Distributionally-Adaptive Variational Meta Learning for Brain Graph Classification

Jing Du[1], Guangwei Dong[1], Congbo Ma[1], Shan Xue[1], Jia Wu[1], Jian Yang[1], Amin Beheshti[1], Quan Z. Sheng[1], and Alexis Giral[2]

[1] Macquarie University, Sydney, Australia
[2] Systemethix, Sydney, Australia
{jing.du, congbo.ma, emma.xue, jia.wu, jian.yang, amin.beheshti, michael.sheng}@mq.edu.au, guangwei.dong1@hdr.mq.edu.au
alexis.giral@systemethix.com.au

Abstract. Recent developments in Graph Neural Networks (GNNs) have shed light on understanding brain networks through innovative approaches. Despite these innovations, the significant costs associated with data collection and the challenges posed by data drift in real-world scenarios present substantial challenges for models dependent on large datasets to capture brain activity features. To address these issues, we propose the **D**istributionally-**A**daptive Variational **M**eta **L**earning (**DAML**) framework, designed to equip the model with rapid adaptability to varying distributions by meta learning-driven minimization of discrepancies between subject sets. Initially, we employ a graph encoder with the message-passing strategy to generate precise brain graph representations. Subsequently, we implement a distributionally-adaptive variational meta learning approach to functionally simulate data drift across subject sets, utilizing variational layers for parameterization and the adaptive alignment method to reduce discrepancies. Through comprehensive experiments on three real-world datasets with both small-data and standard regimes against various baselines, our DAML model demonstrates the state-of-the-art performance across all metrics, underscoring its efficiency and potential within limited data.

Keywords: Adaptive Learning · Brain Graph Classification · Data Drift.

1 Introduction

Neuroimaging modalities capture intricate graphical representations and reveal complex networks of connections among brain regions. Specifically, the brain is segmented into distinct regions of interest (ROIs), treated as nodes in graphs. The inter-regional connectivity then quantifies the strength of edges between different brain regions, relying on various metrics such as the Pearson Correlation Coefficient [22]. Such intrinsic integration of information forms the brain network, inherently aligning with the capabilities of Graph Neural Networks (GNNs) [8,3]. Accordingly, various graph-based models are applied, further highlighting the advance of GNNs as a compelling choice in brain network analysis for both medical and scientific communities [27].

To best capture patterns inherent in brain graphs, recent advancements in brain network analysis focus on enhancing classification performance by capturing accurate graph representations [10,15,16] and identifying unique substructures [14,32,5,17,26] within brain networks. For instance, BrainNetCNN [10] employs Edge-to-Edge, Edge-to-Node, and Node-to-Graph layers to form a hierarchical structure for comprehensive brain graph representations. Li et al. [15] connect structural and functional connectivity using a graph encoder-decoder framework. Similarly, BrainTGL [16] utilizes GNNs for spatial representation while integrating temporal aspects via LSTM blocks. For sub-structure identification, BrainGNN [14] employs top-k loss and group-level consistency loss to underscore the significance of ROIs. IBGNN [5] highlights disorder-related structures with an interpretable generator. Moreover, recent works like HSGPL [26], BrainUSL [32], and BraGCL[17] apply contrastive learning to uncover distinctive structural and connective elements within brain networks.

Despite the potential demonstrated by the aforementioned approaches in the presence of sufficient brain information, significant fluctuation of model performance arises in scenarios with limited supervised data. The collection of brain network data is inherently resource-intensive, involving high costs and the use of sophisticated technical equipment. Although researchers attempt to tackle this problem by employing meta-learning methods [31,34,25], the *data drift* problem [20,6], marked by a discrepancy between training and testing (or even real-world) data, remains inadequately considered. This discrepancy becomes particularly critical in models trained on relatively small datasets, which often share different data distributions from those encountered in real-world scenarios [24]. Crucially, such differences are inherent to the training process and may lead to significant disparities in the analysis and interpretation of brain networks [19], emphasizing the necessity to incorporate distribution-specific strategies.

To tackle the challenge, we are inspired by the concept of Neural Processes (NPs) [7] for their unique meta-learning ability to optimize discrepancy functionally. NPs stand out by using variational techniques to form mappings between varied inputs and outputs, adeptly reconstructing intrinsic distributions with meta-learning procedures. Unlike traditional models that often overlook intrinsic correlations, NPs excel by fast parameter adaptation. Inspired by NP-related works, we provide a refined measurement of distributional discrepancy by approximating the functional distance using variational methods, marking a significant advancement over previous methodologies in adaptation ability.

In this paper, we propose a **D**istributionally-**A**daptive Variational **M**eta **L**earning for brain graph classification, named **DAML**. DAML first employs a graph encoder to aggregate both node and edge information, generating comprehensive graph-level brain network representations. By regarding data drift as the functional discrepancy between continuous functions, we encode distributions using variational methods for reconstructing distributional functions before and after observation. Through adaptively refining the alignment between distributions, DAML acknowledges data drift, enhancing adaptability to real-world scenarios with controlled uncertainty. Evaluated across real-world datasets, our

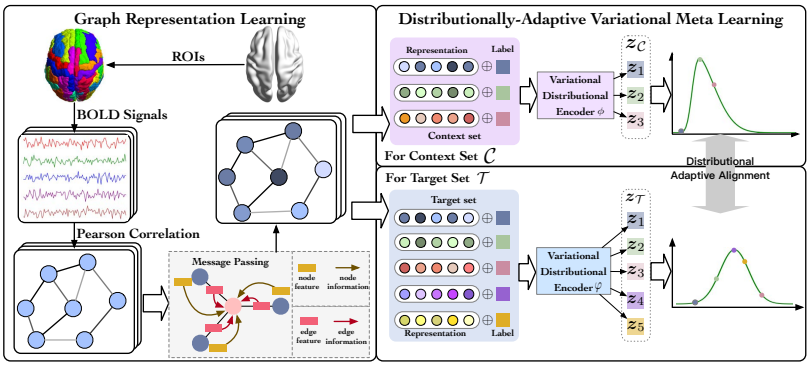

Fig. 1: The DAML Framework.

method outperforms the existing benchmarks under standard and constrained small-data scenarios, providing a novel framework for brain graph analysis.

2 Problem Definition

For each subject s, a brain graph $g_s = \{N, E\}$ is constructed, where the node set $N = \{n_i\}_{i=0}^{|N|}$ represents the brain network with $|N|$ ROIs. The feature matrix $\mathbf{X} = \{\mathbf{x}_i\}_{i=0}^{|N|}$ represents the features of ROIs and is derived from the correlation matrix. The edge set $E = \{e_{ij}\} = \{< n_i, n_j >\}$ $(n_i, n_j \in N, i \neq j)$ signifies the connections between regions, with $e_{ij} = 1$ indicating a connection and $e_{ij} = 0$ otherwise. Derived from nodes N and edges E, the weighted adjacency matrix $\mathbf{A} \in \mathbb{R}^{|N|\times|N|}$ encodes ROI connectivity, with a_{ij} quantifying the connection strength between ROIs. Each subject has a label y_s, indicating the state of the subject. Given the brain network dataset $\mathcal{D} = \{\mathcal{G}, \boldsymbol{y}\}$ with brain graphs $\{(g_1, y_1), (g_2, y_2), ..., (g_m, y_m)\}$ of m subjects, we aim to provide accurate and fast-adaptive classification through distributionally-adaptive methods.

3 Methodology

Fig. 1 shows the framework of DAML, consisting of two main modules: 1) *Graph Representation Learning* (GRL) and 2) *Distributionally-Adaptive Variational Meta Learning* (DAVM). The GRL module incorporates a novel graph encoder to generate comprehensive brain graph representations by providing detailed insights into brain connectivity to capture complex patterns. Subsequently, the DAVM module tackles data drift by simulating it as the discrepancies between subject sets. Here, *Variational Distribution Encoder* is first employed to encode the relationships between data representations and their labels into the latent variable $\boldsymbol{z}$, leveraging variational methods for precise parameterization. The subsequent *Adaptive Alignment* minimizes the differences between sets, ensuring the adaptability to new datasets with varying distributions.

3.1 Graph Representation Learning

A robust process for generating brain graph representations is crucial for classification. Here, we introduce the *graph encoder*, designed to capture and preserve detailed information from brain signals [30,23], thereby laying the groundwork for accurate brain graph classification. Our methodology advances node updates in brain network graphs by incorporating neighboring nodes and edges as follows:

$$\mathbf{h}_i^{(l+1)} \leftarrow \begin{cases} \text{LeakyReLU}(\sum_{j \in N(i)} f(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, a_{ij})), & \text{when } l \geq 1, \\ \text{LeakyReLU}(\sum_{j \in N(i)} f(\mathbf{x}_i, \mathbf{x}_j, a_{ij})), & \text{when } l = 0. \end{cases} \tag{1}$$

where $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^d$ are the initial features of node i and j. $\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)} \in \mathbb{R}^h$ denote the representations of node i and j at layer l. a_{ij} represents the weight of connection between node i and node j. $f(\cdot) \in \mathbb{R}^d$ is our defined message-passing strategy, activated by the LeakyReLU function to avoid gradient vanishing [9]:

$$f(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, e_{ij}) = \mathbf{W}_f \left[\mathbf{h}_i^{(l)} \| \mathbf{h}_j^{(l)} \| e_{ij}\right] + \mathbf{b}_f \tag{2}$$

Our encoder introduces weighted edges into node representations, preserving topology information and deepening our comprehension of connectivity patterns [10,33]. Following, the brain graph representation of subject s is formalized as the concatenation of nodes,

$$\mathbf{r}_s = \text{Reshape}(\overline{\mathbf{h}}_1^{(L)} \| \overline{\mathbf{h}}_2^{(L)} \| \cdots, \| \overline{\mathbf{h}}_N^{(L)}) \tag{3}$$

where $\mathbf{r}_s \in \mathbb{R}^{|N| \times d}$, and $\overline{\mathbf{h}}_i^{(L)} \in \mathbb{R}^d$ is the normalized representation of node i at the final L-th layer.

3.2 Distributionally-Adaptive Variational Meta Learning

Deriving brain graph representations for each subject, we then aim to facilitate our model with the capability for rapid adaptation using meta-learning methods.

Variational Distribution Encoder. Central to our model, we hypothesize the existence of a latent global variable $\boldsymbol{z}$, which represents brain graph patterns of subjects. To facilitate optimization akin to adaption from training to the real-world scenario, we organize brain graph data into a context set $\mathcal{C} = \{\mathbf{r}_1, \cdots, \mathbf{r}_{|\mathcal{C}|}\}$ and target set $\mathcal{T} = \{\mathbf{r}_1, \cdots, \mathbf{r}_{|\mathcal{T}|}\}$ ($\mathcal{C} \in \mathcal{T}$), simulating the different sets of experimental ($\mathcal{C}$) and real-world scenarios ($\mathcal{T}$). The target set $\mathcal{T}$ includes $\mathcal{C}$'s samples plus additional ones from the same batch, simulating experimental and real-world conditions. Drawing on principles from NPs [11] and VAEs [12], we assume $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ as a Gaussian variable, resulting in distinct Gaussian posteriors for both sets:

$$\boldsymbol{z} \sim \begin{cases} \mathcal{N}(\boldsymbol{\mu}_{\mathcal{C}}, \boldsymbol{\Sigma}_{\mathcal{C}}), \text{ for context set } \mathcal{C} \\ \mathcal{N}(\boldsymbol{\mu}_{\mathcal{T}}, \boldsymbol{\Sigma}_{\mathcal{T}}), \text{ for target set } \mathcal{T} \end{cases} \tag{4}$$

The approximation of posterior means and variances of $\boldsymbol{z} \in \mathbb{R}^h$ is achieved by computing sufficient statistics for the context set and target set. Specifically, we employ the reparameterization tricks [12] to calculate the $\boldsymbol{z}$,

$$\begin{aligned} \boldsymbol{z}_{\mathcal{C}} = \boldsymbol{\mu}_{\mathcal{C}} + \boldsymbol{\Sigma}_{\mathcal{C}} \odot \epsilon_{\mathcal{C}} \text{ with } \epsilon_{\mathcal{C}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \\ \text{where } \tilde{\boldsymbol{\mu}}_{\mathcal{C}} = g_{\boldsymbol{\mu}_{\mathcal{C}}}(\mathbf{R}_C), \tilde{\boldsymbol{\Sigma}}_{\mathcal{C}} = g_{\boldsymbol{\Sigma}_{\mathcal{C}}}(\mathbf{R}_C) \end{aligned} \tag{5}$$

$$\begin{aligned} \boldsymbol{z}_{\mathcal{T}} = \boldsymbol{\mu}_{\mathcal{T}} + \boldsymbol{\Sigma}_{\mathcal{T}} \odot \epsilon_{\mathcal{T}} \text{ with } \epsilon_{\mathcal{T}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \\ \text{where } \tilde{\boldsymbol{\mu}}_{\mathcal{T}} = g_{\boldsymbol{\mu}_{\mathcal{T}}}(\mathbf{R}_T), \tilde{\boldsymbol{\Sigma}}_{\mathcal{T}} = g_{\boldsymbol{\Sigma}_{\mathcal{T}}}(\mathbf{R}_T) \end{aligned} \tag{6}$$

Here, $\mathbf{R}_{\mathcal{C}} = \{[\mathbf{r}_i \| y_i]\}_{i \in \mathcal{C}}$ and $\mathbf{R}_{\mathcal{T}} = \{[\mathbf{r}_j \| y_j]\}_{i \in \mathcal{T}}$, connecting brain graph representations with their labels. The parameters of global latent variables $\boldsymbol{z}_{\mathcal{C}}, \boldsymbol{z}_{\mathcal{T}} \in \mathbb{R}^h$ are captured using four different 2-layer Multilayer Perceptrons (MLPs), $g_{\boldsymbol{\mu}_{\mathcal{C}}}, g_{\boldsymbol{\Sigma}_{\mathcal{C}}}, g_{\boldsymbol{\mu}_{\mathcal{T}}}, g_{\boldsymbol{\Sigma}_{\mathcal{T}}} \in \mathbb{R}^h$, structured as

$$g(\cdot) = W_2(\text{ReLU}(W_1 \cdot + b_1)) + b_2 \tag{7}$$

Adaptive Alignment. Two posterior distributions, $\mathcal{N}(\boldsymbol{\mu}_{\mathcal{C}}, \boldsymbol{\Sigma}_{\mathcal{C}})$ and $\mathcal{N}(\boldsymbol{\mu}_{\mathcal{T}}, \boldsymbol{\Sigma}_{\mathcal{T}})$, differ when various sets are applied. To address this, our method focuses on adaptively minimizing these distributional discrepancies through symmetric Jensen-Shannon (JS) divergence [18], facilitating iterative alignment of distributions to generalize with the diverse and complex adaptation process between two different distributions [4]. Each iteration of alignment is formulated as:

$$\mathcal{L}_d = D_{JS}(p_{\mathcal{C}}, p_{\mathcal{T}}) = \frac{1}{2}(D_{KL}(p_{\mathcal{C}}, p_{\mathcal{T}}) + D_{KL}(p_{\mathcal{T}}, p_{\mathcal{C}})) \tag{8}$$

where $p_{\mathcal{C}}, p_{\mathcal{T}}$ are short for $\mathcal{N}(\boldsymbol{\mu}_{\mathcal{C}}, \boldsymbol{\Sigma}_{\mathcal{C}}), \mathcal{N}(\boldsymbol{\mu}_{\mathcal{T}}, \boldsymbol{\Sigma}_{\mathcal{T}})$. $D_{KL}(\cdot, \cdot)$ is Kullback–Leibler divergence [21]. Although $\mathcal{C}$ and $\mathcal{T}$ originated from the same dataset, the small sample sizes and inherent randomness introduce statistical biases, simulating data drift settings between two datasets. Through iteratively minimizing the JS divergence between distributions with varying subject groups, we adaptively simulate the data drift situation in DAML, enabling adjustment to varying datasets.

3.3 Objective Functions

For classification, our model employs a 3-layer MLP architecture, applying the softmax activation function at the output layer, represented as:

$$\hat{y}_i = \text{softmax}\left(W_5 \cdot \sigma\left(W_4 \cdot \sigma\left(W_3 \cdot \boldsymbol{r}_i\right)\right)\right) \tag{9}$$

where $\sigma(\cdot)$ denotes LeakyReLU activation function. The bias is omitted for clarity. The classification objective function $\mathcal{L}_c$ is determined using cross-entropy loss $\mathcal{L}_c$. Coupled with the adaptive alignment loss $\mathcal{L}_d$, the objective function is

$$\mathcal{L} = \mathcal{L}_c + \mathcal{L}_d = \frac{1}{m}\sum_{i=1}^{m}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] + D_{JS}(p_{\mathcal{C}}, p_{\mathcal{T}}) \tag{10}$$

Table 1: Overall performance(%). * indicates that the performance improvement of DAML over the runner-up model is statistically significant, verified by a paired t-test (significance threshold p< 0.05).

Datasets	ABIDE			PPMI			HIV		
	Accuracy	AUC	F1	Accuracy	AUC	F1	Accuracy	AUC	F1
GCN	58.78±1.99	57.10±3.27	53.54±7.05	65.82±8.58	64.05±7.77	54.33±7.72	67.00±10.43	68.22±5.96	64.43± 11.43
GAT	62.57±1.98	62.6±2.35	61.67±2.62	67.64±10.58	65.17±11.11	56.50±12.24	68.43±11.61	69.06±13.26	65.43±11.61
BrainGNN	61.80±1.26	57.73±1.41	58.78±1.18	73.68±1.35	60.90±3.97	52.43±2.86	68.54±5.71	66.81±7.11	61.22±11.47
BrainNetCNN	61.18±1.97	68.09±2.77	60.19±1.88	75.50±6.68	64.73±5.30	52.54±5.31	51.43±6.43	65.31±7.74	43.11±4.28
DGCNN	57.64±1.23	58.10±1.42	56.61±1.81	69.25±8.24	50.00±4.71	44.21±0.08	60.22±3.00	54.79±4.71	46.70±8.66
IBGNN	58.09±3.06	58.25±4.38	55.79±4.04	74.81±1.62	65.22±7.65	57.50±2.95	50.00±6.39	41.12±4.74	48.34±7.53
BrainUSL	61.75±3.48	61.38±3.50	60.93±3.64	64.50±5.24	50.40±6.02	49.91±5.11	60.85±3.29	57.79±3.31	57.51±3.45
MetaGCN	62.41±4.06	67.69±2.98	61.87±2.09	72.70±3.50	64.47±5.99	57.33±2.41	68.57±7.28	69.39±6.58	67.12±8.30
Graph Encoder	62.85±3.35	69.45±3.68	62.45±3.54	76.87±2.11	67.59±3.21	57.54±4.45	64.29±7.29	69.80±6.60	62.30±7.67
DAML*	**63.84**±2.18	**70.30**±3.83	**63.34**±2.51	**77.87**±2.08	**68.33**±2.58	**58.76**±2.53	**70.00**±4.57	**73.88**±5.61	**68.66**±6.37

4 Experiments

4.1 Datasets, Comparison Methods, and Experimental Settings

We evaluate DAML and all baselines on 3 real-world datasets: ABIDE[3], PPMI[4], and HIV[5] datasets. Three datasets have 871, 718, and 70 subjects, respectively, simulating both standard and limited scenarios. ABIDE is preprocessed into 90 ROIs using Automated Anatomical Labeling (AAL) [28], and the correlations between ROIs are measured using Pearson correlation coefficient [2] of the BOLD signals. PPMI and HIV are preprocessed according to [5]. In each batch, we form $\mathcal{C}$ by selecting a subset of samples. $\mathcal{T}$includes C's samples plus additional ones from the same batch. During training, we compute their Gaussian distribution parameters and measure the distributional divergence between C and T's distributions. This metric is incorporated into the loss function. This process simulates differences between training and real-world data, enabling rapid adaptation of DAML. To comprehensively evaluate the efficiency and adaptability of our proposed DAML, we compare it against 8 methods, including two GNNs (GCN [13] and GAT [29]) and five brain-related models (BrainGNN [14], BrainNetCNN [10], IBGNN [5], DGCNN, and BrainUSL [32]). Moreover, to prove the efficiency of our model in the meta-learning aspect, we additionally implement a meta-learning framework based on GCN following [31], named MetaGCN [31]. For all baselines, we utilize the optimal parameters as reported in their respective papers. In our DAML model, the optimal node embedding size d and hidden size h are 16 and 256, respectively, with the optimal group size of 16 for the DAVM process. The best negative slope of the LeakyReLU activation function is set as 0.2. Our model is developed using Pytorch and Pytorch Geometrics frameworks, and we employ the Adam optimizer for training. To ensure a comprehensive evaluation, 5-fold cross-validation with 70 training epochs is used in experiments.

[3] `https://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html`
[4] `https://www.ppmi-info.org/`
[5] `https://github.com/HennyJie/IBGNN`

4.2 Overall Performance Analysis

The overall performance is shown in Table 1. The best results are **bolded**, and the runner-up baselines are underlined. Generally, our proposed DAML achieves the best performance in three datasets. Our findings are reported as follows:

1) Node aggregation matters. Node aggregation enhances the model performance, as evidenced by the superior results of *graph encoder* on the ABIDE and PPMI datasets compared to most baselines, underscoring the effectiveness of our proposed node aggregation procedure in a standard setting. Specifically, the graph encoder outperforms BrainNetCNN by margins of 1.67%, 1.36%, and 2.26% across three metrics in the ABIDE dataset. Furthermore, it showcases its capability with a notably high AUC of 76.87% in the PPMI dataset, outperforming IBGNN by 2.37%. The results show the graph encoder's proficiency in capturing distinct subject features, laying a solid foundation for DAML.

2) Meta-learning methods outperform other ones. While the graph encoder demonstrates superiority in the PPMI and ABIDE datasets, a significant performance decline is observed in the HIV dataset. Particularly, IBGNN, despite offering competitive results in the PPMI dataset, falls to a mere 41.12% AUC in the HIV dataset, illustrating the difficulties encountered in scenarios with limited data. Conversely, MetaGCN and DAML show remarkable adaptability in limited data settings. MetaGCN performs well with 68.57%, 69.89%, and 66.12% across three metrics in the HIV dataset, surpassing all baselines. GAT, which provides the second-best Accuracy in ABIDE, falls behind the MetaGCN in HIV. Moreover, our DAML sets a new standard by achieving the best results across all metrics and datasets, further illustrating the exceptional stability and performance of our design in small-data learning contexts.

3) DAVM outperforms traditional ones. As aforementioned, DAML not only performs the best in the small-data HIV dataset but also achieves the best in standard ABIDE and PPMI datasets. Compared with MetaGCN, our proposed method improves Accuracy, AUC, and F1 with 1.43%, 3.99%, and 2.54% in HIV, respectively. Moreover, DAML achieves the best results in F1 score with 63.34% and 58.76% in ABIDE and PPMI, outperforming all baselines. DAML even provides an improvement of 0.99% and 1.00% in Accuracy compared to the graph encoder, empirically proving the strength of the DAVM method. Since DAML models the Gaussian for both the context and target sets, it allows us to quantify the likelihood that test samples are being modeled, serving as a measure of uncertainty and enhancing DAML's interpretability.

4) Accurate classification when facing difficult cases. We utilize BrainNet Viewer to visualize two subjects. Fig. 2a shows prominent connections between the right hippocampus (HIP.R) and the right fusiform gyrus (FFG.R) in seronegative control. In contrast, Fig. 2b shows a weaker connection in HIV patients at the same sparsity level. Patients exhibit atypical connections involving the left hippocampus (HIP.L), the left middle occipital gyrus (MOG.L), the right postcentral gyrus (PoCG.R), and the right angular gyrus (ANG.R) — connections not found in seronegative controls. These findings align with [1] and can be identified by DAML.

(a) Subject 1 - Seronegative Controls (b) Subject 2 - HIV patients

Fig. 2: ROI connections of two HIV subjects with 15% sparsity.

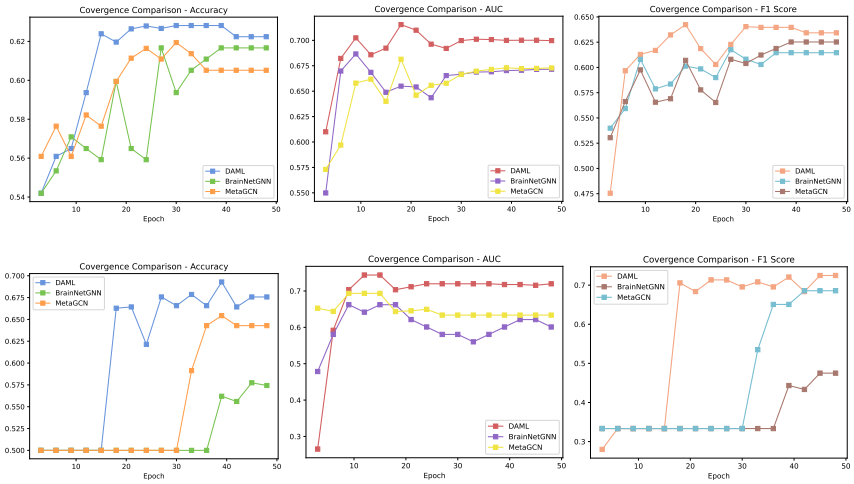

Fig. 3: Convergence speed on ABIDE (up) and HIV (down) Dataset.

4.3 Adaptation Analysis

We further compare our proposed model with two baselines, BrainNetCNN and MetaGCN, to illustrate the advantage of DAML in fast-adaptation ability. Here, we take 80% subjects for training, and all models are tested on the remaining 20% subjects every 3 epochs. Due to page limitation, we only show the results of ABIDE and HIV in Fig. 3. In the ABIDE dataset, the three models demonstrate similar convergence trends, with DAML outperforming the baselines across all metrics. BrainNetCNN, MetaGCN, and DAML experience sharp improvements, followed by slight fluctuations in Accuracy. The F1 score and AUC present more significant fluctuations for each model. For instance, three models exhibit a significant performance drop between the 15th and 25th epochs, followed by a marked improvement in subsequent epochs. Notably, DAML consistently surpasses other baseline models in Accuracy, AUC, and F1 score, showcasing not only superior performance but also promising convergence speed with sufficient

data. In the HIV dataset, we observe that meta-learned models (MetaGCN and DAML) are superior to traditional models both in adaptation speed and results. For example, MetaGCN and DAML provide improvements of 4.67% and 13.42% in Accuracy and 13.42% in F1 score compared to BrainNetCNN. Moreover, DAML provides the fastest convergence speed, achieving the best results on the F1 score at the 18th epoch, superior to MetaGCN (42nd epoch) and BrainNetCNN (45th epoch). These results show that DAML provides robust improvements and improves the convergence speed with both standard and limited data, demonstrating the advancement of distributionally-adaptive methods.

5 Conclusion and Future Work

We propose the Distributionally-Adaptive Meta Learning (DAML) framework to address the data drift problem in brain graph classification. We first introduce the graph representation learning module to generate detailed brain graph representations. The distributionally-adaptive variational meta learning method is then introduced to simulate data drift between datasets from a functional perspective and minimize these discrepancies through adaptive alignment techniques. Our extensive experiments across both small-data and standard regimes demonstrate the DAML model's exceptional performance. For future work, we aim to achieve transferable representation learning among different datasets.

6 Acknowledgement

This work was supported by the Australian Research Council Project with No. LP210301259.

7 Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this publication.

References

1. Aili, X., Wang, W., Zhang, A., Jiao, Z., Li, X., Rao, B., Li, R., Li, H.: Rich-club analysis of structural brain network alterations in hiv positive patients with fully suppressed plasma viral loads. Frontiers in Neurology **13**, 825177 (2022)
2. Benesty, J., Chen, J., Huang, Y., Cohen, I.: Pearson correlation coefficient. In: Cohen, I., Huang, Y., Chen, J., Benesty, J. (eds.) Noise Reduction in Speech Processing, pp. 1–4. Springer (2009)
3. Bessadok, A., Mahjoub, M.A., Rekik, I.: Graph neural networks in network neuroscience. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(5), 5833–5848 (2022)

4. Tachet des Combes, R., Zhao, H., Wang, Y.X., Gordon, G.J.: Domain adaptation with conditional distribution matching and generalized label shift. Advances in Neural Information Processing Systems **33**, 19276–19289 (2020)
5. Cui, H., Dai, W., Zhu, Y., Li, X., He, L., Yang, C.: Interpretable graph neural networks for connectome-based brain disorder analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 375–385 (2022)
6. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. ACM Computing Surveys (CSUR) **46**(4), 1–37 (2014)
7. Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D.J., Eslami, S., Teh, Y.W.: Neural processes. arXiv preprint arXiv:1807.01622 (2018)
8. Hamilton, W.L., Ying, R., Leskovec, J.: Representation learning on graphs: Methods and applications. arXiv preprint arXiv:1709.05584 (2017)
9. Jiang, T., Cheng, J.: Target recognition based on cnn with leakyrelu and prelu activation functions. In: 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC). pp. 718–722. IEEE (2019)
10. Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., Zwicker, J.G., Hamarneh, G.: Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. Neuroimage **146**, 1038–1049 (2017)
11. Kim, H., Mnih, A., Schwarz, J., Garnelo, M., Eslami, A., Rosenbaum, D., Vinyals, O., Teh, Y.W.: Attentive neural processes. In: International Conference on Learning Representations (2018)
12. Kingma, D., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
13. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
14. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L.H., Ventola, P., Duncan, J.S.: Braingnn: Interpretable brain graph neural network for fmri analysis. Medical Image Analysis **74**, 102233 (2021)
15. Li, Y., Shafipour, R., Mateos, G., Zhang, Z.: Supervised graph representation learning for modeling the relationship between structural and functional brain connectivity. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 9065–9069. IEEE (2020)
16. Liu, L., Wen, G., Cao, P., Hong, T., Yang, J., Zhang, X., Zaiane, O.R.: Braintgl: A dynamic graph representation learning model for brain network analysis. Computers in Biology and Medicine **153**, 106521 (2023)
17. Luo, X., Dong, G., Wu, J., Beheshti, A., Yang, J., Xue, S.: An interpretable brain graph contrastive learning framework for brain disorder analysis. In: The 17th ACM International Conference on Web Search and Data Mining (WSDM'24) (2024)
18. Menéndez, M., Pardo, J., Pardo, L., Pardo, M.: The jensen-shannon divergence. Journal of the Franklin Institute **334**(2), 307–318 (1997)
19. Micevska, S., Awad, A., Sakr, S.: Sddm: an interpretable statistical concept drift detection method for data streams. Journal of Intelligent Information Systems **56**, 459–484 (2021)
20. Moreno-Torres, J.G., Raeder, T., Alaiz-Rodríguez, R., Chawla, N.V., Herrera, F.: A unifying view on dataset shift in classification. Pattern Recognition **45**(1), 521–530 (2012)
21. Pérez-Cruz, F.: Kullback-leibler divergence estimation of continuous distributions. In: 2008 IEEE International Symposium on Information Theory. pp. 1666–1670. IEEE (2008)

22. Saad, Z.S., Glen, D.R., Chen, G., Beauchamp, M.S., Desai, R., Cox, R.W.: A new method for improving functional-to-structural mri alignment using local pearson correlation. Neuroimage **44**(3), 839–848 (2009)
23. Song, S., Song, Y., Luo, C., Song, Z., Kuzucu, S., Jia, X., Guo, Z., Xie, W., Shen, L., Gunes, H.: Gratis: Deep learning graph representation with task-specific topology and multi-dimensional edge features. arXiv preprint arXiv:2211.12482 (2022)
24. Storkey, A., et al.: When training and test sets are different: characterizing learning transfer. Dataset Shift in Machine Learning **30**(3-28), 6 (2009)
25. Su, J., Shen, H., Peng, L., Hu, D.: Few-shot domain-adaptive anomaly detection for cross-site brain images. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(3), 1819–1835 (2024)
26. Tang, H., Ma, G., Guo, L., Fu, X., Huang, H., Zhan, L.: Contrastive brain network learning via hierarchical signed graph pooling model. IEEE Transactions on Neural Networks and Learning Systems (2022, Early Access)
27. Telesford, Q.K., Simpson, S.L., Burdette, J.H., Hayasaka, S., Laurienti, P.J.: The brain as a complex system: using network science as a tool for understanding the brain. Brain Connectivity **1**(4), 295–308 (2011)
28. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. Neuroimage **15**(1), 273–289 (2002)
29. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al.: Graph attention networks. stat **1050**(20), 10–48550 (2017)
30. Withnall, M., Lindelöf, E., Engkvist, O., Chen, H.: Building attention and edge message passing neural networks for bioactivity and physical–chemical property prediction. Journal of Cheminformatics **12**(1), 1–18 (2020)
31. Yang, Y., Zhu, Y., Cui, H., Kan, X., He, L., Guo, Y., Yang, C.: Data-efficient brain connectome analysis via multi-task meta-learning. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 4743–4751 (2022)
32. Zhang, P., Wen, G., Cao, P., Yang, J., Zhang, J., Zhang, X., Zhu, X., Zaiane, O.R., Wang, F.: Brainusl: U nsupervised graph s tructure l earning for functional brain network analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 205–214. Springer (2023)
33. Zhang, S., Tong, H., Xu, J., Maciejewski, R.: Graph convolutional networks: a comprehensive review. Computational Social Networks **6**(1), 1–23 (2019)
34. Zhu, L., Liu, Y., Liu, R., Peng, Y., Cao, J., Li, J., Kong, W.: Decoding multi-brain motor imagery from eeg using coupling feature extraction and few-shot learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering **31**, 4683–4692 (2023)