Investigating the Predictive Reproducibility of Federated Graph Neural Networks using Medical Datasets

Mehmet Yiğit Balık¹, Arwa Rekik ¹,² and Islem Rekik ¹⋆

¹BASIRA Lab, Faculty of Computer and Informatics Engineering, Istanbul Technical University, Istanbul, Turkey (http://basira-lab.com/)
²Faculty of Medicine of Sousse, Sousse, Tunisia

Abstract. Graph neural networks (GNNs) have achieved extraordinary enhancements in various areas including the fields medical imaging and network neuroscience where they displayed a high accuracy in diagnosing challenging neurological disorders such as autism. In the face of medical data scarcity and high-privacy, training such data-hungry models remains challenging. Federated learning brings an efficient solution to this issue by allowing to train models on multiple datasets, collected independently by different hospitals, in fully data-preserving manner. Although both state-of-the-art GNNs and federated learning techniques focus on boosting classification accuracy, they overlook a critical unsolved problem: investigating the reproducibility of the most discriminative biomarkers (i.e., features) selected by the GNN models within a federated learning paradigm. Quantifying the reproducibility of a predictive medical model against perturbations of training and testing data distributions presents one of the biggest hurdles to overcome in developing translational clinical applications. To the best of our knowledge, this presents the first work investigating the reproducibility of federated GNN models with application to classifying medical imaging and brain connectivity datasets. We evaluated our framework using various GNN models trained on medical imaging and connectomic datasets. More importantly, we showed that federated learning boosts both the accuracy and reproducibility of GNN models in such medical learning tasks. Our source code is available at https://github.com/basiralab/reproducibleFedGNN.

Keywords: Graph neural networks · Federated Learning · Reproducibility · Brain connectivity graphs · Predictive medicine

1 Introduction

Over the last years, artificial intelligence (AI) applied to medicine has witnessed exponential growth aiming to ease the diagnostic approach and propel, consequently, the development of personalized treatment strategies. Specifically, advanced deep learning (DL) models such as convolutional neural networks (CNNs)
have achieved a remarkable performance across a variety of medical imaging tasks including segmentation, classification, and registration [1,2]. However, such networks were primarily designed to handle images, thereby failing to generalize to non-euclidean data such as graphs and manifolds [3,4]. Recently, graph neural networks (GNNs) were introduced to solve this problem by designing novel graph-based convolutions [4,5]. A recent review paper [6] demonstrated the merits of using GNNs particularly when applied to brain connectomes (i.e., graphs) across different learning tasks including longitudinal brain graph prediction, brain graph super-resolution and classification for neurological disorder diagnosis. Although promising, GNNs remain deep models which are data-hungry. Faced with the scarcity of medical imaging datasets and their high privacy and sensitivity, they can remain sub-optimal in their performance. In this perspective, federated learning [7] can bring a promising alternative to training GNNs models using decentralized data spread across multiple hospitals while boosting the accuracy of each local GNN model in a fully data-preserving manner. Although increasing the model accuracy through federation seems compelling, there remains a more important goal to achieve which is maximizing the reproducibility of a locally trained model. A model is defined as highly reproducible when its top discriminative features (e.g., biomarkers) remain unchanged against perturbations of training and testing data distributions as well as across other models [8,9,10]. Quantifying the reproducibility of a predictive medical model presents one of the biggest hurdles to overcome in developing translational clinical applications. In fact, this allows identifying the most reproducible biomarkers that can be used in treating patients with a particular disorder. To the best of our knowledge, reproducibility in federated learning remains an untackled problem. [8] proposed the first framework investigating the reproducibility of GNN models. Specifically, the designed RG-Select framework used 5 different state-of-the-art GNN models to identify the most reproducible GNN model for a given connectomic dataset of interest. Although RG-Select solves both GNN reproducibility and non-euclidean data learning problems, it does not address the problem of model reproducibility when learning on decentralized datasets distributed across different hospitals. Undeniably, medical datasets carry information about patients and their medical conditions. Hence, the patient may be identified using such data. Patients have the right to control their personal information and keep it for themselves [11]. Such data must be held private between the patient and their health care workers. For such reasons, federated learning presents a great opportunity to learn without clinical data sharing and while boosting the model accuracy as well as its reproducibility.

We draw inspiration from the seminal work on decentralized learning where [7] proposed a federated averaging algorithm based on training many local models on their local datasets then aggregating the learned models at the server level. Next, the global server broadcasts their learned weights to each local modal for local updates. Several researchers were inspired by federated learning and adapted it to graphs [12,13]. Even though these proposed frameworks managed to boost the local accuracy of local models while handling decentral-
ized data, they overlook the reproducibility of the most discriminative features (i.e., biomarkers). Will federated learning also boost the reproducibility of locally trained GNN models? Here we set out to address this prime question by quantifying the reproducibility of federated local models.

In order to ensure high accuracy, handle decentralized datasets and identify the most reproducible discriminative features, we federate GNN models and quantify their reproducibility by perturbing training and testing medical data distributions through random data splits. Our framework generalizes the seminal work of RG-Select [8] to federated models. Specifically, given a pool of GNN architectures to federate, we aim to identify the most reproducible GNN model across local hospitals and its corresponding biomarkers by quantifying the reproducibility of the global model. The key contributions of our framework are to: (1) Federate the learning of predictive GNN models with application to medical imaging and connectomic datasets. (2) Investigate and quantify the reproducibility of federated GNN models, and (3) identify the most reproducible biomarkers for neurological disorder diagnosis.

2 Proposed Method

In this section, we detail our federated reproducibility quantification framework as illustrated in Fig 1. First, we divide the whole data into \( H \) different subsets. Each subset represents the local data of a particular hospital. Second, we train different GNN models using federated learning trained on each local dataset. Following the training, we extract the top \( K \) discriminative biomarkers (features) identified by each locally trained GNN model. Next, for each hospital, we produce a \textit{hospital-specific GNN-to-GNN reproducibility matrix} where each element denotes the overlap ratio between the extracted top \( K \) biomarker sets by pairs of locally trained GNN models. We then construct the \textit{global reproducibility matrix} by averaging all hospital-specific reproducibility matrices. Finally, we identify the most reproducible GNN model across hospitals in the federation process by identifying the central node with the highest overlap with other nodes in the global average reproducibility matrix. The selected model is then used to identify the most reproducible features.

\textbf{Problem statement.} Given \( H \) hospitals with the local datasets \( D_h = (G_h, Y_h) \) that belong to the \( h^{th} \) hospital, where \( h \in \{1, 2, \ldots, H\} \), let \( D_h \) denote a local dataset including subjects with their diagnostic states/labels (e.g., normal control and disordered). Let \( S \) denote the number of subjects in \( D_h \). \( G_h = \{G_{h,1}, G_{h,2}, \ldots, G_{h,S}\} \) denotes the set of medical data graphs and their labels are denoted by \( Y_h = \{y_{h,1}, y_{h,2}, \ldots, y_{h,N}\} \). Each graph \( G_{h,n} \) is represented by an adjacency matrix \( X_{h,n} \in \mathbb{R}^{N \times N} \) and a label \( y_{h,n} \in \{0,1\} \) where \( N \) denotes the number of brain regions of interest (ROIs) for connectivity datasets or pixels for medical imaging datasets. Note that \( N \) also represents the number of nodes in the corresponding graph.

Given a pool of \( M \) GNNs \( \{GNN_1, GNN_2, \ldots, GNN_M\} \), we are interested in training a GNN model \( GNN_{h,m} : G_h \rightarrow Y_h \) on the local dataset of hospital \( h \). Our aim is to identify the most reproducible biomarkers or features that
Fig. 1: Overview of the proposed framework for quantifying the reproducibility of federated GNN models across decentralized datasets. (A) Medical datasets. We split our dataset into $H$ local datasets. (B) Federated learning with different GNNs. We use $M$ GNN models to identify the most reproducible GNN model during the federation learning. For each local hospital model $GNN_m$ where $m \in \{1, \ldots, M\}$, we extract its top $K$ discriminative features and calculate their overlap ratio with discriminative feature sets selected by other GNN models. (C) Construction of hospital-specific reproducibility matrix. Using the intersections calculated in the previous step, we construct the hospital-specific reproducibility matrix where each element $(i, j)$ denotes the overlap in the top $K$ features identified by the locally trained $GNN_i$ and $GNN_j$. (D) Construction of average global reproducibility matrix across federated models. Using the produced hospital-specific reproducibility matrices, we calculate the average global reproducibility matrix, thereby identifying the most reproducible features across models and hospitals.

discriminate between the two classes. Hence, we extract the top $K$ features $r_{h,m}^K \in \mathbb{R}^K$ learned by the $m^{th}$ local GNN model in the $h^{th}$ hospital, where $m \in \{1, 2, \ldots, M\}$. We calculate the intersection of the extracted local top $K$ features $r_{h,m}^K \cap r_{h,l}^K$, where $m$ and $l$ are the indexes of GNN models in the GNN pool and $h$ is the index of a hospital. In order to calculate the reproducibility matrices, we extract the weights $w_{h,m} \in \mathbb{R}^N$ learned by the $h^{th}$ hospital using the $m^{th}$ GNN architecture.

**Definition 1.** Let $GNN_i$ and $GNN_j$ be two GNN models and let $w_i \in \mathbb{R}^n$ and $w_j \in \mathbb{R}^n$ be their weights, respectively. The top $K$ biomarkers extracted using the weights $w_i, w_j$ are denoted by $r_i^K$ and $r_j^K$, respectively. Reproducibility...
among models \(GNN_i\) and \(GNN_j\) is denoted by \(R^K_{i,j}\) which can be calculated as:

\[
R^K_{i,j} = \frac{|r^K_i \cap r^K_j|}{K}
\]

GNN training mode. Each local data is divided into 3 folds where 2 folds are used for training and the left-out fold is used for validation. We train each local GNN on its local dataset over \(E\) epochs and using \(B\) batches. Both global and local models communicate for \(C\) rounds. In each round, the global model sends a deep copy of the current GNN model to all local hospitals. Each hospital does training using its local data. When the training ends, hospitals send locally updated weights to the central server. The server applies Algorithm 1 on the weights that came from the local models and loads the averaged weights to the global model.

Biomarker selection. We extract the learned weights by each GNN model in order to select the top \(K\) discriminative biomarkers. The extracted weights belong to the last embedding layer of the GNN model [8]. Next, we rank the biomarkers according to the absolute value of their corresponding weights and select the top \(K\) with the highest weights. We use these biomarkers to construct GNN-to-GNN hospital-specific reproducibility matrices.

---

**Algorithm 1** Federated Averaging. \(H\) hospitals indexed by \(h\); \(C\) is the number of communication rounds; \(G\) is the global model

1: **LocalUpdate**\(\leftarrow G\) : // Runs on hospital \(h\)
2: for each epoch \(i\) in \(\{1, \ldots, E\}\) do
3: for each batch \(b\) in \(B\) do
4: \(w \leftarrow w - \eta \nabla l(w; b)\)
5: return \(w\)
6: **Server Executes**:
7: initialize global model \(G\)
8: for each round \(t\) in \(\{1, \ldots, C\}\) do
9: for each hospital \(h\) in \(\{1, \ldots, H\}\) do
10: \(w^h_{t+1} \leftarrow \text{LocalUpdate}(\text{deepCopy}(G))\) // Copy of global model sent to local update
11: \(w_{t+1} \leftarrow \frac{1}{H} \sum_{h=1}^{H} w^h_{t+1}\)

---

GNN-to-GNN reproducibility matrix. Using the top \(K\) biomarkers, the overlap of each pair of GNN models is calculated thereby producing their GNN-to-GNN reproducibility score. This step is executed for each hospital individually to produce a hospital-specific reproducibility matrix. Repeating this operation for all \(H\) hospitals, the average of \(H\) hospital-specific matrices is then calculated, and the average global reproducibility matrix is constructed using Algorithm 2.

**The most reproducible GNN and biomarker selection.** In order to select the most reproducible GNN model, we use the average reproducibility matrix of the \(H\) hospital-specific reproducibility matrices. We consider this matrix as a graph where the GNN models are its nodes. We use the highest node strength
Algorithm 2 AvgRepMatrixConstruction. $W$ weights of all GNNs; $K$ is the threshold value

1: RepMatrixConstruction($W$, $K$):
2: \hspace{1em} for $w_i$ and $w_j$ in $|W|$ do // absolute value of weights is used
3: \hspace{2em} $r^K_i \leftarrow$ Top $K$ features from $w_i$
4: \hspace{2em} $r^K_j \leftarrow$ Top $K$ features from $w_j$
5: \hspace{1em} $R^K_{ij} \leftarrow \frac{|r^K_i \cap r^K_j|}{K}$
6: \hspace{1em} return $R^K$
7: Execute:
8: \hspace{1em} initialize $\bar{R} \in \mathbb{R}^{M \times M}$ with zeros
9: \hspace{1em} for each hospital weights $W_h$ where $h$ in $\{1, \ldots, H\}$ do
10: \hspace{2em} $\bar{R} \leftarrow \bar{R} +$ RepMatrixConstruction($W_h$, $K$)
11: \hspace{1em} $\bar{R} \leftarrow \frac{\bar{R}}{H}$

to identify the most reproducible global federated model (Definition 2). In fact, such a hub GNN node implies a maximal overlap with other GNN models, thereby evidencing its reproducible power. Next, we find the most reproducible $K$ biomarkers with the highest weights learned by the most reproducible GNN model.

**Definition 2.** Given $M$ GNN models to federate, let $R \in \mathbb{R}^{M \times M}$ denote the constructed reproducibility matrix where each element encodes the intersection rate of the top $K$ biomarkers identify by pairs of global GNN models. Let $r_i$ be the $i^{th}$ row of $R$ where $i \in \{1, 2, \ldots, M\}$. The $r_i$ includes the top $K$ biomarkers intersection ratios of $GNN_i$ with all GNN models including itself. Let $s_i$ denote the strength (i.e., score) of $GNN_i$ defined as: $s_i = (\sum_{m=1}^{M} r_{i,m}) - 1$ (minus one is for excluding the relation with itself).

## 3 Results and Discussion

**Evaluation of biomedical image datasets.** We evaluated our federated reproducibility framework on two large-scale biomedical image datasets which are retrieved from MedMNIST\(^1\) public dataset collection [14]. The first biomedical image dataset (PneumoniaMNIST dataset) contains 5856 X-ray images, with a size of $28 \times 28$, and belonging to a normal control class or displaying pneumonia which is a respiratory infection that affects the lungs [15]. Out of the 5856 subjects, we randomly selected 1000 samples with balanced classes (normal and pneumonia). The second dataset (BreastMNIST dataset) contains 780 breast ultrasound images, with the size of $28 \times 28$, belonging to a normal control or diagnosed with malignant breast cancer. We randomly sampled 546 subjects where 399 subjects are labeled as normal and 147 as malignant. We used two different representations of the imaging datasets to feed into the models. In the first representation, we simply fed the original image to the target GNN whereas in

\(^1\) [https://medmnist.com/](https://medmnist.com/)
Fig. 2: Accuracy and reproducibility score comparison of image-based and graph-based representations (biomedical image datasets only). Bar charts show the accuracy comparison for both input representations using DiffPool and GCN models. 1) Graph Input | the column shows the average reproducibility matrices of baseline and federated models when the input type is graph. 2) Image Input | the column shows the average reproducibility matrices of baseline and federated models when the input type is image.
the second representation we converted each image into weighted graph matrix. The weights of connectivity matrix were calculated using absolute differences in intensity between pixel pairs.

**Evaluation of connectomic datasets.** Additionally, we used the Autism Brain Imaging Data Exchange (ABIDE I) public dataset [16] to evaluate our federated reproducibility framework on morphological brain networks [17]. We used the left and right hemisphere brain connectivity datasets of autism spectrum disorder (ASD) and normal controls (NC). These datasets include 300 brain graphs with balanced classes. Both left and right hemispheres are parcelled into 35 regions of interest (ROIs) using Desikan-Killiany Atlas [18] and FreeSurfer [19] software. The connectivity weight encodes the average morphological dissimilarity in cortical thickness between two cortical ROIs as introduced in [17,20].

**Pool of GNNs.** For our federated reproducibility framework, we used 2 state-of-the-art GNNs which are DiffPool [21] and GCN [22]. DiffPool includes a differentiable graph pooling module that is able to generate hierarchical representations of a given graph. Soft cluster assignments learned by DiffPool at each layer of GNN [21] to capture the graph nested modularity. The original aim of GCN is to perform node classification. However, we adapted the original GCN to handle whole-graph-based classification as in [8]. The code of [8] was used to develop our framework.

**Training settings and hyperparameters.** To train models in a federated manner, we divided each dataset into $H = 3$ local (independent) sets. We also divided each local data into 3-folds where two folds are used for training and the left one for testing. We selected all of the learning rates empirically. For DiffPool, the learning rate is set to $10^{-4}$ across all datasets. For GCN, the selected learning rates are $10^{-6}, 10^{-5}, 10^{-5}$ and $5 \times 10^{-6}$ for the datasets PneumoniaMNIST, BreastMNIST, ASD/NC LH and ASD/NC RH, respectively. The threshold value $K$ for the top features is set to 20 in our experiment. The epoch size $E$ is fixed to 100 and batch size $B$ is set to 1. The number of communication rounds $C$ is set to 5.

**Model accuracy and reproducibility evaluation.** We compared our federated reproducibility framework to the non-federated technique (without using Algorithm 1). The comparison was performed for both validation accuracies and average reproducibility matrices storing the intersection ratio of the top $K$ discriminative biomarkers between global GNN models. Fig 3 shows the comparison results of the classification accuracy and reproducibility matrices for two biomedical image datasets and two connectomic datasets. Notably, the classification accuracy was boosted across all datasets for each local model using federation. For the datasets, PneumoniaMNIST, ASD/NC LH and ASD/NC RH, an increase in the GNN reproducibility score is noted. However, a slight decrease was observed when we evaluated our federated reproducibility framework with the BreastMNIST dataset. The results of biomedical image datasets displayed in Fig 3 were obtained when training GNN models on the original

2 https://github.com/basiralab/RG-Select
Fig. 3: **Accuracy and global reproducibility matrix comparison across datasets and GNN models.** Each row in the figure represents individual datasets. The first and second columns are the accuracy comparison results of DiffPool and GCN models, respectively. The third and fourth columns represent the baseline and federated reproducibility matrices, respectively.

Images directly. **Fig. 2** displays the accuracy and reproducibility score comparison of the graph and image representations of the biomedical image datasets.
Interestingly, according to Fig. 2, models performed better in terms of both accuracy and reproducibility when the original images were used without resorting to transforming them into graphs.

Most reproducible connectomic biomarkers. Fig 4 and Fig 5 shows the absolute value of the feature weights learned by the globally most reproducible GNN, which are the averages of the locally learned weights using ASD/NC LH and RH datasets, respectively. We considered the global GNN model rather than the hospital-specific local models to select the most reproducible biomarkers since the most reproducible model may change across hospitals. According to Fig 5, the insula cortex and lingual gyrus are selected as the most reproducible biomarkers for both LH and RH datasets followed by the precuneus and the inferior parietal cortex. In patients presenting with autism, the insula cortex shows an important variation in T1 according to [23]. Such finding embodies the nature of this neurodevelopmental disorder mainly characterized by altered cognitive, emotional and sensory functions. These neurological aspects of the disease are orchestrated by the insular cortex [24] pinpointing further that autism is considered an insula pathology and highlighting the reliability of such biomarker as a fingerprint of the disease [25]. [26] demonstrated a significant relationship between ASD traits and cortical thickness of the lingual gyrus. As a matter of fact, it has been linked to the specific aspect of sensory disturbances in ASD [27]. Regarding the precuneus, the medial part of the posterior parietal lobe, it has been linked to a specific clinical phenotype of ASD which is associated with psychological comorbidities, such as post-traumatic stress disorder. According to [28], the reduction in the precuneus gray matter was correlated with adverse childhood experiences leading to intrusive reexperiencing in adults with ASD. Thus, the precuneus represents a potential biomarker of the disease even more valuable since it could be phenotype-dependant. Furthermore, the almost miror effect discernible by comparing both hemispheres (Fig 3 and 4) might be explained by the heterogeneity of the sample with patients’ age ranging from 5 to 64 years (mean age of onset = 14 years). It pinpoints the evolving aspect of the morphological abnormalities over time going from being primarily left-lateralized to inter-hemispheric differences diminishing progressively when reaching adulthood [29].
Fig. 4: The learned weights of the cortical regions by the most reproducible GNN model for the dataset ASD/NC LH.

Fig. 5: The learned GNN weights of the cortical regions by the most reproducible GNN model for the dataset ASD/NC RH.

Limitations and future directions. Even though we used different datasets to evaluate our federated reproducibility framework, it has several limitations. First, we assumed that each local hospital has almost the same number of samples—which might not be the case for decentralized medical datasets. Second, we only used 2 different GNNs. In our future work, we aim to optimize our hyper-parameters using advanced methods, use an early stopping technique, consider
imbalanced data distributions across hospitals and extend the pool of GNNs to obtain more results for an enhanced comparison and generalizability. Incorporating clinical features of patients such as a detailed assessment of cognition, sensory disturbances and the presence of comorbidities may help add phenotypic value to the already established biomarkers of the ASD in our study.

4 Conclusion

In this paper, we investigated and quantified the reproducibility of GNN models trained in a federated manner. We evaluated our federated reproducibility framework using several medical imaging and connectomic datasets. Our framework aims to calculate the most reproducible biomarkers or features while handling decentralized datasets and boosting the local model accuracies. In this prime work, we showed that federated learning not only increases the performance of locally trained GNN models but also boosts their reproducibility. In our future work, we will investigate federated GNN reproducibility when learning on non-IID clinical datasets and examine other state-of-the-art GNN models.

5 Supplementary material

We provide three supplementary items for reproducible and open science:

1. A 7-min YouTube video explaining how our framework works on BASIRA YouTube channel at https://youtu.be/pnattjt981k.
2. Our code in Python on GitHub at https://github.com/basiralab/reproducibleFedGNN.
3. A GitHub video code demo on BASIRA YouTube channel at https://youtu.be/bG54z0v75U0.

6 Acknowledgements

This work was funded by generous grants from the European H2020 Marie Skłodowska-Curie action (grant no. 101003403, http://basira-lab.com/normnets/) to I.R. and the Scientific and Technological Research Council of Turkey to I.R. under the TUBITAK 2232 Fellowship for Outstanding Researchers (no. 118C288, http://basira-lab.com/reprime/). However, all scientific contributions made in this project are owned and approved solely by the authors.

References

1. Lee, J.G., Jun, S., Cho, Y.W., Lee, H., Kim, G.B., Seo, J.B., Kim, N.: Deep learning in medical imaging: general overview. Korean journal of radiology 18 (2017) 570–584
2. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual review of biomedical engineering 19 (2017) 221
3. WOLTERINK, J., SUK, J.: Geometric deep learning for precision medicine. (KEY ENABLING TECHNOLOGY FOR SCIENTIFIC MACHINE LEARNING) 60
4. Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P.: Geometric deep learning: going beyond euclidean data. IEEE Signal Processing Magazine 34 (2017) 18–42
5. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Philip, S.Y.: A comprehensive survey on graph neural networks. IEEE transactions on neural networks and learning systems 32 (2020) 4–24
6. Bessadok, A., Mahjoub, M.A., Rekik, I.: Graph neural networks in network neuroscience. arXiv preprint arXiv:2106.03535 (2021)
7. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics, PMLR (2017) 1273–1282
8. Nebli, A., Gharsallaoui, M.A., Gürler, Z., Rekik, I., Initiative, A.D.N., et al.: Quantifying the reproducibility of graph neural networks using multigraph data representation. Neural Networks 148 (2022) 254–265
9. Georges, N., Mhiri, I., Rekik, I., Initiative, A.D.N., et al.: Identifying the best data-driven feature selection method for boosting reproducibility in classification tasks. Pattern Recognition 101 (2020) 107183
10. Georges, N., Rekik, I., et al.: Data-specific feature selection method identification for most reproducible connectomic feature discovery fingerprinting brain states. International Workshop on Connectomics in Neuroimaging (2018) 99–106
11. Forcier, M.B., Gallois, H., Mullan, S., Joly, Y.: Integrating artificial intelligence into health care through data access: can the gdpr act as a beacon for policymakers? Journal of Law and the Biosciences 6 (2019) 317
12. Chen, C., Hu, W., Xu, Z., Zheng, Z.: Fedgl: federated graph learning framework with global self-supervision. arXiv preprint arXiv:2105.03170 (2021)
13. He, C., Balasubramanian, K., Ceyani, E., Yang, C., Xie, H., Sun, L., He, L., Yang, L., Yu, P.S., Rong, Y., et al.: Fedgraphnn: A federated learning benchmark system for graph neural networks. In: ICLR 2021 Workshop on Distributed and Private Machine Learning (DPML). (2021)
14. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2: A large-scale lightweight benchmark for 2d and 3d biomedical image classification. arXiv preprint arXiv:2110.14795 (2021)
15. Gereige, R.S., Laufer, P.M.: Pneumonia. Pediatrics in Review 34 (2013) 438–456
16. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry 19 (2014) 659–667
17. Soussia, M., Rekik, I.: Unsupervised manifold learning using high-order morphological brain networks derived from t1-w mri for autism diagnosis. Frontiers in neuroinformatics 12 (2018) 70
18. Fischl, B., Salat, D.H., Van Der Kouwe, A.J., Makris, N., Ségonne, F., Quinn, B.T., Dale, A.M.: Sequence-independent segmentation of magnetic resonance images. Neuroimage 23 (2004) S69–S84
19. Fischl, B.: Freesurfer. Neuroimage 62 (2012) 774–781
20. Mahjoub, I., Mahjoub, M.A., Rekik, I.: Brain multiplexes reveal morphological connectional biomarkers fingerprinting late brain dementia states. Scientific reports 8 (2018) 1–14
21. Ying, R., You, J., Morris, C., Ren, X., Hamilton, W.L., Leskovec, J.: Hierarchical graph representation learning with differentiable pooling. arXiv preprint arXiv:1806.08804 (2018)
22. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
23. Lou, B., Jiang, Y., Li, C., Wu, P.Y., Li, S., Qin, B., Chen, H., Wang, R., Wu, B., Chen, M.: Quantitative analysis of synthetic magnetic resonance imaging in alzheimer’s disease. Frontiers in aging neuroscience 13 (2021) 638731
24. Gasquoine, P.G.: Contributions of the insula to cognition and emotion. Neuropsychology review 24 (2014) 77–87
25. Nomi, J.S., Molnar-Szakacs, I., Uddin, L.Q.: Insular function in autism: Update and future directions in neuroimaging and interventions. Progress in Neuro-Psychopharmacology and Biological Psychiatry 89 (2019) 412–426
26. Gebauer, L., Foster, N.E., Vuust, P., Hyde, K.L.: Is there a bit of autism in all of us? autism spectrum traits are related to cortical thickness differences in both autism and typical development. Research in Autism Spectrum Disorders 13 (2015) 8–14
27. Habata, K., Cheong, Y., Kamiya, T., Shiotsu, D., Omori, I.M., Okazawa, H., Jung, M., Kosaka, H.: Relationship between sensory characteristics and cortical thickness/volume in autism spectrum disorders. Translational psychiatry 11 (2021) 1–7
28. Kitamura, S., Makinodan, M., Matsuoka, K., Takahashi, M., Yoshikawa, H., Ishida, R., Kishimoto, N., Yasuno, F., Yasuda, Y., Hashimoto, R., et al.: Association of adverse childhood experiences and precuneus volume with intrusive reexperiencing in autism spectrum disorder. Autism Research 14 (2021) 1886–1895
29. Khundrakpam, B.S., Lewis, J.D., Kostopoulos, P., Carbonell, F., Evans, A.C.: Cortical thickness abnormalities in autism spectrum disorders through late childhood, adolescence, and adulthood: a large-scale mri study. Cerebral Cortex 27 (2017) 1721–1731