# Hypergraph-Guided Federated Distillation Learning for Efficient and Robust Multi-Center fMRI Data Analysis

Tao Jin[†1], Yidan Xu[†1], Yuhan Gao[1], Xichun Sheng[2], Chenggang Yan[1], Yaoqi Sun[3(✉)], Xiangmin Han[4], and Yue Gao[4]

[1] Hangzhou Dianzi University, 310018, China
jintaobika@gmail.com, yidanxu2024@163.com, yuhangao@hdu.edu.cn, cgyan@hdu.edu.cn

[2] Macao Polytechnic University, China
1002612870@qq.com

[3] Lishui University, 323000, China
sunyq2233@163.com

[4] Tsinghua University, 100084, China
{hanxiangmin, gaoyue}@tsinghua.edu.cn

**Abstract.** Multi-center fMRI data analysis faces significant challenges such as data privacy concerns and data integration issues. Federated learning, as an innovative distributed machine learning approach, enables cross-center collaboration by sharing model parameters instead of raw data. However, existing methods often struggle with improving the robustness and inference efficiency of multi-center fMRI data processing. To address these challenges, we propose a novel hypergraph-guided federated distillation framework(HGFD) for multi-center fMRI data analysis. HGFD utilizes a hypergraph structure to model the spatiotemporal features of brain activity, capturing high-order correlations across brain regions. Furthermore, a hypergraph-based knowledge distillation approach is utilized to transfer high-order structural representations into shallow neural networks, thereby preserving their ability for complex relational inference and significantly enhancing computational efficiency. In the federated learning process, participating centers only need to share the parameters of their shallow neural networks to a central server. Through parameter aggregation, each center's shallow network can learn the high-order structural information of other centers. Experiments on multi-center fMRI dataset demonstrate that the proposed method not only improves the robustness and consistency of fMRI-based prediction tasks but also achieves efficient and accurate predictions while ensuring data privacy.

**Keywords:** Hypergraph · Federated Learning · fMRI Analysis · Knowledge Distillation.

† Contributed equally to this work.

## 1 Introduction

Functional Magnetic Resonance Imaging (fMRI) captures the spatiotemporal characteristics of blood oxygen level-dependent (BOLD) signals, providing a non-invasive means to study human brain functional connectivity networks [1,2]. This advancement has shifted the early detection of Autism Spectrum Disorder (ASD) from traditional behavioral assessments to intelligent diagnosis based on dynamic brain network analysis [3]. However, deep learning-based predictive models for brain disorders face a significant challenge of data scarcity. Medical data sensitivity limits individual institutions to acquiring fewer than 200 ASD samples, which is insufficient for training robust deep neural models.

Federated Learning (FL) [4, 5] emerges as a privacy-preserving solution, enabling collaborative global model training through distributed model parameter exchange without direct data sharing. Frameworks such as FedAvg [6], FedBN [7], FedProx [8], and MOON [9] have demonstrated broad applicability in distributed tasks. Recent years have witnessed a surge of FL-based methodologies [10–13] specifically tailored for neurological disorder analysis, addressing challenges in multi-center data utilization while adhering to stringent privacy constraints.

Brain networks are characterized by complex and higher-order interactions among multiple regions, which traditional graph neural networks often fail to capture. Hypergraph learning offers a natural and effective solution to this problem, as it models high-order relationships more expressively. Foundational work such as [14–16] demonstrates the advantages of hypergraph representations in various tasks. Based on this, recent studies such as $I^2$ HBN [17] and HGTrans [18] have applied hypergraph learning to predict brain disorders, highlighting its potential to capture intricate brain connectivity patterns.

To balance the computational complexity of local models and communication efficiency, we employ knowledge distillation, where a complex teacher model guides a simplified student model. The hypergraph structure effectively captures high-order relationships, enriching the teacher model's capability to represent intricate brain features. Distilling this knowledge into a lightweight student model enables the retention of key semantics while substantially reducing computational and communication overhead. This strategy safeguards data privacy and enhances prediction performance in brain disorder analysis. A hypergraph convolution teacher model (BrainHGNN-Teacher) is designed at the client side to model the nonlinear multi-region coupling relationships in the functional connectivity of ASD patients. Knowledge distillation is performed locally on each client, where a lightweight model (MLP-Student) learns from the BrainHGNN-Teacher via KL divergence. During federated training, only the parameters of the MLP-Student are uploaded to the server for aggregation, thereby reducing communication overhead.

The main contributions are summarized as follows: (1) We propose a novel hypergraph-guided federated distillation framework(HGFD) to capture higher-order correlations in functional brain networks, overcoming the limitations of traditional graph models. (2) Through hypergraph knowledge distillation, we transfer complex information from BrainHGNN-Teacher model into a simpler

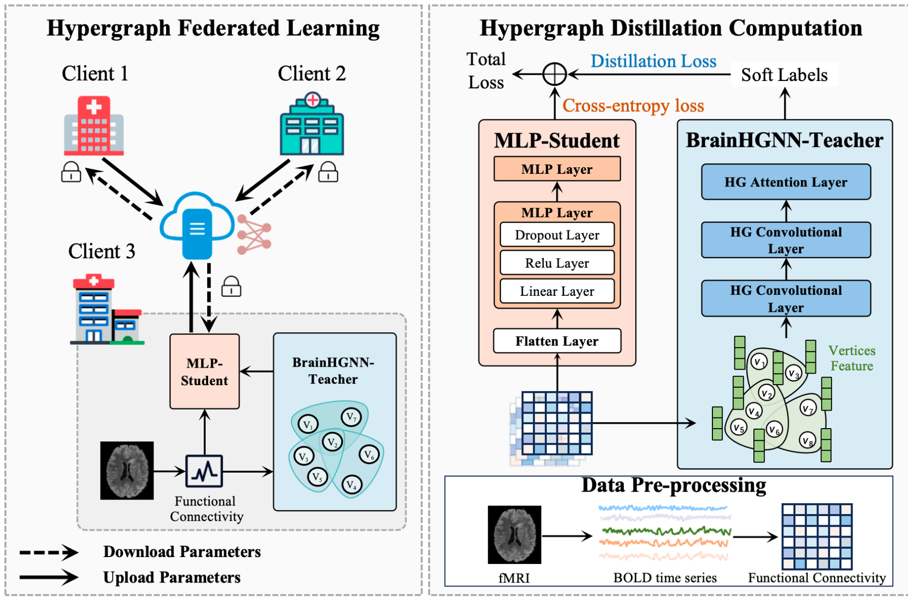


Fig. 1: Pipeline of the proposed method. The left side illustrates the collaborative process between multiple clients and the server; the right side presents the data pre-processing procedure, the specific model architecture within each center, and the model distillation process.

MLP-Student model, reducing computational and communication costs in federated learning. (3) We validate the proposed method on the ABIDE dataset, showing superior performance and optimal results compared to traditional and deep learning methods in a federated framework.

## 2 Methods

The pipeline of the proposed method is illustrated in Fig. 1.

### 2.1 Data Pre-processing

Data pre-processing is performed according to the pipeline described in $I^2HBN$ [17]. We utilize the DPARSF [19] toolbox for pre-processing the raw ABIDE data and adopt the AAL [20] atlas for brain parcellation. The brain space of each individual is then parcellated into 116 regions of interest (ROI). Finally, mean time series are obtained by calculating the processed fMRI data voxels.

### 2.2 Hypergraph Computation

**Hypergraph Definition.** A hypergraph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, which consists of a vertex set $\mathcal{V}$, a hyperedge set $\mathcal{E}$. A hypergraph $\mathcal{G}$ can be described by an incidence matrix $\mathbf{H} \in R^{|\mathcal{V}| \times |\mathcal{E}|}$, where the entries are defined as $\mathbf{H}(v,e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{otherwise} \end{cases}$. For a vertex $v \in \mathcal{V}$, the degree of $v$ is defined as $d(v) = \sum_{e \in \mathcal{E}} \mathbf{H}(v,e)$, and for a hyperedge $e \in \mathcal{E}$, the degree of $e$ is defined as $d(e) = \sum_{v \in \mathcal{V}} \mathbf{H}(v,e)$. $\mathbf{D}_v$ and $\mathbf{D}_e$ are respectively diagonal matrices representing the degrees of $\mathcal{V}$ and $\mathcal{E}$. The initial feature of $\mathcal{V}$ is denoted as $\mathbf{X} = \{x_1, x_2, \cdots, x_N\}, x_i \in \mathbf{R}^C$, where $C$ represents the feature dimension.

**Hypergraph Modeling.** After pre-processing the data, the functional connectivity (FC) between two different brain regions is calculated using the following formula:

$$\mathbf{X}_{FC}(i,j) = \frac{\mathbb{E}[(t_i - \mathbb{E}[t_i])(t_j - \mathbb{E}[t_j])]}{\sigma_{t_i}\sigma_{t_j}}, \tag{1}$$

where $t_i$ and $t_j$ represent the time series data of brain regions $i$ and $j$, respectively. The Pearson correlation value of the $i$-th ROI can be denoted as $\mathbf{X}_{FC}(i) = \{\mathbf{X}_{FC}(i,1), \mathbf{X}_{FC}(i,2), \cdots, \mathbf{X}_{FC}(i,116)\}$.

We regard brain regions as hypergraph vertices, with the Pearson correlation matrix serving as vertex features. We use the K-NN algorithm [21] to connect Each vertex to its top-10 most feature-similar vertices to form hyperedges, thereby constructing the hypergraph structure. The similarity between vertex $v_i$ and vertex $v_j$ is measured by the Euclidean distance between $\mathbf{X}_{FC}(i)$ and $\mathbf{X}_{FC}(j)$, computed as follows:

$$d(v_i, v_j) = \left( \sum_{c=1}^{C} \left( \mathbf{X}_{FC}(i,c) - \mathbf{X}_{FC}(j,c) \right)^2 \right)^{\frac{1}{2}}, \tag{2}$$

where $v_i$ and $v_j$ denote the $i^{th}$ and $j^{th}$ vertices, respectively, and forms the incidence matrix $\mathbf{H_f}$. Finally, we can perform hypergraph computation on the $\mathbf{X}_{FC}$ and $\mathbf{H_f}$ by stacking multiple layers of hypergraph convolution layers [14]:

$$\mathbf{X}^{l+1} = \delta(\mathbf{D}_v^{-1/2}\mathbf{H}\mathbf{D}_e^{-1}\mathbf{H}^\top\mathbf{D}_v^{-1/2}\mathbf{X}^l\Theta^{l+1}), \tag{3}$$

where $\Theta$ is the learnable parameters and $\delta(\cdot)$ is the nonlinear activation function. BrainHGNN-Teacher model consists of two layers of hypergraph convolution layers, enabling the model to capture complex functional relationships between brain regions.

### 2.3 Hypergraph Distilling Learning

**Local update with Hypergraph distillation.** The BrainHGNN-Teacher model is trained using knowledge distillation, where the student model, a simpler MLP network, learns from the more complex teacher model. During the training process, the student model is optimized to minimize the knowledge gap between its output and that of the teacher model, using the Kullback-Leibler (KL) divergence loss:

$$\mathcal{L}_{KL} = \sum_i \mathrm{KL}(p_{\text{teacher}}(i)||p_{\text{student}}(i)), \tag{4}$$

where $p_{\text{teacher}}(i)$ and $p_{\text{student}}(i)$ are the probability distributions generated by the teacher and student models, respectively, for each sample. The teacher model provides high-order semantic guidance to the student model, enabling it to learn more detailed and meaningful features.

The total loss function is defined as follows:

$$\mathcal{L} = \alpha \cdot \mathcal{L}_{CE}(y, f_{stu}(x)) + (1-\alpha) \cdot T^2 \cdot \mathcal{L}_{KL}(f_{tea}(x)/T \| f_{stu}(x)/T), \tag{5}$$

where it consists of two components: $\mathcal{L}_{CE}$, the cross-entropy loss of the student model, which measures the difference from the true labels; and $\mathcal{L}_{KL}$, the KL divergence loss, which measures the discrepancy between the student and teacher models. The temperature coefficient $T$ is used to smooth the output of the teacher model, while $\alpha$ controls the relative importance of two losses in the total loss. This loss function combines hard targets (cross-entropy) and soft targets (KL divergence), enabling the student model to learn both from the true labels and the knowledge distilled from the teacher model.

### 2.4 Hypergraph Federated Learning

**Global update with local model.** We design the hypergraph federated framework based on FedAvg [6]. We set both the local models and the global model to MLP models with the same network structure and the same learning rate. For $k$ local models, we compute their gradient $g_k \nabla F_k(w_t)$, which is the average gradient of local data under the global model $w_t$. Then, we update the global model $w_{t+1} \leftarrow w_t - \eta \sum_{k=1}^{K} \frac{n_k}{n} g_k$ using the gradients of the local models, since $\sum_{k=1}^{K} \frac{n_k}{n} g_k = \nabla f(w_t)$. An equivalent update is given by $\forall k, w_{t+1}^k \leftarrow w_t - \eta g_k$ and then $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$. That is, each local model locally takes one step of gradient descent on the current model using its local data, and the global model then takes a weighted average of the resulting models. We can add more computation to each client by iterating the local update $w^k \leftarrow w^k - \eta \nabla F_k\left(w^k\right)$ multiple times before the averaging step.

In the federated learning setting, the trained teacher and student models are distributed across different client nodes, each using its local data to update the models. The federated aggregation step only requires the transmission of the student model's parameters, significantly reducing the communication overhead. This process ensures privacy while improving the accuracy and robustness of the model for brain disease prediction tasks. The complete procedure is presented in Algorithm. 1.

**Algorithm 1** The hypergraph-guided federated distillation framework(HGFD)

**Input:** $f_t$ : Teacher model, $f_s$ : Student model, $\alpha$ : Loss weight
**procedure** MAIN
  **for** $k = 1$ **to** $n_{fold}$ **do**
    $\theta_s^g \leftarrow \text{Initialize}(f_s)$
    **for** $r = 1$ **to** $R$ **do**
      $\Theta \leftarrow \emptyset$ ▷ Client models pool
      **for each client** $c \in \mathcal{C}$ **do**
        $\theta_t^c \leftarrow \text{TrainTeacher}(f_t, \mathcal{D}_c^{train})$
        $\theta_s^c \leftarrow \text{Clone}(\theta_s^g)$
        $\theta_s^c \leftarrow \text{Distill}(\theta_s^c, \theta_t^c, E, T, \alpha)$
        $\Theta \leftarrow \Theta \cup \{\theta_s^c\}$
      **end for**
      $\theta_s^g \leftarrow \text{FedAvg}(\Theta)$
    **end for**
  **end for**
**end procedure**
**function** DISTILL($\theta_s$, $\theta_t$, $E$, $T$, $\alpha$)
  **for** $e = 1$ **to** $E$ **do**
    Compute: $$\begin{aligned}\mathcal{L} &= \alpha \mathcal{L}_{CE}(f_s(x), y) \\ &+ (1-\alpha) T^2 \mathcal{L}_{KL}(f_s(x)/T \| f_t(x)/T)\end{aligned}$$
    Update $\theta_s$ via $\nabla \mathcal{L}$
  **end for**
  **return** $\theta_s$
**end function**

# 3 Experiments

## 3.1 Experiment Settings

**Dataset and Implementation Details.** We evaluate the HGFD framework on the Autism Brain Imaging Data Exchange (ABIDE) dataset. We partition the ABIDE dataset into three sub-centers, comprising 340 (149 ASD / 191 normal controls), 254 (130 ASD / 124 NC), and 252 (112 ASD / 140 NC) samples, respectively. The data is split into training, validation, and test sets in a ratio of 8:1:1, with five-fold cross-validation using stratified sampling to preserve class balance. The input dimension and intermediate dimensions of the BrainHGNN-Teacher model are 116, 696, and 232, respectively. During the distillation phase, we set T to 3.0, $\alpha$ to 0.5, and utilized 50 training epochs. To mitigate overfitting, given the limited size of the ABIDE dataset, we employed a learning rate of 0.001, a dropout rate of 0.3, weight decay, and an early stopping strategy.

**Evaluation Metrics.** We evaluate the performance of our method using five widely used metrics: Accuracy (ACC), Area Under the Curve (AUC), Specificity (SPE), Sensitivity(SEN), and F1 Score. By integrating these metrics, we are able to conduct a more comprehensive assessment of the FedHGNN framework's performance in the autism brain imaging classification task. This approach ensures that each aspect of performance is effectively considered, particularly in the presence of class imbalance.

### 3.2 Compared Methods

**Baselines.** We conducted experiments to evaluate the performance of our model against eight state-of-the-art methods in both centralized and federated learning settings. The federated learning methods, including Fedavg, FedBN, FedProx, and Moon, improve performance by optimizing parameter aggregation strategies within the federated learning framework. Additionally, we compared our model with more complex FL models, such as FedNI [10], FedBrain [11], OCS-ADA [12], and FS2G [13], which incorporate specialized structures tailored for brain-related tasks and are more suitable for brain task learning scenarios. Finally, we also compared our model's performance with that of an MLP [22] model in a centralized setting.

### 3.3 Experimental Results

The experimental performance of all methods is shown in Table.1. The model performance under the best-performing federated learning settings for each metric is highlighted in bold. The proposed model demonstrated superior performance compared to the baseline methods on four out of the five metrics, with only slightly lower performance on the SEN metric. In summary, the experimental results confirm the effectiveness of the proposed model, demonstrating its superiority in and federated learning environments.

**Ablation experiment.** To further analyze the effectiveness of our model's structure, we conducted a series of experiments to systematically evaluate the contribution of each component to model performance. In the first set of experiments, we investigated the impact of different teacher models within the knowledge distillation framework on the student model's learning effectiveness. Specifically, we compared the hypergraph-based teacher model with traditional models such as BrainNetCNN, BrainNetTransformer(BNT), and BrainGNN. This allowed us to assess the advantages of teacher models based on CNN, graph, and hypergraph structures in brain task learning. As shown in Table.2, the hypergraph model excels in providing higher-order semantic guidance, which enables the student model to better capture complex brain network features, thus enhancing overall model performance.

In another critical experiment, we examined the importance of the knowledge distillation process itself on model performance. We tested the performance of the MLP-Student model alone and the BrainHGNN-Teacher model alone. As shown in Table.2, the performance of using either the student model or the BrainHGNN-Teacher model alone was significantly lower than when both were

Table 1: Comparison of different methods with their performance metrics.

| Method | ACC | AUC | SPE | SEN | F1 |
| --- | --- | --- | --- | --- | --- |
| MLP | $0.63 \pm 0.05$ | $0.67 \pm 0.06$ | $0.73 \pm 0.06$ | $0.64 \pm 0.05$ | $0.68 \pm 0.05$ |
| Fedavg | $0.61 \pm 0.05$ | $0.68 \pm 0.04$ | $0.72 \pm 0.09$ | $0.64 \pm 0.05$ | $0.65 \pm 0.07$ |
| FedBN | $0.61 \pm 0.07$ | $0.59 \pm 0.03$ | $0.55 \pm 0.26$ | $0.60 \pm 0.04$ | $0.52 \pm 0.16$ |
| FedProx | $0.61 \pm 0.05$ | $0.68 \pm 0.04$ | $0.70 \pm 0.05$ | $0.64 \pm 0.06$ | $0.64 \pm 0.05$ |
| Moon | $0.63 \pm 0.05$ | $0.69 \pm 0.05$ | $0.69 \pm 0.06$ | $0.66 \pm 0.05$ | $0.65 \pm 0.04$ |
| FedNI | $0.65 \pm 0.05$ | $0.69 \pm 0.04$ | $0.69 \pm 0.07$ | $0.64 \pm 0.06$ | $0.66 \pm 0.06$ |
| FedBrain | $0.66 \pm 0.04$ | $0.64 \pm 0.03$ | $0.55 \pm 0.06$ | $0.68 \pm 0.03$ | $0.58 \pm 0.03$ |
| OCS-ADA | $0.67 \pm 0.06$ | $0.62 \pm 0.05$ | $0.61 \pm 0.09$ | $\mathbf{0.69 \pm 0.07}$ | $0.62 \pm 0.31$ |
| FS2G | $0.67 \pm 0.03$ | $0.63 \pm 0.05$ | $0.64 \pm 0.06$ | $0.67 \pm 0.05$ | $0.64 \pm 0.04$ |
| HGFD | $\mathbf{0.70 \pm 0.05}$ | $\mathbf{0.69 \pm 0.05}$ | $\mathbf{0.73 \pm 0.11}$ | $0.65 \pm 0.06$ | $\mathbf{0.67 \pm 0.06}$ |

combined. This indicates that neither the student nor the teacher model alone can fully leverage their strengths in feature extraction and learning. When combined, the student model benefits from the higher-order semantics of the teacher model, particularly the complex features extracted from the hypergraph structure. This semantic guidance helps the student model better understand the deeper relationships in the data, capturing more fine-grained features and ultimately achieving better classification performance. The introduction of hypergraph knowledge distillation provides strong theoretical support and empirical validation for the success of our model.

Table 2: Performance Comparison of Different Methods in Ablation experiment

| Method | ACC | AUC | SPE | SEN | F1 |
| --- | --- | --- | --- | --- | --- |
| BrainNetCNN | $0.59 \pm 0.04$ | $0.65 \pm 0.04$ | $0.60 \pm 0.15$ | $0.64 \pm 0.06$ | $0.59 \pm 0.08$ |
| BNT | $0.57 \pm 0.03$ | $0.65 \pm 0.03$ | $0.68 \pm 0.18$ | $0.60 \pm 0.04$ | $0.61 \pm 0.08$ |
| BrainGNN | $0.61 \pm 0.04$ | $0.68 \pm 0.03$ | $0.70 \pm 0.09$ | $0.63 \pm 0.03$ | $0.64 \pm 0.05$ |
| MLP-Student | $0.59 \pm 0.03$ | $0.64 \pm 0.05$ | $0.56 \pm 0.24$ | $\mathbf{0.66 \pm 0.05}$ | $0.56 \pm 0.11$ |
| BrainHGNN-Teacher | $0.62 \pm 0.02$ | $0.66 \pm 0.02$ | $0.70 \pm 0.06$ | $0.64 \pm 0.02$ | $0.65 \pm 0.03$ |
| HGFD | $\mathbf{0.70 \pm 0.05}$ | $\mathbf{0.69 \pm 0.05}$ | $\mathbf{0.73 \pm 0.11}$ | $0.65 \pm 0.06$ | $\mathbf{0.67 \pm 0.06}$ |

## 4 Conclusion

This paper proposes a novel hypergraph-based knowledge distillation model for brain-related tasks in federated learning. Extensive experiments confirm its effectiveness, showing that the hypergraph teacher model outperforms traditional models by providing richer feature representations. The experimental results highlight the importance of knowledge distillation, demonstrating that combining the student and teacher models yields superior performance. This validates their complementary nature and the role of high-order semantic guidance in feature extraction. Our findings underscore the potential of hypergraph knowledge distillation for brain disease classification, offering insights into advanced distillation techniques for medical applications.

**Acknowledgments.** This work was supported by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang Province(2023C01046, 2023C01038), the National Nature Science Foundation of China(U21B2024).

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Stephen M Smith. Overview of fmri analysis. *The British Journal of Radiology*, 77(suppl_2):S167–S175, 2004.
2. David J Heeger and David Ress. What does fmri tell us about neuronal activity? *Nature reviews neuroscience*, 3(2):142–151, 2002.
3. Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H Staib, Pamela Ventola, and James S Duncan. Braingnn: Interpretable brain graph neural network for fmri analysis. *Medical Image Analysis*, 74:102233, 2021.
4. Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, 14(1–2):1–210, 2021.
5. Li Li, Yuxi Fan, Mike Tse, and Kuo-Yi Lin. A review of applications in federated learning. *Computers & Industrial Engineering*, 149:106854, 2020.
6. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.
7. Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, and Qi Dou. Fedbn: Federated learning on non-iid features via local batch normalization. *arXiv preprint arXiv:2102.07623*, 2021.
8. Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems*, 2:429–450, 2020.
9. Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10713–10722, 2021.

10. Liang Peng, Nan Wang, Nicha Dvornek, Xiaofeng Zhu, and Xiaoxiao Li. Fedni: Federated graph learning with network inpainting for population-based disease prediction. *IEEE Transactions on Medical Imaging*, 42(7):2032–2043, 2022.
11. Yi Yang, Han Xie, Hejie Cui, and Carl Yang. Fedbrain: federated training of graph neural networks for connectome-based brain imaging analysis. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2024*, pages 214–225. World Scientific, 2023.
12. Ling-Li Zeng, Zhipeng Fan, Jianpo Su, Min Gan, Limin Peng, Hui Shen, and Dewen Hu. Gradient matching federated domain adaptation for brain image classification. *IEEE transactions on neural networks and learning systems*, 2022.
13. Junbin Mao, Jin Liu, Xu Tian, Yi Pan, Emanuele Trucco, and Hanhe Lin. Towards integrating federated learning with split learning via spatio-temporal graph framework for brain disease prediction. *IEEE Transactions on Medical Imaging*, 2024.
14. Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3558–3565, 2019.
15. Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph learning: Methods and practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2548–2566, 2020.
16. Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. Hgnn+: General hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3181–3199, 2022.
17. Xiangmin Han, Rundong Xue, Shaoyi Du, and Yue Gao. Inter-intra high-order brain network for asd diagnosis via functional mris. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 216–226. Springer, 2024.
18. Xiangmin Han, Jingxi Feng, Heming Xu, Shaoyi Du, and Junchang Li. A hypergraph transformer method for brain disease diagnosis. *Frontiers in Medicine*, 11:1496573, 2024.
19. Chaogan Yan and Yufeng Zang. Dparsf: a matlab toolbox for" pipeline" data analysis of resting-state fmri. *Frontiers in systems neuroscience*, 4:1377, 2010.
20. Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Octave Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. 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.
21. Leif E Peterson. K-nearest neighbor. *Scholarpedia*, 4(2):1883, 2009.
22. Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.