs to model graph-structured brain network data has predominantly been limited to static representations, and there is a notable scarcity of GNN models designed to investigate the dynamic aspects of such networks. For instance, [Azevedo *et al.*, 2020] employed 1-dimensional CNNs for temporal feature extraction and graph convolutional networks (GCNs) for spatial information sharing, and the simple framework showed great potential on both binary and continuous label prediction. [Kim *et al.*, 2021] proposed self-attention and squeeze-excitation for readout, applied successively on GRU-encoded graphs to obtain spatiotemporal embeddings for downstream gender classification. Rather than focusing solely on the gen-

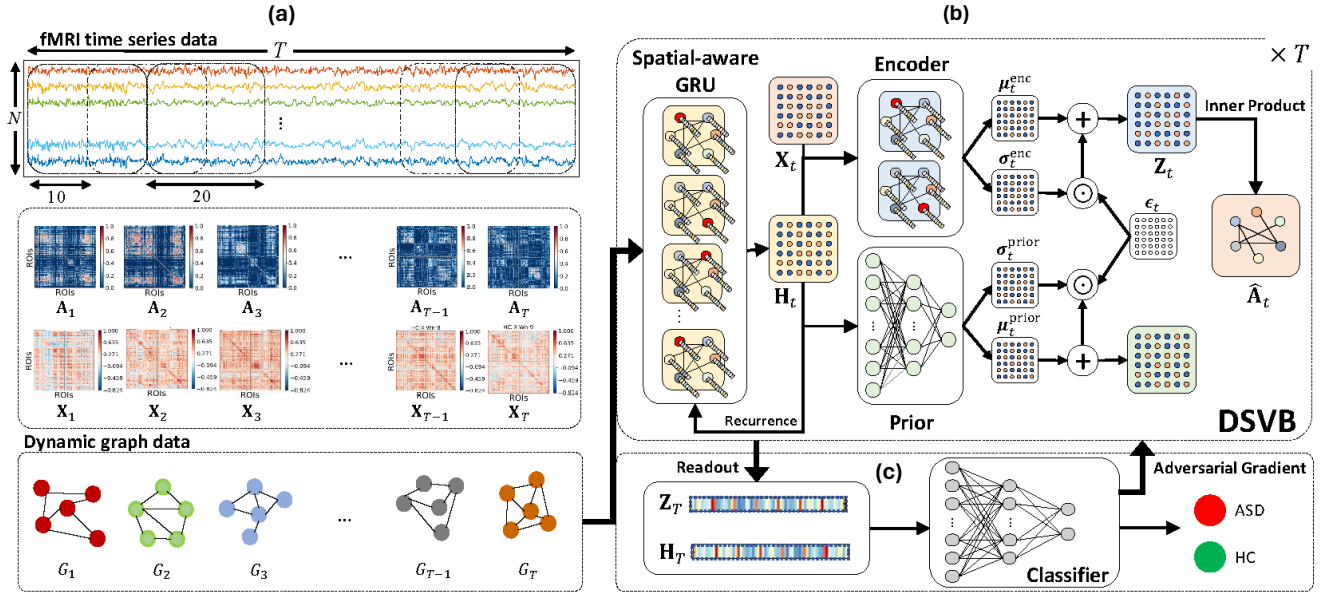


Figure 1: The architecture of the proposed DSVB framework for dynamic functional brain network classification. (a) Dynamic brain network is constructed from sliding window analysis of fMRI time series data. (b) DSVB framework integrating variational Bayes and recurrent layers to learn the sequential stochastic graph latent embeddings. (c) The alterations in latent embeddings readout are used to identify patients with Autism Spectrum Disorder (ASD) through a fully-connected classifier.

der classification problem, further works [Noman *et al.*, 2022; Liu *et al.*, 2023] incorporated GCNs to initially learn spatial-aware graph embeddings of the dynamic FC. This was followed by a long short-term memory (LSTM) model to leverage the temporal dependencies of these learned graph embeddings for subject-level network classification of psychiatric disorders. These methods have predominantly relied on deterministic models, and their compartmentalized, task-specific modules hinder the retention of full spatiotemporal information across layers.

In this paper, a deep spatiotemporal variational Bayes (DSVB) framework is proposed to learn the stochastic dynamic graph embeddings of FC networks, composed of sliding window segments of the resting-state fMRI time-series data. Figure 1 shows an overview of the proposed DSVB framework. Derived from general sequential variational Bayes (VB), our proposed framework provides a stochastic representation modelling solution of using dynamic FC networks for neuropsychiatric disorder diagnosis. Unlike existing GNN-based approaches that learn independent and identically distributed (i.i.d.) graph embeddings, we incorporate a spatial-aware recurrent neural network with attention-based message-passing scheme to model both the complex topological structure and temporal dependencies within a dynamic FC network sequence. Additionally, we introduce an adversarial training strategy to overcome the overfitting issue aggravated by limited training data. Our contributions are highlighted as follows:

1. An end-to-end probabilistic framework, integrating variational Bayes and recurrent GNNs, to learn the sequential stochastic graph latent embedding spaces in a purely unsupervised manner. This accommodates a wider range

of spatiotemporal variability across unseen brain FC networks.

2. A spatial-aware Gated Recurrent Unit (GRU) with attention-based message-passing scheme is incorporated to generate hierarchical latent embeddings to extract underlying non-Euclidean topological structure and rich temporal patterns of dynamic FC networks. The alterations in these latent embeddings are then exploited for downstream ASD identification.
3. An adversarial model perturbation strategy is introduced to apply regularization on the GNN-based graph embedding models during training. This adversarial regularization facilitates learning of a smooth latent embedding space that extrapolates well to unseen dynamic FC networks beyond the training dataset.
4. A nested-stratified 5-fold cross-validation shows that our proposed DSVB framework outperforms state-of-the-art methods in ASD identification from fMRI.

2 Methods

2.1 Dynamic Brain Network Construction

Dynamic brain graphs were constructed from fMRI data using the sliding-window technique. The regions of interest (ROI) fMRI time series was segmented into multiple overlapping blocks using sliding windows of size $L = 20$ with a shift of 10 time steps. To represent the dynamic FC networks, we computed the correlation matrix between ROI time series for each time window. The Ledoit-Wolf (LDW) regularized shrinkage covariance estimator is used to ensure well-conditioned FC estimates.

The dynamic brain network at each time-step t is represented by a graph $G_t \equiv (V, E)$, where $v_i \in V$ represents a particular brain ROI and $e_{ij} \in E$ is the connectivity edge between a pair of nodes v_i and v_j . The topological structure of dynamic brain networks G_t of N (the number of) nodes can be represented by a sequence of time-resolved adjacency matrices $\mathbf{A}_t = [a_{t,ij}] \in \{0, 1\}^{N \times N}$ (for each t), which were constructed by proportional thresholding of the sliding-windowed correlation-based FC matrices to keep the strongest 40% of connections based on absolute correlation value, and setting other connections to zero. Each dynamic graph G_t has associated node features $\mathbf{X}_t = [\mathbf{x}_{t,1}, \dots, \mathbf{x}_{t,N}]^\top \in \mathbb{R}^{N \times D_x}$ where $\mathbf{x}_{t,i} \in \mathbb{R}^{D_x}$ is the feature (column) vector of node v_i . We used the connection weights of each node for $\mathbf{x}_{t,i}$.

2.2 Spatiotemporal Variational Bayes

To implicitly learn the spatiotemporal structure of dynamic FC, we consider the unsupervised learning task of reconstructing the sequential network-derived adjacency matrices $\mathcal{A} = \{\mathbf{A}_t\}_{t=0}^T$ from node feature data $\mathcal{X} = \{\mathbf{X}_t\}_{t=0}^T$, where $T + 1$ is the sequence length. Under the sequential Bayesian framework, this is achieved via maximizing the variational evidence lower bound (ELBO) of conditional adjacency log-likelihood $\log p_\theta(\mathcal{A}|\mathcal{X})$.

Taking into account the difficulty of directly modelling the complex graph log-likelihood, we introduce a sequence of latent embeddings $\mathcal{Z} = \{\mathbf{Z}_t\}_{t=0}^T$ with $\mathbf{Z}_t = [\mathbf{z}_{t,1}, \dots, \mathbf{z}_{t,N}]^\top \in \mathbb{R}^{N \times D_z}$ where $\mathbf{z}_{t,i} \in \mathbb{R}^{D_z}$, and consider an importance decomposition of the ELBO:

$$\mathcal{L}^{\text{ELBO}}(\theta, \vartheta) = \mathbb{E}_{q_\vartheta(\mathcal{Z}|\mathcal{X}, \mathcal{A})} \left[\log \frac{p_\theta(\mathcal{A}, \mathcal{Z}|\mathcal{X})}{q_\vartheta(\mathcal{Z}|\mathcal{X}, \mathcal{A})} \right] \quad (1)$$

where the subscripts θ and ϑ denote GNN parameters that model the generative (prior and decoder) distribution $p_\theta(\mathcal{A}, \mathcal{Z}|\mathcal{X})$ and the approximate posterior (encoder) distribution $q_\vartheta(\mathcal{Z}|\mathcal{X}, \mathcal{A})$ which we can factorize, respectively as

$$\begin{aligned} p_\theta(\mathcal{A}, \mathcal{Z}|\mathcal{X}) &= \prod_{t=0}^T \underbrace{p_\theta(\mathbf{A}_t|\mathbf{Z}_{\leq t}, \mathbf{X}_{< t}, \mathbf{A}_{< t})}_{\text{decoder}} \times \\ &\quad \underbrace{p_\theta(\mathbf{Z}_t|\mathbf{X}_{< t}, \mathbf{A}_{< t}, \mathbf{Z}_{< t})}_{\text{prior}} \\ q_\vartheta(\mathcal{Z}|\mathcal{X}, \mathcal{A}) &= \prod_{t=0}^T \underbrace{q_\vartheta(\mathbf{Z}_t|\mathbf{X}_{\leq t}, \mathbf{A}_{\leq t}, \mathbf{Z}_{< t})}_{\text{encoder}} \end{aligned} \quad (2)$$

Based on the ancestral factorization above, we can then expand (1) to obtain a sequential ELBO (sELBO) as follows:

$$\begin{aligned} \mathcal{L}^{\text{sELBO}}(\theta, \vartheta) &= \\ &\sum_{t=0}^T \left(\mathbb{E}_{q_\vartheta(\mathbf{Z}_t|\mathbf{X}_{\leq t}, \mathbf{A}_{\leq t}, \mathbf{Z}_{< t})} \left[\log p_\theta(\mathbf{A}_t|\mathbf{Z}_{\leq t}, \mathbf{X}_{< t}, \mathbf{A}_{< t}) \right] \right. \\ &\quad \left. - \mathcal{D}^{\text{KL}} \left[q_\vartheta(\mathbf{Z}_t|\mathbf{X}_{\leq t}, \mathbf{A}_{\leq t}, \mathbf{Z}_{< t}) \parallel p_\theta(\mathbf{Z}_t|\mathbf{X}_{< t}, \mathbf{A}_{< t}, \mathbf{Z}_{< t}) \right] \right) \end{aligned} \quad (3)$$

where $\mathbf{X}_{\leq t}$ and $\mathbf{X}_{< t}$ denote the partial sequences up to the t^{th} and $(t-1)^{\text{th}}$ time samples, respectively. \mathcal{D}^{KL} denotes the (positive-valued) Kullback–Leibler divergence (KLD), which measures the statistical discrepancy of the encoder from the prior. The conditional probabilities of (3) encapsulate the underlying topological structure and temporal coherence of the dynamic brain networks. This sELBO (3) then forms the basis of our proposed DSVB framework.

2.3 Recurrent Graph Neural Network

In this subsection, we introduce a model parameterization of the formulated sELBO based on a graph recurrent neural network. Here, the conditional latent prior and approximate posterior on each node v_i are modelled as Gaussian distributions:

$$p_\theta(\mathbf{z}_{t,i}|\mathbf{X}_{< t}, \mathbf{A}_{< t}, \mathbf{Z}_{< t}) = \mathcal{N}(\boldsymbol{\mu}_{t,i}^{\text{prior}}, \boldsymbol{\Sigma}_{t,i}^{\text{prior}}) \quad (4a)$$

$$q_\vartheta(\mathbf{z}_{t,i}|\mathbf{X}_{\leq t}, \mathbf{A}_{\leq t}, \mathbf{Z}_{< t}) = \mathcal{N}(\boldsymbol{\mu}_{t,i}^{\text{enc}}, \boldsymbol{\Sigma}_{t,i}^{\text{enc}}) \quad (4b)$$

with isotropic covariances $\boldsymbol{\Sigma}_{t,i}^{\text{prior}} = \text{Diag}(\boldsymbol{\sigma}_{t,i}^{\text{prior}^2})$, $\boldsymbol{\Sigma}_{t,i}^{\text{enc}} = \text{Diag}(\boldsymbol{\sigma}_{t,i}^{\text{enc}^2})$, where $\text{Diag}(\cdot)$ denotes the diagonal function. The network-wide prior and approximate posterior mean and standard deviation pairs of \mathbf{Z}_t are then obtained as follows:

$$(\boldsymbol{\mu}_t^{\text{prior}}, \boldsymbol{\sigma}_t^{\text{prior}}) = \varphi_\theta^{\text{prior}}(\mathbf{H}_t) \quad (5a)$$

$$(\boldsymbol{\mu}_t^{\text{enc}}, \boldsymbol{\sigma}_t^{\text{enc}}) = \Phi_\vartheta^{\text{enc}}(\varphi_\theta^x(\mathbf{X}_t), \mathbf{H}_t, \mathbf{A}_t) \quad (5b)$$

where the prior model $\varphi_\theta^{\text{prior}}$, data feature model φ_θ^x , and latent feature model φ_θ^z are modelled via fully-connected neural networks (FCNNs); the encoder model $\Phi_\vartheta^{\text{enc}}$ is modelled via the more expressive GNN. The memory-embedding recurrent states $\mathbf{H}_t = [\mathbf{h}_{t,1}, \dots, \mathbf{h}_{t,N}]^\top \in \mathbb{R}^{N \times D_h}$ are obtained as

$$\mathbf{H}_t = \Phi_\theta^{\text{mn}}(\varphi_\theta^x(\mathbf{X}_{t-1}), \varphi_\theta^z(\mathbf{Z}_{t-1}), \mathbf{H}_{t-1}, \mathbf{A}_{t-1}) \quad (6)$$

and the recurrent model Φ_θ^{mn} is modelled as a spatial-aware Gated Recurrent Unit (GRU). By virtue of (6), \mathbf{H}_t thus serves as the memory embeddings for the preceding graphs and embeddings (history path) $\mathbf{Z}_{< t}$, $\mathbf{X}_{< t}$, and $\mathbf{A}_{< t}$. Based on our model parameterization, a deterministic closed-form solution of the KLD in sELBO (3) can be analytically obtained as

$$\begin{aligned} \mathcal{D}^{\text{KL}}(\theta, \vartheta) &= \\ &= \frac{1}{2} \sum_{i=1}^N \sum_{l=1}^{D_z} \left[\frac{\boldsymbol{\sigma}_{t,il}^{\text{enc}^2}}{\boldsymbol{\sigma}_{t,il}^{\text{prior}^2}} - \log \frac{\boldsymbol{\sigma}_{t,il}^{\text{enc}^2}}{\boldsymbol{\sigma}_{t,il}^{\text{prior}^2}} + \frac{(\boldsymbol{\mu}_{t,il}^{\text{enc}} - \boldsymbol{\mu}_{t,il}^{\text{prior}})^2}{\boldsymbol{\sigma}_{t,il}^{\text{prior}^2}} - 1 \right] \end{aligned} \quad (7)$$

where \mathbf{a}_l denotes the l^{th} element of the vector \mathbf{a} .

Subsequently, we model the adjacency matrix decoder $p_\theta(\mathbf{A}_t|\mathbf{Z}_{\leq t}, \mathbf{X}_{< t}, \mathbf{A}_{< t})$ as a Bernoulli distribution, conditioned on the adjacency matrix reconstruction $\hat{\mathbf{A}}_t$ obtained via the following inner product:

$$\hat{\mathbf{A}}_t = \sigma([\mathbf{Z}_t, \mathbf{H}_t][\mathbf{Z}_t, \mathbf{H}_t]^\top) \quad (8)$$

where σ denotes the sigmoid activation. In particular, we allow the adjacency matrix decoder to be conditioned on the entire history path, which suggested the incorporation of the

memory-embedding \mathbf{H}_t in the inner-product decoder, as opposed to the variational graph recurrent neural network [Hajiramezani *et al.*, 2019], which only considered \mathbf{Z}_t . In consideration of (8), the adjacency matrix decoder can be written in the form of a binary cross entropy (BCE) loss as follows:

$$\begin{aligned} \mathcal{L}^{\text{BCE}}(\theta, \vartheta) = & \sum_{i=1}^N \sum_{j=1}^N \left[\mathbf{A}_{t,ij} \log \sigma \left([\mathbf{Z}_t, \mathbf{H}_t]_i^\top [\mathbf{Z}_t, \mathbf{H}_t]_j \right) \right. \\ & \left. + (1 - \mathbf{A}_{t,ij}) \log \left(1 - \sigma \left([\mathbf{Z}_t, \mathbf{H}_t]_i^\top [\mathbf{Z}_t, \mathbf{H}_t]_j \right) \right) \right] \end{aligned} \quad (9)$$

where \mathbf{A}_{ij} and \mathbf{A}_i denote the $(i, j)^{\text{th}}$ element and i^{th} row of the matrix \mathbf{A} , respectively. Hence, this BCE acts as the reconstruction loss between the estimated and the ground-truth adjacency graph edges. In particular, we approximate the first expectation term in (3) via Monte Carlo integration with Monte Carlo samples $\{\mathbf{Z}_t^k, \mathbf{H}_t^k\}_{k=1}^M$, where k is the sample index and M is the number of samples. The approximate posterior latent samples $\mathbf{Z}_t^k \sim q_\vartheta(\mathbf{Z}_t | \mathbf{X}_{\leq t}, \mathbf{A}_{\leq t}, \mathbf{Z}_{< t})$ are obtained via (5b) using the reparameterization $\mathbf{Z}_t^k = \boldsymbol{\mu}_t^{\text{enc},k} + \boldsymbol{\sigma}_t^{\text{enc},k} \odot \boldsymbol{\epsilon}_t^k$, where $\boldsymbol{\epsilon}_t^k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and \odot denotes the Hadamard (element-wise) product. The recurrent state samples \mathbf{H}_t^k are obtained via (6) given the previous time-step \mathbf{Z}_{t-1}^k and \mathbf{H}_{t-1}^k . Substituting (9) and (7) into (3), we establish an unsupervised sELBO loss that underpins the proposed DSVB framework.

In sum, the KLD (7) compels the learnable encoder and prior model distributions (p_θ, q_ϑ) to resemble each other, thus regularize the models from the brain network data $(\mathbf{X}_t, \mathbf{A}_t)$ at current time-step. Furthermore, it promotes emphasis of the learned dynamic latent embeddings on the preceding history path $(\mathbf{X}_{< t}, \mathbf{A}_{< t})$ to incorporate strong temporal coherence, essential for propagating the learnable model distributions forward in time. On the other hand, the BCE (9) and the inner-product decoder (8) enforce the encoder model to construct latent embeddings that closely adhere to the non-Euclidean spatial characteristics of the dynamic brain networks. Together, these losses facilitate an unsupervised learning of generalized spatiotemporal embeddings that accommodate an extensive range of subject-level spatiotemporal variability across dynamic brain networks.

2.4 Attention-based Message Passing and Spatial-aware Gated Recurrent Unit

GNN is a class of message passing neural networks that uses aggregated topological information to construct and update node-level graph embeddings. In our proposed DSVB, we consider a GNN with message passing scheme inspired by the multi-head attention mechanism in the Transformer [Vaswani *et al.*, 2017], of which the node embeddings are updated across layers as follows:

$$\mathbf{f}_i^{(l+1)} = \mathbf{W}_1 \mathbf{f}_i^{(l)} + \sum_{j \in \mathcal{N}(i)} \alpha_{ij} (\mathbf{W}_2 \mathbf{f}_j^{(l)} + \mathbf{W}_5 \mathbf{f}_{ij}) \quad (10)$$

and α_{ij} are the attention coefficients computed as

$$\alpha_{ij} = \text{softmax} \left(\frac{(\mathbf{W}_3 \mathbf{f}_i^{(l)})^\top (\mathbf{W}_4 \mathbf{f}_j^{(l)} + \mathbf{W}_5 \mathbf{f}_{ij})}{\sqrt{D_{l+1}}} \right) \quad (11)$$

with learnable weights $\{\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_5\}$ that project the embeddings from \mathbb{R}^{D_l} to $\mathbb{R}^{D_{l+1}}$, and $\mathbf{f}^{(l)}$ denotes the node embeddings at the l^{th} GNN layer.

It follows from the scaled dot product (11) that the node embeddings \mathbf{f}_i and its neighbors \mathbf{f}_j can be interpreted as the ‘query’ and ‘key’ of a self-attention mechanism. The edge features are added to the key vectors in order to provide an additional source of information for the node. A similarity measure between the query and keys is then calculated via dot product to obtain the coefficients α_{ij} , which act as attention scores in the feature aggregation step of (10). The encoder model (5b) of the DSVB is thus modelled by a two-layered GNN with this message passing scheme, and we let $\mathbf{f}^{(0)} = [\varphi_\theta^x(\mathbf{X}_t), \mathbf{H}_t, \mathbf{A}_t]^\top$ and $\mathbf{f}^{(2)} = [\boldsymbol{\mu}_t^{\text{enc}}, \boldsymbol{\Sigma}_t^{\text{enc}}]^\top$ in (10).

To accurately model spatiotemporal dependencies, we parameterize the recurrent model (6) using a spatial-aware GRU, for which the recurrent states are updated across time-steps as follows:

$$\begin{aligned} \mathbf{S}_t &= \sigma(\Phi_{xz}(\bar{\mathbf{X}}_{t-1}, \mathbf{A}_{t-1}) + \Phi_{hz}(\mathbf{H}_{t-1}, \mathbf{A}_{t-1})) \\ \mathbf{R}_t &= \sigma(\Phi_{xr}(\bar{\mathbf{X}}_{t-1}, \mathbf{A}_{t-1}) + \Phi_{hr}(\mathbf{H}_{t-1}, \mathbf{A}_{t-1})) \\ \tilde{\mathbf{H}}_t &= \tanh(\Phi_{xh}(\bar{\mathbf{X}}_{t-1}, \mathbf{A}_{t-1}) + \Phi_{hh}(\mathbf{R}_t \odot \mathbf{H}_{t-1}, \mathbf{A}_{t-1})) \\ \mathbf{H}_t &= \mathbf{S}_t \odot \mathbf{H}_{t-1} + (1 - \mathbf{S}_t) \odot \tilde{\mathbf{H}}_t \end{aligned} \quad (12)$$

with $\bar{\mathbf{X}}_{t-1} = [\varphi_\theta^x(\mathbf{X}_{t-1}), \varphi_\theta^z(\mathbf{Z}_{t-1})]^\top$. By virtue of the model equations (12), \mathbf{H}_t thus serves as a memory embedding that retains graph-structured temporal information of the preceding latent state sequence $\mathbf{Z}_{< t}$. In comparison to conventional GRU, the FCNN therein is replaced by the set of GNNs $\Phi_{xz}, \Phi_{hz}, \Phi_{xr}, \Phi_{hr}, \Phi_{xh}, \Phi_{hh}$. Similarly, these single-layered GNNs adopt the message passing scheme (10), which allows the modified GRU to simultaneously exploit meaningful spatial structures and temporal dependencies of the dynamic graph-structured data.

2.5 Latent Embeddings for Graph Classification

Due to the recurrent nature of our proposed DSVB, the generated latent embeddings preserve non-Euclidean topological structure and rich temporal patterns which can be further exploited for dynamic graph classification. In combination, these node embeddings entail crucial spatiotemporal information of the entire dynamic graph sequence. To summarize the subject-level dynamic graphs, we apply a flattening operation on the mean of hierarchical embeddings $\mathbf{Z}_T, \mathbf{H}_T$ of the final time-step T to obtain the vectorized readout $\text{vec}([\boldsymbol{\mu}_T^{\text{enc}}, \boldsymbol{\mu}_T^{\text{rec}}]^\top)$, where $\boldsymbol{\mu}_T^{\text{rec}} = \frac{1}{M} \sum_{k=1}^M \mathbf{H}_T^k$.

Subsequently, the flattened readout is fed into the following classification model:

$$\hat{\mathbf{y}} = \varphi_\tau^{\text{classifier}}(\text{vec}[\boldsymbol{\mu}_T^{\text{enc}}, \boldsymbol{\mu}_T^{\text{rec}}]^\top) \quad (13)$$

modelled by multilayer FCNN. A softmax activation is then applied on the obtained logits $\hat{\mathbf{y}} \in \mathbb{R}^2$ to get the predictive probability scores for the final subject-level brain network classification. The predicted class label is thus the one with highest predictive score. The classification loss is taken to be

$$\begin{aligned}
 \mathcal{L}^{\text{DSVB-FCNN}}(\theta, \vartheta, \tau) &= \mathcal{L}^{\text{BCE}}(\theta, \vartheta) + \mathcal{D}^{\text{KL}}(\theta, \vartheta) + \mathcal{L}^{\text{MCE}}(\tau) \\
 &= \frac{1}{M} \sum_{k=1}^M \sum_{t=0}^T \sum_{i=1}^N \sum_{j=1}^N \underbrace{\left[\mathbf{A}_{t,ij} \log \sigma([\mathbf{Z}_t^k, \mathbf{H}_t^k]^\top [\mathbf{Z}_t^k, \mathbf{H}_t^k]_j) - (1 - \mathbf{A}_{t,ij}) \log (1 - \sigma([\mathbf{Z}_t^k, \mathbf{H}_t^k]^\top [\mathbf{Z}_t^k, \mathbf{H}_t^k]_j)) \right]}_{\mathcal{L}^{\text{BCE}}(\theta, \vartheta)} \\
 &\quad + \underbrace{\frac{1}{2} \sum_{t=0}^T \sum_{i=1}^N \sum_{l=1}^{D_z} \left[\frac{\sigma_{t,il}^{\text{enc}2}}{\sigma_{t,il}^{\text{prior}2}} - \log \frac{\sigma_{t,il}^{\text{enc}2}}{\sigma_{t,il}^{\text{prior}2}} + \frac{(\mu_{t,il}^{\text{enc}} - \mu_{t,il}^{\text{prior}})^2}{\sigma_{t,il}^{\text{prior}2}} - 1 \right]}_{\mathcal{D}^{\text{KL}}(\theta, \vartheta)} + \underbrace{\sum_{n=1}^C \mathbf{c}_n \log \frac{\exp(\hat{y}_n)}{\sum_{m=1}^C \exp(\hat{y}_m)}}_{\mathcal{L}^{\text{MCE}}(\tau)}
 \end{aligned} \tag{15}$$

the following multi-class cross entropy:

$$\mathcal{L}^{\text{MCE}}(\tau) = \sum_{n=1}^C \mathbf{c}_n \log \frac{\exp(\hat{y}_n)}{\sum_{m=1}^C \exp(\hat{y}_m)} \tag{14}$$

where $\mathbf{c} \in \{0, 1\}^C$ is the class label and C is the number of classes. Incorporating (14) into the sELBO (3), we obtain the final DSVB-FCNN loss $\mathcal{L}^{\text{DSVB}}$ in (15).

Unlike previous methods, the FCNN classifier (13) is trained jointly with the variational recurrent graph autoencoder (5)-(6) in an end-to-end fashion. Gradients of the classifier are allowed to back-propagate through time and update parameters $\{\theta, \vartheta\}$ of the DSVB models.

2.6 Adversarial Model Regularization

Under a limited amount of training data, deep graph representation learning models are prone to generating node embeddings that aggravate overfitting. Such models typically construct a latent embedding space that overly- or grossly - overfits the limited dataset. Consequently, downstream task-specific (e.g., graph classification) models that are adapted to this data-specific latent space perform poorly on unseen latent embeddings of general graphs. Taking this into consideration, a probabilistic adversarial training strategy is introduced to regularize the latent embedding space constructed by the DSVB models. With the adversarial regularization, the DSVB models are expected to overcome the data overfitting issue by realizing a more inclusive latent embedding space that can be readily extrapolated to unseen data beyond the limited training dataset.

Inspired by the model perturbation strategy in domain adversarial training [Ganin *et al.*, 2016], we train the parameters $\{\theta, \vartheta, \tau\}$ of the DSVB-FCNN loss (15) in adversarial fashion:

$$\begin{aligned}
 (\theta, \vartheta) &= \arg \max_{\theta, \vartheta} \mathcal{L}^{\text{DSVB-FCNN}}(\theta, \vartheta, \tau) \\
 \tau &= \arg \min_{\tau} \mathcal{L}^{\text{MCE}}(\tau)
 \end{aligned} \tag{16}$$

On one hand, the DSVB parameters (θ, ϑ) are optimized to generate hierarchical embeddings $\mathbf{Z}_T, \mathbf{H}_T$ that fool the classifier $\varphi_{\tau}^{\text{classifier}}$. On the other hand, the FCNN classification model parameters τ are optimized to accurately distinguish the class label of the generated latent embeddings. Such an adversarial competition is expected to achieve the Nash equilibrium that generalizes the latent embedding space and prevents model overfitting.

3 Application to Dynamic Brain Networks

3.1 fMRI Dataset & Preprocessing

We used a C-PAC pipeline [Cameron *et al.*, 2013] preprocessed fMRI dataset from the Autism Brain Imaging Data Exchange (ABIDE I) open source database [Di Martino *et al.*, 2009], with a sample of 144 subjects (70 ASD and 74 healthy controls (HC)) resting-state fMRI included in this case study. The inclusion criteria [Plitt *et al.*, 2015] are: males with a full-scale IQ > 80; ages between 11 and 23; and the fMRI acquisition sites are New York University, University of California Los Angeles 1, and University of Utah, School of Medicine. We performed parcellation using Power *et al.*'s brain atlas [Power *et al.*, 2011] to extract the mean time series for a total of $N = 264$ regions of interest (ROIs).

3.2 Graph Neural Network Model Training

The proposed DSVB framework is used to generate the dynamic node embeddings that correspond to the node feature and adjacency matrix sequences. The FCNN classifier is then used to discriminate between the ASD and HC subjects. We train the end-to-end DSVB-FCNN via an Adam optimizer using a learning rate of $1e-5 = 10^{-5}$ with exponential annealing, and a L2 regularization of 0.01 for 400 epochs. The output dimension of the graph-structure GRU is chosen to be 16. The output dimensions of each layer of the GNN encoder are chosen to be 32 and 16, respectively. The output dimensions of the single-layered FCNN data and latent feature models are chosen to be 64 and 8, respectively. The output dimensions of each layer of the FCNN classifier are chosen to be 32 and 2, respectively. These model hyperparameters are selected based on a 5-fold cross-validation.

3.3 Baseline Methods

SVM: The flattened upper triangle of the FC matrix is used as input to SVM. **BrainNetCNN** [Kawahara *et al.*, 2017]: An extension of CNNs to handle brain graph-structured data using special kernels. **GroupINN** [Yan *et al.*, 2019b]: An ensemble of GCNs to learn graph-level latent embedding representations. **ASD-DiagNet** [Eslami *et al.*, 2019]: Dense autoencoder for embedding learning and single-layered perceptron for classification. **Hi-GCN** [Jiang *et al.*, 2020b]: This method uses GroupINN to produce the embedding for all network instances, the learned embeddings are fed to a population-based GCN for classification. **E-HI-GCN** [Li *et*

Type of FC	Classifier	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)	AUC (%)
Static	SVM	56.26 ± 4.51	55.71 ± 3.19	55.05 ± 4.69	55.37 ± 3.97	56.24 ± 3.99
	BrainNetCNN [Kawahara <i>et al.</i> , 2017]	57.04 ± 10.99	44.29 ± 15.29	61.52 ± 15.24	49.03 ± 13.10	56.62 ± 10.86
	ASD-DiagNet [Eslami <i>et al.</i> , 2019]	68.03 ± 2.87	58.57 ± 12.29	74.04 ± 10.60	63.55 ± 4.05	67.62 ± 2.64
	GroupINN [Yan <i>et al.</i> , 2019b]	64.53 ± 9.06	34.29 ± 24.49	67.88 ± 35.09	43.10 ± 25.83	63.76 ± 9.37
	Hi-GCN [Jiang <i>et al.</i> , 2020b]	66.60 ± 6.05	60.00 ± 10.69	68.51 ± 7.19	63.30 ± 6.74	66.38 ± 6.06
	E-Hi-GCN [Li <i>et al.</i> , 2021]	66.60 ± 8.64	54.29 ± 21.48	75.42 ± 14.71	58.31 ± 18.40	58.00 ± 6.38
Dynamic	SVM	63.82 ± 9.41	55.43 ± 7.23	56.66 ± 3.34	55.52 ± 3.61	63.52 ± 8.20
	BrainNetCNN [Kawahara <i>et al.</i> , 2017]	54.49 ± 8.15	57.14 ± 8.34	50.00 ± 14.00	53.33 ± 10.15	54.58 ± 7.03
	GAE-FCNN [Noman <i>et al.</i> , 2022]	66.03 ± 7.14	65.71 ± 14.57	68.92 ± 16.26	64.96 ± 6.37	66.19 ± 7.17
	GAE-LSTM [Noman <i>et al.</i> , 2022]	54.78 ± 6.05	47.14 ± 19.48	71.00 ± 25.12	54.92 ± 17.09	54.53 ± 6.13
	DSVB-FCNN	78.44 ± 2.77	66.67 ± 7.45	89.64 ± 5.90	75.99 ± 4.01	78.94 ± 2.64

Table 1: Performance comparison of our proposed DSVB-FCNN with different static and dynamic baseline classifiers.

Model Specification	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)	AUC (%)
I Graph Autoencoder + Conventional GRU	72.39 ± 9.01	65.00 ± 8.16	78.91 ± 13.04	71.02 ± 9.05	72.65 ± 9.08
II Graph Autoencoder + Spatial-aware GRU	76.63 ± 9.08	73.33 ± 8.16	82.05 ± 11.47	76.74 ± 7.47	76.74 ± 9.32
III Variational Graph Autoencoder + Spatial-aware GRU	75.00 ± 10.06	63.33 ± 6.67	86.94 ± 15.26	72.78 ± 8.90	75.38 ± 10.31
IV Graph Autoencoder + Spatial-aware GRU + Adversarial Training	77.61 ± 5.65	76.67 ± 6.24	79.48 ± 6.45	77.96 ± 5.73	77.65 ± 5.67
V Variational Graph Autoencoder + Spatial-aware GRU + Adversarial Training	78.44 ± 2.77	66.67 ± 7.45	89.64 ± 5.90	75.99 ± 4.01	78.94 ± 2.64

Table 2: Ablation study of our proposed DSVB framework based on a nested-stratified 5-fold cross-validation.

et al., 2021]: An ensemble of HI-GCN, each of which is trained on different sparsity level brain networks. **GAE-FCNN** [Noman *et al.*, 2022]: GAE is used to learn dynamic graph-level latent embeddings, which are then fed to a FCNN for classification. **GAE-LSTM** [Noman *et al.*, 2022]: The learned embeddings of GAE are fed to a LSTM for classification.

4 Results and Discussion

To evaluate the effectiveness of our proposed method, a nested-stratified 5-fold cross-validation was applied for each experiment. The performance of ASD identification is measured using five metrics: accuracy, recall, precision, F1-score, and area under the curve (AUC). Table 1 compares the average performances of the proposed DSVB-FCNN with the existing baseline classifiers based on both dynamic and static FC networks. It shows that our proposed DSVB-FCNN achieves substantial improvement over the baseline classifiers in each metric. The DSVB-FCNN also achieved low standard deviations in all metrics, demonstrating a strong consistency in its performances across the 5-fold cross-validation. In general, these results indicate that the framework facilitates robust brain disorder identification by generating graph latent embeddings that generalize well to unseen dynamic FC networks beyond the limited training dataset.

To provide insights into these performance gains of the DSVB-FCNN, we conduct an ablation study where certain components (i.e., variational Bayes, spatial-aware GRU and adversarial training) of the DSVB framework are removed to highlight their distinctive contributions to the improved dynamic FC classification result. Table 2 shows the average performances of the ablation study based on five DSVB model variants. Model I underperformed in comparison to other spatial-aware GRU-empowered models. This indicates the importance of the proposed spatial-aware GRU in endowing recurrent model’s ability to capture rich spatiotemporal patterns crucial for dynamic network classification.

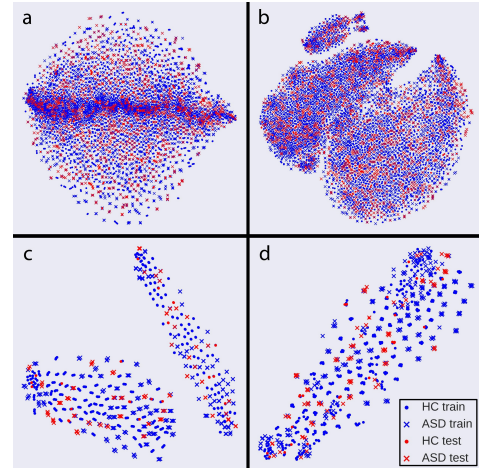


Figure 2: Top figures are t-SNE visualizations on the latent state sequences $\{\mathbf{Z}_t\}_{t=0}^T$ of (a) graph recurrent autoencoder and (b) variational graph recurrent autoencoder. Bottom figures are t-SNE visualizations on the final (time-step) readouts $\text{vec}([\boldsymbol{\mu}_T^{\text{enc}}, \boldsymbol{\mu}_T^{\text{H}}]^T)$ of (c) variational graph recurrent autoencoder and (d) variational graph recurrent autoencoder with adversarial training.

Figures 2a and 2b show the t-distributed Stochastic Neighbor Embedding (t-SNE) visualizations of respective latent state sequences $\{\mathbf{Z}_t\}_{t=0}^T$ from models II and III of the ablation study. It shows that the latent state t-SNEs of model II are more concentrated towards the centre in comparison to the t-SNEs of model III, which is more uniform. This indicates that the VB elements (KLD and reparameterization) facilitate learning of a smooth latent embedding space to better accommodate subject-level spatiotemporal variability across unseen dynamic FC networks. Nevertheless, due to incorporation of the prior model via KLD loss (7), VB-based models generate latent embeddings that are highly accustomed to the spatiotemporal structure of the dynamic FC data. This ag-

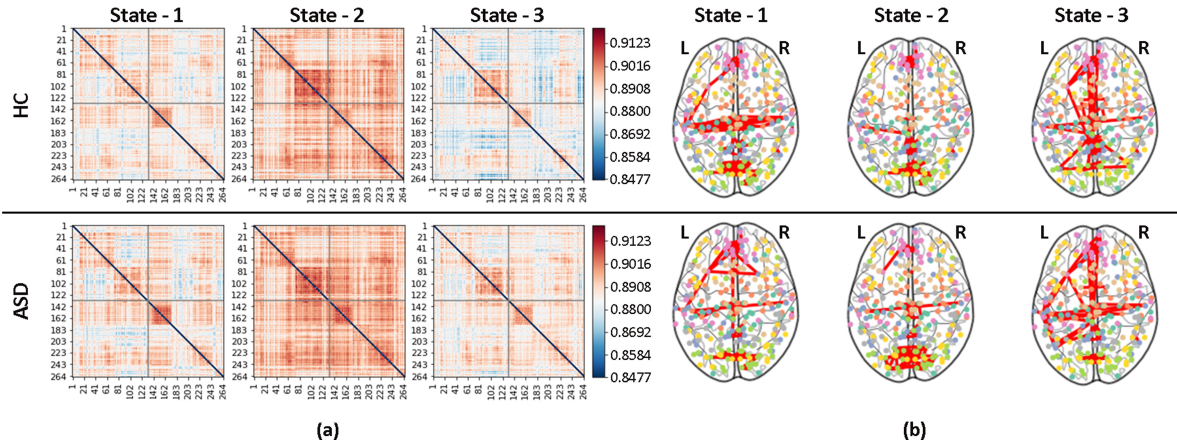


Figure 3: The average connectivity patterns derived from DSVB-learned embeddings. (a) 3 main states derived from k-means clustering of 264 x 264 ROIs connectivity matrices from HC and ASD groups. (b) The corresponding connectome plots for HC and ASD.

gravates model overfitting and increases the chance of false classification, resulting in an inferior performance of model III in Table 2 compared to model II.

Figures 2c and 2d show the t-SNE visualizations of respective readouts $\text{vec}([\mu_T^{\text{enc}}, \mu_T^{\text{rec}}]^T)$ from models III and V. In particular, Figure 2c shows that model III divided the readout t-SNEs into two predominant clusters based on the causal structure of each dynamic FC data. In contrast, t-SNEs of model V in Figure 2d are more regularized, which demonstrates the effectiveness of the proposed adversarial training strategy in learning indiscriminate latent embedding space that generalized well to unseen dynamic FC networks. The coordination between all components of the DSVB framework produces superior ASD identification performance, as shown in Tables 1 and 2.

In addition, we constructed higher-order dynamic FC by correlating the DSVB-learned embeddings z_{it} between pairs of nodes, and examined the dynamic connectivity states in ASD by applying dynamic FC clustering of these dynamic networks. The optimal number of clusters = 3 is determined using the Elbow method [Bholowalia and Kumar, 2014]. Figure 3 shows the estimated connectivity matrices for three different states with the corresponding highest FC connections. By referring to [Power *et al.*, 2011], the corresponding brain functional system can be traced back. The ASD group shows stronger FC connections in visual, default mode, salience networks; whereas higher connectivity is observed in HC for sensory and auditory networks. The difference in FC patterns between ASD and HC is state-dependent. In ASD, visual ROIs show increased interconnections in state-1 and state-2 where default mode and salience are mostly activated in state-1. This is consistent with [Holiga *et al.*, 2019], who also concluded that - compared to HC - ASD has hyperconnectivity in prefrontal and parietal cortices but hypoconnectivity in sensory-motor regions. Figure 4 shows that the latent embedding matrices derived from DSVB are more visually discriminative compared to the raw FC matrices. The ASD embeddings highlight hyperconnectivity within visual networks (ROIs 143 to 173), particularly on the occipital lobe, and this aligns with earlier neu-

rosience discoveries [Keehn *et al.*, 2008; Clery *et al.*, 2013; Matsuoka *et al.*, 2020].

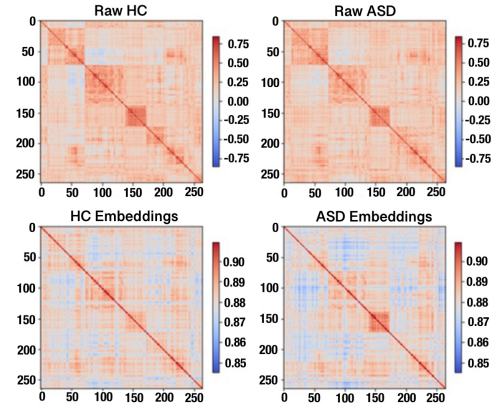


Figure 4: Comparison of the group means of raw FC (264 x 264 ROIs) with the latent embeddings derived from our DSVB framework over time for both HC and ASD groups.

5 Conclusion

We have developed a deep probabilistic spatiotemporal framework based on sequential variational Bayes for dynamic graph representation learning. The proposed DSVB framework incorporates a spatial-aware GRU to capture both topological and temporal alterations across brain networks. A downstream FCNN then leverages the learned graph embeddings to reveal atypical neural connectivity patterns. Evaluation on resting-state fMRI data shows substantial improvements over the existing baseline methods, suggesting potential in neuropsychiatric disorder identification. Despite our current focus on brain disorders, this framework is domain-agnostic and applicable to other dynamic graph-structured data. Moreover, the proposed framework could be extended to handle large graphs by incorporating sparse attention and heterogeneous graphs via cross-modality representations.

Acknowledgements

The work of Junn Yong Loo is supported by Monash University under the SIT Collaborative Research Seed Grants 2024 I-M010-SED-000242. The work of Chee-Ming Ting is supported by the Ministry of Higher Education, Malaysia under the Fundamental Research Grant Scheme FRGS/1/2023/ICT02/MUSM/02/1. Sin-Yee Yap has been based at Monash's Department of Data Science and AI for some of the duration of this work.

Contribution Statement

Sin-Yee Yap, Junn Yong Loo (corresponding author) and Chee-Ming Ting contributed equally to this work.

References

- [Aedo-Jury *et al.*, 2020] Felipe Aedo-Jury, Miriam Schwalm, Lara Hamzehpour, and Albrecht Stroh. Brain states govern the spatio-temporal dynamics of resting-state functional connectivity. *Elife*, 9, 2020.
- [Azevedo *et al.*, 2020] Tiago Azevedo, Luca Passamonti, Pietro Liò, and Nicola Toschi. A deep spatiotemporal graph learning architecture for brain connectivity analysis. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1120–1123. IEEE, 2020.
- [Bholowalia and Kumar, 2014] Purnima Bholowalia and Arvind Kumar. Ebc-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications*, 105(9), 2014.
- [Cameron *et al.*, 2013] Craddock Cameron, Sikka Sharad, Cheung Brian, Khanuja Ranjeet, Ghosh Satrajit, Yan Chaogan, Li Qingyang, Lurie Daniel, Vogelstein Joshua, Burns Randal, Colcombe Stanley, Mennes Maarten, Kelly Clare, Di Martino Adriana, Castellanos Francisco, and Milham Michael. Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (C-PAC). *Front. Neuroinform.*, 7, 2013.
- [Clery *et al.*, 2013] H Clery, F Andersson, F Bonnet-Brilhault, A Philippe, B Wicker, and M Gomot. fmri investigation of visual change detection in adults with autism. *NeuroImage: Clinical*, 2:303–312, 2013.
- [Di Martino *et al.*, 2009] Adriana Di Martino, Kathryn Ross, Lucina Q Uddin, Andrew B Sklar, F Xavier Castellanos, and Michael P Milham. Functional brain correlates of social and nonsocial processes in autism spectrum disorders: an activation likelihood estimation meta-analysis. *Biological psychiatry*, 65(1):63–74, 2009.
- [Eslami *et al.*, 2019] Taban Eslami, Vahid Mirjalili, Alvis Fong, Angela R Laird, and Fahad Saeed. ASD-DiagNet: a hybrid learning approach for detection of autism spectrum disorder using fMRI data. *Front. Neuroinform.*, 13:70, 2019.
- [Filippi *et al.*, 2019] Massimo Filippi, Edoardo G Spinelli, Camilla Cividini, and Federica Agosta. Resting state dynamic functional connectivity in neurodegenerative conditions: A review of magnetic resonance imaging findings. *Front. Neurosci.*, 13:657, 2019.
- [Ganin *et al.*, 2016] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.
- [Hajiramezanali *et al.*, 2019] Ehsan Hajiramezanali, Arman Hasanzadeh, Krishna Narayanan, Nick Duffield, Mingyuan Zhou, and Xiaoning Qian. Variational graph recurrent neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019.
- [Holiga *et al.*, 2019] Štefan Holiga, Joerg F Hipp, Christopher H Chatham, Pilar Garces, Will Spooren, Xavier Liogier D'Ardhuy, Alessandro Bertolino, Céline Bouquet, Jan K Buitelaar, Carsten Bours, Annika Rausch, Marianne Oldehinkel, Manuel Bouvard, Anouck Amestoy, Mireille Caralp, Sonia Gueguen, Myriam Ly-Le Moal, Josselin Houenou, Christian F Beckmann, Eva Loth, Declan Murphy, Tony Charman, Julian Tillmann, Charles Laidi, Richard Delorme, Anita Beggiano, Alexandru Gaman, Isabelle Scheid, Marion Leboyer, Marc-Antoine d'Albis, Jeff Sevigny, Christian Czech, Federico Bolognani, Garry D Honey, and Juergen Dukart. Patients with autism spectrum disorders display reproducible functional connectivity alterations. *Sci. Transl. Med.*, 11(481), 2019.
- [Hyde *et al.*, 2019] Kayleigh K Hyde, Marlena N Novack, Nicholas LaHaye, Chelsea Parlett-Pelleriti, Raymond Anden, Dennis R Dixon, and Erik Linstead. Applications of supervised machine learning in autism spectrum disorder research: a review. *Review Journal of Autism and Developmental Disorders*, 6(2):128–146, 2019.
- [Jiang *et al.*, 2020a] Hao Jiang, Peng Cao, Mingyi Xu, Jinzhu Yang, and Osmar Zaiane. Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction. *Comput. Biol. Med.*, 127:104096, 2020.
- [Jiang *et al.*, 2020b] Hao Jiang, Peng Cao, Mingyi Xu, Jinzhu Yang, and Osmar Zaiane. Hi-GCN: A hierarchical graph convolution network for graph embedding learning of brain network and brain disorders prediction. *Comput. Biol. Med.*, 127:104096, 2020.
- [Kawahara *et al.*, 2017] Jeremy Kawahara, Colin J Brown, Steven P Miller, Brian G Booth, Vann Chau, Ruth E Grunau, Jill G Zwicker, and Ghassan Hamarneh. Brainnetcn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.
- [Keehn *et al.*, 2008] Brandon Keehn, Laurie Brenner, Erica Palmer, Alan J Lincoln, and Ralph-Axel Mueller. Functional brain organization for visual search in asd. *Journal of the International Neuropsychological Society*, 14(6):990–1003, 2008.
- [Khodatars *et al.*, 2021] Marjane Khodatars, Afshin Shoeibi, Delaram Sadeghi, Navid Ghaasemi, Mahboobeh Jafari, Parisa Moridian, Ali Khadem, Roohallah Alizadehsani,

- Assef Zare, Yinan Kong, Abbas Khosravi, Saeid Naha-vandi, Sadiq Hussain, U Rajendra Acharya, and Michael Berk. Deep learning for neuroimaging-based diagnosis and rehabilitation of autism spectrum disorder: A review. *Comput. Biol. Med.*, 139:104949, 2021.
- [Kim *et al.*, 2021] Byung-Hoon Kim, Jong Chul Ye, and Jae-Jin Kim. Learning dynamic graph representation of brain connectome with spatio-temporal attention. *Advances in Neural Information Processing Systems*, 34:4314–4327, 2021.
- [Lee and Kim, 2019] Namgil Lee and Jong-Min Kim. Dynamic functional connectivity analysis of functional MRI based on copula time-varying correlation. *J. Neurosci. Methods*, 323:32–47, 2019.
- [Li *et al.*, 2021] Lanting Li, Hao Jiang, Guangqi Wen, Peng Cao, Mingyi Xu, Xiaoli Liu, Jinzhu Yang, and Osmar Zaiane. TE-HI-GCN: An ensemble of transfer hierarchical graph convolutional networks for disorder diagnosis. *Neuroinform.*, pages 1–23, 2021.
- [Li *et al.*, 2022] Lanting Li, Hao Jiang, Guangqi Wen, Peng Cao, Mingyi Xu, Xiaoli Liu, Jinzhu Yang, and Osmar Zaiane. TE-HI-GCN: An ensemble of transfer hierarchical graph convolutional networks for disorder diagnosis. *Neuroinformatics*, 20(2):353–375, 2022.
- [Liu *et al.*, 2023] Lingwen Liu, Guangqi Wen, Peng Cao, Tianshun Hong, Jinzhu Yang, Xizhe Zhang, and Osmar R. Zaiane. Braintgl: A dynamic graph representation learning model for brain network analysis. *Computers in Biology and Medicine*, 153:106521, 2023.
- [Matsuoka *et al.*, 2020] Kiwamu Matsuoka, Manabu Makinodan, Soichiro Kitamura, Masato Takahashi, Hiroaki Yoshikawa, Fumihiko Yasuno, Rio Ishida, Naoko Kishimoto, Yuka Yasuda, Ryota Hashimoto, et al. Increased dendritic orientation dispersion in the left occipital gyrus is associated with atypical visual processing in adults with autism spectrum disorder. *Cerebral Cortex*, 30(11):5617–5625, 2020.
- [Moridian *et al.*, 2022] Parisa Moridian, Navid Ghassemi, Mahboobeh Jafari, Salam Salloum-Asfar, Delaram Sadeghi, Marjane Khodatars, Afshin Shoeibi, Abbas Khosravi, Sai Ho Ling, Abdulhamit Subasi, Roohallah Alizadehsani, Juan M. Gorriz, Sara A. Abdulla, and U. Rajendra Acharya. Automatic autism spectrum disorder detection using artificial intelligence methods with mri neuroimaging: A review. *Frontiers in Molecular Neuroscience*, 15, 2022.
- [Noman *et al.*, 2022] Fuad Noman, Sin-Yee Yap, Raphaël C.-W. Phan, Hernando Ombao, and Chee-Ming Ting. Graph autoencoder-based embedded learning in dynamic brain networks for autism spectrum disorder identification. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 2891–2895, 2022.
- [Plitt *et al.*, 2015] Mark Plitt, Kelly Anne Barnes, and Alex Martin. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage Clin.*, 7:359–366, 2015.
- [Power *et al.*, 2011] Jonathan D Power, Alexander L Cohen, Steven M Nelson, Gagan S Wig, Kelly Anne Barnes, Jessica A Church, Alecia C Vogel, Timothy O Laumann, Fran M Miezin, Bradley L Schlaggar, and Steven E Petersen. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011.
- [RaviPrakash *et al.*, 2019] Harish RaviPrakash, Arjun Watane, Sachin Jambawalikar, and Ulas Bagci. Deep learning for functional brain connectivity: Are we there yet? In *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics*, pages 347–365. Springer International Publishing, 2019.
- [Tanveer *et al.*, 2020] M Tanveer, B Richhariya, R U Khan, A H Rashid, P Khanna, M Prasad, and C T Lin. Machine learning techniques for the diagnosis of alzheimer’s disease. *ACM Trans. Multimed. Comput. Commun. Appl.*, 16(1s):1–35, 2020.
- [Valenti *et al.*, 2020] Marco Valenti, Maria Chiara Pino, Monica Mazza, Gianvito Panzarino, Claudio Di Paolantonio, and Alberto Verrotti. Abnormal structural and functional connectivity of the corpus callosum in autism spectrum disorders: a review. *Rev. J. Autism. Dev. Disord.*, 7(1):46–62, 2020.
- [Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017.
- [Wang *et al.*, 2021] Qing Wang, Hua-Yun Li, Yun-Da Li, Ya-Ting Lv, Hui-Bin Ma, An-Feng Xiang, Xi-Ze Jia, and Dong-Qiang Liu. Resting-state abnormalities in functional connectivity of the default mode network in autism spectrum disorder: a meta-analysis. *Brain imaging and behavior*, pages 1–10, 2021.
- [Yan *et al.*, 2019a] Yujun Yan, Jiong Zhu, Marlena Duda, Eric Solarz, Chandra Sripada, and Danai Koutra. GroupINN: Grouping-based interpretable neural network for classification of limited, noisy brain data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 772–782. Association for Computing Machinery, 2019.
- [Yan *et al.*, 2019b] Yujun Yan, Jiong Zhu, Marlena Duda, Eric Solarz, Chandra Sripada, and Danai Koutra. GroupINN: Grouping-based interpretable neural network for classification of limited, noisy brain data. In *Proc. of the 25th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining*, pages 772–782, 2019.
- [Zhang *et al.*, 2020] Li Zhang, Mingliang Wang, Mingxia Liu, and Daoqiang Zhang. A survey on deep learning for Neuroimaging-Based brain disorder analysis. *Front. Neurosci.*, 14:779, 2020.
- [Zhou and Zhang, 2021] Haiyu Zhou and Daoqiang Zhang. Graph-In-Graph convolutional networks for brain disease diagnosis. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 111–115, 2021.