



# Multi-modal classification of neurodegenerative disease by progressive graph-based transductive learning



Zhengxia Wang<sup>a,b,f</sup>, Xiaofeng Zhu<sup>b,e</sup>, Ehsan Adeli<sup>b</sup>, Yingying Zhu<sup>b</sup>, Feiping Nie<sup>c</sup>, Brent Munsell<sup>d</sup>, Guorong Wu<sup>b,\*</sup>, for the ADNI and PPMI

<sup>a</sup> Department of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, 400074, PR China

<sup>b</sup> Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA

<sup>c</sup> School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, PR China

<sup>d</sup> Department of Computer Science, College of Charleston, Charleston, SC 29424, USA

<sup>e</sup> Department Computer Science and Information Engineering, Guangxi Normal University, Guilin, 541004, PR China

<sup>f</sup> Department of Automation, Chongqing University, Chongqing, 400044, PR China

## ARTICLE INFO

### Article history:

Received 15 June 2016

Revised 27 January 2017

Accepted 9 May 2017

Available online 13 May 2017

### Keywords:

Graph-based transductive learning (GTL)

Multi-modality

Intrinsic representation

Computer-assisted diagnosis

## ABSTRACT

Graph-based transductive learning (GTL) is a powerful machine learning technique that is used when sufficient training data is not available. In particular, conventional GTL approaches first construct a fixed inter-subject relation graph that is based on similarities in voxel intensity values in the feature domain, which can then be used to propagate the known phenotype data (*i.e.*, clinical scores and labels) from the training data to the testing data in the label domain. However, this type of graph is exclusively learned in the feature domain, and primarily due to outliers in the observed features, may not be optimal for label propagation in the label domain. To address this limitation, a progressive GTL (pGTL) method is proposed that gradually finds an intrinsic data representation that more accurately aligns imaging features with the phenotype data. In general, optimal feature-to-phenotype alignment is achieved using an iterative approach that: (1) refines inter-subject relationships observed in the feature domain by using the learned intrinsic data representation in the label domain, (2) updates the intrinsic data representation from the refined inter-subject relationships, and (3) verifies the intrinsic data representation on the training data to guarantee an optimal classification when applied to testing data. Additionally, the iterative approach is extended to multi-modal imaging data to further improve pGTL classification accuracy. Using Alzheimer's disease and Parkinson's disease study data, the classification accuracy of the proposed pGTL method is compared to several state-of-the-art classification methods, and the results show pGTL can more accurately identify subjects, even at different progression stages, in these two study data sets.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

In the elderly population, neurodegenerative diseases, such as Alzheimer's diseases (AD) and Parkinson's disease (PD), are the most common types of neurological disorders. Because of the progressive nature of these disorders, memory and other mental functions gradually worsen over time, which eventually affects the patients' quality of life (Group, 2004; Reisberg et al., 2008; Thompson et al., 2007). Unfortunately, there is no cure for these neurodegenerative diseases, although treatments include medications and management strategies may improve the quality of life. Therefore, timely and accurate diagnosis of neurodegenerative diseases and

its prodromal stage, *i.e.*, mild cognitive impairment (MCI) for AD, is highly desired in practice. MCI stage can be further categorized into progressive MCI (pMCI) and stable MCI (sMCI). Since an overwhelming amount of literature exists (Mueller et al., 2005; Ohtsuka et al., 2013) that relate neurodegenerative impairments to morphological abnormalities in the brain, MRI studies that reveal the structural abnormalities of the brain, or PET and SPECT studies that reveal the functional abnormalities of the brain have been widely used. Furthermore, methods that combine structural and functional neuroimaging data have been used to guide computer aided diagnosis techniques (Long et al., 2012; Ohtsuka et al., 2013; Prashanth et al., 2014; Rana et al., 2014; Salvatore et al., 2014; Weiner et al., 2013). More specifically, a technique called OPLS (orthogonal partial least squares to latent structures) is used to distinguish subjects with AD and MCI from healthy controls by combining MRI and CSF data (Westman et al., 2012). Joint feature and sample selection

\* Corresponding author.

E-mail addresses: [zxwang619@gmail.com](mailto:zxwang619@gmail.com) (Z. Wang), [grwu@med.unc.edu](mailto:grwu@med.unc.edu) (G. Wu).

methods based on SVM classification model for classification of AD and PD related diseases are also proposed in (Adeli et al., 2016; An et al., 2016). Other machine learning methods, such as kernel learning methods (Liu et al., 2014; Peng et al., 2016), subspace learning methods (Hu et al., 2016; Zhu et al., 2016), random forest (Gray et al., 2013b), deep learning (Liu et al., 2015a) and graph fusion (Tong et al., 2015; Wang et al., 2014a), have also been used to guide the classification of neurodegenerative diseases. However, morphological abnormalities are often subtle when compared to the high inter-subject variations (Zhu et al., 2013). Hence, sophisticated pattern recognition methods are of high demand to accurately identify individuals at different stages of neurodegenerative disease.

On the other hand, medical imaging applications also have various challenges that are related to high feature dimensionality, large data heterogeneity, and the small number of samples with ground-truth labels (e.g., diagnosis scores). Furthermore, even if a large number of labeled samples exist, it is very difficult to identify a computational model that will work well with the entire set of data due to large inter-subject variations across individuals. Transductive learning is a semi-supervised learning (SSL) method, which is recently emerged in the machine learning domain, introducing a strategy halfway between supervised and unsupervised learning schemes to improve classification performance by exploring the relationship between both labeled and unlabeled samples (Adeli-Mosabbeh and Fathy, 2015; Joachims, 2003; Zhou and Burges, 2007; Zhu et al., 2005). Here, the labeled samples are used to guide the transductive learning, while the unlabeled samples are used to maintain the intrinsic geometric structure of the observed samples. In particular, the graph-based SSL takes advantage of computational efficiency and representational ease for the medical imaging data. Because of the graph structures, it is more efficient to integrate different types of data for better explanations of the clinical outcomes (Kim et al., 2013). Since graph is usually used to describe the data manifold, most of the proposed transductive learning methods fall to the category of graph-based transductive learning (Blum and Chawla, 2001; Zhou et al., 2004; Zhu et al., 2005).

Graph-based transductive learning is widely used in image retrieval, image segmentation, data clustering and classification (Huang et al., 2014; Liu and Chang, 2009; Wang et al., 2014a; Zhang et al., 2015). For example, a fast and robust graph-based transductive learning method was proposed in (Zhang et al., 2015) by using a minimum tree cut, which was designed for large-scale web-spam detection and interactive image segmentation. Also, graph-based transductive learning methods have been investigated with great success in medical imaging area (Gao et al., 2015; Kim et al., 2013; Tong et al., 2015), since it can overcome the above difficulties by taking advantage of the data representation on unlabeled testing subjects. In the current state-of-the-art methods, each subject, regardless of being labeled or unlabeled, is often treated as a graph node. Then two subjects are connected by an edge in the graph if they both show similar morphological patterns. Using these connections, the labels can be propagated throughout the graph until all latent labels are determined. Typically, there are two separate steps in graph-based transductive learning methods: (1) construct the graph, where the vertices represent the labeled and unlabeled samples and the edges reflect the similarity degree between two connected samples (Zhu et al., 2005); and (2) propagate labels from labeled samples to unlabeled samples. Many current label propagation strategies have been proposed to determine the latent labels of testing subjects based on the inter-subject relationships encoded in the graph (Wang and Tsotsos, 2016; Zhang et al., 2015).

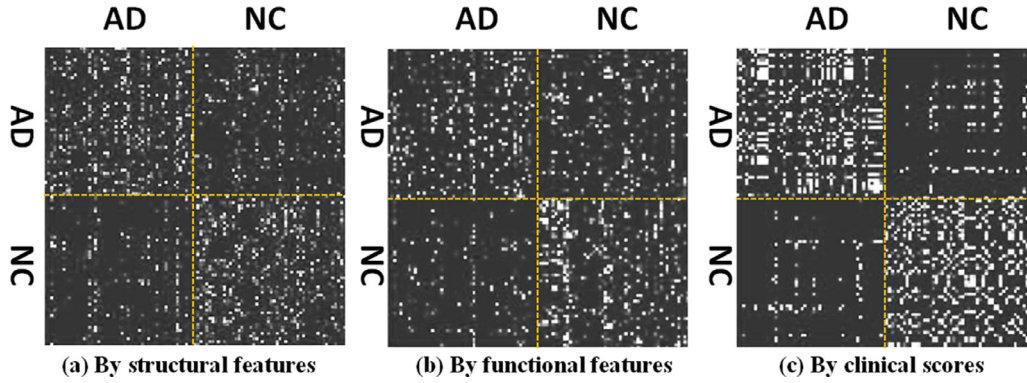
The basic assumption of current methods is that the graph constructed in the observed feature domain represents the real data

distribution and can be transferred to guide label propagation. However, this assumption usually does not hold, since the distribution of examples in the feature space does not necessarily cluster into groups as defined by the clinical scores and labels (Braak and Braak, 1995). Although the clinical scores and labels are different, they are highly correlated since the diagnosis is drawn upon the clinical score. Meanwhile, we believe the intrinsic data representation should be close or reflect the characteristic of the clinical score. Due to lack of ground truth, the underlying clinical score distribution used to validate the learned intrinsic data representation. As an example, Fig. 1(a) shows the affinity matrix of 51 AD and 52 NC subjects using the ROI-based features extracted from each MR image, where red dots and blue dots denote the high and low inter-subject similarities, respectively. Since the clinical data (e.g., MMSE and CDR scores (Thompson et al., 2007)) are more relevant with clinical labels, we use these clinical scores to construct another affinity matrix, as shown in Fig. 1(c). It is apparent that the data representations using imaging features and clinical scores are completely different. Thus, it is not guaranteed that the learned graph from the affinity matrix in Fig. 1(a) can effectively guide the classification of AD and NC subjects. More critically, the affinity matrix using observed image features is not even necessarily optimal in the feature domain, due to possible imaging noises and outlier subjects. In the literature, many studies have taken advantage of multi-modal information to improve discrimination power of transductive learning. However, the graphs from different modalities might also be different, as shown in the affinity matrices using structural image features from MR images (Fig. 1(a)) and functional image features from PET images (Fig. 1(b)). Although recent graph diffusion technique (Wang et al., 2014a) is effective in finding a common graph from multiple graphs, as shown in Fig. 1, it is hard to find a combination for the graphs in Figs. 1(a) and (b) that can be similar to the graph in Fig. 1(c), which is more related with the final classification task.

To solve these issues, we propose a progressive graph-based transductive learning method to learn the intrinsic data representation for optimal label propagation. Specifically, the intrinsic data representation should be (a) in consensus with inter-subject relationships constructed by imaging features extracted from different modalities, (b) aligned with the clinical labels or scores, and (c) verified on the training data for label propagation. To that end, we simultaneously (1) refine the data representation (inter-subject graph) in the feature domain, (2) find the intrinsic data representation based on the constructed graphs on both multi-modal imaging data and the clinical labels of entire subject set (including known labels on training subjects and the tentatively-determined labels on testing subjects), and (3) propagate the clinical labels from training subjects to testing subjects, following the latest learned intrinsic data representation. Promising results have been achieved in identify subjects with neurodegenerative disease on two neurodegenerative databases (i.e., Alzheimer's disease (AD) and Parkinson's disease (PD)), each with two modality images (such as MR and PET/SPECT).

The rest of this paper is organized as follows. Section 2 presents our proposed progressive graph-based transductive learning method. After that, we apply the method to the two real brain neurodegenerative imaging databases (ADNI and PPMI datasets<sup>1</sup>), and present the comparison results to validate the advantages of our method in Section 3. Finally, we conclude our method in Section 4.

<sup>1</sup> Alzheimer's disease Neuroimaging Initiative (ADNI), and Parkinson's Progressive Markers initiative (PPMI)



**Fig. 1.** Affinity matrices using structural image features (a), functional image features (b), and clinical scores (c). Bright dots and dark dots indicate the high and low inter-subject similarities, respectively.

## 2. Method

Suppose we have  $N$  subjects  $\{I_1, \dots, I_p, I_{p+1}, \dots, I_N\}$ , which sequentially consist of  $P$  training subjects and  $Q$  ( $Q = N - P$ ) testing subjects. For  $P$  training subjects, the clinical labels  $\mathbf{F}_P = [\mathbf{f}_p]_{p=1, \dots, P}$  are known, where each  $\mathbf{f}_p \in [0, 1]^C$  is a binary coding vector indicating the clinical label from  $C$  classes. Our goal is to jointly determine the latent labels for  $Q$  testing subjects based on a set of their continuous likelihood vectors  $\mathbf{F}_Q = [\mathbf{f}_q]_{q=p+1, \dots, N}$ , where each element in the vector  $\mathbf{f}_q$  indicates the likelihood of the  $q$ -th subject belonging to one of  $C$  classes. For convenience, we concatenate  $\mathbf{F}_P$  and  $\mathbf{F}_Q$  into a single label matrix  $\mathbf{F} = [\mathbf{F}_P \mathbf{F}_Q]$ .

### 2.1. Graph-based transductive learning on single-modal imaging data

Graph-based transductive learning learns over both labeled and unlabeled samples, aiming to harness the structure of entire data representation to improve the prediction of the latent labels. For clarity, we first extract single-modality image features from each subject  $I_i$  ( $i = 1, \dots, N$ ), denoted as  $\mathbf{x}_i$ . Using measurement from each modality, a graph  $G = (V, E)$  can be constructed to model the relations among the  $N$  subjects, where the nodes  $V$  correspond to  $N$  subjects and the edges  $E$  are weighted by the similarities between linked subjects. In the conventional graph-based transductive learning methods, the inter-subject relationships are computed based on feature similarity, which is encoded in an  $N \times N$  feature affinity matrix  $\mathbf{A}$ . Each element  $a_{ij}$  ( $a_{ij} \geq 0$ ,  $i, j = 1, \dots, N$ ) in  $\mathbf{A}$  represents the feature affinity degree between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . Therefore, the graph construction can be divided into two steps: **graph topology definition and edge weight computation**.

For the graph topology definition, current methods can be classified into two categories (de Sousa et al., 2013; Zhu et al., 2005): 1) Using the fully-connected graph. A fully-connected graph is created with edges between all pairs of nodes. Similar nodes have larger edge weights between them. In these methods, usually the weights of a fully-connected graph can be simply learned, but the computational cost is relatively high. 2) Using sparse graph. The  $k$ -nearest neighbor ( $k$ NN) graph and  $\epsilon$ -neighborhood ( $\epsilon$ NN) graph are both the sparse graphs, in which each node connects to only a few nodes. Sparse graphs are computationally fast and can often provide good empirical performance. However, the neighborhood relationship changes with the change of hyperparameters. Hence, for the sake of generalizability, the  $k$ NN graphs are constructed for each modality in this paper.

The most direct way to compute the weight matrix  $\mathbf{A}$  (by defining the edge weight between each pair of nodes) is based on a given similarity measure; in practice, it generally redefines the weight matrix  $\mathbf{A}$  by using different measures for better interpretability. **Binary weighting** is the simplest method for assign-

ing edge weights, which is to set  $\mathbf{A} = \mathbf{E}$  directly (where  $\mathbf{E}$  is a binary matrix, indicating if there is an edge between each pair of the nodes). Obviously, such a scheme cannot provide any extra information beyond the graph topology. **RBF (Gaussian) kernel** is one of the most common methods to assign edge weights for a graph. RBF kernel computes the similarity between  $\mathbf{x}_i$  and  $\mathbf{x}_j$  by  $a_{ij} = \exp(-d(\mathbf{x}_i, \mathbf{x}_j)^2 / 2\sigma^2)$ , where  $d(\mathbf{x}_i, \mathbf{x}_j)$  is a pair-wise similarity measure. For instance, pair-wise Euclidean distance can be used here. In addition,  $\sigma$  is a scale parameter (de Sousa et al., 2013; Zhu et al., 2005). In practice, one can employ any meaningful measure for defining edge weights, such as **mutual information** that has been successfully applied to brain and gene network modeling or detecting non-linear relationships (Plis et al., 2014). Additionally, some post-processing and optimization processes can also be used to efficiently weight the edges. These processes are often referred to as **graph learning methods** (Nie et al., 2014; Wang et al., 2014a). Without loss of generality, we select RBF kernel as the similarity measure to define the pair-wise affinity degree  $a_{ij}$  as:

$$a_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2}\right) \quad (1)$$

where  $\sigma$  is the scale controlling the exponential penalty strength of Euclidean distance between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . Based on the affinity matrix  $\mathbf{A}$ , conventional methods determine the latent label for each testing subject  $I_q$  by solving a classic graph learning problem (Golub and Van Loan, 2012; Nocedal and Wright, 2006):

$$\hat{\mathbf{F}}_q = \arg \min_{\mathbf{F}_q} \sum_{i,j=1}^N \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 a_{ij}. \quad (2)$$

As shown in Fig. 1, the affinity matrix  $\mathbf{A}$  might not be closely related with the intrinsic data representation in the label domain. Therefore, it is necessary to find a hidden data representation which aligns with the clinical labels, rather than solely using the affinity matrix constructed based on imaging features. However, initially, labels on the testing subjects are not determined yet. In order to solve this chicken-and-egg dilemma, we propose to iteratively optimize the data representation of each observed imaging data and align the refined imaging data representations to a common space for reflecting the intrinsic data representation of phenotype data.

### 2.2. Progressive graph-based transductive learning

Instead of relying on the affinity matrix  $\mathbf{A}$ , we propose to find an intrinsic data representation  $\mathbf{T} = [t_{ij}]_{i,j=1, \dots, N}$  which is more relevant than using affinity matrix  $\mathbf{A}$  to guide the label propagation in Eq. (2). Therefore, the problem of determining the latent label

for each testing subject  $I_q$  becomes:

$$\arg \min_{\mathbf{F}_q, \mathbf{T}} \sum_{i,j=1}^N (\|\mathbf{f}_i - \mathbf{f}_j\|_2^2 t_{ij}) \text{ s.t. } t_{ij} \geq 0, \mathbf{t}'_i \mathbf{1} = 1. \quad (3)$$

where  $t_{ij}$  ( $t_{ij} \geq 0, i, j = 1, \dots, N$ ) denotes the latent intrinsic inter-subject relationship between subject  $I_i$  and  $I_j$ . Since the clinical labels on the testing subjects are unknown, joint optimization of latent clinical label  $\mathbf{F}_q$  and hidden intrinsic data representation  $\mathbf{T}$  in Eq. 3 is an ill-posed problem. In order to turn the energy function to a well-posed problem, we require that the latent intrinsic data representation  $\mathbf{T}$  should respect the affinity matrix  $\mathbf{A}$  as follows:

$$\arg \min_{\mathbf{F}_q, \mathbf{T}} \sum_{i,j=1}^N \left( \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 t_{ij} + \lambda \|a_{ij} - t_{ij}\|_2^2 \right) \text{ s.t. } t_{ij} \geq 0, \mathbf{t}'_i \mathbf{1} = 1. \quad (4)$$

where  $a_{ij}$  is computed by Eq. 1.  $\lambda$  is the parameter controlling the influence of affinity matrix  $\mathbf{A}$  on the estimation of  $\mathbf{T}$  (intrinsic data representation). Since the affinity degree  $a_{ij}$  is computed based on the observed imaging data  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , possible noisy/outlier features could bring a series of unrealistic feature similarities. In order to suppress the influence of noisy/outlier imaging features, we propose to estimate the optimal imaging data representation  $\mathbf{S} = [s_{ij}]_{N \times N}$  based on the observed imaging features, where the regularization term is enforced on  $s_{ij}$ :

$$\arg \min_{\mathbf{S}} \sum_{i,j=1}^N \left\{ \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij} + \eta s_{ij}^2 \right\} \text{ s.t. } s_{ij} \geq 0, \mathbf{s}'_i \mathbf{1} = 1. \quad (5)$$

where  $\eta$  is the scalar controlling the strength of regularization term. Although the optimization of inter-subject relationship  $s_{ij}$  (in Eq. 5) and the calculation of affinity value  $a_{ij}$  (in Eq. 1) are both driven by the imaging features, the optimized inter-subject relationship  $s_{ij}$  is more robust than  $a_{ij}$  to the deteriorated imaging features. Hence, the edge weights are learned in the optimization process.

By replacing the affinity degree  $a_{ij}$  with the optimal inter-subject relationship  $s_{ij}$ , we jointly optimize the intrinsic data representation  $\mathbf{T}$ , imaging data representation  $\mathbf{S}$ , and the latent clinical label  $\mathbf{F}_Q$  in the following energy function:

$$\arg \min_{\mathbf{S}, \mathbf{T}, \mathbf{F}} \sum_{i,j=1}^N \left\{ \mu \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 t_{ij} + \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij} + \eta s_{ij}^2 + \lambda \|s_{ij} - t_{ij}\|_2^2 \right\} \text{ s.t. } s_{ij} \geq 0, \mathbf{s}'_i \mathbf{1} = 1, t_{ij} \geq 0, \mathbf{t}'_i \mathbf{1} = 1, \mathbf{F} = [\mathbf{F}_p \mathbf{F}_Q] \quad (6)$$

where  $\mu$  is the scalar balancing the data fitting terms from two different domains (i.e., the first and second terms in Eq. (6)). Note,  $s_{ij}$  and  $t_{ij}$  are required to be 0. The sum of inter-subject similarity degree of subject  $I_i$  to all other subjects equals to 1, i.e.,  $\mathbf{s}'_i \mathbf{1} = 1$  and  $\mathbf{t}'_i \mathbf{1} = 1$ , where  $\mathbf{s}_i$  and  $\mathbf{t}_i$  denote for the  $i$ -th column vectors of matrices  $\mathbf{S}$  and  $\mathbf{T}$ , respectively.

### 2.3. Progressive graph-based transductive learning on multi-modal imaging data

Recently, multi-modal neuroimaging data become more and more popular. For example, ADNI dataset provides a wide spectrum of neuroimaging data, which includes MR images and PET images. In order to improve the classification accuracy, we go one step further to extend our progressive graph-based transductive learning by fully using the complementary information in multi-modal data.

Suppose we have  $M$  modalities. For each subject  $I_i$ , we can extract multi-modal image features  $\mathbf{x}_i^m$ ,  $m = 1, \dots, M$ . For the  $m$ -th modality, we can optimize the imaging data representation  $\mathbf{S}^m$  of

imaging data  $\{\mathbf{x}_i^m\}_{i=1, \dots, N}$ . As shown in Figs. 1(a) and (b), the data representations across different modalities could be different. Thus, we require the intrinsic data representation  $\mathbf{T}$  to be close to all  $\mathbf{S}^m$ . To that end, we extend our above pGTL method from the single-modal to the multi-modal scenario:

$$\arg \min_{\mathbf{S}^m, \mathbf{T}, \mathbf{F}} \sum_{i,j=1}^N \left\{ \mu \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 t_{ij} + \sum_{m=1}^M \left[ \|\mathbf{x}_i^m - \mathbf{x}_j^m\|_2^2 s_{ij}^m + \eta (s_{ij}^m)^2 + \lambda \|s_{ij}^m - t_{ij}\|_2^2 \right] \right\} \text{ s.t. } s_{ij}^m \geq 0, (\mathbf{s}^m)' \mathbf{1} = 1, t_{ij} \geq 0, \mathbf{t}'_i \mathbf{1} = 1, \mathbf{F} = [\mathbf{F}_p \mathbf{F}_Q]. \quad (7)$$

The intuition behind Eq. 7 is that the label propagation is steered by the hidden intrinsic data representation  $\mathbf{T}$ . The criteria for obtaining reasonable estimation of  $\mathbf{T}$  are: (1)  $\mathbf{T}$  should be close to all imaging data representations  $\mathbf{S}^m$  estimated from the observed imaging features  $\{\mathbf{x}_i^m\}$  (as shown in the last term in Eq. (7)), which eventually makes  $\mathbf{T}$  act as a common space for  $\mathbf{S}^1, \dots, \mathbf{S}^M$ ; and (2) the label propagation results should be in consensus with the labels on the known subjects (the first term in Eq. (7)) such that the intrinsic data representation is essentially aligned with the phenotype data. It is apparent that our energy function describes a highly dynamic system since the variables are all correlated to each other. In the following, we give the optimization solution to Eq. 7, which falls into a divide-and-conquer scenario.

### 2.4. Optimization

Fortunately, our proposed energy function in Eq. (7) is convex with respect to each of the variables  $\mathbf{S}^m$ ,  $\mathbf{T}$ , and  $\mathbf{F}$ . Thus, we can alternatively optimize one set of variables at a time by fixing other sets of variables. The optimization for each sub-problem is detailed below.

#### 2.4.1. Estimation of imaging data representation $\mathbf{S}^m$ for each modality

Removing the unrelated terms w.r.t.  $\mathbf{S}^m$  in Eq. (7), the optimization of  $\mathbf{S}^m$  falls to the following objective function:

$$\arg \min_{\mathbf{S}^m} \sum_{i,j=1}^N \left\{ \|\mathbf{x}_i^m - \mathbf{x}_j^m\|_2^2 s_{ij}^m + \eta (s_{ij}^m)^2 + \lambda \sum_{i,j=1}^N \|s_{ij}^m - t_{ij}\|_2^2 \right\} \text{ s.t. } s_{ij} \geq 0, (\mathbf{s}^m)' \mathbf{1} = 1, t_{ij} \geq 0, \mathbf{t}'_i \mathbf{1} = 1. \quad (8)$$

Since Eq. (8) is independent of variables  $i$  and  $j$ , we further reformulate Eq. (8) in the vector form as below:

$$\min_{\mathbf{s}_i^m} \left\| \mathbf{s}_i^m + \frac{\mathbf{d}_i}{2r_1} \right\|_2^2 \text{ s.t. } s_{ij} \geq 0, \mathbf{s}'_i \mathbf{1} = 1. \quad (9)$$

where  $\mathbf{d}_i = [d_{ij}]_{j=1, \dots, N}$  is a column vector with each  $d_{ij} = \|\mathbf{x}_i^m - \mathbf{x}_j^m\|_2^2 - 2\lambda t_{ij}$ , and  $r_1 = \eta + \lambda$ . As shown in the Appendix, Eq. (9) has a closed-form solution. After we solve each  $\mathbf{s}_i^m$ , we can obtain the imaging data representation matrix  $\mathbf{S}^m$ .

#### 2.4.2. Estimation of intrinsic data representation $\mathbf{T}$

Fixing  $\mathbf{S}^m$  and  $\mathbf{F}$ , the objective function w.r.t.  $\mathbf{T}$  reduces to:

$$\arg \min_{\mathbf{T}} \sum_{i,j=1}^N \mu \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 t_{ij} + \lambda \sum_{m=1}^M \sum_{i,j=1}^N \left( \|s_{ij}^m - t_{ij}\|_2^2 \right), \text{ s.t. } s_{ij} \geq 0, (\mathbf{s}^m)' \mathbf{1} = 1, t_{ij} \geq 0, \mathbf{t}'_i \mathbf{1} = 1. \quad (10)$$

Similarly, we can reformulate Eq. (10) by solving each  $\mathbf{t}_i$  at a time:

$$\arg \min_{\mathbf{t}_i} \left\| \mathbf{t}_i + \frac{\mathbf{h}_i}{2r_2} \right\|_2^2 \text{ s.t. } t_{ij} \geq 0, \mathbf{t}'_i \mathbf{1} = 1. \quad (11)$$

where  $\mathbf{h}_i = [h_{ij}]_{j=1, \dots, N}$  is a vector with each element  $h_{ij} = \mu \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 - 2\lambda \sum_{m=1}^M s_{ij}^m$ , and  $r_2 = M\lambda$  is a scalar. Similar to the solution for Eq. (9), the problem in Eq. (11) can also be solved using a closed-form solution. After solving each  $\mathbf{t}_i$ , we can obtain the affinity matrix  $\mathbf{T}$ .

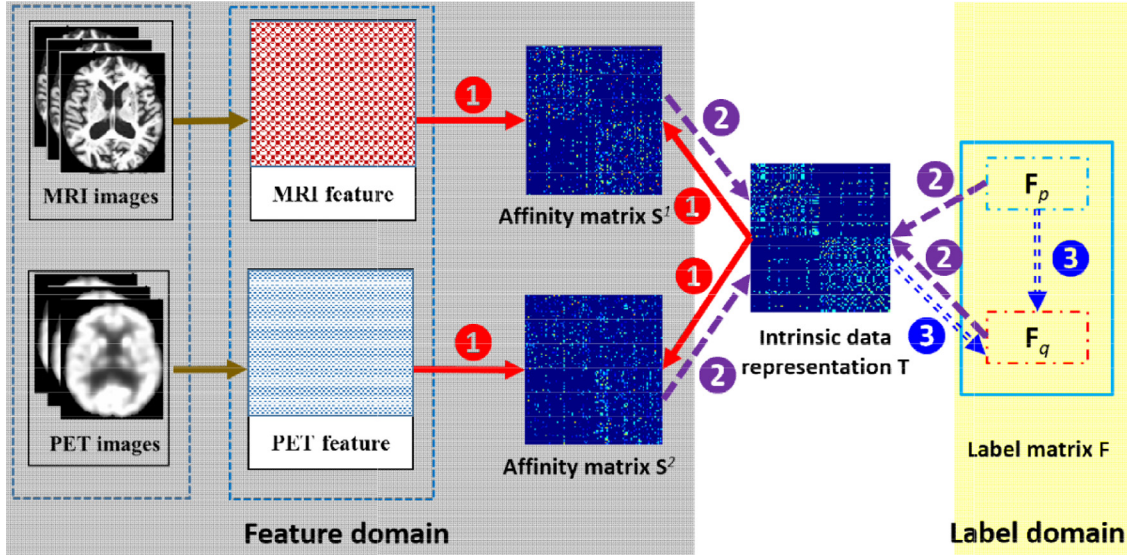


Fig. 2. The dynamic procedure of the proposed pGTL method. See text for details.

#### 2.4.3. Updating of latent labels $F_Q$ on testing subjects

Given both  $S^w$  and  $T$ , the objective function for the latent label  $F_Q$  can be derived from Eq. (3) as below:

$$\min_{\mathbf{F}} \sum_{i,j=1}^N \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 t_{ij}. \quad (12)$$

Eq. (12) is equal to the following problem:

$$\min_{\mathbf{F}} \text{trace}(\mathbf{F}\mathbf{L}'\mathbf{F}') = \min_{\mathbf{F}_Q} \text{trace}(\mathbf{F}\mathbf{L}'\mathbf{F}') \quad (13)$$

where  $\text{trace}(\cdot)$  denotes the matrix trace operator.  $\mathbf{L} = \text{diag}(\mathbf{T}) - (\mathbf{T}' + \mathbf{T})/2$  is the Laplacian matrix of  $\mathbf{T}$  ( $\text{diag}(\mathbf{T})$  denotes for the diagonal matrix of  $\mathbf{T}$ ).  $\mathbf{F}_p$  is with known clinical labels. By differentiating Eq. (13) w.r.t.  $\mathbf{F}$  and letting the gradient equal to zero such as  $\mathbf{L}\mathbf{F} = 0$ , we can obtain the following equation:

$$\begin{bmatrix} \mathbf{L}_{pp} & \mathbf{L}_{pq} \\ \mathbf{L}_{qp} & \mathbf{L}_{qq} \end{bmatrix} \begin{bmatrix} \mathbf{F}_p \\ \mathbf{F}_q \end{bmatrix} = 0, \quad (14)$$

where  $\mathbf{L}_{pp}$ ,  $\mathbf{L}_{pq}$ ,  $\mathbf{L}_{qp}$ , and  $\mathbf{L}_{qq}$  denote the top-left, top-right, bottom-left, and bottom-right blocks of  $\mathbf{L}$ . The solution for  $\mathbf{F}_Q$  can be obtained by  $\hat{\mathbf{F}}_Q = -(\mathbf{L}_{qq})^{-1}\mathbf{L}_{qp}\mathbf{F}_p$ .

The solution to the optimization problem in Eq. (7) is briefly summarized as follows.

**Algorithm 1:** Progressive transductive learning on multi-modal imaging data.

**Input:** Imaging data  $\{x_i^m\}_{m=1, \dots, M, i=1, \dots, N}$ , labels of labeled data  $\mathbf{F}_p \in R^{P \times C}$ , parameters  $\eta$ ,  $\lambda$  and  $\mu$ .

**Output:** Predicted labels of unlabeled data  $\mathbf{F}_Q \in R^{Q \times C}$ .

Compute the Euclidean distance between samples in each modality; Initialize  $\mathbf{S}^m$  using the affinity matrix  $\mathbf{A}^m$ , and initialize  $\mathbf{T}$  by letting

$$\mathbf{T} = \sum_{m=1}^M \mathbf{S}^m;$$

Initialize  $\mathbf{F}_Q = \{0\}^{Q \times C}$ .

**while** not converged **do**

1. Update  $\mathbf{F}_Q$ , which is obtained by  $\mathbf{F}_Q = -(\mathbf{L}_{qq})^{-1}\mathbf{L}_{qp}\mathbf{F}_p$  and  $\mathbf{L}$  is the Laplacian matrix of  $\mathbf{T}$ .

2. Update each imaging data representation  $\mathbf{S}^m$  in a column by column manner, where the optimization of each column vector  $\mathbf{s}_i^m$  in the matrix  $\mathbf{S}^m$  is shown in Eq. (9) and Appendix.

3. Update affinity matrix  $\mathbf{T}$  in a column by column manner, where the optimization of each column vector  $\mathbf{t}_i^m$  in the matrix  $\mathbf{T}^m$  is shown in Eq. (11) and Appendix.

**end while**

**Discussion.** Taking MRI and PET modalities as example, Fig. 2 illustrates the optimization of Eq. (7) by alternating the following three steps. (1) Estimate each imaging data representation  $\mathbf{S}^m$ , which depends on the observed imaging features  $\{\mathbf{x}_i^m\}$  and the currently estimated intrinsic data representation  $\mathbf{T}$  (red arrows); (2) Estimate the intrinsic data representation  $\mathbf{T}$ , which requires the estimations of both  $\mathbf{S}^1$  and  $\mathbf{S}^2$  and also the known clinical labels in the label domain (purple arrows); (3) Update the latent labels  $\mathbf{F}_Q$  on the testing subjects which needs guidance from the learned intrinsic data representation  $\mathbf{T}$  (blue arrows). It is apparent that the intrinsic data representation  $\mathbf{T}$  links the feature domain and label domain, which eventually leads to the dynamic graph learning model.

### 3. Experiments

In this study, we use two popular brain neurodegenerative databases *i.e.*, the Alzheimer's disease neuroimaging initiative (ADNI) database (<http://adni.loni.ucla.edu>) (Mueller et al., 2005) and the Parkinson's progression marker initiative (PPMI) database (<http://www.ppmi-info.org/data>) (Marek et al., 2011), to compare our proposed method with some state-of-the-art methods, *i.e.*, Support Vector Machine (SVM) (Suykens and Vandewalle, 1999), Safe Semi-Supervised Support Vector Machine (S4VM) (Li and Zhou, 2015), wellSVM (Li et al., 2013), supervised Joint Classification and Regression (JCR) (Wang et al., 2011), Canonical Correlation Analysis (CCA) based SVM (Thompson, 2005), Multi-Kernel SVM (MK-SVM) (Gönen and Alpaydin, 2011), and Graph-based Transductive learning (GTL) (Zhu et al., 2003). Specifically, the brief introduction of each of these comparison methods is given as follows:

- SVM: Support Vector Machine is a parametrically kernel-based supervised learning method, which maps the data into a higher dimensional input space and constructs an optimal separating hyperplane in this space. In our experiment, we use linear kernel.
- S4VM: Safe Semi-Supervised Support Vector Machine is a semi-supervised learning approach that does not significantly reduce learning performance when unlabeled data are used. This method uses multiple low-density separators to approximate the ground-truth decision boundary and maximizes the improvement in performance of inductive SVMs for any candidate separator. S4VM is semi-supervised learning method that guar-

**Table 1**  
Parameters and their explanations and respective ranges in competing methods.

Method	Parameters	Range
<b>SVM</b>	Regularization parameter controlling the margin	$[10^{-3}, 10^{-3}]$
<b>S4VM</b>	Weight for the hinge loss of labeled and unlabeled instance	$[10^{-2}, 10^2]$
<b>wellSVM</b>	Regularization parameter for labeled and unlabeled data	$[10^{-2}, 10^2]$
<b>JCR</b>	Regularization parameter	$[10^{-5}, 10^3]$
<b>MK-SVM</b>	Weight to blend two kernels	$[0.1, 0.9]$
<b>GTL</b>	Exponential decay factor $\sigma$ in computing affinity degree (Eq. 1)	$[2^{-5}, 2^5]$
<b>pGTL</b>	$\mu, \eta, \lambda$ , regularizing and balancing parameters in Eq. 7	$[10^3, 10^{-3}]$

antes the performance improvement using unlabeled data will be maximized (Li and Zhou, 2015).

- **wellSVM**: wellSVM is a semi-supervised method via a novel label generation strategy (Li et al., 2013). It is focused on the problem of learning from weakly labeled data, where labels of the training examples are incomplete. This method assumes different weakly labeled scenarios; including (i) semi-supervised learning, where labels are partially known; (ii) multi-instance learning, where labels are implicitly known; and (iii) clustering, in which labels are completely unknown. In this paper we use the first case, *i.e.* semi-supervised learning, to compare with our proposed method.
- **JCR**: This sparse joint classification and regression method utilizes the sparse regularization to perform imaging biomarker selection and learn a sparse matrix under a unified framework that integrates both heterogeneous and homogenous tasks (Wang et al., 2011). In this paper we obtain the classification results using this unified framework that integrate both label and clinical score information.
- **CCA-SVM**: Canonical correlation analysis is used to find the mappings for aligning two distributions of sets of multivariate variables (vectors), which makes the correlation between the projected variables to be mutually maximized (Thompson, 2005) after mapping. Then, we train the SVM classifier based on the projected features.
- **MK-SVM**: Multi-Kernel SVM method adequately utilizes the particular characteristic of each source and provides more possibility to choose suitable kernels or their weighted combination especially for the data from multiple heterogeneous sources (Gönen and Alpaydın, 2011). Each input has a kernel, and in this work a combination kernel (*i.e.*, weighted sum of all kernels) is used to classify.
- **GTL**: Graph-based transductive learning method is a semi-supervised learning method. The affinity matrix is constructed only in the feature domain but fixed in label propagation. In this experiment, we use the code of classic graph-based learning method in (Zhu et al., 2003).

For single-modality case, we only compare our proposed pGTL method with SVM method and GTL methods. For multiple-modality case, SVM, S4VM, wellSVM, JCR, CCA-SVM, MK-SVM, and GTL methods are used to compare with our pGTL method. Specifically, we apply SVM, S4VM, wellSVM, JCR, GTL methods to multi-modal imaging data by concatenating the feature vectors from all modalities into a single feature vector.

### 3.1. Experiments setting

**Evaluation measurements.** We evaluate the classification performance on four binary classification tasks: 1) AD vs NC, 2) MCI (Mild Cognitive Impairment) vs NC, 3) pMCI (progressive MCI) vs sMCI (stable MCI), and 4) PD vs NC. A set of quantitative measurements, such as Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), Positive Predictive Value (PPV), Negative Predictive Value (NPV) and Mean Predictive Value (MPV) are used to compare the classification performance of the competing methods in the experi-

ments. **Validation strategy.** Specifically, we follow a 10-fold cross-validation strategy, in which for each testing fold, the nine other folds are used to train the models. This is repeated for all ten existing folds and the performance score are averaged over these ten runs to illustrate reliable non-over-fitted results. In order to narrow down the factors affecting the classification performance, no feature selection step is included. **Parameter settings.** For all competing methods, the best parameters are selected through an inner 5-fold cross-validation on the training data using a grid-search strategy. The important parameters (along with their explanations and the respective ranges) used in each classification method are summarized in Table 1.

### 3.2. Experimental results on Alzheimer's disease

#### 3.2.1. Subjects and image preprocessing

In this study, we consider subjects with both MRI and PET modalities available in the ANDI database. As a result, we have 93 AD subjects, 202 MCI subjects, and 101 NC subjects. Specially, 55 pMCI subjects (who converted from MCI to AD in last 36 months) and 63 sMCI subjects (who didn't not convert to AD in both 24 months and 36 months) are included in pMCI vs sMCI classification. Each subject has both MR and 18-Fluoro-DeoxyGlucose PET (FDG-PET) images. The demographics of the subjects are detailed in Table 2.

For each subject, we first align the PET image to the MR image space. Then we remove both the skull and cerebellum from MR image, and segment MR image into white matter, gray matter and cerebrospinal fluid (Wang et al., 2014b; Zhang et al., 2011). Next, we parcellate each subject image into 93 ROIs (Regions of Interest) by registering the template (with manual annotation of 93 ROIs) to the subject image domain. Of note, these 93 ROIs cover important cortical and sub-cortical regions in human brain. Finally, the gray matter volume and the mean PET intensity in each ROI are used and form a 186-dimensional feature vector.

#### 3.2.2. Experimental results of classification performance

The classification performance by SVM, S4VM, wellSVM, JCR, CCA-SVM, MK-SVM, GTL and our proposed method are evaluated in three classification tasks (AD vs NC, MCI vs NC, and pMCI vs sMCI), respectively. Each task is conducted in both single-modal (MRI or PET) and multi-modal (MRI and PET) scenarios separately. Our proposed pGTL method achieves better classification performance compared to the other counterpart methods. Specifically, Table 3 shows the classification performance of the competing methods in the classification AD and NC. Our proposed pGTL method shows the best classification accuracies of 88.6%, 87.3% and 92.6% by using MRI, PET and (MRI+PET), respectively. Moreover, the performance improvements of classification accuracy over the second best counterpart method are 1.8% when using MRI only, 0.3% when using PET only, and 2.1% when using MRI+PET, respectively. Similarly, our proposed method achieves the best classification accuracy in MCI vs NC and pMCI vs sMCI tasks, as shown in Table 4 and Table 5, respectively.

The comparisons with recently published state-of-the-art methods are reported in Table 6. It summarizes the subject informa-

**Table 2**  
Demographic information of the subjects from the ADNI dataset. (SD: standard deviation).

	Female/male	Age (mean $\pm$ SD)[min-max]	Education (mean $\pm$ SD)[min-max]
AD (93)	36/57	75.38 $\pm$ 7.4 [55–88]	14.66 $\pm$ 3.2 [4–20]
MCI (202)	66/136	75.06 $\pm$ 7.1 [55–88]	15.71 $\pm$ 2.9 [7–20]
NC (101)	39/62	75.82 $\pm$ 4.8 [62–86]	15.82 $\pm$ 3.2 [7–20]
pMCI (55)	20/35	75.04 $\pm$ 6.7 [57–88]	16.00 $\pm$ 2.6 [12–20]
sMCI (63)	18/45	76.48 $\pm$ 6.7 [61–86]	15.46 $\pm$ 3.0 [7–20]

**Table 3**  
Comparison of AD/NC classification performance of the competing methods.

Modal	Method	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	MPV (%)
<b>MRI</b>	<i>S4VM</i>	86.1 $\pm$ 0.91	81.3 $\pm$ 1.55	90.4 $\pm$ 0.96	89.3 $\pm$ 0.88	84.9 $\pm$ 1.01	85.9 $\pm$ 0.92
	<i>JCR</i>	80.5 $\pm$ 1.57	78.4 $\pm$ 1.57	82.4 $\pm$ 3.02	81.6 $\pm$ 2.82	81.5 $\pm$ 1.25	80.4 $\pm$ 1.44
	<i>wellSVM</i>	86.5 $\pm$ 6.25	80.0 $\pm$ 12.6	91.8 $\pm$ 5.16	88.9 $\pm$ 6.12	85.6 $\pm$ 8.20	85.9 $\pm$ 6.73
	<i>SVM</i>	86.5 $\pm$ 1.38	82.2 $\pm$ 1.52	90.4 $\pm$ 1.55	89.6 $\pm$ 1.49	85.7 $\pm$ 1.32	85.7 $\pm$ 1.37
	<i>GTL</i>	86.8 $\pm$ 0.22	83.6 $\pm$ 1.25	89.8 $\pm$ 0.93	89.2 $\pm$ 0.90	86.6 $\pm$ 0.99	86.6 $\pm$ 0.26
	<i>pGTL</i>	<b>88.6 <math>\pm</math> 1.69</b>	<b>86.6 <math>\pm</math> 2.14</b>	<b>90.5 <math>\pm</math> 1.50</b>	<b>90.3 <math>\pm</math> 1.66</b>	<b>88.8 <math>\pm</math> 1.74</b>	<b>88.5 <math>\pm</math> 1.68</b>
<b>PET</b>	<i>S4VM</i>	86.1 $\pm$ 1.26	85.3 $\pm$ 1.61	86.9 $\pm$ 1.89	86.8 $\pm$ 2.01	87.7 $\pm$ 1.39	86.1 $\pm$ 1.28
	<i>JCR</i>	83.5 $\pm$ 2.12	80.3 $\pm$ 3.15	86.4 $\pm$ 2.49	85.6 $\pm$ 2.81	83.9 $\pm$ 2.23	83.4 $\pm$ 2.13
	<i>wellSVM</i>	87.0 $\pm$ 4.83	84.4 $\pm$ 7.77	<b>89.1 <math>\pm</math> 5.75</b>	86.7 $\pm$ 6.36	87.8 $\pm$ 5.76	86.8 $\pm$ 4.95
	<i>SVM</i>	86.0 $\pm$ 1.70	84.3 $\pm$ 2.69	87.6 $\pm$ 1.57	87.4 $\pm$ 1.53	86.9 $\pm$ 2.05	85.9 $\pm$ 1.73
	<i>GTL</i>	85.0 $\pm$ 1.20	83.6 $\pm$ 2.21	86.3 $\pm$ 0.65	86.3 $\pm$ 0.67	86.2 $\pm$ 1.77	85.0 $\pm$ 1.23
	<i>pGTL</i>	<b>87.3 <math>\pm</math> 1.47</b>	<b>86.9 <math>\pm</math> 2.20</b>	87.6 $\pm$ 1.93	<b>87.5 <math>\pm</math> 1.82</b>	<b>88.9 <math>\pm</math> 1.72</b>	<b>87.3 <math>\pm</math> 1.45</b>
<b>MRI+PET</b>	<i>S4VM</i>	88.8 $\pm$ 1.30	87.2 $\pm$ 2.16	90.3 $\pm$ 1.08	90.0 $\pm$ 1.41	89.3 $\pm$ 2.00	88.7 $\pm$ 1.27
	<i>JCR</i>	87.7 $\pm$ 1.73	86.7 $\pm$ 3.18	88.7 $\pm$ 1.67	88.9 $\pm$ 1.63	88.9 $\pm$ 2.08	87.7 $\pm$ 1.79
	<i>wellSVM</i>	90.5 $\pm$ 5.50	87.8 $\pm$ 8.19	92.7 $\pm$ 7.17	91.4 $\pm$ 8.13	90.6 $\pm$ 5.97	90.3 $\pm$ 5.55
	<i>SVM</i>	86.7 $\pm$ 1.42	85.5 $\pm$ 2.05	87.9 $\pm$ 1.54	87.7 $\pm$ 1.61	87.9 $\pm$ 1.78	86.7 $\pm$ 1.44
	<i>CCA-SVM</i>	89.1 $\pm$ 1.57	87.6 $\pm$ 2.02	90.5 $\pm$ 1.25	90.4 $\pm$ 1.38	89.5 $\pm$ 1.81	89.1 $\pm$ 1.57
	<i>MK-SVM</i>	90.0 $\pm$ 1.03	89.1 $\pm$ 1.53	90.7 $\pm$ 1.28	90.7 $\pm$ 1.13	90.8 $\pm$ 1.24	89.9 $\pm$ 1.01
	<i>GTL</i>	88.2 $\pm$ 1.08	86.7 $\pm$ 1.84	89.6 $\pm$ 1.23	89.3 $\pm$ 1.29	89.1 $\pm$ 1.37	88.2 $\pm$ 1.07
	<i>pGTL</i>	<b>92.6 <math>\pm</math> 0.65</b>	<b>92.2 <math>\pm</math> 1.34</b>	<b>92.9 <math>\pm</math> 1.37</b>	<b>92.9 <math>\pm</math> 1.36</b>	<b>93.3 <math>\pm</math> 1.20</b>	<b>92.5 <math>\pm</math> 0.71</b>

**Table 4**  
Comparison of MCI/NC classification performance of the competing methods.

Modal	Method	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	MPV (%)
<b>MRI</b>	<i>S4VM</i>	66.3 $\pm$ 1.28	80.1 $\pm$ 2.13	38.7 $\pm$ 5.85	72.5 $\pm$ 1.59	49.9 $\pm$ 2.10	59.4 $\pm$ 2.23
	<i>JCR</i>	65.4 $\pm$ 2.17	81.4 $\pm$ 3.86	33.4 $\pm$ 8.00	71.5 $\pm$ 2.10	46.9 $\pm$ 6.39	57.4 $\pm$ 3.07
	<i>wellSVM</i>	68.8 $\pm$ 4.53	73.2 $\pm$ 4.52	60.0 $\pm$ 7.66	78.6 $\pm$ 3.65	52.9 $\pm$ 6.13	66.6 $\pm$ 5.03
	<i>SVM</i>	68.7 $\pm$ 1.39	85.2 $\pm$ 1.16	35.9 $\pm$ 4.91	72.9 $\pm$ 1.35	59.6 $\pm$ 3.18	59.6 $\pm$ 2.19
	<i>GTL</i>	69.4 $\pm$ 1.14	79.8 $\pm$ 3.89	<b>48.7 <math>\pm</math> 9.10</b>	<b>77.0 <math>\pm</math> 2.66</b>	56.2 $\pm$ 4.47	56.3 $\pm$ 2.82
	<i>pGTL</i>	<b>70.7 <math>\pm</math> 0.81</b>	<b>86.3 <math>\pm</math> 2.40</b>	39.6 $\pm$ 6.08	75.0 $\pm$ 1.42	<b>58.7 <math>\pm</math> 4.08</b>	<b>62.9 <math>\pm</math> 1.99</b>
<b>PET</b>	<i>S4VM</i>	68.2 $\pm$ 1.27	83.9 $\pm$ 1.65	36.9 $\pm$ 2.17	73.0 $\pm$ 0.88	51.9 $\pm$ 4.70	60.4 $\pm$ 1.33
	<i>JCR</i>	66.6 $\pm$ 1.39	78.7 $\pm$ 1.46	42.3 $\pm$ 3.16	73.3 $\pm$ 1.16	50.8 $\pm$ 3.19	60.5 $\pm$ 1.69
	<i>wellSVM</i>	68.2 $\pm$ 5.76	80.5 $\pm$ 14.2	43.6 $\pm$ 23.3	75.1 $\pm$ 6.25	56.5 $\pm$ 18.3	62.0 $\pm$ 7.43
	<i>SVM</i>	66.5 $\pm$ 1.14	83.1 $\pm$ 2.84	33.5 $\pm$ 6.27	71.9 $\pm$ 1.29	50.2 $\pm$ 3.69	58.3 $\pm$ 3.69
	<i>GTL</i>	69.9 $\pm$ 0.83	80.1 $\pm$ 1.63	<b>49.8 <math>\pm</math> 4.07</b>	<b>77.0 <math>\pm</math> 1.44</b>	57.0 $\pm$ 2.73	64.9 $\pm$ 1.48
	<i>pGTL</i>	<b>72.5 <math>\pm</math> 0.76</b>	<b>85.7 <math>\pm</math> 1.52</b>	46.1 $\pm$ 3.93	76.7 $\pm$ 0.96	<b>63.3 <math>\pm</math> 3.59</b>	<b>65.9 <math>\pm</math> 1.44</b>
<b>MRI+PET</b>	<i>S4VM</i>	69.5 $\pm$ 2.17	83.9 $\pm$ 2.46	40.6 $\pm$ 4.37	74.2 $\pm$ 1.63	57.8 $\pm$ 3.57	62.3 $\pm$ 2.53
	<i>JCR</i>	67.8 $\pm$ 1.62	78.5 $\pm$ 3.02	46.5 $\pm$ 3.96	74.8 $\pm$ 1.30	52.7 $\pm$ 2.49	62.5 $\pm$ 1.64
	<i>wellSVM</i>	70.6 $\pm$ 4.05	86.8 $\pm$ 3.98	38.2 $\pm$ 10.3	73.9 $\pm$ 3.21	58.9 $\pm$ 10.3	62.5 $\pm$ 5.27
	<i>SVM</i>	69.2 $\pm$ 1.40	84.2 $\pm$ 1.19	39.1 $\pm$ 5.04	73.6 $\pm$ 1.37	57.6 $\pm$ 3.12	61.6 $\pm$ 2.22
	<i>CCA-SVM</i>	70.0 $\pm$ 2.15	81.7 $\pm$ 2.08	46.8 $\pm$ 4.97	75.4 $\pm$ 1.76	56.1 $\pm$ 4.21	64.2 $\pm$ 2.69
	<i>MK-SVM</i>	72.6 $\pm$ 1.98	72.2 $\pm$ 2.26	<b>72.9 <math>\pm</math> 2.07</b>	74.0 $\pm$ 1.76	72.7 $\pm$ 2.40	72.6 $\pm$ 1.98
	<i>GTL</i>	71.9 $\pm$ 0.94	<b>92.8 <math>\pm</math> 0.93</b>	29.9 $\pm$ 2.31	72.8 $\pm$ 0.68	67.9 $\pm$ 3.00	61.3 $\pm$ 1.19
	<i>pGTL</i>	<b>78.9 <math>\pm</math> 1.80</b>	85.5 $\pm$ 2.19	66.3 $\pm$ 3.39	<b>83.8 <math>\pm</math> 1.34</b>	<b>71.6 <math>\pm</math> 4.16</b>	<b>75.9 <math>\pm</math> 2.02</b>

tion, imaging modality, and average classification accuracy by using state-of-the-art methods. These comparison methods represent various machine learning techniques. Since the classification are not reported between pMCI and sMCI groups in (Gray et al., 2013b; Liu et al., 2015c; Peng et al., 2016; Tong et al., 2015), between MCI and NC groups in (Trzepak et al., 2014), we do not include the classification results and use ‘—’ instead. Our method achieves higher classification accuracy than both random forest and graph fusion methods, even though those two methods use additional CSF and genetic information.

Deep learning approach in (Liu et al., 2015b) learns feature representation in a layer-by-layer manner. Thus, it is time consuming to re-train the deep neural-network from scratch. Instead, our proposed method only uses handcrafted features for classification. It is noteworthy that we can complete the classification on a new

dataset (including grid-search for parameter tuning) within three hours on a regular PC (8 CPU cores and 16GB memory), which is much more economic than massive training cost in (Liu et al., 2015b). Complementary information in multi-modal data can help improve the classification performance; therefore, in order to find the intrinsic data representation, we combine our proposed pGTL with multi-modal information.

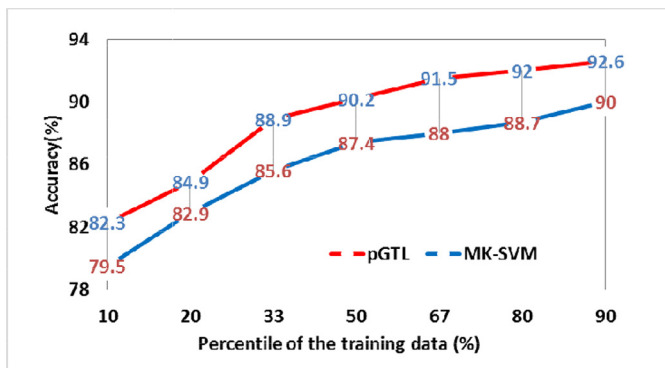
Besides, we also evaluate the classification performance *w.r.t.* the number of training samples using AD vs. NC classification as example, as shown in Fig. 3. It is clear that (1) our proposed method always has higher classification accuracy than MK-SVM methods; and (2) all methods can improve the classification accuracy as the number of training samples increases. It is worth noting that our proposed method achieves large improvement against MK-SVM, when only 33% of data is used as training samples. The

**Table 5**  
Comparison of pMCI/sMCI classification performance of the competing methods.

Modal	Method	ACC (%)	SEN (%)	SPE (%)	PPV (%)	NPV (%)	MPV (%)
<b>MRI</b>	<i>S4VM</i>	54.4 ± 2.28	47.2 ± 6.43	60.9 ± 4.87	51.8 ± 4.41	57.5 ± 2.91	54.4 ± 2.53
	<i>JCR</i>	56.8 ± 1.75	51.9 ± 4.20	61.0 ± 2.80	54.5 ± 3.71	59.9 ± 1.89	56.5 ± 1.95
	<i>wellSVM</i>	60.9 ± 7.48	46.0 ± 13.5	73.3 ± 8.61	58.7 ± 10.4	62.3 ± 6.29	59.7 ± 7.77
	<i>SVM</i>	57.1 ± 2.62	47.9 ± 6.34	65.1 ± 4.70	55.4 ± 3.40	59.6 ± 3.20	56.5 ± 2.66
	<i>GTL</i>	63.2 ± 3.31	53.1 ± 5.48	72.1 ± 2.69	62.1 ± 5.11	64.7 ± 3.17	63.2 ± 3.52
	<i>pGTL</i>	<b>65.8 ± 2.06</b>	<b>61.6 ± 5.88</b>	<b>69.8 ± 5.42</b>	<b>67.4 ± 3.36</b>	<b>70.4 ± 3.78</b>	<b>65.7 ± 2.03</b>
<b>PET</b>	<i>S4VM</i>	60.1 ± 2.66	49.3 ± 5.80	69.5 ± 3.00	59.5 ± 5.59	62.1 ± 2.62	59.4 ± 2.95
	<i>JCR</i>	66.3 ± 3.04	61.9 ± 4.58	70.0 ± 5.52	65.1 ± 4.77	69.5 ± 2.79	65.9 ± 2.94
	<i>wellSVM</i>	68.2 ± 12.3	<b>68.0 ± 13.9</b>	68.3 ± 18.3	66.8 ± 16.8	<b>72.1 ± 13.2</b>	68.2 ± 12.0
	<i>SVM</i>	64.8 ± 2.43	52.7 ± 4.26	75.6 ± 4.10	68.2 ± 4.61	65.3 ± 1.63	64.1 ± 2.37
	<i>GTL</i>	67.7 ± 1.27	57.8 ± 2.85	76.5 ± 2.33	68.7 ± 2.86	67.8 ± 1.37	67.7 ± 1.31
	<i>pGTL</i>	<b>69.7 ± 2.06</b>	52.6 ± 4.41	<b>84.4 ± 3.61</b>	<b>79.3 ± 3.89</b>	68.4 ± 2.00	68.5 ± 2.21
<b>MRI+PET</b>	<i>S4VM</i>	63.1 ± 2.35	49.9 ± 1.10	74.4 ± 3.43	63.4 ± 2.39	63.3 ± 1.81	62.1 ± 2.12
	<i>JCR</i>	66.6 ± 4.15	61.8 ± 5.06	70.8 ± 4.38	67.0 ± 5.30	68.4 ± 4.28	66.3 ± 4.19
	<i>wellSVM</i>	69.1 ± 9.77	<b>72.0 ± 13.9</b>	66.7 ± 13.6	65.2 ± 12.9	74.9 ± 11.1	69.3 ± 9.76
	<i>SVM</i>	68.6 ± 2.14	59.8 ± 13.7	76.8 ± 11.7	74.1 ± 11.3	69.3 ± 5.15	68.3 ± 2.64
	<i>CCA-SVM</i>	67.4 ± 1.20	42.6 ± 3.97	88.7 ± 3.42	77.5 ± 5.08	64.3 ± 0.79	65.7 ± 1.36
	<i>MK-SVM</i>	68.0 ± 1.63	43.0 ± 3.18	<b>89.8 ± 1.36</b>	<b>81.2 ± 2.48</b>	65.1 ± 1.46	66.4 ± 1.95
	<i>GTL</i>	69.7 ± 1.70	60.6 ± 3.28	77.9 ± 3.28	71.9 ± 3.30	69.4 ± 1.54	69.7 ± 1.80
	<i>pGTL</i>	<b>76.7 ± 1.76</b>	66.8 ± 3.09	85.0 ± 3.26	80.8 ± 3.37	<b>76.9 ± 1.77</b>	<b>75.9 ± 1.88</b>

**Table 6**  
Comparison with the classification accuracies reported in the literatures (%), ‘—’ represents the results are not reported in these papers.

Method	Subject information	Modality	AD/NC	MCI/NC	p/sMCI
Modal-fusion1 (Westman et al., 2012)	96AD+162MCI+111NC	MRI+CSF	91.8	77.6	68.5
Modal-fusion2 (Trzepacz et al., 2014)	20pMCI+30sMCI	MRI+PET	—	—	76
HMFSS (An et al., 2016)	165AD+342MCI+195NC	MRI+SNP	90.8	77.6	<b>78.3</b>
Kernel learning (Peng et al., 2016)	49AD+93MCI+47NC	MRI+PET	92.3	76.4	—
Feature-trans (Zhu et al., 2015)	198AD+403MCI+229NC	MRI-HOG+MRI-ROI	89.9	75.2	72.1
Random forest (Gray et al., 2013a)	37AD+75MCI+35NC	MRI+PET+CSF+Genetic	89.0	74.6	—
Graph fusion (Tong et al., 2015)	35AD+75MCI+77NC	MRI+PET+CSF+Genetic	91.8	79.5	—
Deep learning (Liu et al., 2015b)	85AD+169MCI+77NC	MRI+PET	91.4	<b>82.1</b>	—
Our method	99AD+202MCI+101NC	MRI+PET	<b>92.6</b>	78.6	76.7



**Fig. 3.** Classification accuracy as a function of the number of training samples used.

reason is that the supervised methods require a sufficiently large number of samples (with labels) for training a robust classifier. Otherwise, the classification performance decreases rapidly. On the contrary, our proposed pGTL method can alleviate this issue by leveraging the data distribution of both labeled and unlabeled data. Since the training samples with known labels are expensive to collect in medical imaging area, this experiment indicates the potential of our method in current neuroimaging studies.

To illustrate the representation of our method, confusion matrix is also introduced. Confusion matrix, also known as error matrix, is a specific table layout that allows visualization of the performance of an algorithm (Hay, 1988). In confusion matrix, each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class. The use of confusion matrix makes it easy to see if the system is confusing two classes.

### 3.3. Experimental results on Parkinson's disease

#### 3.3.1. Subject information and image preprocessing

Recently, a major initiative, the Parkinson Progression Marker Initiative (PPMI) (PPMI, 2011), was developed to identify and validate PD progression markers. Abundant imaging data from the enrolled PD subjects at the earliest detectable stage of disease significantly enhances the potential to both identify PD imaging markers and develop computer-assisted diagnosis system for neuroprotective interventions (Beitz, 2014; Jankovic, 2008; Stern and Siderowf, 2010). PD subjects in the PPMI study are just diagnosed and unmediated. The healthy/normal control subjects are both age- and gender-matched with the PD patients. In this research, we use 369 PD and 165 NC subjects, each with both MRI and SPECT modalities.

For MR images, a T1-weighted, 3D sequence (e.g., MPRAGE or SPGR) is acquired for each subject using 3T SIEMENS MAGNETOM Trio Tim syngo scanners. The T1-weighted images were acquired for 176 sagittal slices with the following parameters: repetition time = 2300 ms, echo time = 2.98 ms, flip angle = 9°, and voxel size = 1 × 1 × 1 mm<sup>3</sup>. All the MR images were preprocessed by skull stripping (Wang et al., 2014b), cerebellum removal, and then segmented into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) tissues (Lim and Pfefferbaum, 1989). The AAL atlas (Tzourio-Mazoyer et al., 2002), parcellated with 90 predefined regions of interest (ROI), was registered using HAMMER (Shen and Davatzikos, 2002) to each subject's native space. We further added 8 more ROIs to the atlas in the basal ganglia and brainstem regions, which are clinically important ROIs for PD. These 8 ROIs are 'superior cerebellar peduncle', 'midbrain', 'pons' and 'medulla oblongata' in the brainstem, along with 'substantia nigra' (left and right) and 'red nucleus' (left and right). We then computed WM, GM and CSF tissue volumes in each of these 98 ROIs as features.



n = 194		Predicted AD	Predicted NC	
Actual AD	TP = 86	FN = 7	93	
Actual NC	FP = 7	TN = 94	101	
	93	101		

a. AD vs NC classification

n = 303		Predicted MCI	Predicted NC	
Actual MCI	TP = 172	FN = 30	202	
Actual NC	FP = 34	TN = 67	101	
	206	97		

b. MCI vs NC classification

n = 118		Predicted pMCI	Predicted sMCI	
Actual pMCI	TP = 37	FN = 18	55	
Actual sMCI	FP = 9	TN = 54	63	
	46	72		

c. pMCI vs sMCI classification

n = 534		Predicted PD	Predicted NC	
Actual PD	TP = 359	FN = 10	369	
Actual NC	FP = 4	TN = 161	165	
	363	171		

d. PD vs NC classification

Fig. 4. Confusion matrix of classification results for proposed pGTL method.

To acquire SPECT images, the  $^{123}\text{I}$ -ilflupane neuroimaging radiopharmaceutical biomarker was injected, which binds to the dopamine transporters in the striatum. Brain images were then acquired. To process these images, the PPMI study has performed attenuation correction on the SPECT images, along with a standard 3D 6.0 mm Gaussian filter. Then, the images were normalized to standard Montreal Neurological Institute (MNI) space. Next, the transaxial slice with the highest striatal uptake was identified and the 8 hottest striatal slices around this slice were averaged, to generate a single slice image. On the averaged slice, the four caudate and putamen (left and right) ROIs, which are in the striatum brain region, were labeled and considered as target ROIs. The occipital cortex region was also segmented and used as a reference ROI. Count densities for the regions were used to calculate the striatal binding ratios (SBRs), which were used as morphological signatures for SPECT images.

### 3.3.2. Experimental results of classification performance

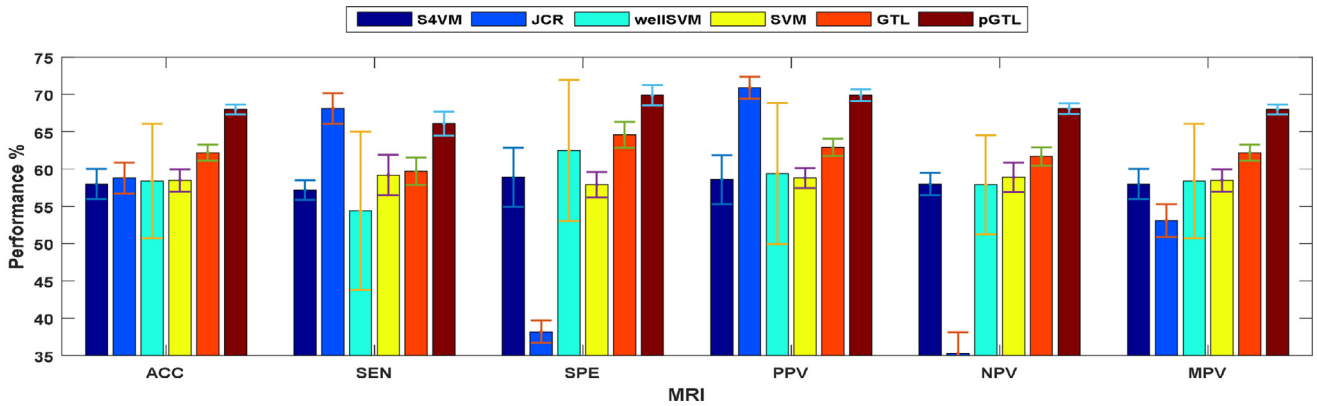
We randomly select 165 subjects out of 369 PD subjects to evaluate the classification performance with another 165 NC subjects. This is used to make the data balanced. Moreover, to prevent any unintended bias in the results, the process of random selection is repeated 5 times, and the average value of the 5 times of reputation is used as the final result as shown in Fig. 5.

In the single-modal MR image based classification of PD and NC subjects, the proposed method achieves the accuracy of 68.0%. Compared to other competing methods (S4VM, JCR, wellSVM, SVM and GTL) that achieve the accuracies of 58.0%, 58.8%, 58.4%, 58.5% and 62.2%, respectively, our proposed method improved by 10% over S4VM. For the case of using only SPECT images, the improvements of classification accuracy achieved by our pGTL method are less significant over other two methods (such as 95.4% by S4VM, 94.2% by JCR, 95.3% by wellSVM, 94.9% by SVM, 95.9% by GTL, and 96.6% by our pGTL), due to the high sensitivity of features from SPECT images. In multi-modal (MRI+SPECT) classification scenario, the overall classification accuracies are 92.9% by S4VM, 82.2% by JCR, 87.2% by wellSVM, 88.5% by SVM, 90.3% by CCA-SVM, 94.2% by MK-SVM, 85.1% by GTL, and 97.4% by our proposed pGTL method. It is apparent that our proposed pGTL method has achieved the highest classification performance in both single- and multi-modal classification scenarios. Confusion matrix about classification performance for PD vs NC is showed in Fig. 4(d).

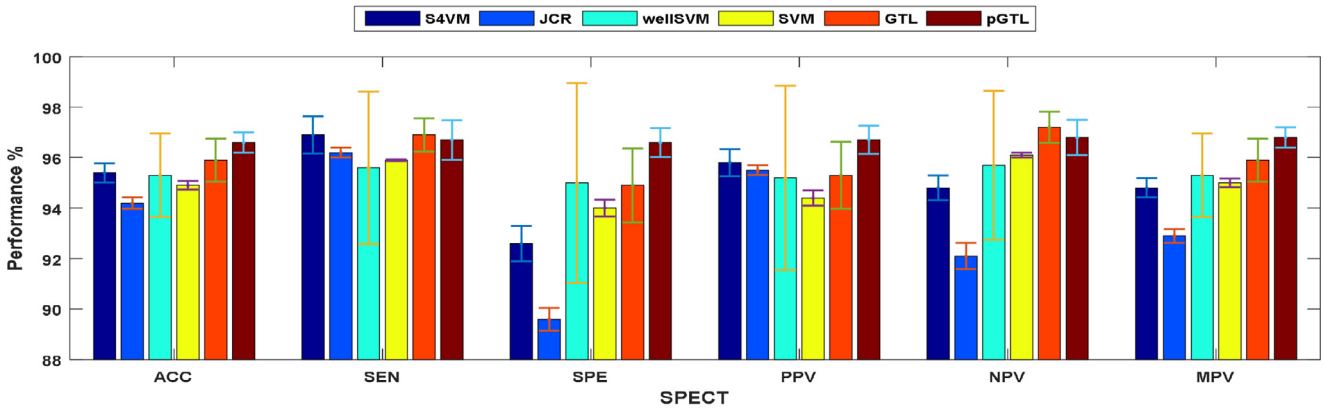
Since the SPECT image provides only four features, high-sensitivity morphological patterns are nominated by the overwhelming less-discriminative imaging features from MRI. Thus, the overall classification accuracy of the competing methods (except pGTL) using both MRI and SPECT data are lower than only using SPECT data, indicating high importance of using the state-of-the-art multi-modal classification method to combine the powers of different modalities. It is noteworthy that, although our proposed method does not learn the weights for different modalities, the learning process of finding the intrinsic data representation can adaptively adjust the effect of different modalities. On the other hand, CCA-SVM and MK-SVM methods can find *either* maximum correlation or suitable weighted kernel between different imaging modalities, thus improving the classification accuracies up to 90.3% by CCA-SVM and 94.2% by MK-SVM, respectively. Compared to CCA-SVM and MK-SVM, our proposed pGTL method uses data representation of unlabeled samples to guide the classification in a semi-supervised manner, which is very effective in alleviating the issue of small sample size. Thus, our proposed pGTL method can achieve the highest classification accuracy in classifying PD and NC by using both MRI and SPECT data.

### 3.4. Discussion

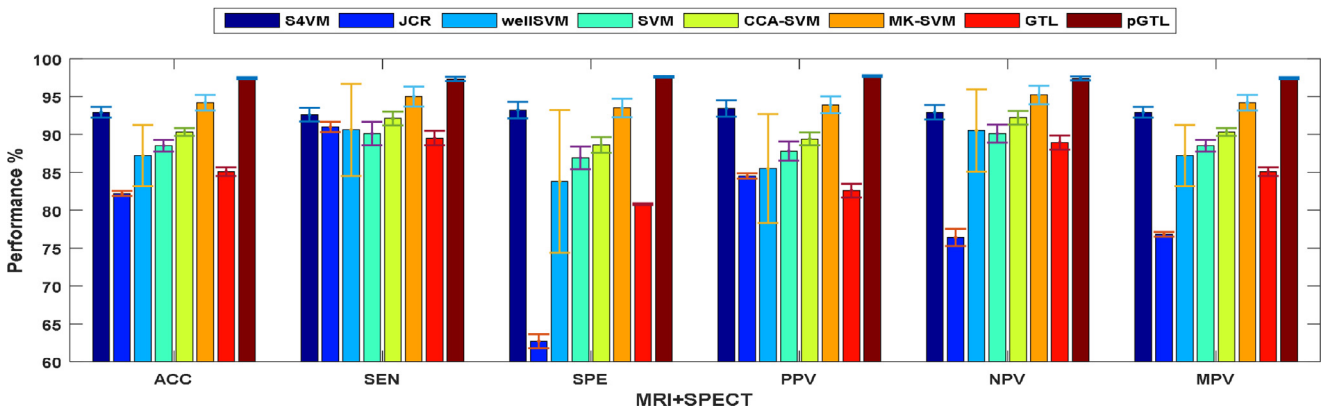
Feature extraction and data representation are always the very important steps in many classification tasks. Specifically, in medical imaging applications, deficiency in the imaging devices will be reflected as noisy or redundant features for the latter processes, which will reduce the overall learning performance of the classification system. **Feature selection** aims to choose a small subset of the relevant features from the original ones according to certain relevance evaluation criteria. This usually leads to better performance, lower computational cost and better model interpretability (Tang et al., 2014). **One possible strategy** is to integrate the classic feature selection and our graph-based transductive classification, where the input to our method will be the optimized features, instead of features extracted from the whole brain. To verify the effectiveness of our proposed method, we use the selected features from MRI reported in (Adeli et al., 2016), and then combine them with the features from SPECT for obtaining even better performance (ACC: 97.5%). **Furthermore**, we can simultaneously select the best features and also learn the data representation by intro-



(a) MRI classification performance



(b) SPECT classification performance



(c) Combination of MRI and PSECT classification performance

Fig. 5. Comparison of PD/NC classification performance of the competing methods.

ducing an additional variable for measuring the importance of each observed feature. However, in this paper, we focused mainly on the graph-learning strategy, since feature selection schemes have been widely explored in the literature. It is important to note that our proposed method can learn the importance of each feature, through looking into the graph weights and regularizing the optimization objective to enforce the selection of a compact set of features. This is a direction for our future work.

Lastly, biomarkers from different modalities provide complementary information, which is very useful for neurodegenerative disease diagnosis. However, it is clear that different modalities should be weighted differently. For example, the imaging features from SPECT in Section 3.3 have high-sensitivity morphologi-

cal patterns; when SPECT features are weighted equally with less-discriminative imaging features from MRI, the classification performance of multi-modalities will be reduced, which can be seen in Fig. 4. In our method, we can adaptively learn a weight for each graph during the optimization. However, this will lead to some additional parameters to optimize in our proposed method. Hence, in the current implementation, we treat each imaging modality equally. In the future, we will try to adopt a strategy similar to Auto-weighted Multiple Graph Learning (AMGL) framework in our method to learn a set of weights automatically for all the graphs. This process will not need any additional parameters (Nie et al., 2016).

#### 4. Conclusion

Here we presented a novel pGTL method that can accurately identify the different neurodegenerative stages for wide range of subjects, when applied to multi-modal imaging data. Compared to the conventional methods, the proposed method seeks to identify an intrinsic data representation that is simultaneously learned from the observed imaging features while also being validated on the training data with known phenotype labels. Since the learned intrinsic data presentation is more relevant to phenotype label propagation, the pGTL approach has shown promising results when performing AD vs NC, MCI vs NC, pMCI vs sMCI, and PD vs NC classification tasks when compared to several the state-of-the-art supervised and semi-supervised machine learning methods.

#### Acknowledgments

Parts of the data used in preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.ucla.edu>) and Parkinson's Progressive Markers initiative (PPMI) database (<http://www.ppmi-info.org>). The investigators within the ADNI and PPMI contributed to the design and implementation of ADNI and PPMI and/or provided data but did not participate in analysis or writing this paper. A complete listing of ADNI investigators can be found at: <http://adni.loni.ucla.edu/wp-content/uploads/howtoapply/ADNIAcknowledgementList.pdf>.

This work was supported in part by National Institutes of Health (NIH) grants (HD081467, EB006733, EB008374, EB009634, MH100217, AG041721, AG049371, AG042599, CA140413). Zhengxia Wang was supported in part by the National Natural Science Foundation of China (61273021), Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ1500501).

#### Appendix A

**Lemma 1.** Eq. (9) has a closed form solution.

For each  $i$ , the objective function in problem (8) is equal to the one in problem (9). The Lagrangian function of problem (9) is as

follows (Duchi et al., 2008):

$$\frac{1}{2} \left\| s_i + \frac{d_i}{2r_1} \right\|_2^2 - \eta (s_i'1 - 1) - \beta_i' s_i \quad (15)$$

where  $\eta$  and  $\beta_i \geq 0$  are the Lagrange multiplier to be determined. Differentiating with respect to  $s_{ij}$  and comparing to zero gives the optimality condition. And, according to KKT conditions (Boyd and Vandenberghe, 2004), we have the following equations:

$$\begin{cases} \forall j, s_{ij} + \frac{d_{ij}}{2r_1} - \eta - \beta_{ij} = 0 \\ \forall j, s_{ij} \geq 0 \\ \forall j, s_{ij}\beta_{ij} = 0 \\ \forall j, \beta_{ij} \geq 0 \end{cases} \quad (16)$$

The complementary slackness KKT condition implies that, whenever  $s_{ij} > 0$ , we must have  $\beta_{ij} = 0$ , so  $s_{ij} = -\frac{d_{ij}}{2r_1} + \eta$ . If  $s_{ij} \geq 0$ , it can be verified that the optimal solution  $s_{ij}$  should be

$$s_{ij} = \left( -\frac{d_{ij}}{2r_1} + \eta \right)_+ \quad (17)$$

where  $(a)_+ = \max(0, a)$  is the positive part of the variable  $a$ .

Therefore, the remaining problem is the estimation of  $\eta$  in Eq. 18. From Lemma 1 in (Duchi et al., 2008), suppose that  $d_{i1}, d_{i2}, \dots, d_{iN}$  are ordered from small to large. If the optimal  $s_i$  has only  $k$  nonzero elements, then according to Eq. (17), we know  $s_{ik} > 0$  and  $s_{i,k+1} = 0$ . Therefore, we have

$$\begin{cases} -\frac{d_{ik}}{2r_1} + \eta > 0 \\ -\frac{d_{i,k+1}}{2r_1} + \eta \leq 0 \end{cases} \quad (18)$$

According to Eq. (18) and the constraint  $s_i'1 = 1$ , we have

$$\sum_{j=1}^k \left( -\frac{d_{ij}}{2r_1} + \eta \right) = 1 \Rightarrow \eta = \frac{1}{k} + \frac{1}{2kr_1} \sum_{j=1}^k d_{ij} \quad (19)$$

After we solve each  $s_i^m$ , we can obtain the affinity matrix  $S^m$ . The convergence of our algorithm is  $O(n \log n)$  (Duchi et al., 2008).

#### Appendix B

Table 7.

Table 8.

**Table 7**  
IDs of the ADNI subjects.

Categories	ID of subjects
AD(93)	1257, 221, 929, 1341, 547, 653, 316, 1339, 1354, 786, 3, 10, 53, 183, 712, 720, 699, 1161, 1205, 991, 1263, 286, 682, 213, 343, 642, 1109, 219, 543, 1171, 1307, 850, 1254, 836, 1056, 321, 554, 147, 400, 1037, 889, 1281, 1283, 1285, 341, 577, 760, 1001, 627, 1368, 1391, 1044, 474, 1371, 1373, 1379, 535, 690, 730, 565, 1090, 1164, 1397, 1402, 149, 470, 492, 1144, 743, 747, 1062, 777, 1157, 374, 979, 370, 891, 1221, 431, 754, 1382, 167, 216, 266, 740, 1409, 1430, 1201, 1290, 497, 438, 841, 1041
NC(101)	610, 484, 498, 731, 751, 842, 862, 67, 419, 420, 2, 5, 8, 16, 21, 23, 637, 1133, 502, 575, 359, 43, 55, 97, 883, 647, 14, 96, 130, 985, 1063, 74, 120, 843, 845, 866, 618, 95, 734, 741, 48, 555, 576, 672, 813, 1023, 327, 454, 467, 262, 898, 1002, 779, 818, 934, 768, 1099, 315, 311, 312, 386, 363, 489, 526, 171, 90, 352, 533, 534, 47, 967, 1013, 173, 416, 360, 648, 657, 506, 680, 259, 230, 245, 272, 500, 522, 863, 778, 232, 1200, 123, 319, 283, 301, 459, 686, 972, 1194, 1195, 1197, 1202, 1203
MCI(202)	1074, 1122, 222, 546, 1224, 675, 1130, 101, 128, 293, 344, 414, 698, 1030, 1199, 161, 422, 904, 326, 362, 861, 1282, 634, 917, 932, 1033, 1165, 1175, 240, 325, 860, 1120, 1186, 1275, 354, 590, 1028, 1092, 57, 80, 142, 155, 141, 178, 424, 626, 544, 924, 961, 1351, 1394, 1393, 1400, 256, 408, 461, 485, 914, 1038, 1073, 1215, 1218, 1318, 1384, 294, 214, 718, 978, 511, 513, 567, 723, 906, 33, 204, 292, 997, 656, 673, 748, 945, 976, 1135, 1240, 150, 377, 552, 566, 1078, 1421, 282, 314, 407, 446, 549, 598, 679, 721, 1010, 1260, 1411, 1412, 1418, 1420, 1423, 1425, 1346, 389, 621, 919, 464, 941, 957, 1007, 1217, 1265, 1294, 1299, 1211, 1380, 746, 909, 1357, 641, 531, 1188, 1314, 1398, 1417, 160, 51, 54, 291, 551, 880, 958, 1034, 892, 930, 995, 1154, 950, 1114, 1343, 378, 410, 1103, 1106, 1118, 361, 1243, 1315, 1322, 708, 709, 865, 1077, 112, 394, 925, 1032, 1210, 1419, 1427, 135, 138, 188, 200, 205, 225, 227, 258, 608, 715, 770, 947, 1043, 1406, 1407, 1408, 1204, 1246, 285, 289, 783, 409, 987, 695, 158, 443, 481, 669, 722, 800, 825, 994, 1414, 1426, 1245, 1378, 1295, 1311
pMCI(56)	54, 57, 101, 128, 141, 161, 204, 214, 222, 240, 256, 258, 289, 294, 325, 344, 394, 461, 511, 549, 567, 675, 695, 708, 723, 860, 861, 892, 904, 906, 930, 941, 947, 978, 987, 997, 1007, 1010, 1033, 1077, 1130, 1135, 1217, 1240, 1243, 1282, 1295, 1299, 1311, 1393, 1394, 1398, 1412, 1423, 1427
sMCI(63)	33, 142, 150, 158, 178, 188, 200, 225, 285, 291, 414, 464, 481, 544, 546, 598, 608, 621, 626, 634, 656, 673, 679, 698, 709, 715, 718, 746, 748, 770, 783, 800, 914, 919, 925, 932, 950, 961, 1028, 1032, 1034, 1103, 1114, 1118, 1120, 1122, 1165, 1175, 1186, 1211, 1215, 1218, 1246, 1260, 1314, 1378, 1380, 1384, 1414, 1417, 1418, 1419, 1421

**Table 8**  
IDs of the PPMI subjects.

Categories	ID of subjects
PD(369)	3001, 3002, 3006, 3010, 3012, 3014, 3018, 3020, 3021, 3023, 3024, 3026, 3027, 3028, 3051, 3052, 3054, 3056, 3059, 3060, 3061, 3062, 3066, 3067, 3068, 3076, 3077, 3078, 3080, 3081, 3083, 3086, 3088, 3089, 3102, 3105, 3107, 3108, 3111, 3113, 3116, 3118, 3119, 3120, 3122, 3123, 3124, 3125, 3126, 3127, 3128, 3129, 3130, 3131, 3132, 3134, 3150, 3154, 3166, 3167, 3168, 3173, 3174, 3175, 3176, 3178, 3179, 3181, 3182, 3184, 3185, 3190, 3251, 3252, 3253, 3254, 3267, 3268, 3269, 3272, 3275, 3278, 3279, 3280, 3281, 3282, 3284, 3285, 3288, 3290, 3305, 3307, 3308, 3309, 3311, 3314, 3321, 3322, 3323, 3325, 3327, 3328, 3332, 3352, 3354, 3359, 3360, 3364, 3365, 3366, 3367, 3371, 3372, 3373, 3374, 3375, 3376, 3377, 3378, 3380, 3383, 3385, 3386, 3387, 3392, 3406, 3407, 3409, 3413, 3415, 3417, 3418, 3419, 3420, 3421, 3422, 3423, 3429, 3430, 3432, 3433, 3434, 3435, 3436, 3440, 3443, 3444, 3445, 3446, 3448, 3451, 3454, 3455, 3459, 3461, 3462, 3467, 3469, 3470, 3471, 3472, 3473, 3475, 3476, 3482, 3500, 3501, 3502, 3504, 3505, 3506, 3507, 3514, 3516, 3522, 3528, 3530, 3532, 3536, 3540, 3542, 3552, 3556, 3557, 3558, 3559, 3564, 3567, 3574, 3575, 3577, 3584, 3585, 3586, 3587, 3588, 3589, 3591, 3592, 3593, 3601, 3603, 3604, 3605, 3606, 3607, 3608, 3609, 3612, 3616, 3617, 3621, 3622, 3625, 3628, 3629, 3630, 3631, 3632, 3633, 3634, 3638, 3650, 3653, 3654, 3657, 3659, 3660, 3661, 3664, 3665, 3666, 3700, 3702, 3710, 3752, 3753, 3757, 3758, 3760, 3762, 3763, 3764, 3770, 3771, 3775, 3776, 3777, 3778, 3780, 3781, 3787, 3788, 3789, 3800, 3802, 3808, 3814, 3815, 3818, 3819, 3822, 3823, 3824, 3825, 3826, 3827, 3828, 3829, 3830, 3831, 3832, 3833, 3834, 3835, 3837, 3838, 3863, 3866, 3868, 3869, 3870, 3900, 3903, 3904, 3905, 3910, 3911, 3914, 3916, 3951, 3953, 3954, 3957, 3958, 3960, 3961, 3962, 3963, 3964, 3970, 3972, 4001, 4005, 4006, 4012, 4013, 4019, 4020, 4021, 4022, 4024, 4025, 4026, 4027, 4029, 4030, 4033, 4034, 4035, 4037, 4038, 4051, 4052, 4054, 4055, 4056, 4057, 4058, 4059, 4061, 4065, 4069, 4070, 4071, 4072, 4073, 4074, 4075, 4076, 4077, 4078, 4091, 4092, 4093, 4094, 4096, 4098, 4099, 4101, 4102, 4103, 4106, 4107, 4108, 4109, 4110, 4111, 4112, 4113, 4114, 4115, 4117, 4121, 4122, 4123, 4126, 4135, 4136
NC(165)	3000, 3004, 3008, 3011, 3013, 3016, 3029, 3053, 3055, 3057, 3064, 3069, 3071, 3072, 3073, 3074, 3075, 3104, 3106, 3112, 3114, 3115, 3151, 3156, 3157, 3160, 3161, 3165, 3169, 3171, 3172, 3188, 3191, 3257, 3260, 3264, 3270, 3271, 3274, 3276, 3277, 3300, 3301, 3310, 3316, 3318, 3320, 3350, 3353, 3355, 3357, 3358, 3361, 3362, 3368, 3369, 3370, 3389, 3390, 3405, 3410, 3411, 3414, 3424, 3428, 3450, 3452, 3453, 3457, 3464, 3466, 3468, 3478, 3479, 3480, 3481, 3503, 3515, 3517, 3518, 3519, 3521, 3523, 3525, 3526, 3527, 3541, 3544, 3551, 3554, 3555, 3563, 3565, 3569, 3570, 3571, 3572, 3600, 3611, 3614, 3615, 3619, 3620, 3624, 3627, 3635, 3636, 3637, 3651, 3656, 3658, 3662, 3668, 3750, 3756, 3759, 3765, 3767, 3768, 3769, 3779, 3803, 3804, 3805, 3806, 3807, 3811, 3812, 3813, 3816, 3817, 3850, 3851, 3852, 3853, 3854, 3855, 3857, 3859, 3901, 3907, 3908, 3917, 3950, 3952, 3955, 3959, 3965, 3966, 3967, 3968, 3969, 4004, 4010, 4018, 4032, 4067, 4079, 4090, 4100, 4104, 4105, 4116, 4118, 4139

## References

- Adeli-Mosabbab, E., Fathy, M., 2015. Non-negative matrix completion for action detection. *Image Vision Comput.* 39, 38–51.
- Adeli, E., Shi, F., An, L., Wee, C.-Y., Wu, G., Wang, T., Shen, D., 2016. Joint feature-sample selection and robust diagnosis of Parkinson's disease from MRI data. *NeuroImage*.
- An, L., Adeli, E., Liu, M., Zhang, J., Shen, D., 2016. Semi-supervised hierarchical multimodal feature and sample selection for Alzheimer's disease diagnosis. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 79–87.
- Beitz, J., 2014. Parkinson's disease: a review. *Front Biosci.* 6, 65–74.
- Blum, A., Chawla, S., 2001. Learning from labeled and unlabeled data using graph mincuts.
- Boyd, S., Vandenberghe, L., 2004. *Convex Optimization*. Cambridge University Press.
- Braak, H., Braak, E., 1995. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol. Aging* 16, 271–278.
- de Sousa, C.A.R., Rezende, S.O., Batista, G.E., 2013. Influence of graph construction on semi-supervised learning. In: *Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 160–175.
- Duchi, J., Shalev-Shwartz, S., Singer, Y., Chandra, T., 2008. Efficient projections onto the  $l_1$ -ball for learning in high dimensions. In: *Proceedings of the 25th International Conference on Machine Learning*. ACM, pp. 272–279.
- Gao, Y., Adeli-M, E., Kim, M., Giannakopoulos, P., Haller, S., Shen, D., 2015. Medical image retrieval using multi-graph learning for MCI diagnostic assistance. In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Springer, pp. 86–93.
- Golub, G.H., Van Loan, C.F., 2012. *Matrix Computations*. JHU Press.
- Gönen, M., Alpaydm, E., 2011. Multiple kernel learning algorithms. *J. Mach. Learn. Res.* 12, 2211–2268.
- Gray, K., Aljabar, P., Heckemann, R., Hammers, A., Rueckert, D., 2013a. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage* 65, 167–175.
- Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D., Initiative, A.S.D.N., 2013b. Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage* 65, 167–175.
- Group, P.S., 2004. Levodopa and the progression of Parkinson's disease. *N Engl J Med* 2004, 2498–2508.
- Hay, A., 1988. The derivation of global estimates from a confusion matrix. *Int. J. Remote Sens.* 9, 1395–1398.
- Hu, C., Sepulcre, J., Johnson, K.A., Fakhri, G.E., Lu, Y.M., Li, Q., 2016. Matched signal detection on graphs: theory and application to brain imaging data classification. *NeuroImage* 125, 587–600.
- Huang, L., Liu, Y., Liu, X., Wang, X., Lang, B., 2014. Graph-based active semi-supervised learning: A new perspective for relieving multi-class annotation labor, multimedia and expo (ICME). In: *2014 IEEE International Conference on. IEEE*, pp. 1–6.
- Jankovic, J., 2008. Parkinson's disease: clinical features and diagnosis. *J. Neurol. Neurosurg. Psychiatry* 79, 368–376.
- Joachims, T., 2003. Transductive learning via spectral graph partitioning. *ICML* 290–297.
- Kim, D., Kim, S., Risacher, S.L., Shen, L., Ritchie, M.D., Weiner, M.W., Saykin, A.J., Nho, K., 2013. A graph-based integration of multimodal brain imaging data for the detection of early mild cognitive impairment (E-MCI). *MICCAI*.
- Li, Y.-F., Tsang, I.W., Kwok, J.T., Zhou, Z.-H., 2013. Convex and scalable weakly labeled SVMs. *J. Mach. Learn. Res.* 14, 2151–2188.
- Li, Y.-F., Zhou, Z.-H., 2015. Towards making unlabeled data never hurt. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 175–188.
- Lim, K.O., Pfefferbaum, A., 1989. Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter. *J. Comput. Assist. Tomogr.* 13, 588–593.
- Liu, F., Zhou, L., Shen, C., Yin, J., 2014. Multiple kernel learning in the primal for multimodal Alzheimer's disease classification. *Biomed. Health Inf., IEEE J.* 18, 984–990.
- Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M.J., 2015a. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Trans. Biomed. Eng.* 62, 1132–1140.
- Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M.J., 2015b. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Trans. Biomed. Eng.* 62, 1132–1141.
- Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., Feng, D., Fulham, M.J., 2015c. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *Biomed. Eng., IEEE Trans.* 62, 1132–1140.
- Liu, W., Chang, S.-F., 2009. Robust multi-class transductive learning with graphs. *IEEE Conference on Computer Vision and Pattern Recognition*.
- Long, D., Wang, J., Xuan, M., Gu, Q., Xu, X., Kong, D., Zhang, M., 2012. Automatic classification of early Parkinson's disease with multi-modal MR imaging. *PLoS One* 7, e47714.
- Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kiebertz, K., Flagg, E., Chowdhury, S., 2011. The Parkinson progression marker initiative (PPMI). *Prog. Neurobiol.* 95, 629–635.
- Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L., 2005. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin. N. Am.* 15, 869–877.
- Nie, F., Li, J., Li, X., 2016. Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification. *International Joint Conferences on Artificial Intelligence*.
- Nie, F., Wang, X., Huang, H., 2014. Clustering and projected clustering with adaptive neighbors. In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 977–986.
- Nocedal, J., Wright, S., 2006. *Numerical Optimization*. Springer Science & Business Media.

- Ohtsuka, C., Sasaki, M., Konno, K., Koide, M., Kato, K., Takahashi, J., Takahashi, S., Kudo, K., Yamashita, F., Terayama, Y., 2013. Changes in substantia nigra and locus coeruleus in patients with early-stage Parkinson's disease using neuromelanin-sensitive MR imaging. *Neurosci. Lett.* 541, 93–98.
- Peng, J., An, L., Zhu, X., Jin, Y., Shen, D., 2016. Structured sparse kernel learning for imaging genetics based Alzheimer's disease diagnosis. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 70–78.
- Plis, S.M., Sui, J., Lane, T., Roy, S., Clark, V.P., Potluru, V.K., Huster, R.J., Michael, A., Sponheim, S.R., Weisend, M.P., 2014. High-order interactions observed in multi-task intrinsic networks are dominant indicators of aberrant brain function in schizophrenia. *Neuroimage* 102, 35–48.
- PPMI, P.P.M.I., 2011. The Parkinson Progression Marker Initiative (PPMI). *Prog. Neurobiol.* 95, 629–635.
- Prashanth, R., Roy, S.D., Mandal, P.K., Ghosh, S., 2014. Automatic classification and prediction models for early Parkinson's disease diagnosis from SPECT imaging. *Expert Syst. Appl.* 41, 3333–3342.
- Rana, B., Juneja, A., Saxena, M., Gudwani, S., Kumaran, S., Behari, M., Agrawal, R., 2014. A machine learning approach for classification of Parkinson's disease and controls using T1-weighted MRI. *Mov. Disord.* 29, S88–S89.
- Reisberg, B., Ferris, S.H., Kluger, A., Franssen, E., Wegiel, J., de Leon, M.J., 2008. Mild cognitive impairment (MCI): a historical perspective. *Int. Psychogeriatr.* 20, 18–31.
- Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopez, M., Arabia, G., Morelli, M., Gilardi, M., Quattrone, A., 2014. Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and progressive supranuclear palsy. *J. Neurosci. Methods* 222, 230–237.
- Shen, D., Davatzikos, C., 2002. HAMMER: hierarchical attribute matching mechanism for elastic registration. *Med. Imaging, IEEE Trans.* 21, 1421–1439.
- Stern, M., Siderowf, A., 2010. Parkinson's at risk syndrome: can Parkinson's disease be predicted? *Mov. Disord.* 25, S89–S93.
- Suykens, J.A., Vandewalle, J., 1999. Least squares support vector machine classifiers. *Neural Process. Lett.* 9, 293–300.
- Tang, J., Alelyani, S., Liu, H., 2014. Feature selection for classification: a review. *Data Classification: Algorithms and Applications* 37.
- Thompson, B., 2005. Canonical correlation analysis. *Encycl. Stat. Behav. Sci.*
- Thompson, P.M., Hayashi, K.M., Dutton, R.A., Chiang, M.-C., MD, A.D.L., Sowell, E.R., Zubicaray, G.d., Becker, J.T., Lopez, O.L., Aizenstein, H.J., Toga, A.W., 2007. Tracking Alzheimer's Disease. *Ann. N.Y. Acad. Sci.* 1097, 198–214.
- Tong, T., Gray, K., Gao, Q., Chen, L., Rueckert, D., 2015. Nonlinear Graph Fusion for Multi-modal Classification of Alzheimer's Disease. *MICCAI, Munich Germany*.
- Trzepacz, P.T., Yu, P., Sun, J., Schuh, K., Case, M., Witte, M.M., Hochstetler, H., Hake, A., Initiative, A.s.D.N., 2014. Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to Alzheimer's dementia. *Neurobiol. Aging* 35, 143–151.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., 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.
- Wang, B., Mezlini, A.M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haihe-Kains, B., Goldenberg, A., 2014a. Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* 11, 333–337.
- Wang, B., Tsotsos, J., 2016. Dynamic label propagation for semi-supervised multi-class multi-label classification. *Pattern Recognit.* 52, 75–84.
- Wang, H., Nie, F., Huang, H., Risacher, S., Saykin, A.J., Shen, L., 2011. Identifying AD-sensitive and cognition-relevant imaging biomarkers via joint classification and regression. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 115–123.
- Wang, Y., Nie, J., Yap, P.-T., Li, G., Shi, F., Geng, X., Guo, L., Shen, D., Initiative, A.s.D.N., 2014b. Knowledge-guided robust MRI brain extraction for diverse large-scale neuroimaging studies on humans and non-human primates. *PLoS One* 9, e77810.
- Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., 2013. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimers Dementia* 9, e111–e194.
- Westman, E., Muehlboeck, J.-S., Simmons, A., 2012. Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *Neuroimage* 62, 229–238.
- Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., Initiative, A.s.D.N., 2011. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867.
- Zhang, Y., Huang, K., Geng, G., Liu, C., 2015. MTC: A Fast and Robust Graph-Based Transductive Learning Method. *Neural Netw. Learn. Syst. IEEE Trans.* 26, 1979–1991.
- Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Schölkopf, B., 2004. Learning with local and global consistency. *Adv. Neural Inf. Process. Syst.* 16, 321–328.
- Zhou, D., Burges, C.J., 2007. Spectral clustering and transductive learning with multiple views. In: *Proceedings of the 24th international conference on Machine learning*. ACM, pp. 1159–1166.
- Zhu, X., Ghahramani, Z., Lafferty, J., 2003. Semi-supervised learning using gaussian fields and harmonic functions. *ICML* 912–919.
- Zhu, X., Lafferty, J., Rosenfeld, R., 2005. Semi-supervised learning with graphs. *Carnegie Mellon University, Language Technologies Institute, School of Computer Science*.
- Zhu, X., Perry, G., Smith, M.A., Wang, X., 2013. Abnormal mitochondrial dynamics in the pathogenesis of Alzheimer's disease. *J. Alzheimers Dis.* 33, S253–S262.
- Zhu, X., Suk, H.-I., Lee, S.-W., Shen, D., 2016. Subspace regularized sparse multi-task learning for multiclass neurodegenerative disease identification. *IEEE Trans. Biomed. Eng.* 63, 607–618.
- Zhu, X., Suk, H.-I., Zhu, Y., Thung, K.-H., Wu, G., Shen, D., 2015. Multi-view classification for identification of Alzheimer's disease. In: *International Workshop on Machine Learning in Medical Imaging*. Springer, pp. 255–262.