FedGST

Mao, Junbin; Lin, Hanhe; Tian, Xu; Pan, Yi; Liu, Jin

Published in:
Proceedings of 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

DOI:
10.1109/BIBM58861.2023.10385983

Publication date:
2024

Document Version
Peer reviewed version

FedGST: Federated Graph Spatio-Temporal Framework for Brain Functional Disease Prediction

Junbin Mao
School of Computer Science and Engineering, Central South University
Changsha, China
maojunbin@csu.edu.cn

Hanhe Lin
School of Science and Engineering University of Dundee
Dundee, United Kingdom
hlin001@dundee.ac.uk

Xu Tian
School of Computer Science and Engineering, Central South University
Changsha, China
tianxu@csu.edu.cn

Yi Pan
Shenzhen Institute of Advanced Technology Chinese Academy of Sciences
Shenzhen, China
yi.pan@siat.ac.cn

Jin Liu*
School of Computer Science and Engineering, Central South University
Changsha, China
liujin06@csu.edu.cn

***Abstract*—Currently, most medical institutions face the challenge of training a unified model using fragmented and isolated data to address disease prediction problems. Although federated learning has become the recognized paradigm for privacy-preserving model training, how to integrate federated learning with fMRI temporal characteristics to enhance predictive performance remains an open question for functional disease prediction. To address this challenging task, we propose a novel Federated Graph Spatio-Temporal (FedGST) framework for brain functional disease prediction. Specifically, anchor sampling is used to process variable-length time series data on local clients. Then dynamic functional connectivity graphs are generated via sliding windows and Pearson correlation coefficients. Next, we propose an InceptionTime model to extract temporal information from the dynamic functional connectivity graphs on the local clients. Finally, the hidden activation variables are sent to a global server. We propose a UniteGCN model on the global server to receive and process the hidden activation variables from clients. Then, the global server returns gradient information to clients for backpropagation and model parameter updating. Client models aggregate model parameters on the local server and distribute them to clients for the next round of training. We demonstrate that FedGST outperforms other federated learning methods and baselines on ABIDE-1 and ADHD200 datasets.**

***Index Terms*—Spatio-Temporal, Federated Learning, Brain Functional Disease, Graph Learing**

I. Introduction

Brain function diseases such as Autism Spectrum Disorder (ASD) and Attention Deficit and Hyperactivity Disorder (ADHD) significantly impair patients' cognition, communication, and behavior, posing certain challenges to their families and society [1]. Recently, the rapid development of Functional Magnetic Resonance Imaging (fMRI) provides technical support for the diagnosis of functional brain diseases. Researchers use similarity metrics between blood oxygen signals in brain regions of fMRI to construct Functional Connectivity (FC) [2]. The flattened FC are then used to explore brain network abnormalities in patients.

Since FC fluctuates over time, it is important to use dynamic FC to predict functional brain disorders [3], [4]. Spatio-temporal graph networks [4], due to their powerful modeling ability for multivariate time series, make them effective for time series forecasting and classification, even in medical-related fields. So far, disease prediction based on spatio-temporal networks has achieved significant progress [5]–[7]. For example, Campbell et al. [5] proposed a Dynamic Brain Graph Structure Learning (DBGSL) network, which constructs dynamic graph adjacency matrices based on spatial embeddings of brain regions learned from blood oxygen signals using sliding windows. The classification performance and interpretability are further improved by incorporating temporal attention and learnable edge sparsity. Tiago et al. [7] proposed a novel deep neural network architecture that combines Graph Neural Network (GNN) and Temporal Convolutional Network (TCN) to learn from spatial and temporal information in rs-fMRI data in an end-to-end manner, which was validated on the UK Biobank dataset. The research by Pati et al. [8] demonstrates that training robust and accurate models requires a large amount of data, and the diversity of the data affects the model's ability to generalize to "out-of-sample" cases. To address this issue, data from many sites have to be collected and trained together, a method called centralized learning.

However, the existing state-of-the-art spatio-temporal prediction models are not applicable for real-world applications as data are not always publicly accessible due to data privacy issues. Instead, by sharing model parameters from different sites, Federated Learning (FL) allow models to achieve performance similar to those who trained with centralized learning. Recently, some approaches w.r.t. FL and Split Learning (SL) [9] have been proposed. For example, Peng et al. [10] proposed FedNI, a disease prediction model based on federated graph neural network embeddings, which uses GAN to train and generate missing nodes and edges to repair local networks, then uses FL to train a global GCN node classifier, validated

on ABIDE-1 and ADNI datasets. The accuracy reached 66.7% and 75.8%, respectively. Meng et al. [11] proposed a model of Cross-Node Federated Graph Neural Networks (CNFGNN) for spatio-temporal data modeling based on SL, which decomposes temporal and spatial dependencies using encoder-decoder models on each client, extracts temporal features locally using Gate Recurrent Unit (GRU), and captures spatial dependencies across clients on the server using GNN. Thapa et al. [12] combined FL and SL to propose SplitFed, which first splits the ResNet18 model into two parts, with clients and server holding one part each, clients learn hidden features of input images and send to server model, server then returns gradients to client models for parameter update, finally client models perform one round of parameter aggregation and dispatch to individual clients for next round of training.

Considering the spatio-temporal dependence of multi-site data, we propose Federated Graph Spatio-Temporal (FedGST) framework, the main contributions of FedGST can be summarized as follows:

1) We propose a novel FedGST framework. It introduces a split learning mechanism to deploy a temporal model on the client side and a spatial model on the server side, which enhances the models' decoupling capability.
2) We proposed to use the InceptionTime network to extract the temporal information of dynamic FC.
3) We propose a United Graph Convolutional Network (UniteGCN) which learns the differences between multiple graph convolution methods through cross-layer attention, and fuses the output features for disease prediction.

II. Methodology

A. Overview

The framework of our proposal Federated Spatio-Temporal Network (FedGST) is shown in Fig. 1. It is mainly divided into a temporal model on the client side and a spatial model on the server side. Specifically, various clients use the temporal model InceptionTime to extract temporal information from the subject's dynamic brain functional networks and send the latent variables to the server. The spatial model on the server performs forward propagation and returns the gradients to the client for updating the model parameters. The client aggregates the model parameters on the local server and distributes them to each client for the next training.

B. Client Temporal Model

1) Data Preprocessing: Since the length of fMRI data collected at each site varies, we propose an anchor point sampling algorithm. By specifying the number of anchor points M, M will be uniformly distributed in the fMRI time series. The sampling stride is automatically determined by the ratio of time series length to M. Each sampling interval will generate a mean value, so the time series will have a length of M. Then multiple time series segments are extracted through sliding window operations. Finally, dynamic FC is obtained by calculating functional connections FC between brains using Pearson Correlation. The formula for calculating the interval length of anchor point sampling is as follows:

$$L_{\text{sample}} = \left\lfloor \frac{L_{\text{roi}}}{M} \right\rfloor + \varepsilon, \begin{cases} \varepsilon = 1 \ , if \ \ \lfloor \frac{L_{\text{roi}}}{M} \rfloor \ \% \ 2 = 0 \\ \varepsilon = 2 \ , if \ \ \lfloor \frac{L_{\text{roi}}}{M} \rfloor \ \% \ 2 = 1, \end{cases} \tag{1}$$

where L_{sample} represents the sampling interval length, L_{roi} represents the length of the brain region time series, $\lfloor \rfloor$ represents the round down operation.

Let $X_j = \{x_1, x_2, x_3, \ldots x_M\}$ be the brain region time series with a length of M, we apply a sliding window to it to generate dynamic FC. Let the window size be L_{inter}, the step size be L_{step}, then the i-th window can be represented as:

$$W_i = \{x_{i-1} * L_{\text{step}} + 1, \ldots, x_{i-1} * L_{\text{step}} + L_{\text{inter}}\}, \tag{2}$$

where i slides from 1 to $\lfloor N - w)/L_{\text{step}} \rfloor + 1$. For the brain region time series in each window W, we calculate the Pearson correlation coefficients to obtain a FC graph:

$$\text{FC}\left(r_i, r_j\right) = \frac{E\left(r_i r_j\right) - E\left(r_i\right) E\left(r_j\right)}{\sqrt{E\left(r_i^2\right) - E^2\left(r_i\right)} \sqrt{E\left(r_j^2\right) - E^2\left(r_j\right)}}, \tag{3}$$

where $\text{FC}\left(r_i, r_j\right)$ represents the FC between brain regions r_i and r_j, r_i and r_j are the time series of brain regions i and j, respectively. E is the mathematical expectation.

2) InceptionTime: For the dynamic FC $\mathbf{X}_{\text{dfc}} \in \mathbb{R}^{N \times T \times F}$ obtained through preprocessing, where N is the number of samples, T is the number of time points, and F is the number of FC. We input $\mathbf{X}_{\text{dfc}} \in \mathbb{R}^{N \times T \times F}$ into the InceptionTime model. As shown in Fig. 1, the InceptionTime model consists of three InceptionTime blocks, a global pooling layer, a fully connected layer and a softmax layer. The InceptionTime Blocks are connected in a residual manner.

Fig. 2 illustrates the internal details of the InceptionTime Block. The first major component of the Inception module is called the bottleneck layer. This layer performs an operation of sliding f filters of length 1 with a stride of 1. This will convert the time series from $\mathbf{X}$ with F dimensions to $\mathbf{X}$ with f dimensions, thereby significantly reducing the dimension of the time series as well as the complexity of the model and alleviating the overfitting problem on small datasets. The second major component of the Inception module is sliding multiple filters of different lengths over the same input time series simultaneously. For example, in Fig. 2, three different convolutions with lengths $l \in (4, 8, 16)$ are applied to the input $\mathbf{X}$. In addition, to make our model invariant to small perturbations, we introduce another parallel MaxPooling operation, followed by a bottleneck layer to reduce the dimension. Finally, the outputs of each independent parallel convolution and max pooling are concatenated to form the output $\mathbf{X}$.

$$X' \leftarrow \text{Softmax}(W \cdot \text{GAP}(\sum_{i}^{3} X + g(X))), \tag{4}$$

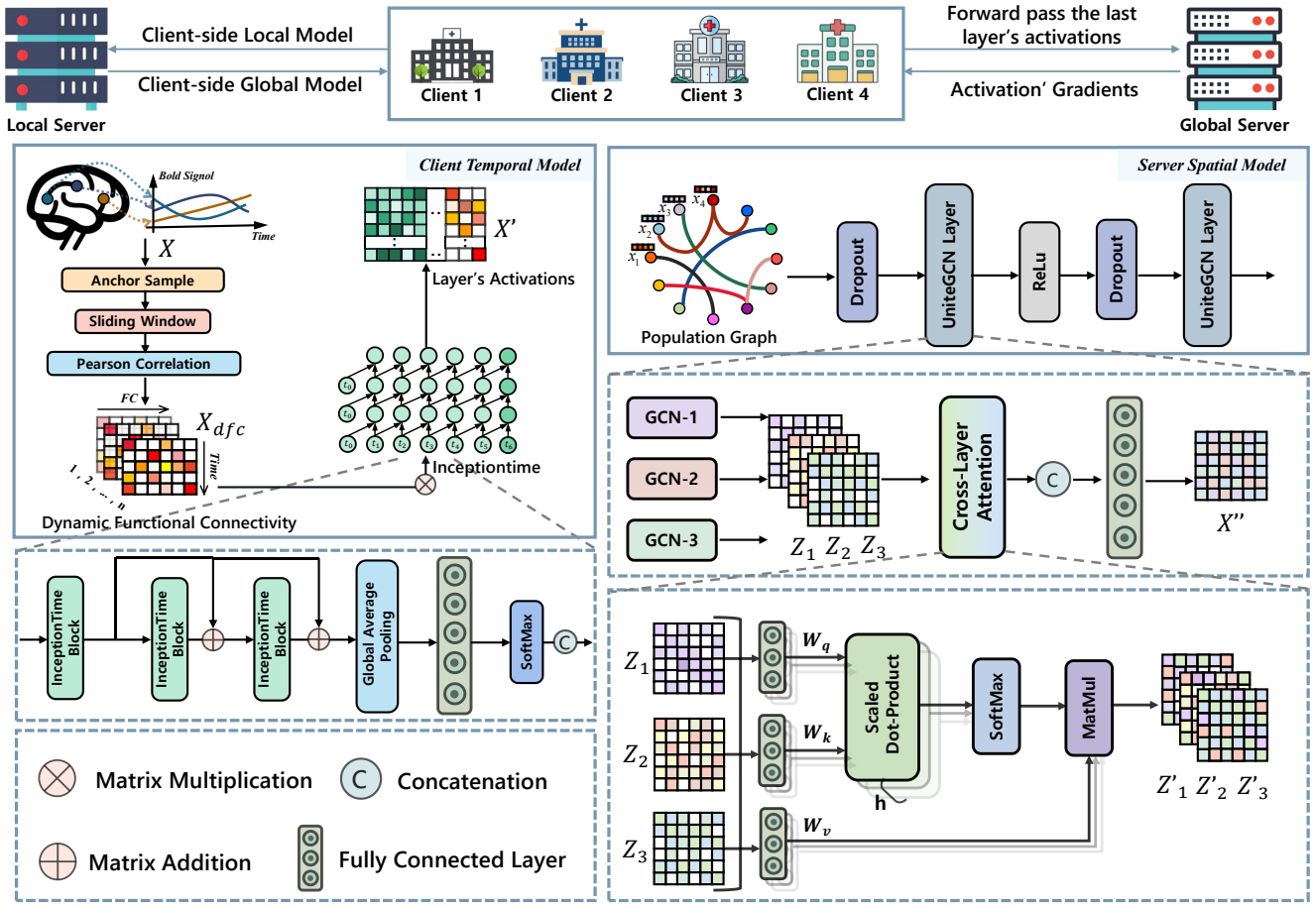

Fig. 1. The network architecture of the proposed FedGST. **Client Temporal Model**: the fMRI data are processed through anchor sampling, sliding window, and Pearson Correlation to generate the dynamic FC. Activation $\mathbf{X}'$ is then generated by the InceptionTIme model. **Server Spatial Model**: the server accepts activations from each client and uses UniteGCN to extract spatial information from the samples for disease prediction.

where $g(X)$ is the output of the InceptionTime Block, $X + g(X)$ represents the three residual networks, GAP is global average pooling, W is the learnable parameter of the fully connected layer. Finally, the fMRI temporal information $X' \in \mathbb{R}^{N \times d}$ representing each sample is obtained from the InceptionTime model outputs.

C. Server Spatial Model

1) UniteGCN: The central server receives the last layer activation variables $\mathbf{X}'$ from the client model, generates the population graph G of the subjects through the following formula,

$$G = \frac{\left(\mathbf{X}'_i\right)^T \mathbf{X}'_j}{2 \left\|\mathbf{X}'_i\right\| \left\|\mathbf{X}'_j\right\|} + 0.5, \tag{5}$$

where $\mathbf{X}'_i$ represents the temporal information of the i-th sample.

Then inputs population graph G and activation variables $\mathbf{X}'$ into the spatial model UniteGCN. UniteGCN consists of two UniteGCN Layers. The internal details of UniteGCN Layer are shown in Fig. 1. Specifically, $\mathbf{X}'$ first go through Dropout to enhance model generalization, and then uses three different graph convolution methods, GraphConv [13], SAGEConv [14] and TransformerConv [15], to aggregate and learn information from $\mathbf{X}'$ to obtain outputs Z_1, Z_2, Z_3. Next, cross-layer

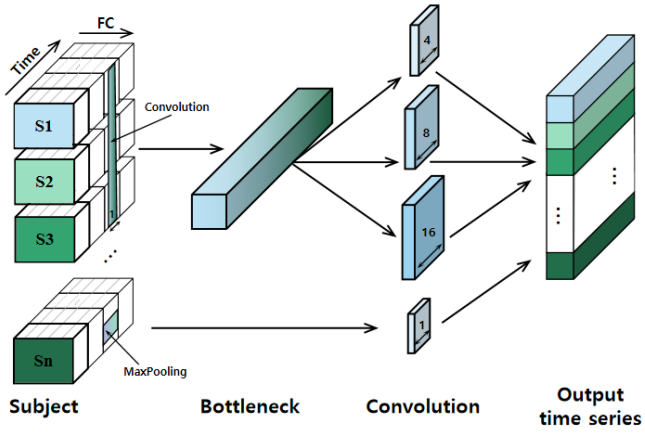

Fig. 2. The network architecture of the InceptionTime block.

attention is used to calculate the attention scores between Z_1, Z_2, Z_3 in pairs, and multiply them with Z_1, Z_2, Z_3 to obtain Z_1', Z_2', Z_3'. The formula for cross-layer attention can be expressed as follows:

$$Q_1, Q_2, Q_3 = W_q \cdot (Z_1, Z_2, Z_3), \tag{6}$$

$$K_1, K_2, K_3 = W_k \cdot (Z_1, Z_2, Z_3), \tag{7}$$

$$V_1, V_2, V_3 = W_v \cdot (Z_1, Z_2, Z_3). \tag{8}$$

Z_1, Z_2, Z_3 generate Q, K, V through three learnable parameters W_q, W_k, W_v respectively, and then calculate the attention scores between each other. The formula is shown as follows:

$$\mathbf{P}_{n,ij} = \frac{\exp\left[q_n^i \cdot (k_n^j)^\top / \tau\right]}{\sum_{j=1}^{Z} \exp\left[q_n^i \cdot (k_n^j)^\top / \tau\right]}, \tag{9}$$

where $\mathbf{P}_{n,ij}$ indicates how much attention Z_i pays to Z_j in the n-th sample, that is, their dependence relationship. τ is a proportional factor that controls the hardness of attention, usually set to the square root of $\sqrt{d}$. Then, after normalizing $\mathbf{P}_{n,ij}$ with sofamax, it is multiplied with V_1, V_2, V_3 to obtain the output results Z_1', Z_2', Z_3' after cross-layer attention of different convolution methods. Finally, concat is used to concatenate Z_1', Z_2', Z_3' and pass them through a fully connected layer to learn features and obtain the output $\mathbf{X}''$.

The second UniteGCNLayer sets the output dimension of $\mathbf{X}''$ to 2 to obtain the probability of disease prediction, and then calculates the loss using cross-entropy loss, the formula is shown below:

$$\mathcal{L} = -\sum_{i=1}^{N} \left(y_i \ln\left(\hat{x}_i\right) + \left(1 - y_i\right) \ln\left(1 - \hat{x}_i\right)\right), \tag{10}$$

where y_i is the label of the i-th sample, $\hat{x}_i$ is its corresponding predicted probability score. Finally, the global server sends the gradients to the client for backpropagation and executes model parameter updates.

III. Experiments and Results

A. Dataset and Experimental Settings

We evaluate our method FedGST using two datasets (ABIDE-1 and ADHD200). For ABIDE-1, we use the identical sample size as in the study [16], which consisted of a total of 871 subjects, including 468 normal controls (NC) and 403 ASD patients. For ADHD200, there are a total of 775 subjects, consisting of 491 normal controls and 284 ADHD patients. We perform preprocessing operations using Dpabi [17]. The main processes include time point correction, head motion correction, skull stripping from structural images, alignment between structural MRI (sMRI) and fMRI, normalization to MNI space, etc.

We use four common metrics to evaluate the performance of our model: accuracy (ACC), area under the curve (AUC), specificity (SPE), and sensitivity (SEN). We use k-fold cross-validation (k=5) and report the mean and standard deviation of the experimental results after running 5 times.

The hyperparameter settings for the experiments are as follows: learning rate = 0.001, weight decay = 0.0005, number of epochs = 150, number of clients = 4.

B. Comparison with some existing methods

We compare our method with state-of-the-art federated learning methods of the same type and some baseline methods. Specifically, we include four advanced federated learning methods: MoE [18], SFL [12], CNFGNN [11] and FedNI [10], as well as four FL-based baseline methods: FedMLP, FedGCN, FedSage, and LSTN+GCN.

The experimental results on the ABIDE-1 dataset are shown in Table 1. It can be seen that FedGST outperforms on three metrics: ACC, AUC, and SEN, with only FedSage surpassing FedGST on SPE. However, FedGST has the smallest standard deviation, indicating that FedGST is more stable than FedSage.

The experimental results on the ADHD200 dataset are shown in Table 1. It can be seen that FedGST outperforms on three metrics: ACC, AUC, and SPE, with only FedSage surpassing FedGST on SEN. The accuracy performance of the four advanced methods is generally higher than that of the baseline methods, except for SFL. This may be because SFL processes the brain functional networks into image inputs and is unable to effectively extract features. In contrast, Table 1 shows that the accuracy of the four advanced methods is generally weaker than the four baseline methods. We assume this may be related to the distribution characteristics of the data and the imbalance between the two datasets.

In Table 1, the four advanced methods generally outperform the four baseline methods in terms of accuracy. This may be because the hyperparameters used in the experiments happened to be well-suited for the advanced methods. This interesting phenomenon also indicates significant differences between the ABIDE-1 and ADHD200 datasets. This could arise from differences in biomarkers between the two disorders, including variations in disease-related functional brain regions.

C. Ablation study

To assess the effectiveness of various modules in FedGST, we performed ablation experiments on the InceptionTime and UniteGCN modules.

The ablation results on ABIDE-1 are shown in Table 2. LSTM+GCN is the backbone network, representing the basic temporal model + spatial model, with an accuracy of 65.81%. On the basis of backbone, we replace LSTM with InceptionTime model, and the accuracy is improved by 0.7%. Then, we replace GCN with UniteGCN, and the accuracy improved by 3.02% to 69.53%, proving that UniteGCN is more effective than the standard GCN.

The ablation results on ADHD are shown in Table 3. LSTM+GCN is the backbone network, with an accuracy of 67.76%. On the basis of backbone, we replace LSTM with InceptionTime, resulting in an improvement in accuracy by 2.37%. Then, we replace GCN with UniteGCN, and the accuracy is improved by 1.14% to 71.27%, also proving UniteGCN is more effective than standard GCN.

TABLE I
THE PERFORMANCE COMPARISON OF SOTA METHODS.

Method	ASD vs NC				ADHD vs NC			
	ACC (%)	AUC (%)	SEN (%)	SPE (%)	ACC (%)	AUC (%)	SEN (%)	SPE (%)
FedMLP	66.63 ± 3.24	68.60 ± 3.94	72.81 ± 10.49	59.47 ± 11.23	66.05 ± 3.83	64.04 ± 2.59	56.88 ± 10.49	75.41 ± 5.71
FedGCN	65.93 ± 3.48	67.59 ± 4.17	73.71 ± 6.43	56.65 ± 5.21	66.84 ± 1.53	62.75 ± 3.98	40.79 ± 11.96	82.41 ± 7.58
FedSage	66.51 ± 3.65	67.92 ± 4.00	67.93 ± 6.04	**65.00 ± 8.33**	66.18 ± 4.21	69.24 ± 3.93	**63.38 ± 9.88**	67.05 ± 12.03
LSTM+GCN	63.84 ± 1.74	66.20 ± 2.82	69.83 ± 6.19	56.85 ± 9.34	67.11 ± 3.79	69.80 ± 4.56	49.69 ± 10.31	77.36 ± 1.76
MoE	65.70 ± 3.87	67.89 ± 4.03	74.51 ± 8.07	55.53 ± 7.59	68.42 ± 2.20	66.93 ± 2.97	49.24 ± 11.82	78.82 ± 7.80
SFL	62.67 ± 2.64	60.31 ± 3.69	70.68 ± 9.80	53.57 ± 10.03	64.21 ± 0.89	54.60 ± 6.15	8.62 ± 7.79	**95.55 ± 4.21**
CNFGNN	63.95 ± 1.84	63.88 ± 3.60	69.75 ± 3.38	57.46 ± 5.33	68.82 ± 2.34	66.89 ± 4.31	48.34 ± 5.19	80.59 ± 1.22
FedNI	65.81 ± 6.97	65.85 ± 8.90	**78.93 ± 11.87**	51.20 ± 26.29	69.74 ± 3.09	66.89 ± 2.44	50.23 ± 7.47	80.83 ± 3.33
FedGST	**69.53 ± 1.45**	**73.35 ± 2.17**	74.89 ± 6.65	63.32 ± 8.35	**71.27 ± 1.75**	**70.79 ± 7.92**	60.80 ± 8.58	79.31 ± 5.52

TABLE II
THE PERFORMANCE OF MODULE ABLATION OF FEDGST ON ABIDE-1.

InceptionTime	UniteGCN	ACC (%)	AUC (%)	SEN (%)	SPE (%)	F1 (%)
		65.81 ± 3.44	65.16 ± 2.85	74.03 ± 4.19	56.28 ± 4.89	69.92 ± 3.13
✓		66.51 ± 2.48	70.07 ± 3.01	70.91 ± 10.06	61.27 ± 11.55	69.21 ± 3.55
✓	✓	**69.53 ± 1.45**	**73.35 ± 2.17**	**74.89 ± 6.65**	**63.32 ± 8.35**	**72.43 ± 2.09**

TABLE III
THE PERFORMANCE OF MODULE ABLATION OF FEDGST ON ADHD200.

InceptionTime	UniteGCN	ACC (%)	AUC (%)	SEN (%)	SPE (%)	F1 (%)
		67.76 ± 2.73	69.64 ± 4.39	47.98 ± 10.22	79.33 ± 3.72	51.87 ± 6.84
✓		70.13 ± 2.59	70.21± 3.36	48.74 ± 10.79	**82.46 ± 6.75**	53.95 ± 6.82
✓	✓	**71.27 ± 1.75**	**70.79 ± 3.92**	60.80 ± 8.58	79.31 ± 5.52	**61.45 ± 7.89**

From the ablation results on both datasets, we can see that the two proposed modules are all effective, and can be combined together nicely to produce good results. Additionally, on ABIDE-1, the performance improvements from InceptionTime are limited to only 0.7%, suggesting that InceptionTime is slightly better than LSTM. The improvements from UniteGCN on ABIDE-1 are much higher than on ADHD200, differing by 1.88%. This not only demonstrates the efficacy of UniteGCN but also further validates the differences between the ABIDE-1 and ADHD200 datasets.

IV. CONCLUSION

In this study, we propose a novel framework called Federated Graph Spatio-Temporal (FedGST) for brain functional diseases. The experimental results demonstrate that our method outperforms existing approaches, and ablation studies validate the effectiveness of each component. There is a limitation in this study. Since the sample sizes vary across sites, models cannot be trained independently using each site's data. Therefore, we mix and average data from all sites when assigning it to clients for training, which does not accurately simulate real-world conditions.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 62172444 and U22A2041, in part by the Shenzhen Science and Technology Program under Grant KQTD20200820113106007, in part by the Natural Science Foundation of Hunan Province under Grant 2022JJ30753, in part by the Central South University Innovation-Driven Research Pro-gramme under Grant 2023CXQD018, and in part by the High Performance Computing Center of Central South University.

REFERENCES

[1] E. Canario, D. Chen, and B. Biswal, "A review of resting-state fmri and its use to examine psychiatric disorders," *Psychoradiology*, vol. 1, no. 1, pp. 42–53, 2021.

[2] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, and D. Rueckert, "Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease," *Medical image analysis*, vol. 48, pp. 117–130, 2018.

[3] C. Chang and G. H. Glover, "Time–frequency dynamics of resting-state brain connectivity measured with fmri," *Neuroimage*, vol. 50, no. 1, pp. 81–98, 2010.

[4] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 01, 2020, pp. 914–921.

[5] A. Campbell, A. G. Zippo, L. Passamonti, N. Toschi, and P. Lio, "Dbgsl: Dynamic brain graph structure learning," *arXiv preprint arXiv:2209.13513*, 2022.

[6] T. Azevedo, L. Passamonti, P. Lio, and N. Toschi, "Towards a predictive spatio-temporal representation of brain data," *arXiv preprint arXiv:2003.03290*, 2020.

[7] T. Azevedo, A. Campbell, R. Romero-Garcia, L. Passamonti, R. A. Bethlehem, P. Lio, and N. Toschi, "A deep graph neural network architecture for modelling spatio-temporal dynamics in resting-state functional mri data," *Medical Image Analysis*, vol. 79, p. 102471, 2022.

[8] S. Pati, U. Baid, B. Edwards, M. Sheller, S.-H. Wang, G. A. Reina, P. Foley, A. Gruzdev, D. Karkada, C. Davatzikos *et al.*, "Federated learning enables big data for rare cancer boundary detection," *Nature communications*, vol. 13, no. 1, p. 7346, 2022.

[9] O. Gupta and R. Raskar, "Distributed learning of deep neural network over multiple agents," *Journal of Network and Computer Applications*, vol. 116, pp. 1–8, 2018.

[10] L. Peng, N. Wang, N. Dvornek, X. Zhu, and X. Li, "Fedni: Federated graph learning with network inpainting for population-based disease prediction," *IEEE Transactions on Medical Imaging*, 2022.

[11] C. Meng, S. Rambhatla, and Y. Liu, "Cross-node federated graph neural network for spatio-temporal data modeling," in *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2021, pp. 1202–1211.

[12] C. Thapa, P. C. M. Arachchige, S. Camtepe, and L. Sun, "Splitfed: When federated learning meets split learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, 2022, pp. 8485–8493.

[13] C. Morris, M. Ritzert, M. Fey, W. L. Hamilton, J. E. Lenssen, G. Rattan, and M. Grohe, "Weisfeiler and leman go neural: Higher-order graph neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 4602–4609.

[14] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.

[15] Y. Shi, Z. Huang, S. Feng, H. Zhong, W. Wang, and Y. Sun, "Masked label prediction: Unified message passing model for semi-supervised classification," *arXiv preprint arXiv:2009.03509*, 2020.

[16] A. Abraham, M. P. Milham, A. Di Martino, R. C. Craddock, D. Samaras, B. Thirion, and G. Varoquaux, "Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example," *NeuroImage*, vol. 147, pp. 736–745, 2017.

[17] C.-G. Yan, X.-D. Wang, X.-N. Zuo, and Y.-F. Zang, "Dpabi: data processing & analysis for (resting-state) brain imaging," *Neuroinformatics*, vol. 14, pp. 339–351, 2016.

[18] X. Li, Y. Gu, N. Dvornek, L. H. Staib, P. Ventola, and J. S. Duncan, "Multi-site fmri analysis using privacy-preserving federated learning and domain adaptation: Abide results," *Medical Image Analysis*, vol. 65, p. 101765, 2020.