International Journal of Semantic Computing
Vol. 13, No. 1 (2019) 25–44

DOI: 10.1142/S1793351X19400026

Identification of Discriminative Subnetwork from fMRI-Based Complete Functional Connectivity Networks

Shah Muhammad Hamdi[*,‖], Yubao Wu[*,**], Rafal Angryk[*,††],
Lisa Crystal Krishnamurthy[†,‡,§,‡‡] and Robin Morris[‡,¶,§§]

[*]*Department of Computer Science*
Georgia State University, Atlanta, GA 30302, USA

[†]*Center for Visual and Neurocognitive Rehabilitation*
Atlanta VAMC, Decatur GA 30030, USA

[‡]*Center for Advanced Brain Imaging*
Georgia State University and
Georgia Institute of Technology, Atlanta GA 30302, USA

[§]*Department of Physics and Astronomy*
Georgia State University, Atlanta GA 30302, USA

[¶]*Department of Psychology*
Georgia State University, Atlanta GA 30302, USA
[‖]*shamdi1@gsu.edu*
[**]*ywu28@gsu.edu*
[††]*rangryk@gsu.edu*
[‡‡]*lkrishnamurthy@gsu.edu*
[§§]*robinmorris@gsu.edu*

The comprehensive set of neuronal connections of the human brain, which is known as the human connectomes, has provided valuable insight into neurological and neurodevelopmental disorders. Functional Magnetic Resonance Imaging (fMRI) has facilitated this research by capturing regionally specific brain activity. Resting state fMRI is used to extract the functional connectivity networks, which are edge-weighted complete graphs. In these complete functional connectivity networks, each node represents one brain region or Region of Interest (ROI), and each edge weight represents the strength of functional connectivity of the adjacent ROIs. In order to leverage existing graph mining methodologies, these complete graphs are often made sparse by applying thresholds on weights. This approach can result in loss of discriminative information while addressing the issue of biomarkers detection, i.e. finding discriminative ROIs and connections, given the data of healthy and disabled population. In this work, we demonstrate a novel framework for representing the complete functional connectivity networks in a threshold-free manner and identifying biomarkers by using feature selection algorithms. Additionally, to compute meaningful representations of the discriminative ROIs and connections, we apply tensor decomposition techniques. Experiments on a fMRI dataset of neurodevelopmental reading disabilities show the highly interpretable nature of our approach in finding the biomarkers of the diseases.

Keywords: fMRI-based brain network; discriminative subgraph; feature selection; tensor decomposition.

1. Introduction

Enriched by neuroimaging technologies such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Electroencephalography (EEG), and Diffusion Tensor Imaging (DTI), *brain informatics* has been playing a key role in the investigation of neurological and neurodevelopmental disorders. The complexity, heterogeneity, and scarcity of brain informatics data pose a great challenge to data mining. Among the data collection modalities, functional MRI (fMRI) is a popular one, which measures the functional activity of different brain regions. Resting state fMRI is used to construct the functional connectivity network, a complete, edge-weighted graph, where the nodes represent brain regions (ROIs) and the edge-weights represent similarity/dissimilarity of two ROIs in their Blood Oxygenation Level Dependent (BOLD) time series. The BOLD time series is thought to be related to the aggregate neuronal activity within the ROI over the scan period. The analysis of functional connectivity networks has facilitated the early diagnosis of several neurological and neurodevelopmental diseases such as Alzheimer's [1], Schizophrenia [2], Bipolar disorder [3], Attention-deficit/hyperactivity disorder (ADHD) [4], Autism [5], and Dyslexia [6].

Given a set of functional connectivity networks and associated healthy/disabled class labels, one of the most important research questions asked by the neuroscientist community is, "How can one find biomarkers that distinguish two groups?" Data mining people consider the problem as binary classification, and mostly give two types of solutions — graph classification and tensor decomposition.

Because of the advances of graph mining algorithms, most supervised learning studies transform the complete functional connectivity networks into sparse graphs by thresholding and binarizing [1, 3, 7– 9]. Then, they extract structure-based features such as degree, clustering coefficient, and PageRank score of each node and/or subgraph-based features such as gSpan-based frequent subgraphs [10] to construct a feature space. While structure-based features are hand-engineered and require a lot of domain knowledge, subgraph-based features are computationally very expensive. Another disadvantage of thresholded and binarized graphs is that they can lose some discriminative information. Figure 1 shows a motivating example, where two complete functional connectivity networks of different class labels are shown. Each network has four nodes and six weighted edges, where signed weights represent positive/negative correlation of the BOLD time series of the adjacent nodes (ROIs). To create the sparse graphs, the edges are positively thresholded and binarized. As a result, both representations fail to find the discriminative edges $(1, 2)$ and $(3, 4)$, which distinguish two classes by a big difference in same-signed weights.

This approach also cannot weight the edges of the discriminative subgraph in terms of discrimination score. On the contrary, while thresholding and binarization are not applied and the complete graphs are represented by fixed-length vectors (since each complete graph has the same number of nodes and edges), simple feature selection scheme such as thresholding on absolute difference of edge weights and

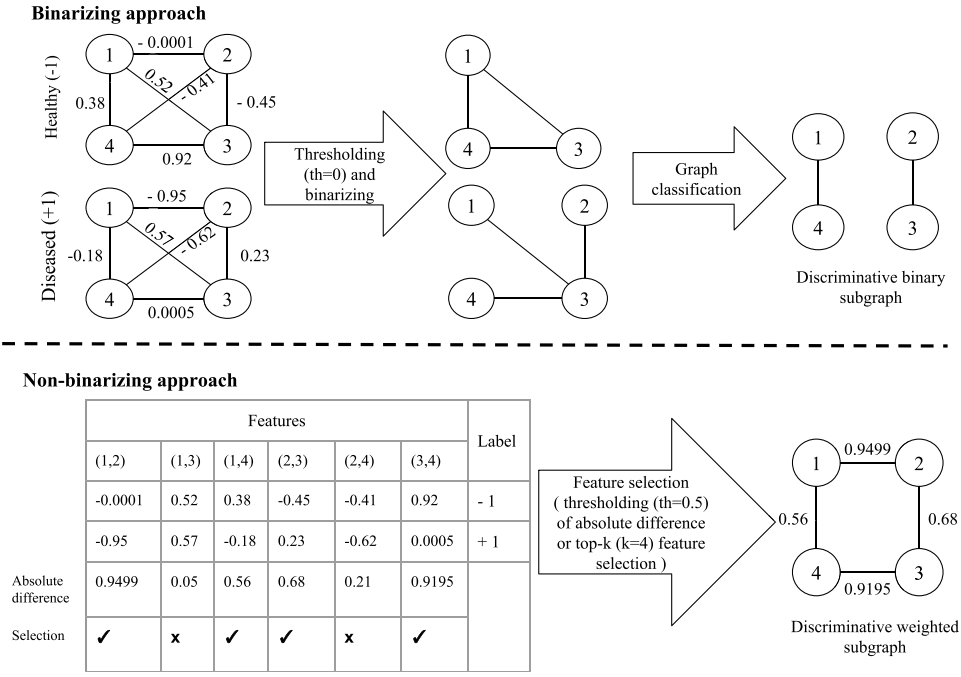


Fig. 1. Non-binarizing approach can result in producing more significant biomarkers.

selecting the edges (features) with top-k absolute difference of weights, can produce more discriminative edges, e.g. $(1,2)$, $(2,3)$, $(3,4)$ and $(4,1)$ along with their discrimination scores. The unsupervised feature selection approach (shown for simplicity in this example) also uses *a priori* knowledge such as k in top-k feature selection, but still, this *a priori* knowledge is less expensive than the *a priori* knowledge about the sparsity of the graph [11].

Multi-dimensional arrays or tensor-based approaches stack the adjacency matrices of the thresholded functional connectivity networks to construct one third-order tensor and decompose the tensor into one subject factor matrix and two identical ROI factor matrices (since the adjacency matrices make partially symmetric tensor) [12–16]. Each column of the factor matrices represents one latent feature of different entities (subjects or ROIs). After tensor decomposition, the subject factor matrix is used for classifier training and testing. While this approach may result in high classification accuracy, the latent features, i.e. the columns of the subject factor matrix are difficult to interpret because the relationship of these latent features with the discriminative ROIs and connections are not yet established [12].

In this work, we address the problem of graph thresholding by using vector representation and applying feature selection algorithms to find discriminative features, which eventually become the edges of the discriminative subgraph. We also address the issue of interpretability of tensor decomposition-based approach by

constructing the tensor with ROI-connection-based incidence matrices of the induced discriminative subgraphs of the complete functional networks of the subjects, instead of the adjacency matrices of the thresholded functional connectivity networks.

The main contributions of this work are:

(1) Transforming the fMRI-based biomarker detection problem from the graph space to vector space, i.e. solving the problem with functional connectivity vectors instead of thresholded graphs.
(2) Using feature selection algorithms for constructing the discriminative subgraph, whose nodes and edges represent the biomarkers.
(3) Proposing a novel tensor construction scheme with the computed discriminative subgraphs so that tensor decomposed factor matrices can be easily manipulated to determine the relationship between the biomarkers and diagnostic groups of subjects.

The rest of the paper is organized as follows. In Sec. 2, we discuss the related work. We present the framework for finding discriminative connections and ROIs from fMRI-based functional connectomes using feature selection algorithms in Sec. 3. Representation learning of the discriminative ROIs and connections is discussed in Sec. 4. In Sec. 5, we present the experimental findings. Finally, we discuss our future work and conclude the paper in Sec. 6.

2. Related Work

Data mining research on fMRI-based functional connectivity data can be divided into two categories: discriminative subgraph-based graph classification and discriminative latent feature-based tensor decomposition. In this section, we discuss both approaches and distinguish our approach from them.

Graph classification-based approaches: In these approaches, binary or weighted graphs are generated by applying a threshold on functional connectivity matrices so that all the nodes of the graphs are connected. Then, structure-based and/or subgraph pattern-based features are extracted to represent the graphs in feature space. Wee *et al.* [1] used functional connectivity matrices extracted from fMRI and DTI modalities to distinguish Mildly Cognitive Impaired (MCI) subjects from healthy controls (MCI is the early stage of Alzheimer's disease). Their approach is an example of a structure-based graph classification approach. Given a functional network, weighted local clustering coefficient of each node is calculated, and the graph is represented by a vector consisting of these local connectivity measures. Then SVM classifier is applied to this vector space. This model gives a ranking of the ROIs in terms of how well they are clustered with respect to other ROIs, which provides a good step towards the interpretability of the biomarkers. Jie *et al.* [9] presented another structure-based graph classification approach, where they used Weisfeiler–Lehman graph kernel [17] for computing the global connectivity features of each

graph, which are used along with a local connectivity feature of weighted local clustering coefficient of each node. Considering the thresholded edge weights as the probabilities of the link between two nodes, Kong *et al.* [8] presented a discriminative subgraph feature selection algorithm based on dynamic programming to compute the probability distribution of the discrimination scores for each subgraph pattern. In some neuroimaging studies of neurological and neurodevelopmental disorders, there may be additional clinical, serologic, and cognitive measures data from each subject. Cao *et al.* [7] presented a discriminative subgraph mining algorithm for brain networks which leverages such multiple side views-based data.

Tensor decomposition-based approaches: Tensor decomposition is used to extract the latent discriminative features of each subject. In [12], tensors are formed by stacking the non-negative connectivity matrices of all subjects. The resulting tensor is decomposed with several constraints such as symmetry of the factor matrix representing the ROI space and orthogonality of the factor matrix representing the subject space in order to maximize the discrimination among the subjects of different classes. In [14], the time-sliced non-negative connectivity matrices are used to create the tensors in order to discover the latent factors of the time windows. Both studies [12] and [14] modeled the problem as constrained CANDECOMP/PARAFAC (CP) decomposition and used Alternating Direction Method of Multipliers (ADMM) as the optimization framework.

Although tensor decomposition-based approaches represent each functional connectivity network in latent feature space, graph mining-based approaches can find meaningful discriminative patterns or biomarkers, i.e. discriminative ROIs and connections. However, the imposed threshold-based sparsity make these approaches lose some discriminative information. Moreover, some of these approaches rely on the exhaustive enumeration of the induced subgraphs, which can be computationally very expensive in terms of both space and time [15, 16].

The model we present in this paper does not use thresholds for imposing sparsity. By leveraging the completeness and equal number of nodes of the functional connectivity networks, our model enumerates the edges, and represent each functional connectivity network as fixed-length vector instead of enumerating the subgraphs. The edges whose weights are maximally dependent on the class label can be found by applying feature selection algorithms such as Fisher and minimal Redundancy Maximal Relevance (mRMR) criterion. The selected edges (features) form the discriminative subgraph. Therefore, our model for fMRI-based biomarker detection does a paradigm shift from frequent/discriminative subgraph mining to feature selection.

3. Finding Discriminative Subgraph from Complete Functional Connectivity Networks

In this section, we discuss how to represent a complete functional connectivity network using a fixed-length functional connectivity vector. Although any feature

selection algorithm can be plugged into our framework, for simplicity and brevity, we discuss two well-known feature selection algorithms from the literature — Fisher [18], and mRMR [19], which can be used to find discriminative connections from vector represented complete functional connectivity networks.

3.1. *Functional connectivity vector*

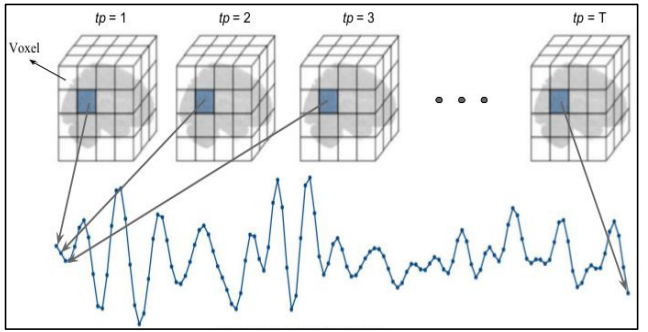
In its raw form, fMRI data is four dimensional. The scanner captures a sequence of whole brain volumes of the subject in a regular time interval (Fig. 2(a)). Therefore, three dimensions represent spatial information and one dimension represents temporal information. In fMRI scans, a voxel is the unit of the brain volume. After a series of preprocessing steps, the BOLD time series of each voxel is found as the change of activation over a time period (Fig. 2(b)). The voxels can be grouped into regions of interest (ROIs), where the regions can be defined manually by the neuroscientists targeting a specific-disease relevant area or by some standard brain atlas such as Harvard–Oxford Atlas for whole brain analysis. A time series is found for each ROI, which is the mean of the time series across the voxels of that ROI. One representation of the fMRI data is the multivariate time series, where the variables are represented by the ROIs (Fig. 2(c)). From these ROI time series, pairwise Pearson correlation coefficients are calculated. Then, Fisher's r-to-z transformation is applied on the elements of the correlation matrix to improve the normality of the correlation coefficients as

$$z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right),$$

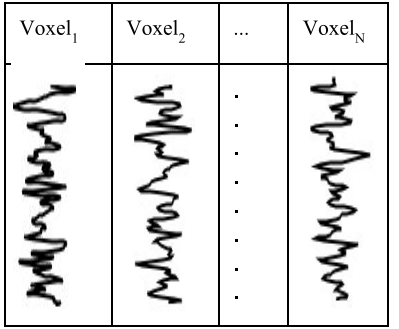
where r is the Pearson correlation coefficient. This z-map is used as another data representation, and called the functional connectivity matrix (Fig. 2(d)). The functional connectivity matrix is symmetric and can be considered as the adjacency matrix of a complete graph (Fig. 2(e)), where the entries of the matrix define the edge weights. A complete graph with n nodes has $e = n(n-1)/2$ edges. By considering each edge as a feature, the complete functional connectivity network of a subject is represented by an e-dimensional vector (Fig. 2(f)). Functional connectivity vector is found by flattening the upper/lower triangular portion of the functional connectivity matrix. A dataset of n_s subjects, where each subject's functional connectivity network has n nodes with e edges is represented by $\{(X^{(i)}, y^{(i)})\}_{i=1}^{n_s}$, where $X^{(i)} \in \mathbb{R}^e$ and $y^{(i)} \in \{+1, -1\}$. The feature space is denoted by $X = \{x_1, x_2, \ldots, x_e\}$.

3.2. *Mining discriminative subgraph by feature selection algorithms*

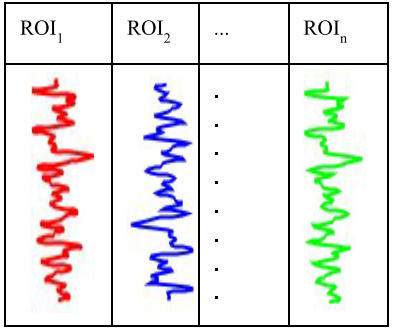
After extracting the functional connectivity vectors from the complete functional connectivity networks, we get a high-dimensional feature space. Since each feature represents an edge or connection between two ROIs, selection of the edge features that are most statistically relevant to the class label results in a subgraph that can distinguish healthy and disabled classes. In the resultant discriminative subgraph,



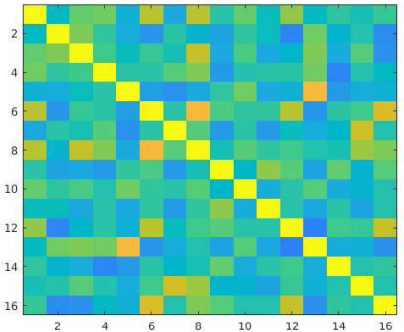
(a) BOLD time series extraction of the voxels by sequentially capturing whole brain volume over a period of time



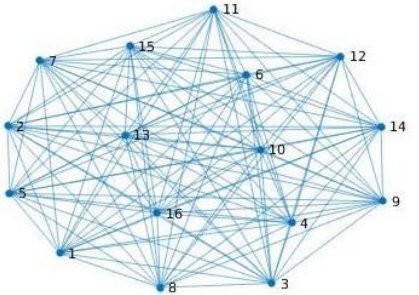
(b) Multivariate time series (voxel-level)



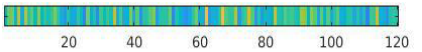
(c) Multivariate time series (ROI-level)



(d) Functional connectivity matrix found after pairwise Pearson correlation calculation on the ROI-based multivariate time series



(e) Complete functional connectivity network of 16 nodes and 120 edges



(f) Functional connectivity vector found by flattening the upper/lower triangular portion of functional connectivity matrix; each feature represents an edge of the complete graph

Fig. 2. Extraction of functional connectivity vector.

each edge is assigned a weight, which is the score assigned to its corresponding feature by a feature selection algorithm.

Feature selection algorithms can be divided into three categories — filters, wrappers, and embedding methods [20]. Filter methods use the intrinsic property of the data and rank the features before feeding the reduced feature space into a classifier for learning. Wrappers use the classifier performance to evaluate the feature subset. Wrapper models tend to give better results, but they are classifier dependent and computationally more expensive than filters. Embedding methods inject the feature selection process into the learning step of the classifier.

Depending on whether the label information is used in feature selection, filter methods are divided into supervised (e.g. Fisher and mRMR), and unsupervised (e.g. maximum variance and Laplacian score) approaches. Since we consider supervised case in this work, i.e. biomarker detection from labeled functional connectivity networks, we propose supervised filter-based feature selection algorithms. Whether the edges of the discriminative subgraph depend on each other in distinguishing the classes, two variants of supervised filter-based feature selection algorithms can be applied — univariate and multivariate.

3.2.1. *Univariate feature selection*

Univariate feature selection algorithms consider each feature independently and ignore any correlation between them. Each edge of the discriminative subgraph selected by such algorithms is individually discriminating and independent with other selected edges. Fisher scoring is an example of univariate feature ranking. Fisher score or F-score is the ratio of the inter-class distance and intra-class distance of a given feature [18]. Fisher score of a feature x_j is defined as

$$F(x_j) = \frac{\sum_{k=1}^{c} n_k(\mu_j^k - \mu_j)^2}{(\sigma_j)^2},$$

where c denotes the number of classes, n_k denotes the number of samples of class k, μ_j and σ_j denote the mean and standard deviation of feature x_j, and μ_j^k denotes the mean of kth class, corresponding to the feature x_j.

3.2.2. *Multivariate feature selection*

Multivariate feature selection algorithms select features whose combination ensures higher discrimination ability, although the individual discrimination ability of the selected features might be poor. The edges of the discriminative subgraph selected by such algorithms are jointly discriminating. An example of multivariate feature selection is mRMR [19]. mRMR selects features sequentially and the selected features are expected to have maximal relevance with the class labels while having minimal redundancy with already selected features. If $f-1$ features are already selected by mRMR and the already selected feature set is denoted by S_{f-1}, then the algorithm

finds the fth feature from the set $X - S_{f-1}$ by optimizing the following condition:

$$\max_{x_j \in X - S_{f-1}} \left[I(x_j; y) - \frac{1}{f-1} \sum_{x_i \in S_{f-1}} I(x_j; x_i) \right],$$

where y is the class label variable and $I(p; q)$ is the mutual information between two random variables p and q. Mutual information between p and q is defined in terms of their probability density function $\Pr(p)$, $\Pr(q)$, and $\Pr(p, q)$.

$$I(p; q) = \iint \Pr(p, q) \log \frac{\Pr(p, q)}{\Pr(p)\Pr(q)} dpdq.$$

The mRMR score of the fth selected feature x_j is found by

$$\text{mRMR\_score}(x_j) = I(x_j; y) - \frac{1}{f-1} \sum_{x_i \in S_{f-1}} I(x_j; x_i).$$

After scoring each feature by univariate/multivariate feature selection algorithm, and selecting the top-k features by cross-validation, the selected features provide the edges of the discriminative subgraph, where the feature scores become the weights of the edges. In the discriminative subgraph, the nodes (ROIs) and the edges (connections) can be considered as the discriminative patterns or biomarkers of a neurological or neurodevelopmental disease. Visualization of the edge-weighted discriminative subgraph can help neuroscientists differentiate different types of neurological disease, and to identify brain areas for targeted intervention.

4. Representation Learning for the Biomarkers

Given a set of biomarkers, i.e. discriminative ROIs and connections, how does a subject's functional connectivity network characterize them? Good representation of biomarkers facilitates the computation of the discriminating ability of them on the functional connectivity network of a subject. We address this issue by tensor decomposition technique. In this section, we discuss the procedure of tensor construction, tensor decomposition techniques, and utilizing the resultant biomarker factor matrices for computing the biomarker characteristics on healthy and disabled subjects. The important notations used in this section are summarized in Table 1.

Table 1. Notations.

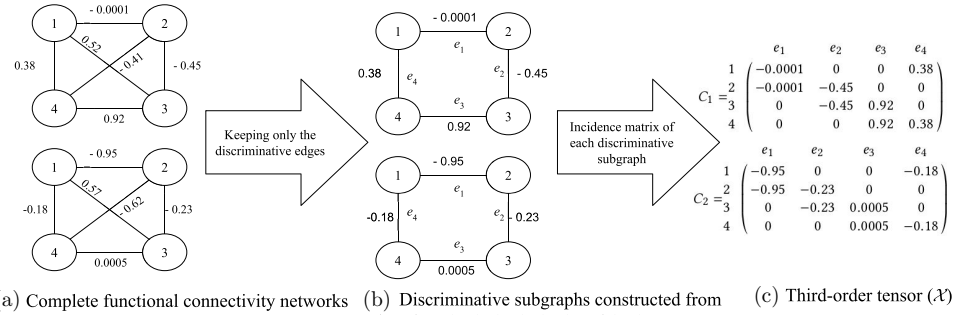
Symbol	Definition	Symbol	Definition
n	Number of nodes in each complete graph	$\boldsymbol{\chi}$	Tensor
e	Number of edges in each complete graph	$\mathbf{S}$	Subject factor matrix
n_s	Number of subjects in the dataset	$\mathbf{N}$	Discriminative nodes factor matrix
n_{dn}	Number of discriminative nodes in the dataset	$\mathbf{E}$	Discriminative edges factor matrix
n_{de}	Number of discriminative edges in the dataset	$\mathbf{I}_k$	Identity matrix of size k
$\mathbf{C}$	Incidence matrix	P, Q, R	Ranks of tensor decomposition

4.1. *Tensor construction*

A third-order tensor is constructed by stacking the weighted incidence matrices of the discriminative subgraphs induced in the complete functional connectivity network of each subject. Three modes of the tensor are the subjects, discriminative nodes, and discriminative edges. The set of discriminative edges, which is computed by a feature selection algorithm after vectorizing all the complete functional connectivity networks of the dataset, can be used to construct the discriminative subgraph of each subject. Incidence matrix of the discriminative subgraph has a row for each discriminative node and a column for each discriminative edge. The weighted incidence matrix $\mathbf{C} \in \mathbb{R}^{n_{dn} \times n_{de}}$ is defined as

$$\mathbf{C}_{i,j} = \begin{cases} w_j, & \text{if node } i \text{ is end-node of edge } j \\ 0, & \text{otherwise} \end{cases}$$

for $1 \leq i \leq n_{dn}$ and $1 \leq j \leq n_{de}$. The weighted incidence matrices of the discriminative subgraph of all subjects are stacked to construct a third-order tensor $\mathcal{X} \in \mathbb{R}^{n_s \times n_{dn} \times n_{de}}$. An example is shown in Fig. 3. In this example, the functional connectivity networks are the same as the functional connectivity networks of Fig. 1, and the discriminative edges are $(1,2)$, $(2,3)$, $(3,4)$, and $(4,1)$.



(a) Complete functional connectivity networks (b) Discriminative subgraphs constructed from the edges (1,2), (2,3), (3,4) and (4,1) (c) Third-order tensor ($\mathcal{X}$)

Fig. 3. Constructing third-order tensor from induced discriminative subgraph of the complete functional connectivity networks.

4.2. *Tensor decomposition*

Tensor decomposition is performed to acquire the latent factor-based representations of the objects defined in each mode of a higher order tensor. Two widely used tensor decomposition techniques are CP decomposition and Tucker decomposition [21]. In this subsection, we briefly discuss them in terms of our *subjects-discriminative nodes-discriminative edges*-based tensor.

4.2.1. *CP decomposition*

CP decomposition factorizes the tensor into a sum of rank one tensors. Given a third order tensor $\mathcal{X} \in \mathbb{R}^{n_s \times n_{dn} \times n_{de}}$, CP decomposition factorizes the tensor as follows:

$$\mathcal{X} \approx \sum_{r=1}^{R} \mathbf{s}_r \circ \mathbf{n}_r \circ \mathbf{e}_r = [\![\mathbf{S}, \mathbf{N}, \mathbf{E}]\!].$$

Here, $\mathbf{o}$ denotes the outer product of the vectors. R is a positive integer and also called the tensor rank. $\mathbf{s}_r, \mathbf{n}_r$, and $\mathbf{e}_r$ are vectors, where $\mathbf{s}_r \in \mathbb{R}^{n_s}$, $\mathbf{n}_r \in \mathbb{R}^{n_{dn}}$, and $\mathbf{e}_r \in \mathbb{R}^{n_{de}}$ for $r = 1, 2, 3, \ldots, R$. After stacking those vectors, we can get the factor matrices $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_R]$, $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_R]$, and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_R]$. We show a visualization of the CP decomposition of our *subjects-discriminative nodes-discriminative edges*-based tensor in Fig. 4. Each row of a factor matrix is a

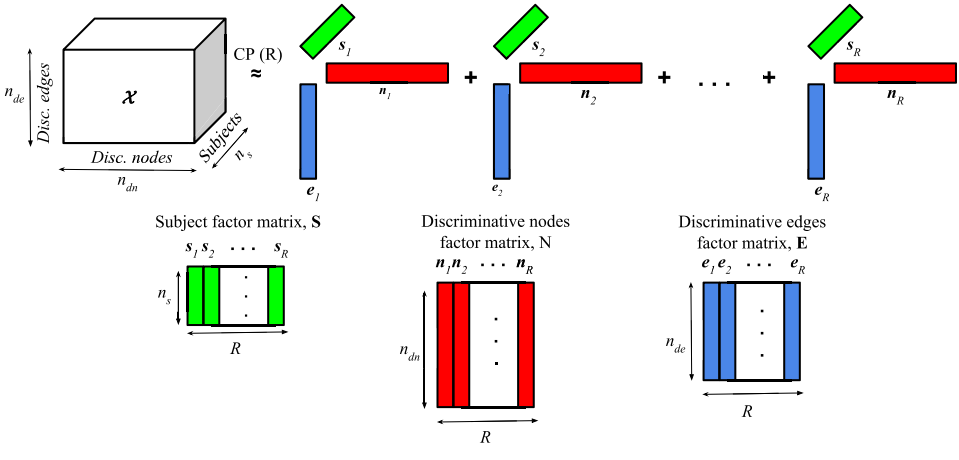


Fig. 4. CP decomposition of the *subjects-discriminative nodes-discriminative edges*-based tensor.

R-dimensional representation of an object, i.e. subject, discriminative node, or discriminative edge. In CP decomposition, there are no imposed orthogonality constraints for the factor matrices. Nevertheless, we can compute the impact of the discriminative node j on subject i by the inner product:

$$\mathbf{S}(i,:) * \mathbf{N}(j,:)^T.$$

Similarly, we can compute the impact of the discriminative edge j on subject i by the inner product:

$$\mathbf{S}(i,:) * \mathbf{E}(j,:)^T.$$

4.2.2. *Tucker decomposition*

Tucker decomposition is a form of higher-order Principal Component Analysis (PCA). A tensor is decomposed into a core tensor, which is multiplied by a matrix along its each mode. Tucker decomposition of a third-order tensor $\mathcal{X} \in \mathbb{R}^{n_s \times n_{dn} \times n_{de}}$ is given by

$$\mathcal{X} \approx \mathcal{G} \times_1 \mathbf{S} \times_2 \mathbf{N} \times_3 \mathbf{E} = [\![\mathcal{G}; \mathbf{S}, \mathbf{N}, \mathbf{E}]\!]. \tag{1}$$

Here, $\times_n$ denotes mode-n tensor product. $\mathbf{S} \in \mathbb{R}^{n_s \times P}$, $\mathbf{N} \in \mathbb{R}^{n_{dn} \times Q}$, and $\mathbf{E} \in \mathbb{R}^{n_{de} \times R}$ are the factor matrices. These factor matrices can be thought as the principal components along each mode. $\mathcal{G} \in \mathbb{R}^{P \times Q \times R}$ is the core tensor and the elements of this tensor represents the interaction between those principal components. We show a visualization of the Tucker decomposition of our *subjects-discriminative nodes-discriminative edges*-based tensor in Fig. 5.

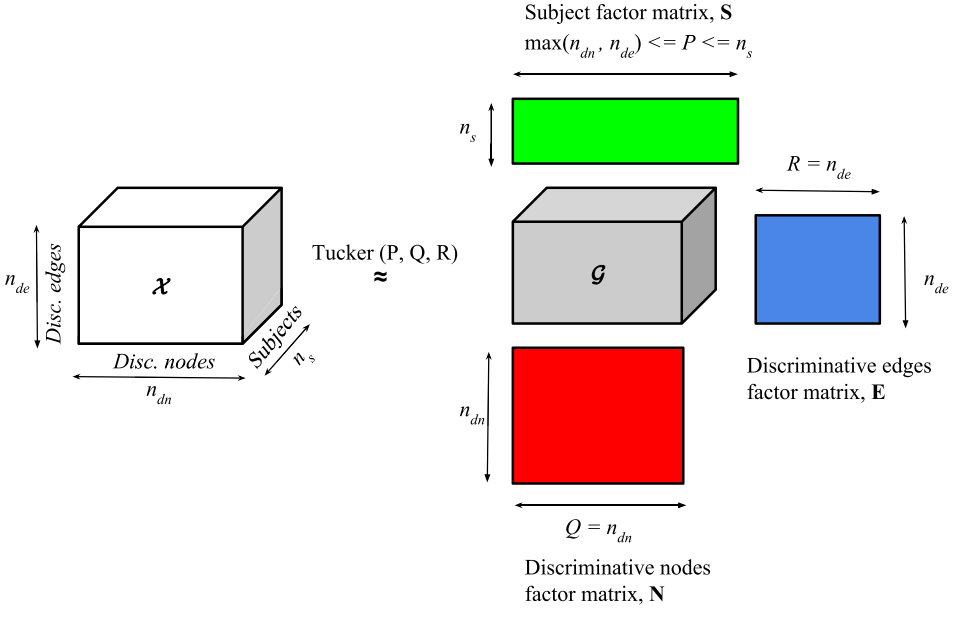


Fig. 5. Tucker decomposition of the *subjects-discriminative nodes-discriminative edges*-based tensor.

The row-wise projection of the subject factor matrix onto biomarker factor matrix requires each biomarker factor matrix to be orthogonal. Additionally, the columns (features) of the subject factor matrix must be pairwise orthogonal, to ensure meaningful representation of each row (subject). If $n_s > n_{dn}$ and $n_s > n_{de}$, by setting $\max(n_{dn}, n_{de}) \leq P \leq n_s$, $Q = n_{dn}$, and $R = n_{de}$, Tucker decomposition assumes the follwing constraints are held.

- $\mathbf{N}\mathbf{N}^T = \mathbf{N}^T\mathbf{N} = \mathbf{I}_{n_{dn}}$, i.e. discriminative nodes factor matrix, $\mathbf{N}$ is an orthogonal matrix.
- $\mathbf{E}\mathbf{E}^T = \mathbf{E}^T\mathbf{E} = \mathbf{I}_{n_{de}}$, i.e. discriminative edges factor matrix, $\mathbf{E}$ is an orthogonal matrix.
- $\mathbf{S}^T\mathbf{S} = \mathbf{I}_P$, i.e. columns of the subject factor matrix $\mathbf{S}$ are pairwise orthogonal.
- $\mathbf{S}\mathbf{S}^T \neq \mathbf{I}_{n_s}$, i.e. rows of the subject factor matrix $\mathbf{S}$ are not pairwise orthogonal.

Now, we can compute the impact of the discriminative node j on subject i by the inner product:

$$\mathbf{S}(i, 1 : n_{dn}) * \mathbf{N}(j, :)^T.$$

Similarly, we can compute the impact of the discriminative edge j on subject i by the inner product:

$$\mathbf{S}(i, 1 : n_{de}) * \mathbf{E}(j, :)^T.$$

Both CP and Tucker decompositions can be solved by Alternating Least Squares (ALS) optimization. After a random initialization of all factor matrices, ALS updates one factor matrix while keeping the other two fixed until convergence. The details of ALS optimization for CP and Tucker decomposition can be found in [21].

Now, we summarize our proposed framework, which combines the feature selection-based discriminative subgraph mining method with the tensor decomposition-based representation learning of the biomarkers.

(1) First, we represent the complete functional connectivity networks of the dataset as fixed-length functional connectivity vectors.
(2) We apply univariate/multivariate filter-based supervised feature selection algorithm for finding top-k discriminative features. By incrementally selecting features, and feeding the reduced feature space to a classifier, while train and test examples are selected by a predefined cross-validation scheme, we can observe the classification accuracies found from the selected feature sets. The feature set resulting in the best cross-validation accuracy is considered to be the edge set of the discriminative subgraph, where the edges are weighted by their corresponding feature scores.
(3) After we find the discriminative subgraph of the dataset, we compute the characterizing value of its nodes and edges for a given subject by decomposing the third-order tensor, whose three modes represent subjects, discriminative nodes, and discriminative edges.

5. Experimental Evaluation

In this section, we evaluate our framework on a real-world dataset. We first discuss our data collection and preprocessing steps. Then, we visualize the discriminative subgraph of that dataset and compare the biomarkers' differentiation between the

healthy and disabled subjects. Finally, we show the effect of different feature selection algorithms on different classifiers.

We used *Feature Selection Toolbox* [20] for mRMR and Fisher-based feature selection and *Tensor Toolbox* [22] for computing CP and Tucker decomposition with ALS optimization, and ran all experiments in MATLAB 2017a.

5.1. *Data collection*

In our experimental evaluation, the target neurodevelopmental disorder is reading disability. For the study, we used preprocessed resting-state fMRI scans of 27 adult subjects, 12 labeled as *struggling* readers (below-average reading test scores), and 15 labeled as *typical* (average or above on the reading test). Because of the specific experimental setting and preprocessing requirements of this study, the number of collected samples is small. We summarize the preprocessing steps as follows:

(1) Removal of time-locked physiological noise.
(2) Slice timing correction.
(3) Bulkhead motion correction.
(4) Reorientation.
(5) Echo-planar Imaging (EPI)-based distortion correction.
(6) Linear and nonlinear coregistration to template space.
(7) Resampling to EPI voxel size.
(8) Removal of physiological noise by detrending time-shifted respiratory volume per time.
(9) Removal of local white matter BOLD time series.
(10) Skull-stripping.
(11) Low-pass filtering in the range of 0.001–0.1 Hz.
(12) Smoothing with Gaussian kernel (FWHM 6 mm).

The details of the preprocessing steps for this study can be found in [23]. The preprocessing steps are done using AFNI [24], FSL [25], and FreeSurfer [26]. We considered 16 ROIs based on *a priori* reading research [27]. In Fig. 6, we show the tags, names and visualizations of these ROIs in left and right hemisphere of the brain.

After preprocessing, the number of voxels in one brain volume is $53 \times 63 \times 45$, while the repetition time (time between capturing two whole brain volumes) is 2 s. Finally, using Conn [28], the ROI-based multivariate time series are extracted. The length of each ROI time series is 125. Then, we calculate the Pearson correlation matrix of the ROI-based multivariate time series. Finally, we compute the functional connectivity matrix by Fisher's r-to-z transformation on the elements of the Pearson correlation matrix. Because of 16 ROIs, the resultant functional connectivity matrix has size 16×16. The functional connectivity matrix is the weighted adjacency matrix of the complete functional connectivity network. Since each complete functional connectivity network has 16 nodes and $16 * (16 - 1)/2 = 120$ weighted edges, the functional connectivity vector of each subject is 120 dimensional.

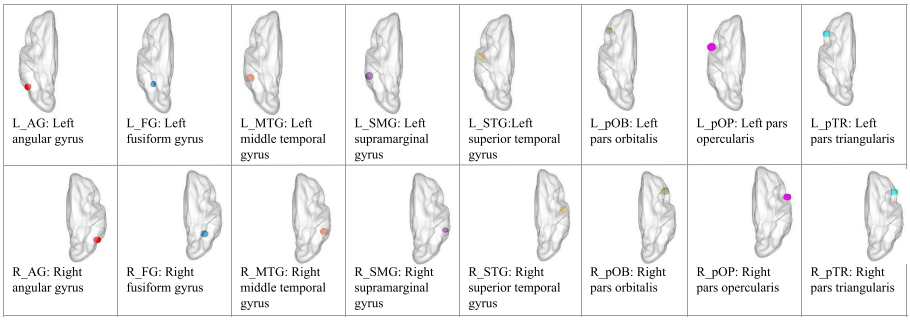


Fig. 6. Visualization of 16 ROIs in left and right hemisphere.

5.2. *Generation and visualization of discriminative subgraph*

After representing the complete functional connectivity networks of the dataset as 120-dimensional functional connectivity vectors, we apply mRMR and Fisher-based feature selection algorithm to select k discriminative features. The discriminative subgraph is found after a cross-validation-based experiment. Because of the small number of examples in our dataset, we used leave-one-out (LOO) train/test splitting strategy. LOO is a special case of K-fold cross-validation while K is the number of examples. At ith iteration of LOO, all examples except the ith one are used as training examples and the ith example is used as the test example. For this cross-validation-based experiment, we used nearest neighbor classifier with Euclidean distance measure. At kth iteration, we first select k features by mRMR/Fisher scoring, and report the mean accuracy of the nearest neighbor classifier over the iterations of LOO. In Fig. 7, we show the accuracy of nearest neighbor classifier after

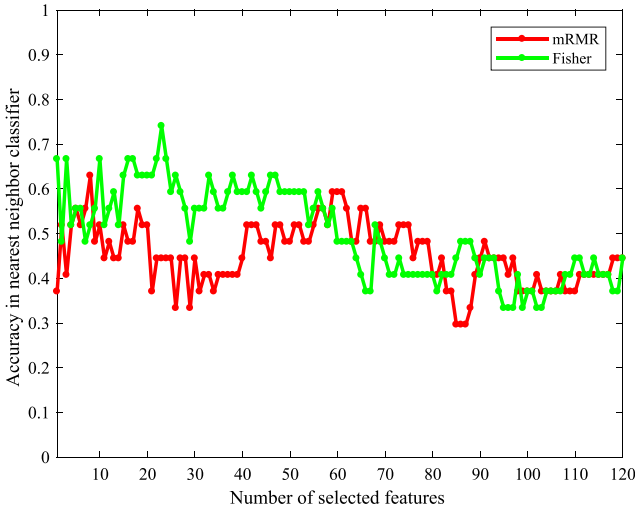


Fig. 7. Cross-validation experiment for finding the best k features.

incrementally selecting features. Since Fisher-based top-23 selected features give maximum LOO accuracy, i.e. 74%, we get $k = 23$, and consider these 23 features as the edges of the discriminative subgraph. Each edge of the discriminative subgraph is assigned a weight, which is the Fisher score of its corresponding feature. In Fig. 8, we show the discriminative subgraph for our dataset, where edge weights are visualized by the thickness of the edge. It is easy to see that the connections (*R_AG*, *R_STG*) and (*R_SMG*, *R_MTG*) has maximum discriminating ability.

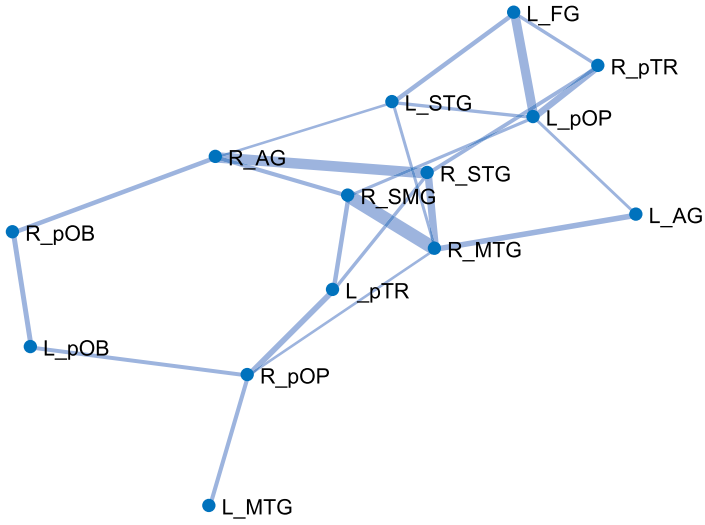


Fig. 8. Discriminative subgraph using Fisher-selected 23 edges.

5.3. *Discriminating ability of the biomarkers*

This experiment is done to verify the discrimination ability of the selected biomarkers. First, Fisher-selected top 23 features are used to construct the discriminative subgraph. Then, a third-order tensor is constructed by stacking the incidence matrices of the induced discriminative subgraph of the complete functional connectivity networks. After construction of the tensor, CP and Tucker decomposition are performed separately. As a result of tensor decomposition, three factor matrices representing the subjects, discriminative nodes, and discriminative edges are found. Then, the rows of the subject factor matrix are divided into typical and struggling classes. Inner product (cosine similarity) of each typical subject representation and discriminative node representation is calculated and mean similarity of the typical subjects with the discriminative nodes are calculated. The similar calculation is done to find the mean similarity of the typical subjects with the discriminative edges. For the struggling class, mean similarity of the struggling subjects to the discriminative nodes, and mean similarity of the struggling subjects to the discriminative edges are calculated. In the discriminative subgraph (Fig. 8), there are 14 discriminative nodes

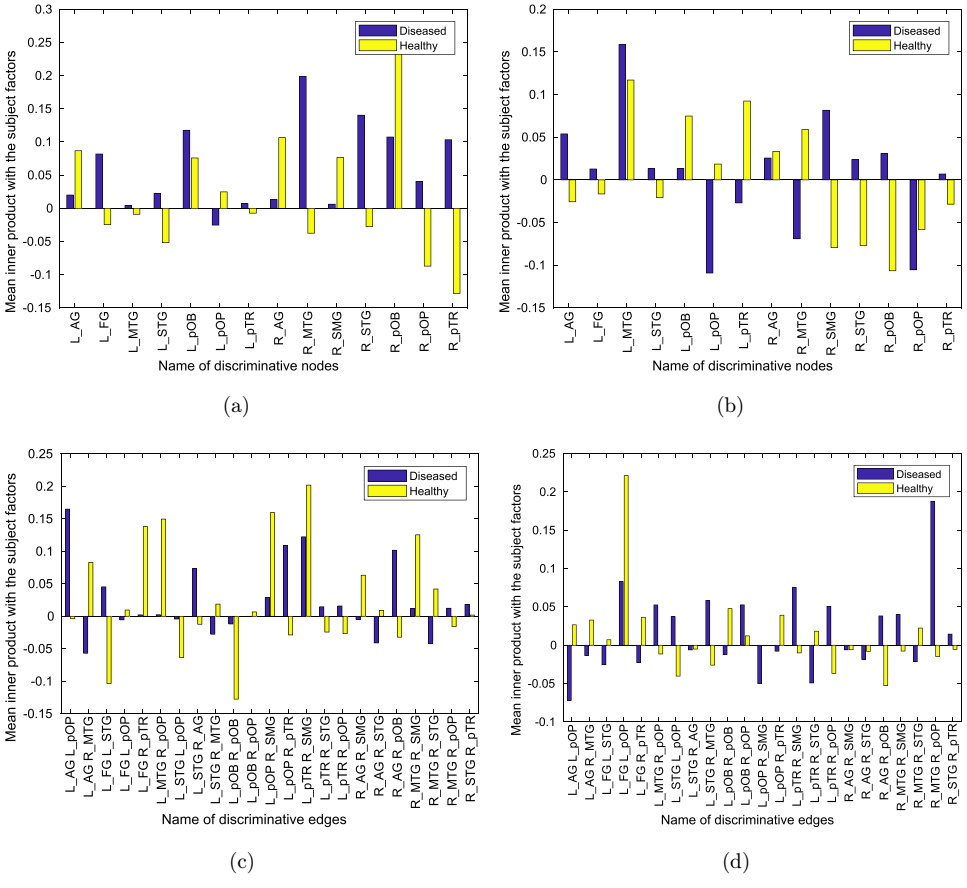
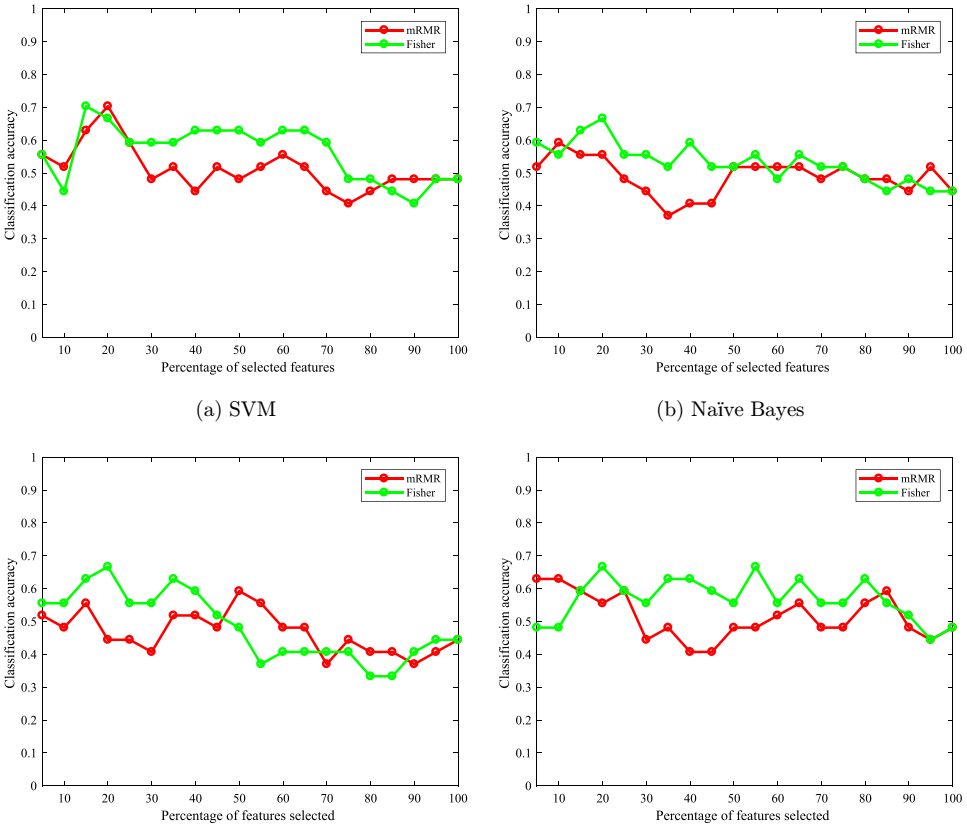


Fig. 9. CP and Tucker decomposition-based representations of the discriminative nodes and edges. (a) Discrimination made by CP-based node factors, (b) discrimination by Tucker-based node factors, (c) discrimination by CP-based edge factors and (d) discrimination by Tucker-based edge factors.

and 23 discriminative edges. From Fig. 9, we see that almost all the discriminative nodes and edges show a high difference in mean similarity with the typical and struggling class. Some less discriminative patterns are also found by this verification, such as the node *L_MTG* and the edge (*R_STG*, *R_PTR*). Figure 8 also supports the fact that the edge (*R_STG*, *R_PTR*) has comparatively smaller Fisher score than other edges.

5.4. *Performance in different classifiers*

The selected features by mRMR and Fisher algorithms are tested using four classifiers — Support Vector Machine (SVM) with radial basis function kernel, Naïve Bayes, knn (number of neighbors = 1), and knn (number of neighbors = 10). We observe that in most of the cases, Fisher-based feature sets result in higher



(a) SVM

(b) Naïve Bayes

(c) knn (number of neighbors = 1)

(d) knn (number of neighbors = 10)

Fig. 10. LOO accuracies of different classifiers after selecting features in mRMR and Fisher methods.

classification accuracy than mRMR-based feature sets (Fig. 10). Around top-20% Fisher-selected features give maximum accuracy in most of the cases. The better discrimination ability of Fisher-selected top-20% features in most of the classifiers also supports the selection of 23 features ($\approx 20\%$ of the total number of features) for constructing the discriminative subgraph for this dataset.

6. Conclusion

In this work, we find discriminative patterns, i.e. ROIs and connections from fMRI-based complete functional connectivity networks that can act as biomarkers for certain neurological and neurodevelopmental diseases such as reading disabilities in adults. Instead of representing the complete functional connectivity networks as thresholded graphs and using computationally expensive frequent/discriminative subgraph mining algorithms, we represent each complete functional connectivity

network by threshold-free connectivity vector and apply univariate/multivariate feature selection algorithms for finding the important features, which eventually become the edges of the discriminative subgraph. We used tensor decomposition-based approaches for finding meaningful representations of the discriminative ROIs and connections.

In the future, we aim to extend this idea in the presence of dual functional connectivity networks from different modalities such as fMRI, EEG, and DTI for each subject, and also in the presence of multiple side-view-based data such as immunologic, serologic, and clinical measures of the subjects. Other multivariate time series data, such as flare prediction data [29], where the features represent different solar weather parameters, can be represented by correlation matrices, and feature selection and tensor decomposition-based approaches can be used for increasing interpretability of the models. Therefore, we plan to extend our method to other types of multivariate time series data to test its applicability on a larger scale.

References

[1] C.-Y. Wee, P.-T. Yap, D. Zhang, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang and D. Shen, Identification of mci individuals using structural and functional connectivity networks, *Neuroimage* **59**(3) (2012) 2045–2056.

[2] V. D. Calhoun, P. K. Maciejewski, G. D. Pearlson and K. A. Kiehl, Temporal lobe and default hemodynamic brain modes discriminate between schizophrenia and bipolar disorder, *Human Brain Mapping* **29**(11) (2008) 1265–1275.

[3] B. Cao, L. Zhan, X. Kong, S. Y. Philip, N. Vizueta, L. L. Altshuler and A. D. Leow, Identification of discriminative subgraph patterns in fmri brain networks in bipolar affective disorder, in *Proc. Int. Conf. Brain Informatics and Health*, 2015, pp. 105–114.

[4] A. dos Santos Siqueira, C. E. Biazoli Junior, W. E. Comfort, L. A. Rohde and J. R. Sato, Abnormal functional resting-state networks in ADHD: Graph theory and pattern recognition analysis of fmri data, *BioMed Res. Int.* **2014** (2014).

[5] L. Q. Uddin, K. Supekar, C. J. Lynch, A. Khouzam, J. Phillips, C. Feinstein, S. Ryali and V. Menon, Salience networkbased classification and prediction of symptom severity in children with autism, *JAMA Psych.* **70**(8) (2013) 869–879.

[6] E. S. Finn, X. Shen, J. M. Holahan, D. Scheinost, C. Lacadie, X. Papademetris, S. E. Shaywitz, B. A. Shaywitz and R. T. Constable, Disruption of functional networks in dyslexia: A whole-brain, data-driven analysis of connectivity, *Biol. Psychiatry* **76**(5) (2014) 397–404.

[7] B. Cao, X. Kong, J. Zhang, S. Y. Philip and A. B. Ragin, Mining brain networks using multiple side views for neurological disorder identification, in *Proc. IEEE Int. Conf. Data Mining*, 2015, pp. 709–714.

[8] X. Kong, P. S. Yu, X. Wang and A. B. Ragin, Discriminative feature selection for uncertain graph classification, in *Proc. SIAM Int. Conf. Data Mining*, 2013, pp. 82–93.

[9] B. Jie, D. Zhang, W. Gao, Q. Wang, C.-Y. Wee and D. Shen, Integration of network topological and connectivity properties for neuroimaging classification, *IEEE Trans. Biomed. Eng.*, 2014, pp. 576–589.

[10] X. Yan and J. Han, Span: Graph-based substructure pattern mining, in *Proc. IEEE Int. Conf. Data Mining*, 2002, pp. 721–724.

[11] Y. Zhu and I. Cribben, Graphical models for functional connectivity networks: Best methods and the autocorrelation issue,*bioRxiv*, 2017, 128–488.

[12] B. Cao, L. He, X. Wei, M. Xing, P. S. Yu, H. Klumpp and A. D. Leow, t-bne: Tensor-based brain network embedding, in *Proc. SIAM Int. Conf. Data Mining*, 2017, pp. 189–197.

[13] B. Cao, L. He, X. Kong, S. Y. Philip, Z. Hao and A. B. Ragin, Tensor-based multi-view feature selection with applications to brain diseases, in *Proc. IEEE Int. Conf. Data Mining*, 2014, pp. 40–49.

[14] B. Cao, C.-T. Lu, X. Wei, S. Y. Philip and A. D. Leow, Semi-supervised tensor factorization for brain network analysis, in *Proc. Joint European Conf. Machine Learning and Knowledge Discovery in Databases*, 2016, pp. 17–32.

[15] S. M. Hamdi, B. Aydin, S. Filali Boubrahimi, R. Angryk, L. Crystal Krishnamurthy and R. Morris, Biomarker detection from fMRI-based complete functional connectivity networks, in *Proc. IEEE Int. Conf. Artificial Intelligence and Knowledge Engineering*, 2018, pp. 17–24.

[16] S. M. Hamdi, Y. Wu, S. Filali Boubrahimi, R. Angryk, L. Crystal Krishnamurthy and R. Morris, Tensor decomposition for neurodevelopmental disorder prediction, in *Proc. Int. Conf. Brain Informatics*, 2018, pp. 339–348.

[17] N. Shervashidze, P. Schweitzer, E. J. v. Leeuwen, K. Mehlhorn and K. M. Borgwardt, Weisfeiler-Lehman graph kernels, *J. Machine Learning Res.* **12** (2011) 2539–2561.

[18] Q. Gu, Z. Li and J. Han, Generalized fisher score for feature selection, arXiv:1202.3725, 2012.

[19] H. Peng, F. Long and C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(8) (2005) 1226–1238.

[20] G. Roffo, S. Melzi, U. Castellani and A. Vinciarelli, Infinite latent feature selection: A probabilistic latent graph-based ranking approach, in *Proc. IEEE Int. Conf. Computer Vision*, 2017, pp. 1226–1238.

[21] T. G. Kolda and B. W. Bader, Tensor decompositions and applications, *SIAM Review* **51**(3) (2009) 455–500.

[22] B. W. Bader and T. G. Kolda, Matlab tensor toolbox version 2.6 (2015), Available online, February 2015. [online]. Available: http://www.sandia.gov/∼tgkolda/Tensor-Toolbox/index-2.6.html.

[23] V. Krishnamurthy, L. C. Krishnamurthy, D. M. Schwam, A. Ealey, J. Shin, D. Greenberg and R. D. Morris, Retrospective correction of physiological noise: Impact on sensitivity, specificity, and reproducibility of resting-state functional connectivity in a reading network model, *Brain Connectivity* **8**(2) (2018) 94–105.

[24] R. W. Cox, Afni: Software for analysis and visualization of functional magnetic resonance neuroimages, *Comput. Biomedical Res.* **29**(3) (1996) 162–173.

[25] J. Ashburner, G. Barnes, C. Chen, J. Daunizeau, G. Flandin, K. Friston, S. Kiebel, J. Kilner, V. Litvak, R. Moran *et al.*, SPM12 Manual, Wellcome Trust Centre for Neuroimaging, London, UK, 2014.

[26] B. Fischl, Freesurfer, *Neuroimage* **62**(2) (2012) 774–781.

[27] A. Martin, M. Schurz, M. Kronbichler and F. Richlan, Reading in the brain of children and adults: A meta-analysis of 40 functional magnetic resonance imaging studies, *Human Brain Mapping* **36**(5) (2015) 1963–1981.

[28] S. Whitfield-Gabrieli and A. Nieto-Castanon, Conn: A functional connectivity toolbox for correlated and anticorrelated brain networks, *Brain Connectivity* **2**(3) (2012) 125–141.

[29] S. M. Hamdi, D. Kempton, R. Ma, S. F. Boubrahimi and R. A. Angryk, A time series classification-based approach for solar flare prediction, in *Proc. IEEE Int. Conf. Big Data*, 2017, pp. 2543–2551.