Positive Definite Wasserstein Graph Kernel for Brain Disease Diagnosis

Kai Ma, Xuyun Wen, Qi Zhu, and Daoqiang Zhang(✉)

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, Nanjing 211106, China

dqzhang@nuaa.edu.cn

Abstract. In brain functional networks, nodes represent brain regions while edges symbolize the functional connections that enable the transfer of information between brain regions. However, measuring the transportation cost of information transfer between brain regions is a challenge for most existing methods in brain network analysis. To address this problem, we propose a graph sliced Wasserstein distance to measure the cost of transporting information between brain regions in a brain functional network. Building upon the graph sliced Wasserstein distance, we propose a new graph kernel called sliced Wasserstein graph kernel to measure the similarity of brain functional networks. Compared to existing graph methods, including graph kernels and graph neural networks, our proposed sliced Wasserstein graph kernel is positive definite and a faster method for comparing brain functional networks. To evaluate the effectiveness of our proposed method, we conducted classification experiments on functional magnetic resonance imaging data of brain diseases. Our experimental results demonstrate that our method can significantly improve classification accuracy and computational speed compared to state-of-the-art graph methods for classifying brain diseases.

Keywords: Graph kernel · Brain functional network · Brain diseases · Classification · Wasserstein distance

1 Introduction

Brain functional networks characterize functional interactions of human brain, where brain regions correspond to nodes and functional interactions between brain regions are considered as edges. Brain functional networks are widely utilized to classify brain diseases, including Alzheimer's disease [13], attention deficit hyperactivity disorder (ADHD) [7], major depressive disorder [4] and schizophrenia [26]. In these studies, various network characteristics, e.g., degree, clustering coefficient [17], ordinal pattern [16,27] are utilized to represent brain functional network and then used to calculate network measurements for classify brain diseases. However, these network characteristics are built on local connections (e.g., node degree) and ignore the transfer of information between brain regions and the global geometric information in brain functional networks.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 168–177, 2023.
https://doi.org/10.1007/978-3-031-43904-9_17

Recently, Wasserstein distance has attracted broad attention in image processing [6], computer vision [18] and neural network [1]. The Wasserstein distance, also known as the optimal transport distance or Earth Mover's distance [28], was originally proposed to investigate the translocation of masses [8]. In mathematics, the Wasserstein distance is used to quantify the dissimilarity between two probability distributions based on a given ground metric [11,28]. Graph data can be represented as probability distributions using graph or network embedding methods [20]. Graph kernels based on the Wasserstein distance can capture the global geometric information of graphs, making them useful for measuring the similarities between graphs, including brain networks. For example, in the graph kernel based on Wasserstein distance [19], Wasserstein distance is used to measure the similarity of bag-of-vectors between two graphs where the eigen-decompositions of adjacency matrix are used to represent the graphs as bag-of-vectors. In Wasserstein Weisfeiler-Lehman graph kernels [22], the Weisfeiler-Lehman scheme is used to generate node label sequences, and the Wasserstein distance is utilized to measure the similarity of label sequences between two graphs. In the optimal transport based ordinal pattern tree kernel [16], optimal transport distance (i.e., Wasserstein distance) is used to measure the similarity of ordinal pattern tree in brain functional networks and is then applied to classify brain diseases. However, these graph kernels based on Wasserstein distance ignore the transfer of information between brain regions, and their positive definition is hardly guaranteed.

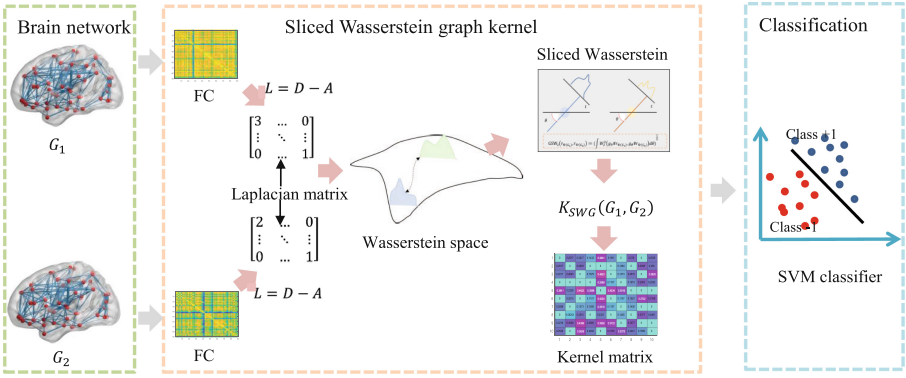

Fig. 1. Illustration of the sliced Wasserstein graph kernel. FC is functional connection matrix.

To tackle these problems, we develop a sliced Wasserstein graph kernel to measure the similarity of brain functional networks. Firstly, we use Laplacian embedding as a feature projection function to project each brain functional network into a set of points in Euclidean space. Then, we calculate the eigen-decomposition on these embeddings and acquire the Laplacian eigenvalues and eigenvectors for a brain functional network. On Laplacian eigenvectors, each row

of Laplacian eigenvectors is the representation of a node in brain functional network. We utilize the Wasserstein distance to measure the cost of transporting information between nodes based on these Laplacian eigenvectors. At last, we calculate one-dimensional empirical measures for the points in Euclidean space and calculate sliced Wasserstein graph kernel. We apply the proposed sliced Wasserstein graph kernel to support vector machine (SVM) classifier [14] for brain disease diagnosis. Figure 1 presents the schematic diagram of the proposed framework with each network representing a specific subject. Specifically, our work has following advantages:

- We provide a new approach for investigating the transfer of information between brain regions with sliced Wasserstein distance.
- The proposed sliced Wasserstein graph kernel is positive definite and a faster method for comparing brain functional networks.
- The proposed sliced Wasserstein graph kernel can improve classification accuracy compared to state-of-the-art graph methods for classifying brain diseases.

2 Methods

2.1 Data and Preprocessing

The functional network data used in the experiments are based on three datasets of brain diseases: ADHD[1], autistic spectrum disorder (ASD)[2], and early mild cognitive impairment (EMCI)[3]. The ADHD dataset consists of 121 ADHD patients and 101 normal controls (NCs). The ASD dataset includes 36 ASD patients and 38 NCs. The EMCI dataset includes 56 EMCI patients and 50 NCs. These brain network data were generated using resting-state functional magnetic resonance imaging (rs-fMRI) data [24]. The rs-fMRI data underwent several preprocessing steps, including brain skull removal, motion correction, temporal pre-whitening, spatial smoothing, global drift removal, slice time correction, and bandpass filtering. Following this, the entire cortical and subcortical structures of the brain were subdivided into 90 brain regions for each subject, based on the Automated Anatomical Labeling atlas. The linear correlation between the mean time series of a pair of brain regions was then calculated to measure the functional connectivity. Finally, a 90×90 fully-connected weighted functional network was constructed for each subject. In this work, we remove the negative connections from brain functional networks.

2.2 Sliced Wasserstein Graph Kernel

Sliced Wasserstein graph kernel is used to measure the similarity of paired brain functional networks. In this subsection, we firstly introduce graph sliced Wasserstein distance and then use it to calculate sliced Wasserstein graph kernel.

[1] http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline.
[2] http://fcon-1000.projects.nitrc.org/indi/abide/.
[3] http://adni.loni.usc.edu/.

Throughout the paper, we will refer to brain functional network when mentioning the graph, unless noted otherwise.

In image processing, computer vision, and graph comparison, many efficient algorithms of machine learning are available in Euclidean space [10]. We define graph sliced Wasserstein distance in Euclidean space. In other words, Euclidean space is as metric space, i.e., $M \subseteq \mathbb{R}^d$, and we use Euclidean distance as the ground distance, i.e., $d(x, y) = |x - y|$, when defining Wasserstein and graph sliced Wasserstein distances. Here, we provide the definition of the Wasserstein distance in Euclidean space.

Let r and c be two probability measures on $\mathbb{R}^d$. The Wasserstein distance between r and c is defined as

$$W_p(r, c) := (inf_{\gamma \in \Gamma(r,c)} \int_{\mathbb{R}^d \times \mathbb{R}^d} (x - y)^p d\gamma(x, y))^{\frac{1}{p}} \tag{1}$$

where $p \in [1, \infty]$ and $\Gamma(r, c)$ denotes the set of all transportation plans of r and c.

The sliced Wasserstein kernel on Wasserstein distance has been studied in [2,10]. In this paper, we extend sliced Wasserstein distance to graph domain and design graph sliced Wasserstein distance and sliced Wasserstein graph kernel.

Graph Sliced Wasserstein Distance: Given two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, the feature projection function $\Phi : \mathcal{G} \to \mathbb{R}^{n \times d}$ projects graph G_1 or G_2, $G_1, G_2 \in \mathcal{G}$ into a set of n points (i.e., $\Phi(G_1), \Phi(G_2) \in \mathbb{R}^{n \times d}$) in d-dimensional Euclidean space. Let $r_{\Phi(G_1)}$ and $c_{\Phi(G_2)}$ denote probability measures on d-dimensional feature representation of graph G_1 and G_2. The graph sliced Wasserstein distance between G_1 and G_2 is defined as

$$D^{\Phi}_{GSW}(G_1, G_2) = GSW_2^2(r_{\Phi(G_1)}, c_{\Phi(G_2)}) \tag{2}$$

where Φ denotes the d-dimensional feature projection. $r_{\Phi(G_1)}$ and $c_{\Phi(G_2)}$ are the probability measures of graph G_1 and G_2, respectively, in d-dimensional Euclidean space. The 2-sliced Wasserstein distance is defined as

$$GSW_2(r_{\Phi(G_1)}, c_{\Phi(G_2)}) := (\int_{\mathbb{R}^d \times \mathbb{R}^d} W_2^2(g_{\theta\#} r_{\Phi(G_1)}, g_{\theta\#} c_{\Phi(G_2)}) d\theta)^{\frac{1}{2}} \tag{3}$$

where $g_{\theta\#} r_{\Phi(G_1)}$ and $g_{\theta\#} c_{\Phi(G_2)}$ are the one-dimensional projections of the measure $r_{\Phi(G_1)}$ and $c_{\Phi(G_2)}$. θ is a one dimensional absolutely continuous positive probability density function.

Theorem 1. *Graph sliced Wasserstein distance $D^{\Phi}_{GSW} : \mathcal{G} \times \mathcal{G} \to \mathbb{R}^+$ is distance metric.*

Proof. According to [10], W_2^2 is distance metric which is symmetric, nonnegativity, identity of indiscernibles, and triangle inequality. D^{Φ}_{GSW} is an integral of W_2^2 terms. Hence, D^{Φ}_{GSW} is also a distance metric and satisfies nonnegativity, symmetry, identity of indiscernibles, and triangle inequality.

Algorithm 1. Compute sliced Wasserstein graph kernel

Input: Two graphs G_1, G_2, parameter λ
Output: kernel value $K_{SWG}(G_1, G_2)$
$X_{G_1} \leftarrow \Phi(G_1)$; %Compute feature representations of graph G_1 and G_2%
$X_{G_2} \leftarrow \Phi(G_2)$;
$r \leftarrow r(X_{G_1})$; %Compute probability measure on feature representations%
$c \leftarrow c(X_{G_2})$
$D^{\Phi}_{GSW}(G_1, G_2) \leftarrow GSW_2^2(r, c)$ %Compute graph sliced Wasserstein distance
$K_{SWG}(G_1, G_2) \leftarrow e^{-\lambda D^{\Phi}_{GSW}(G_1, G_2)}$;

According to the definition of graph sliced Wasserstein distance in Eq.(3), we can find the thought behind graph sliced Wasserstein distance is to achieve the one-dimensional representations for the probability measures on d-dimensional feature representation of graph. Then, the distance between two input probability measures is computed as a function on Wasserstein distance of their corresponding one-dimensional representations.

Feature Projection: We use Laplacian embedding as a feature projection function to project a graph into a set of points in Euclidean space. Then, we calculate the eigen-decomposition on these embeddings and acquire the Laplacian eigenvalues and eigenvectors. On Laplacian eigenvectors, we construct the points in Euclidean space where each row of Laplacian eigenvectors is a node representation.

In Eq.(3), $r_{\Phi(G_1)}$ and $c_{\Phi(G_2)}$ are the feature projection of graph G_1 and G_2 by using Laplacian embedding. Assume that $X_v = \{x_i^v\}_{i=1}^n$ and $Y_v = \{y_i^v\}_{j=1}^n$ is the Laplacian eigenvectors of graph G_1 and G_2. We reformulate the graph sliced Wasserstein distance as a sum rather than an integral. Following the work of [9], we calculate the graph sliced Wasserstein distance by sorting the samples and calculate the L_2 distance between the sorted samples. The graph sliced Wasserstein distance between r and c can be approximated from their node representation which is defined as

$$GSW_2(r_{\Phi(G_1)}, c_{\Phi(G_2)}) \approx (\frac{1}{L}\sum_{l=1}^{L}\sum_{n=1}^{N}|g_{\theta_l}(x^v_{i[n]}) - g_{\theta_l}(y^v_{j[n]})|^2)^{\frac{1}{2}} \tag{4}$$

where $i[n]$ and $j[n]$ are the indices of sorted $\{g_{\theta_l}(x_i^v)\}_{i=1}^N$ and $\{g_{\theta_l}(y_i^v)\}_{j=1}^N$, $g_{\theta_l}(x^v_{i[n]}) =< \theta_l, x^v_{i[n]} >$.

By combining graph sliced Wasserstein distance and feature projection on graphs, we can construct a new graph kernel called sliced Wasserstein graph kernel which can be used to measure the similarity between the paired graphs.

Sliced Wasserstein Graph Kernel: Given two graphs $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ and graph sliced Wasserstein distance on them (i.e., $D^{\Phi}_{GSW}(G_1, G_2)$). We define the sliced Wasserstein graph (SWG) kernel as

$$K_{SWG}(G_1, G_2) = e^{-\lambda D^{\Phi}_{GSW}(G_1, G_2)} \tag{5}$$

Obviously, sliced Wasserstein graph kernel is a case of Laplacian kernel. The procedure of calculating SWG kernel is described in Algorithm 1. According the theorems in [10], one-dimensional Wasserstein space is a flat space. The graph sliced Wasserstein distance is induced from one-dimensional Wasserstein distance. Hence, the graph sliced Wasserstein distance is isometric. SWG kernel based on graph sliced Wasserstein distance is positive definite.

Theorem 2. *Sliced Wasserstein graph kernel is positive definite and differentiable for all* $\lambda > 0$.

Proof. The sliced Wasserstein graph kernel is the extension of sliced Wasserstein kernel on graphs. According to [2,10], sliced Wasserstein kernel is positive definite. Hence, sliced Wasserstein graph kernel is also positive definite.

2.3 Sliced Wasserstein Graph Kernel Based Learning

We use the image processing method described in the data preprocessing to analyze the rs-fMRI data for all subjects and create a brain functional network for each subject. In these brain networks, brain regions are represented as nodes, while the functional connections between paired brain regions are represented as edges. After constructing the brain functional networks for all subjects, we compute the sliced Wasserstein graph kernel using Eq.(5) and apply SVM for disease classification.

3 Experiments

3.1 Experimental Setup

In the experiments, we compare our proposed method with the state-of-the-art graph methods including graph kernels and graph neural networks. Graph kernels include Weisfeiler-Lehman subtree (WL-ST) kernel [21], Weisfeiler-Lehman shortest path (WL-SP) kernel [21], random walk (RW) kernel [23], Wasserstein Weisfeiler-Lehman (WWL) kernel [22], GraphHopper (GH) kernel [5], depth-first-based ordinal pattern (DOP) kernel [15], optimal transport based ordinal pattern tree (OT-OPT) kernel [16]. Graph neural networks include DIFFPOOL [25] and BrainGNN [12]. SVM [3] as the final classifier is exploited to conduct the classification experiment. We perform the leave-one-out cross-validation for all the classification experiments. In the experiments, uniform weight λ is chosen from $\{10^{-2}, 10^{-1}, \cdots, 10^{2}\}$ and the tradeoff parameter C in the SVM is selected from $\{10^{-3}, 10^{-2}, \cdots, 10^{3}\}$.

Table 1. Comparison of different methods on three classification tasks

Method	ADHD vs. NCs		ASD vs. NCs		EMCI vs. NCs	
	ACC(%)	AUC	ACC(%)	AUC	ACC(%)	AUC
WL-ST kernel [21]	70.13	0.682	72.35	0.697	69.41	0.653
WL-SP kernel [21]	68.25	0.664	67.55	0.643	63.18	0.607
RW kernel [23]	69.78	0.643	70.11	0.691	66.87	0.641
WWL kernel [22]	73.63	0.688	72.26	0.715	71.13	0.695
GH kernel [5]	67.56	0.639	71.36	0.674	65.56	0.649
DOP kernel [15]	73.28	0.715	78.68	0.774	75.47	0.737
OT-OPT kernel [16]	76.42	0.738	78.38	0.778	80.36	0.784
BrainGNN [12]	76.46	0.741	82.14	0.801	81.71	0.809
DIFFPOOL [25]	74.63	0.727	80.50	0.798	82.18	0.812
SWG kernel (Proposed)	**78.83**	**0.762**	**90.54**	**0.863**	**85.44**	**0.823**

3.2 Classification Results

We compare the proposed SWG kernel with the state-of-the-art graph kernels on three classification tasks, i.e., ADHD vs. NCs, ASD vs. NCs and EMCI vs. NCs classification. Classification performance is evaluated by accuracy (ACC) and area under receiver operating characteristic curve (AUC). The classification results are shown in Table 1. From Table 1, we can find that our proposed method achieves the best performance on three tasks. For instance, the accuracy achieved by our method is respectively 78.83%, 90.54%, and 85.44% in ADHD vs. NCs, ASD vs. NCs and EMCI vs. NCs classification, which is better than the second best result obtained by BrainGNN and DIFFPOOL. This demonstrates that the proposed SWG kernel is good at distinguishing the patients with brain diseases (i.e., ADHD, ASD, and EMCI) from NCs, compared with the state-of-the-art graph kernels and graph neural networks.

We select four graph methods, including DOP kernel, OT-OPT kernel, BrainGNN and DIFFPOOL, whose classification accuracies are superior to those of other graph methods, except for the SWG kernel. We have recorded the computational time required by these methods, as shown in Table 2. The results in

Table 2. Computational time (in seconds) of different methods.

Method	ADHD vs. NCs	ASD vs. NCs	EMCI vs. NCs
DOP kernel [15]	$\geq$100	$\geq$100	$\geq$100
OT-OPT kernel [16]	$\geq$100	$\geq$100	$\geq$100
BrainGNN [12]	$\geq$100	$\geq$100	$\geq$100
DIFFPOOL [25]	$\geq$100	$\geq$100	$\geq$100
SWG kernel (Proposed)	**83.27**	**8.17**	**15.29**

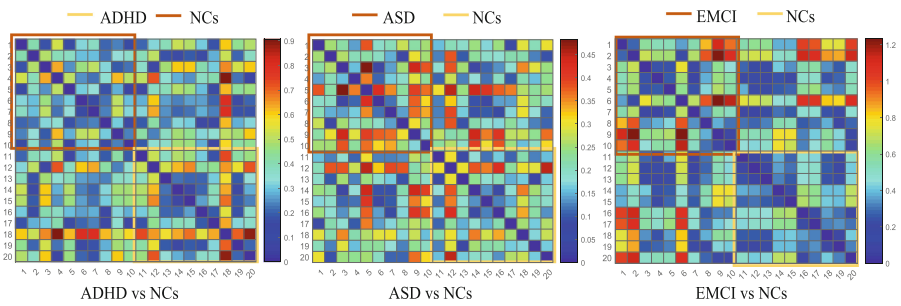

Fig. 2. Distance matrix heatmap for ADHD, ASD and EMCI.

Table 2 demonstrate that our proposed method can enhance computational efficiency as compared to other graph kernels and graph neural networks for the classification of brain diseases.

3.3 Analysis on Wasserstein Distance

The Wasserstein distance was initially proposed to examine mass translocation [28]. In this subsection, we utilize sliced Wasserstein distance to investigate the transportation cost of brain functional networks and brain regions (i.e., nodes). We respectively selected 10 patients and 10 NCs from each dataset, and calculated sliced Wasserstein distances for all selected subjects. We generated a matrix heatmap of the sliced Wasserstein distances for ADHD, ASD, and EMCI, as shown in Fig. 2. The results in Fig. 2 indicate that the sliced Wasserstein distances of patients with brain diseases (e.g., ADHD, ASD, and EMCI) are greater than those of NCs. These results suggest that the transportation cost of information transfer in the brain for patients is higher than that of NCs.

We use Eq.(4) to calculate the sliced Wasserstein distances between brain regions for ADHD, ASD and EMCI and then conduct statistical analysis on these distances. We identify three important brain regions where the sliced Wasserstein distances of patients with brain diseases significantly differed from those of NCs. In ADHD, important brain regions involve right paracentral lobule (PCL.R), left parahippocampal gyrus (PHG.L), and right angular gyrus (ANG.R). In ASD, important brain regions involve left superior frontal gyrus-medial orbital (ORBsupmed.L), right posterior cingulate gyrus (PCG.R), and right superior temporal gyrus (STG.R). In EMCI, important brain regions include right supplementary motor area (SMA.R), right superior occipital gyrus (SOG.R), and right inferior temporal gyrus (ITG.R).

4 Conclusion

In this paper, we propose a sliced Wasserstein graph kernel to measure the similarities between a pair of brain functional networks. We use this graph kernel

to develop a classification framework of brain functional network. We perform the classification experiments in the brain functional network data including ADHD, ASD, and EMCI constructed from fMRI data. The results indicate that our proposed method outperforms the existing state-of-the-art graph kernels and graph neural networks in classification tasks. In computational speed, the proposed method is faster than latest graph kernels and graph neural networks.

Acknowledgement. This work was supported by the National Natural Science Foundation of China (Nos. 62136004, 62276130, 61732006, 62050410348), Jiangsu Funding Program for Excellent Postdoctoral Talent, and also by the Key Research and Development Plan of Jiangsu Province (No. BE2022842).

References

1. Akbari, A., Awais, M., Fatemifar, S., Kittler, J.: Deep order-preserving learning with adaptive optimal transport distance. IEEE Trans. Pattern Anal. Mach. Intell. **45**, 313–328 (2022)
2. Carriere, M., Cuturi, M., Oudot, S.: Sliced Wasserstein kernel for persistence diagrams. In: International Conference on Machine Learning, pp. 664–673. PMLR (2017)
3. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**, 1–27 (2011)
4. Fee, C., Banasr, M., Sibille, E.: Somatostatin-positive gamma-aminobutyric acid interneuron deficits in depression: cortical microcircuit and therapeutic perspectives. Biol. Psychiat. **82**(8), 549–559 (2017)
5. Feragen, A., Kasenburg, N., Petersen, J., De Bruijne, M., Borgwardt, K.M.: Scalable kernels for graphs with continuous attributes. In: Advances in Neural Information Processing Systems, pp. 216–224 (2013)
6. Ge, Z., Liu, S., Li, Z., Yoshie, O., Sun, J.: Ota: optimal transport assignment for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 303–312 (2021)
7. Hartmut, H., Thomas, H., Moll, G.H., Oliver, K.: A bimodal neurophysiological study of motor control in attention-deficit hyperactivity disorder: a step towards core mechanisms? Brain **4**, 1156–1166 (2014)
8. Kantorovitch, L.: On the translocation of masses. Manage. Sci. **5**(1), 1–4 (1958)
9. Kolouri, S., Nadjahi, K., Simsekli, U., Badeau, R., Rohde, G.: Generalized sliced Wasserstein distances. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
10. Kolouri, S., Zou, Y., Rohde, G.K.: Sliced Wasserstein kernels for probability distributions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5258–5267 (2016)
11. Le, T., Yamada, M., Fukumizu, K., Cuturi, M.: Tree-sliced variants of Wasserstein distances. In: Advances in Neural Information Processing Systems (2019)
12. Li, X., et al.: BrainGNN: interpretable brain graph neural network for fMRI analysis. Med. Image Anal. **74**, 102233 (2021)
13. Liu, M., Zhang, J., Adeli, E., Shen, D.: Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis. IEEE Trans. Biomed. Eng. **66**(5), 1195–1206 (2019)

14. Ma, K., Huang, S., Wan, P., Zhang, D.: Optimal transport based pyramid graph kernel for autism spectrum disorder diagnosis. Pattern Recogn. 109716 (2023)
15. Ma, K., Huang, S., Zhang, D.: Diagnosis of mild cognitive impairment with ordinal pattern kernel. IEEE Trans. Neural Syst. Rehabil. Eng. **30**, 1030–1040 (2022)
16. Ma, K., Wen, X., Zhu, Q., Zhang, D.: Optimal transport based ordinal pattern tree kernel for brain disease diagnosis. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention. MICCAI 2022. LNCS, vol. 13433, pp. 186–195. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_18
17. Ma, K., Yu, J., Shao, W., Xu, X., Zhang, Z., Zhang, D.: Functional overlaps exist in neurological and psychiatric disorders: a proof from brain network analysis. Neuroscience **425**, 39–48 (2020)
18. Ma, Z., Wei, X., Hong, X., Lin, H., Qiu, Y., Gong, Y.: Learning to count via unbalanced optimal transport. In: Association for the Advancement of Artificial Intelligence, vol. 35, pp. 2319–2327 (2021)
19. Nikolentzos, G., Meladianos, P., Vazirgiannis, M.: Matching node embeddings for graph similarity. In: Association for the Advancement of Artificial Intelligence (2017)
20. Peng, C., Wang, X., Pei, J., Zhu, W.: A survey on network embedding. IEEE Trans. Knowl. Data Eng. **31**(5), 833–852 (2019)
21. Shervashidze, N., Schweitzer, P., Jan, E., Leeuwen, V., Borgwardt, K.M.: Weisfeiler-Tehman graph kernels. J. Mach. Learn. Res. **12**(3), 2539–2561 (2011)
22. Togninalli, M., Ghisu, E., Llinares-López, F., Rieck, B., Borgwardt, K.: Wasserstein weisfeiler-lehman graph kernels. In: Advances in Neural Information Processing Systems, pp. 6439–6449 (2019)
23. Vishwanathan, S.V.N., Schraudolph, N.N., Kondor, R., Borgwardt, K.M.: Graph kernels. J. Mach. Learn. Res. **11**(2), 1201–1242 (2010)
24. Wang, M.L., Shao, W., Hao, X.K., Zhang, D.Q.: Machine learning for brain imaging genomics methods: a review. Mach. Intell. Res. **20**(1), 57–78 (2023)
25. Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., Leskovec, J.: Hierarchical graph representation learning with differentiable pooling. In: Advances in Neural Information Processing Systems, vol. 31 (2018)
26. Yu, Q., Sui, J., Kiehl, K.A., Pearlson, G.D., Calhoun, V.D.: State-related functional integration and functional segregation brain networks in schizophrenia. Schizophr. Res. **150**(2), 450–458 (2013)
27. Zhang, D., Huang, J., Jie, B., Du, J., Tu, L., Liu, M.: Ordinal pattern: a new descriptor for brain connectivity networks. IEEE Trans. Med. Imaging **37**(7), 1711–1722 (2018)
28. Zhao, P., Zhou, Z.: Label distribution learning by optimal transport. In: Association for the Advancement of Artificial Intelligence (2018)