

Semi-Supervised Graph Learning for Brain Disease Identification

Kunpeng Zhang, Yining Zhang*, Xueyan Liu

School of Mathematics Science, Liaocheng University, Liaocheng, China

Email: *zhangyining@lcu.edu.cn

How to cite this paper: Zhang, K.P., Zhang, Y.N. and Liu, X.Y. (2023) Semi-Supervised Graph Learning for Brain Disease Identification. *Journal of Applied Mathematics and Physics*, 11, 1846-1859.

<https://doi.org/10.4236/jamp.2023.117119>

Received: June 18, 2023

Accepted: July 15, 2023

Published: July 18, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Using resting-state functional magnetic resonance imaging (fMRI) technology to assist in identifying brain diseases has great potential. In the identification of brain diseases, graph-based models have been widely used, where graph represents the similarity between patients or brain regions of interest. In these models, constructing high-quality graphs is of paramount importance. Researchers have proposed various methods for constructing graphs from different perspectives, among which the simplest and most popular one is Pearson Correlation (PC). Although existing methods have achieved significant results, these graphs are usually fixed once they are constructed, and are generally operated separately from downstream task. Such a separation may result in neither the constructed graph nor the extracted features being ideal. To solve this problem, we use the graph-optimized locality preserving projection algorithm to extract features and the population graph simultaneously, aiming in higher identification accuracy through a task-dependent automatic optimization of the graph. At the same time, we incorporate supervised information to enable more flexible modelling. Specifically, the proposed method first uses PC to construct graph as the initial feature for each subject. Then, the projection matrix and graph are iteratively optimized through graph-optimization locality preserving projections based on semi-supervised learning, which fully employs the knowledge in various transformation spaces. Finally, the obtained projection matrix is applied to construct the subject-level graph and perform classification using support vector machines. To verify the effectiveness of the proposed method, we conduct experiments to identify subjects with mild cognitive impairment (MCI) and Autism spectrum disorder (ASD) from normal controls (NCs), and the results showed that the classification performance of our method is better than that of the baseline method.

Keywords

Graph Learning, Mild Cognitive Impairment, Autism Spectrum Disorder

1. Introduction

The global incidence of brain diseases, such as autism spectrum disorder and Alzheimer's disease, has been increasing in recent years, which poses a huge burden to both families and society [1]. Up to now, researchers have not found a strategy that can fully treat most of the brain disorders, and it is especially important to diagnose and intervene brain disorders in their early stages, which can greatly alleviate the progression of the disease [2] [3] [4].

Utilizing resting-state functional magnetic resonance imaging (rs-fMRI) to study brain activity during rest has become an independent field of research, due to its non-invasive, operable, and easily shareable advantages, as well as its potential to serve as the biomarker for psychiatric disorders [5]. Due to the ubiquity of graphs as a common data structure, graph-based modeling has gained significant [6] [7] [8]. In recent years, using graphs to study brain activity has become increasingly popular, and the construction of high-quality graphs is the key to the graph modeling process. However, the presence of artifacts, structured noise, and our limited understanding of the brain make the construction and evaluation of graphs in the context of brain diseases a challenging and unresolved problem [9] [10]. Additionally, obtaining excellent labeling in medical data is often expensive and time-consuming process, leading to a scarcity of labeled data. To address this issue, semi-supervised learning has emerged as an effective approach, allowing for the training of generalizable models using a small amount of labeled data and a large amount of unlabeled data [11] [12] [13]. In this paper, we primarily focus on studying the estimation of edges in subject-level graphs within the framework of semi-supervised learning.

In the exploration of graph estimation methods, various approaches have been proposed and attempted. One popular method is the utilization of similarity measures, such as Pearson's correlation (PC) [14], partial correlation [15] [16], Cosine similarity [17], and others. These similarity measures offer convenience and efficiency in constructing graphs directly. However, they often overlook important factors such as data distribution, which can result in suboptimal outcomes. As a result, researchers are increasingly inclined towards data-driven methods for graph learning, such as metric learning [18] [19], which aim to capture underlying patterns in the data. In addition, some methods incorporate regularized extensions [9] [20] [21] [22] to enhance graph learning performance by reducing node dominance. Of course, these regularization constraints require prior knowledge that fits the scene, leading to more realistic graph, example include SR [23], group sparsity [24], modularity [20] [25], and so on. The combination of data fitting term and regularized term can be unified within the matrix-regularized graph learning framework [26], providing a flexible platform to incorporate various graph learning methods.

However, the graphs constructed by these methods are fixed and often not linked to downstream tasks, which can result in both unsatisfactory constructed graph and extracted features, affecting the final performance and effect. In addi-

tion, when using certain methods, a considerable amount of manual adjustments and optimization are required to match the constructed graph with downstream tasks. To address this limitation, we propose a novel adaptive graph estimation method for brain diseases that jointly learns the graph and projection matrix [27]. Specifically, we first construct the graph using PC as the initial feature of each subject. In addition, we employ locality preserving projections [28] to construct the initial graph, then we label connections between same-class subjects as 1 and between different-class subjects as 0 by constraining the loss function. Furthermore, we iteratively optimize the learning of the projection matrix and graph, maximizing the utilization of these information in different transformation spaces. Finally, we leverage the learned projection matrix to construct subject-level adaptive graph and utilize the support vector machine (SVM) for classification. To verify the effectiveness of the method, we conducted experiments on the MCI and ASD datasets to identify subjects with diseases from the healthy control group. The experimental results show that our method significantly outperforms the baseline methods in terms of classification performance.

In the subsequent sections, this paper will further elaborate on the methods, experiments, and conclusions in order. The method section begins by introducing the data source and describing the preprocessing steps. It then proceeds to review the existing baseline works before presenting our proposed method, including the underlying motivation, model, and algorithm. Moving on to the experiment section, we conducted identification experiments on MCI and ASD to validate the feasibility of our method and compare it with the baseline methods. Finally, in the conclusion section, a concise summary of the paper is provided, along with a discussion on potential future research directions.

2. Materials and Methods

2.1. Data Preparation

This study utilized two benchmark databases to validate the effectiveness of the proposed method, including the publicly available ADNI database and ABIDE database with rs-fMRI data. For the ADNI database, we selected 137 subjects (68 MCIs and 69 NCs) and preprocessed the data according to the most recent study [29]. For the ABIDE database, we used data from 184 participants (79 ASD and 105 NC) from the largest site (NYU) and preprocessed the data using the Data Processing Assistant for Resting-State fMRI (DPARSF) [30]. In **Table 1**, we present some clinical and demographic characteristics of the subjects, such as gender, age, etc. Please note that the subject information shown in **Table 1** conforms to the general inclusion/exclusion criteria of the ADNI dataset, which can be briefly summarized as follows: 1) NC subjects: Mini-Mental State Examination (MMSE) scores between 20 and 30 (inclusive), Clinical Dementia Rating (CDR) of 0, non-depressed, non-MCI, and non-demented; 2) MCI subjects: MMSE scores between 24 and 30 (inclusive), memory impairment, CDR of 0.5, no significant impairment in other cognitive domains, intact activities of daily living, and non-dementia.

Table 1. Demographic and clinical information of subjects in the ADNI and ABIDE datasets. Values are reported as mean \pm standard deviation. M/F: Male/Female; MMSE: Mini-Mental Examination; GCDR: Global Clinical Dementia Rating; FIQ: Full-Scale Intelligence Quotient; VIQ: Verbal Intelligence Quotient; PIQ: Performance Intelligence Quotient.

| Datasets | Class | Gender (M/F) | Age (Years) | MMSE | GCDR | FIQ | VIQ | PIQ |
|----------|-------|--------------|-------------------|------------------|-----------------|-------------------|-------------------|-------------------|
| ADNI | MCI | 39/29 | 76.50 \pm 13.50 | 26.77 \pm 1.23 | 0.48 \pm 0.02 | - | - | - |
| | NC | 17/52 | 71.50 \pm 14.50 | 28.85 \pm 1.15 | 0 | - | - | - |
| ABIDE | ASD | 68/11 | 18.58 \pm 11.45 | - | - | 107.92 \pm 3.15 | 105.81 \pm 1.23 | 108.81 \pm 2.10 |
| | NC | 79/26 | 19.13 \pm 11.85 | - | - | 113.15 \pm 2.45 | 113.13 \pm 1.15 | 115.07 \pm 2.08 |

For the preprocessing pipeline of each subject in ADNI dataset, the scanning time was 7 min, corresponding to 140 volumes. To address errors caused by magnetic field and signal instability, we removed the first 10 volumes. Then, we performed motion correction and calculated frame-wise displacement (FD) based on head motion parameters, and the subjects with more than 2.5 min of FD larger than 0.5 mm were excluded from the dataset. To reduce the influence of signal from head motion parameters, white matter, and cerebrospinal fluid (CSF), we adopted nuisance regression based on the Friston 24-parameter model. Subsequently, the corrected images were registered to the standard Montreal Neurological Institute (MNI) space and underwent spatial smoothing with band-pass temporal filtering (0.015 - 0.150 Hz). Finally, we divided the brain into 116 regions of interest (ROIs) based on AAL atlas [31]. It is worth noting that we used the AAL atlas for ROIs segmentation mainly because of its popularity and simplicity.

For the ABIDE dataset, all fMRI data were acquired on a standard echo-planar imaging sequence using a clinical routine 3.0 Tesla Allegra scanner. 184 subjects (including 79 ASDs and 105 NCs) from the largest site (*i.e.*, NYU) were used in our study. The imaging parameters are as follows: TR/TE is 2000/15 ms with 180 volumes, the number of slices is 33, and the slice thickness is 4.0 mm. The preprocessing pipeline mainly consisted of four steps: 1) volume slice and head motion correction, 2) interference signal regression (including ventricle, white matter signal, and the high-order effects of head motion described by the Friston 24-parameter model), 3) registration to MNI space, and 4) temporal filtering (0.01 - 0.10 Hz). Finally, the brain was segmented into 116 ROIs based on the AAL atlas.

2.2. Baseline Method

2.2.1. Pearson Correlation

The Pearson correlation is the simplest and widely adopted method for constructing graph, and in this study, we used it as a baseline task to develop our work. Assuming the features of each subject is represented by $x_i \in \mathbb{R}^{t \times 1}$, where t is the number of the features. The graph edge values based on PC can be defined by the following formula:

$$w_{ij}^{(PC)} = \frac{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^\top (\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^\top (\mathbf{x}_i - \bar{\mathbf{x}}_i)} \sqrt{(\mathbf{x}_j - \bar{\mathbf{x}}_j)^\top (\mathbf{x}_j - \bar{\mathbf{x}}_j)}} \quad (1)$$

where $\bar{\mathbf{x}}_i \in R^{1 \times l}$ is the mean vector of \mathbf{x}_i .

If \mathbf{x}_i is centered by $\mathbf{x}_i - \bar{\mathbf{x}}_i$ and standardized by $\sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^\top (\mathbf{x}_i - \bar{\mathbf{x}}_i)}$ denoted as \mathbf{x}_i , then the above formula can be expressed as an inner product form:

$$W^{(PC)} = X^\top X \quad (2)$$

here, $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r] \in R^{l \times r}$ is the preprocessed data matrix.

2.2.2. Spares Representation

SR is one of the commonly-used methods for calculating partial correlation. The mathematical model for SR is expressed as follows:

$$\min_{w_{ij}} \sum_{i=1}^n \left(\left\| \mathbf{x}_i - \sum_{i \neq j} w_{ij} \mathbf{x}_j \right\|^2 + \delta \sum_{i \neq j} |w_{ij}| \right), \quad (3)$$

$$\text{s.t. } w_{ii} = 0, \forall i = 1, \dots, n$$

which can be further rewritten by the following matrix form:

$$\min_W \|X - XW\|_F^2 + \delta \|W\|_1, \quad (4)$$

$$\text{s.t. } w_{ii} = 0, \forall i = 1, \dots, n$$

where $\|X - XW\|_F^2$ is a data fitting term for capturing the partial correction information, $\|W\|_1$ is an l_1 -regularized term for obtaining sparse solutions of W , and δ is a regularization parameter for controlling the balance between these two terms. Note that the constraint $w_{ii} = 0$ is used here to avoid the trivial solution (*i.e.*, $W = I$, the identity matrix) by implicitly removing x_i from X .

2.3. The Proposed Method for Graph Construction

2.3.1. Motivation

As mentioned earlier, the quality of the graph we construct is crucial for successful classification task. However, due to the lack of fundamental facts, we can only explore the real graph through research. Although previous methods have achieved good results in constructing graph, they are usually fixed once constructed and are generally separated from subsequent downstream task. This separation may lead to suboptimal graph, and extracted features are not ideal. Moreover, some methods cannot naturally integrate supervised information, which may limit application of the method in certain areas such as medicine. In **Figure 1**, we summarize the basic motivations and ideas.

To address the aforementioned issues, this paper utilizes subject-level graphs in the alternating optimization process, allowing the projection matrix to fully utilize the potential relationships between subjects in the graph, and synchronously learn with graph optimization. Such projection matrix constructs the population graph that merges the subsequent feature extraction and is closely linked

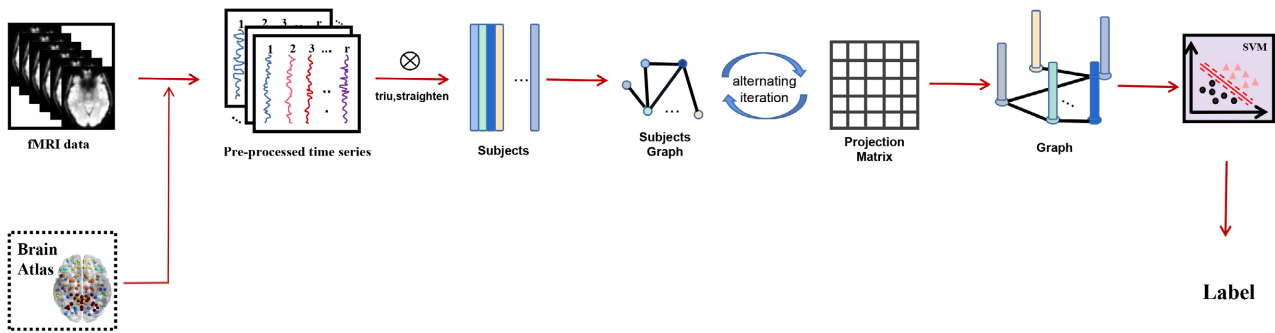


Figure 1. Most popular method (*i.e.*, PC) use the inner product of the preprocessed data matrix to obtain the graph without considering distribution of the data. In contrast, our method obtains the adaptive graph that fits the data by alternately optimising the projection matrix and graph. First, we perform PC to construct graph as the initial feature for each subject. Then, we construct a subject-level graph and use the locally preserved projection algorithm to alternately optimize the graph and projection matrix. Finally, we calculate the adaptive graph using the projection matrix, and apply the node features to classify subjects through the support vector machine classifier (SVM). It is worth noting that we used the AAL atlas for ROIs segmentation mainly because of its popularity and simplicity.

to the downstream task. At the same time, in order to balance the uniformity (or smoothness) and locality of the graph, we regularize the objective function by adding an entropy term. Then, we introduce a pairwise constraint (in the form of must-link/cannot-link constraints) to integrate supervised information into the joint optimization objective, for the construction of subject-level graphs and calculation of the projection matrix, thereby achieving more flexible modeling. This will be detailed in the following sections on algorithms and models.

2.3.2. Model and Algorithm

Given a set of data points $X = [x_1, x_2, \dots, x_r]$, $x_i \in R^D$, D is the number of upper triangular elements in the data calculated using Pearson correlation. The paired-constraints linking constraint

$M = \{(x_i, x_j) | x_i \text{ and } x_j \text{ belong to the same class}\}$ and the paired-constraints non-linking constraint $C = \{(x_i, x_j) | x_i \text{ and } x_j \text{ belong to different class}\}$ cannot be linked. The goal of semi-supervised GoLPP is to optimize both the graph and projection directions within a unified framework while maintaining paired constraints. The objective function is defined as follows:

$$\min_{W, S_{ij}} \frac{\sum_{i,j=1}^n \|W^T x_i - W^T x_j\|^2 S_{ij}}{\sum_{j=1}^n \|W^T x_i\|^2} + \varphi \sum_{i,j=1}^n S_{ij} \ln(S_{ij}/\alpha),$$

$$\text{s.t. } \sum_{j=1}^n S_{ij} > 0, \forall i, j = 1, \dots, n$$

$$S_{ij} = 1, \text{ if } (x_i, x_j) \in M, \forall i, j = 1, \dots, n \tag{5}$$

$$S_{ij} = 0, \text{ if } (x_i, x_j) \in C, \forall i, j = 1, \dots, n$$

$$S_{ij} > 0, \text{ otherwise}$$

Although the objective function is not convex, it can be easily solved through

Alternating Optimization (AO). The iterative process includes the following two main steps:

Step 1: Initialize the weights for the graph. For example, a simple initialization method is shown below:

$$S_{ij} = \begin{cases} 1, & \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in M \\ 0, & \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in C \\ \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / \sigma), & \text{otherwise} \end{cases} \quad (6)$$

Then, the optimal projection matrix is obtained using a generalized eigenvalue problem. In fact, the solution obtained in this way is not precise, as it involves ratio tracing and ratio tracking problems, which are beyond the scope of our main focus. For more details, please refer to [32].

Step 2: Revise $W = \bar{W}$ and minimize the objective function, and we can obtain the following problem:

$$\begin{aligned} \min_{S_{ij}} \quad & \sum_{i,j=1}^n \|W^T \mathbf{x}_i - W^T \mathbf{x}_j\|^2 S_{ij} + \varphi \sum_{i,j=1}^n S_{ij} \ln(S_{ij}/\alpha), \\ \text{s.t.} \quad & \sum_{j=1}^n S_{ij} > 0, \forall i, j = 1, \dots, n \\ & S_{ij} = 1, \text{ if } (\mathbf{x}_i, \mathbf{x}_j) \in M, \forall i, j = 1, \dots, n \\ & S_{ij} = 0, \text{ if } (\mathbf{x}_i, \mathbf{x}_j) \in C, \forall i, j = 1, \dots, n \\ & S_{ij} > 0, \text{ otherwise} \end{aligned} \quad (7)$$

In fact, we only need to consider the following problem first, and then add the part of equality constraints for S_{ij} :

$$\begin{aligned} \min_{S_{ij}} \quad & \sum_{i,j=1}^n \|W^T \mathbf{x}_i - W^T \mathbf{x}_j\|^2 S_{ij} + \varphi \sum_{i,j=1}^n S_{ij} \ln(S_{ij}/\alpha), \\ \text{s.t.} \quad & \sum_{j=1}^n S_{ij} > 0, \forall i, j = 1, \dots, n \\ & S_{ij} > 0, \text{ otherwise} \end{aligned} \quad (8)$$

where $(\mathbf{x}_i, \mathbf{x}_j) \notin M$ and $(\mathbf{x}_i, \mathbf{x}_j) \notin C$. Then, letting

$\frac{\partial L}{\partial S_{ij}} = \|W^T \mathbf{x}_i - W^T \mathbf{x}_j\|^2 + \varphi(\ln(S_{ij}/\alpha) + 1) = 0$, We can obtain:

$$S_{ij} = \alpha \cdot \exp(-1) \cdot \exp\left(-\|W^T \mathbf{x}_i - W^T \mathbf{x}_j\|^2 / \varphi\right) \quad (9)$$

α is a positive parameter used only to obtain a more compact and desirable solution under the new constraint. When the distance between two samples approaches 0, we expect the weight S_{ij} to approach 1. We set $\alpha = e$, then we get:

$$S_{ij} = \exp\left(-\|W^T \mathbf{x}_i - W^T \mathbf{x}_j\|^2 / \varphi\right) \quad (10)$$

Finally, when we add the equality constraints, we have:

$$\bar{S}_{ij} = \begin{cases} 1, & \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in M \\ 0, & \text{if } (\mathbf{x}_i, \mathbf{x}_j) \in C \\ \exp\left(-\|W^T \mathbf{x}_i - W^T \mathbf{x}_j\|^2 / \sigma\right), & \text{otherwise} \end{cases} \quad (11)$$

Finally, we update S_{ij} using \bar{S}_{ij} and continue to iterate until convergence.

We use this approach to continuously update the construction of the graph and projection matrix, so as to learn important information between subjects in different transformation spaces, such as subject category information. This information is implicitly contained in the projection matrix, so that when the projection matrix is used to map data in the future, it can better distinguish between subjects. In addition, it implies the possibility that two samples belong to the same class. Specifically, if two samples come from the same class, the edge weight is 1; if they come from different classes, it is $\exp\left(-\|W^T \mathbf{x}_i - W^T \mathbf{x}_j\|^2 / \sigma\right)$, if they are unlabeled, then the edge weight lies between 0 and 1. Specifically, when two samples are very close to each other, this value tends to 1; conversely, when samples are far apart from each other, it tends to 0.

3. Experiments and Results

3.1. Experimental Setting

We evaluated our proposed method on the MCI and ASD datasets using a leave-one-out (LOO) cross-validation strategy. The performance was measured using four metrics: accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under the receiver operating characteristic curve (AUC). They intuitively display the overall recognition accuracy, patient recognition accuracy, normal recognition accuracy and confidence level of the model. Here are the definitions for these metrics:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (12)$$

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (14)$$

here, TP represents the number of true positive subjects predicted, FN represents the number of false negative subjects predicted, and similarly, TN and FP represent the number of their corresponding experimental subjects.

3.2. Competing Method

We compared the proposed the performance of our proposed method with with the baseline methods: PC, SR, the unsupervised form of our method (un-ours) and GCN. Next, we will explain some parameter settings. PC has no parameters,

the sparse parameter of SR is selected as the optimal parameter in the range of $[2^{-4}, 2^{-2}, 2^0, 2^2, 2^4]$, and feature extraction uses t-test with a p-value set to 0.01. GCN performs node classification, and the initial graph is constructed using PC. The design of GCN is the same as the original paper [33], with the following partly parameter settings: lr is 0.01, weight decay is $5e-4$, and dropout is 0.5. In our method, the parameter settings are the same for unsupervised and semi-supervised. The regularization parameter defaults to 1, and the maximum iteration defaults to 100.

It is worth noting that our method does not perform feature extraction using t-test because the projection matrix learned during the process of learning the graph can map the data to the optimal position without the need for feature extraction. At the same time, we use thresholding to ensure a certain level of sparsity in the brain constructed by our method, and we discuss the impact of different levels of sparsity on our method. In our method, To simulate the lack of labeled data, we assume that 10% of the training samples are labeled and used to create must-link/cannot-link constraints depending on whether these samples belong to the same class. In this paper, we use leave-one-out cross-validation to split the data. Finally, we use SVM to predict the labels and calculate four evaluation metrics, namely accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the curve (AUC).

3.3. Result

As shown in **Table 2**, we have listed the results obtained by each method in both datasets. Also, we visualized the results obtained by the comparison method and the proposed method in **Figure 2**, showing the disease classification results of ASD in the left graph and the disease classification results of MCI in the right graph. We can obtain the following results, where our proposed model clearly

Table 2. Classification results under four performance indicators for five different methods.

| Dataset | Method | ACC | SEN | SPE | ROC |
|---------|---------|---------------|---------------|---------------|---------------|
| ASD | PC | 61.41% | 58.23% | 63.81% | 69.83% |
| | SR | 66.85% | 46.84% | 81.90% | 68.61% |
| | GCN | 67.39% | 65.82% | 68.57% | 67.20% |
| | Un-ours | 69.02% | 62.03% | 74.29% | 73.14% |
| | Ours | 70.11% | 63.29% | 75.24% | 73.16% |
| MCI | PC | 78.10% | 77.94% | 78.26% | 85.68% |
| | SR | 84.67% | 85.29% | 84.06% | 91.86% |
| | GCN | 85.40% | 86.76% | 84.06% | 85.41% |
| | Un-ours | 85.40% | 83.82% | 86.96% | 89.37% |
| | Ours | 86.13% | 83.82% | 88.41% | 89.49% |

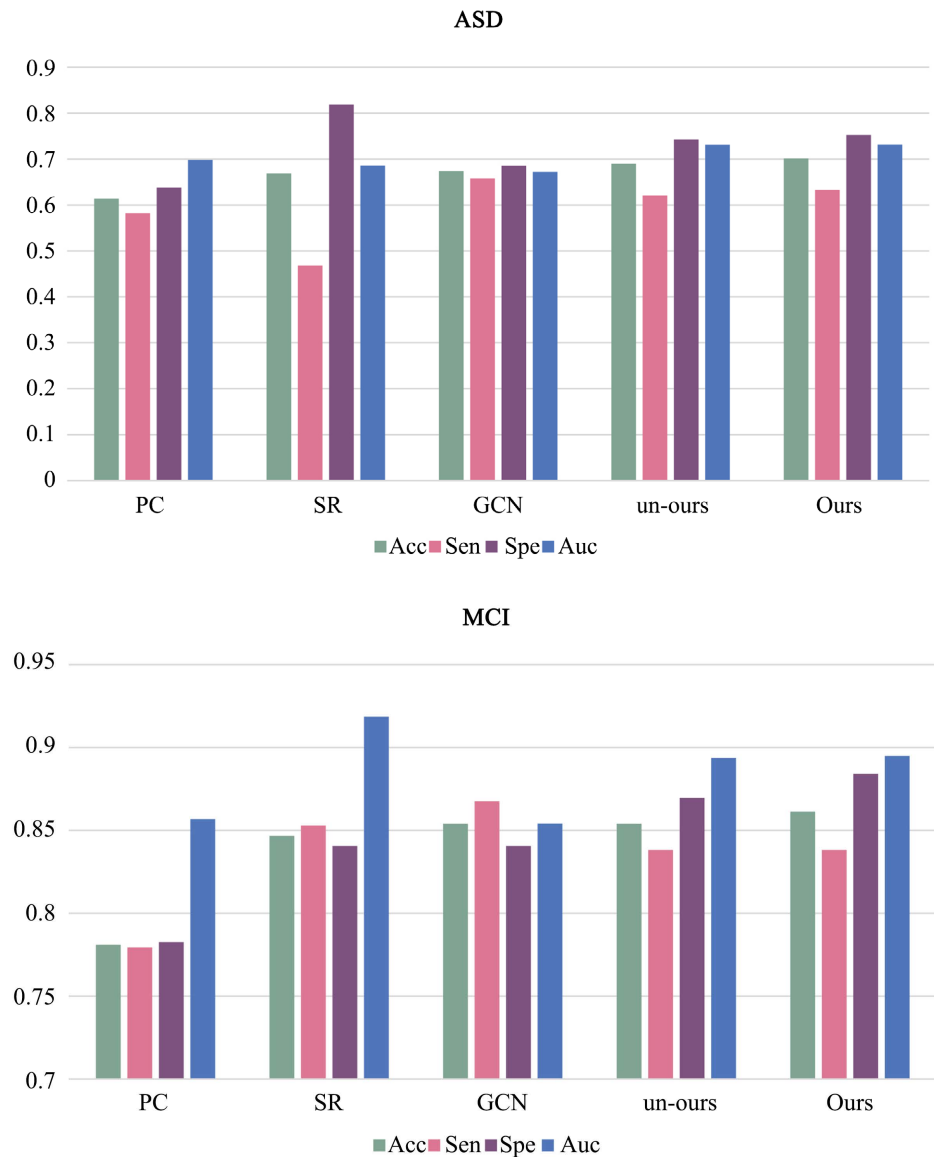


Figure 2. The ASD and MCI classification results based on five methods (*i.e.*, PC, SR, GCN, Unsupervised ours and Ours) based on the four performance indicators (*i.e.*, ACC, SEN, SPE, and AUC) using LOOCV.

outperforms the comparison method, which illustrates the benefit of considering the co-optimization of the projection matrix of the graph and feature extraction during the construction of the graph. In addition, the inclusion of appropriate supervision information also contributes to the performance improvement. In **Figure 3** we also compare the graph constructed by our method with different degrees of sparsity, and the result is that the graph no longer needs to be sparse on the ASD dataset, it has reached the optimum, and it is basically no longer needed to be sparse on the MCI dataset, which shows that the projection matrix learned by our method is able to map the data to the learned optimal space without generating similar to the PC due to noise and other reasons The weak connection due to noise, etc.

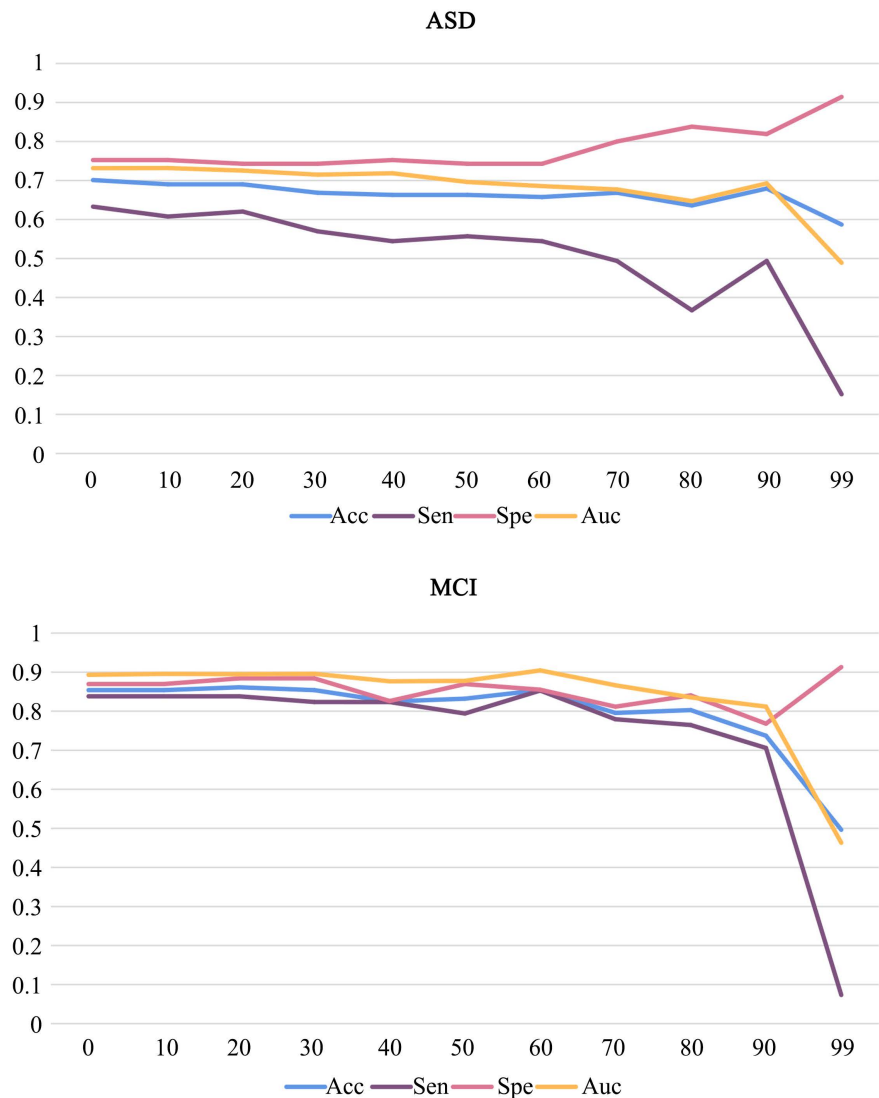


Figure 3. Implement sparseness operations on our method, with abscissa representing sparsity and 0 indicating no sparsity. It can be seen that our method does not require sparse operations, indicating that there is basically no weak connection similar to that of a PC due to noise, etc.

4. Conclusion

In this paper, we construct task-dependent adaptive graph in a semi-supervised scene using a relaxed constrained graph-optimized holdout projection, which fits well into the context of the lack of labelling in medicine. We make use of the projection matrix to map the data to the learned best space, and use the node features in the graph to classify by SVM. The experimental results show that the method achieves 86.13% and 70.11% correct rates for the MCI and ASD recognition tasks based on the leave-one-out method, respectively, and outperforms the baseline method in terms of ACC and SPE. This further illustrates the importance of task-dependent adaptive graph and reliable supervised information to improve the generalization of subsequent classifier. Our model has very few

hyperparameters, which avoids the difficulty of parameter selection, but on the other hand, it also reduces the flexibility of the model. In addition, heterogeneity between data from different sites is also a limitation. In the future, we will try to design more powerful and adaptive graph learning methods to overcome the problems of hyperparameter selection and site heterogeneity.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Zhang, Y., Guo, P., Ma, Z., Lu, P., Kebebe, D. and Liu, Z. (2021) Combination of Cell-Penetrating Peptides with Nanomaterials for the Potential Therapeutics of Central Nervous System Disorders: A Review. *Journal of Nanobiotechnology*, **19**, Article No. 255. <https://doi.org/10.1186/s12951-021-01002-3>
- [2] Vismara, L.A. and Rogers, S.J. (2010) Behavioral Treatments in Autism Spectrum Disorder: What Do We Know? *Annual Review of Clinical Psychology*, **6**, 447-468. <https://doi.org/10.1146/annurev.clinpsy.121208.131151>
- [3] Dawson, G., Jones, E.J., Merkle, K., Venema, K., Lowy, R., Faja, S., Kamara, D., Murias, M., Greenson, J., Winter, J., *et al.* (2012) Early Behavioral Intervention Is Associated with Normalized Brain Activity in Young Children with Autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, **51**, 1150-1159. <https://doi.org/10.1016/j.jaac.2012.08.018>
- [4] Alzheimer's Association (2019) 2019 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia*, **15**, 321-387. <https://doi.org/10.1016/j.jalz.2019.01.010>
- [5] Bijsterbosch, J., Smith, S.M. and Beckmann, C. (2017) An Introduction to Resting State fMRI Functional Connectivity. Oxford University Press, Oxford.
- [6] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C. and Sun, M. (2020) Graph Neural Networks: A Review of Methods and Applications. *AI Open*, **1**, 57-81. <https://doi.org/10.1016/j.aiopen.2021.01.001>
- [7] Xie, Y., Xu, Z., Zhang, J., Wang, Z. and Ji, S. (2022) Self-Supervised Learning of Graph Neural Networks: A Unified Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 2412-2429. <https://doi.org/10.1109/TPAMI.2022.3170559>
- [8] Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S. and Liu, H. (2021) Graph Learning: A Survey. *IEEE Transactions on Artificial Intelligence*, **2**, 109-127. <https://doi.org/10.1109/TAI.2021.3076021>
- [9] Jiang, X., Zhang, L., Qiao, L. and Shen, D. (2019) Estimating Functional Connectivity Networks via Low-Rank Tensor Approximation with Applications to MCI Identification. *IEEE Transactions on Biomedical Engineering*, **67**, 1912-1920. <https://doi.org/10.1109/TBME.2019.2950712>
- [10] Lurie, D.J., Kessler, D., Bassett, D.S., Betzel, R.F., Breakspear, M., Kheilholz, S., Kucyi, A., Liegeois, R., Lindquist, M.A., McIntosh, A.R., *et al.* (2020) Questions and Controversies in the Study of Time-Varying Functional Connectivity in Resting fMRI. *Network Neuroscience*, **4**, 30-69. https://doi.org/10.1162/netn_a_00116
- [11] Van Engelen, J.E. and Hoos, H.H. (2020) A Survey on Semisupervised Learning. *Machine Learning*, **109**, 373-440. <https://doi.org/10.1007/s10994-019-05855-6>

- [12] Zhou, Z.-H. and Zhou, Z.-H. (2021) Semi-Supervised Learning. In: Zhou, Z.-H., Ed., *Machine Learning*, Springer, Berlin, 315-341. https://doi.org/10.1007/978-981-15-1967-3_13
- [13] Song, Z., Yang, X., Xu, Z. and King, I. (2022) Graph-Based Semi-Supervised Learning: A Comprehensive Review. *IEEE Transactions on Neural Networks and Learning Systems*. <https://doi.org/10.1109/TNNLS.2022.3155478>
- [14] Biswal, B., Zerrin Yetkin, F., Haughton, V.M. and Hyde, J.S. (1995) Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar MRI. *Magnetic Resonance in Medicine*, **34**, 537-541. <https://doi.org/10.1002/mrm.1910340409>
- [15] Marrelec, G., Krainik, A., Duffau, H., Pelegrini-Issac, M., Lehericy, S., Doyon, J. and Benali, H. (2006) Partial Correlation for Functional Brain Interactivity Investigation in Functional MRI. *Neuroimage*, **32**, 228-237. <https://doi.org/10.1016/j.neuroimage.2005.12.057>
- [16] Salvador, R., Suckling, J., Schwarzbauer, C. and Bullmore, E. (2005) Undirected Graphs of Frequency-Dependent Functional Connectivity in Whole Brain Networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 937-946. <https://doi.org/10.1098/rstb.2005.1645>
- [17] Erkan, G. and Radev, D.R. (2004) Lexrank: Graph-Based Lexical Centrality as Saliency in Text Summarization. *Journal of Artificial Intelligence Research*, **22**, 457-479. <https://doi.org/10.1613/jair.1523>
- [18] Yang, L. and Jin, R. (2006) Distance Metric Learning: A Comprehensive Survey. Michigan State University, East Lansing.
- [19] Weinberger, K.Q. and Saul, L.K. (2009) Distance Metric Learning for Large Margin nearest Neighbor Classification. *Journal of Machine Learning Research*, **10**, 207-244.
- [20] Qiao, L., Zhang, H., Kim, M., Teng, S., Zhang, L. and Shen, D. (2016) Estimating Functional Brain Networks by Incorporating a Modularity Prior. *Neuroimage*, **141**, 399-407. <https://doi.org/10.1016/j.neuroimage.2016.07.058>
- [21] Li, W., Wang, Z., Zhang, L., Qiao, L. and Shen, D. (2017) Remodeling Pearson's Correlation for Functional Brain Network Estimation and Autism Spectrum Disorder Identification. *Frontiers in Neuroinformatics*, **11**, Article No. 55. <https://doi.org/10.3389/fninf.2017.00055>
- [22] Zhang, Y., Jiang, X., Qiao, L. and Liu, M. (2021) Modularity Guided Functional Brain Network Analysis for Early-Stage Dementia Identification. *Frontiers in Neuroscience*, **15**, Article ID: 720909. <https://doi.org/10.3389/fnins.2021.720909>
- [23] Lee, H., Lee, D.S., Kang, H., Kim, B.-N. and Chung, M.K. (2011) Sparse Brain Network Recovery under Compressed Sensing. *IEEE Transactions on Medical Imaging*, **30**, 1154-1165. <https://doi.org/10.1109/TMI.2011.2140380>
- [24] Varoquaux, G., Gramfort, A., Poline, J.-B. and Thirion, B. (2010) Brain Covariance Selection: Better Individual Functional Connectivity Models Using Population Prior. *NIPS10: Proceedings of the 23rd International Conference on Neural Information Processing Systems*, Volume 2, 2334-2342.
- [25] Sporns, O. and Betzel, R.F. (2016) Modular Brain Networks. *Annual Review of Psychology*, **67**, 613-640. <https://doi.org/10.1146/annurev-psych-122414-033634>
- [26] Qiao, L., Zhang, L., Chen, S. and Shen, D. (2018) Data-Driven Graph Construction and Graph Learning: A Review. *Neurocomputing*, **312**, 336-351. <https://doi.org/10.1016/j.neucom.2018.05.084>
- [27] Zhang, L. and Qiao, L. (2017) A Graph Optimization Method for Dimensionality Reduction with Pairwise Constraints. *International Journal of Machine Learning and*

-
- Cybernetics*, **8**, 275-281. <https://doi.org/10.1007/s13042-014-0321-6>
- [28] He, X. and Niyogi, P. (2003) Locality Preserving Projections. *NIPS'03: Proceedings of the 16th International Conference on Neural Information Processing Systems*, Vancouver, 8-13 December 2003, 153-160.
- [29] Zhou, Y., Zhang, L., Teng, S., Qiao, L. and Shen, D. (2018) Improving Sparsity and Modularity of High-Order Functional Connectivity Networks for MCI and ASD Identification. *Frontiers in Neuroscience*, **12**, Article No. 959. <https://doi.org/10.3389/fnins.2018.00959>
- [30] Yan, C.-G., Wang, X.-D., Zuo, X.-N. and Zang, Y.-F. (2016) Dpabi: Data Processing & Analysis for (Resting-State) Brain Imaging. *Neuroinformatics*, **14**, 339-351. <https://doi.org/10.1007/s12021-016-9299-4>
- [31] Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B. and Joliot, M. (2002) Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *Neuroimage*, **15**, 273-289. <https://doi.org/10.1006/nimg.2001.0978>
- [32] Wang, H., Yan, S., Xu, D., Tang, X. and Huang, T. (2007) Trace Ratio vs. Ratio Trace for Dimensionality Reduction. 2007 *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, 17-22 June 2007, 1-8. <https://doi.org/10.1109/CVPR.2007.382983>
- [33] Kipf, T.N. and Welling, M. (2016) Semi-Supervised Classification with Graph Convolutional Networks.