

# Progressive Graph-Based Transductive Learning for Multi-modal Classification of Brain Disorder Disease

Zhengxia Wang<sup>1,2</sup>, Xiaofeng Zhu<sup>1</sup>, Ehsan Adeli<sup>1</sup>, Yingying Zhu<sup>1</sup>,  
Chen Zu<sup>1</sup>, Feiping Nie<sup>3</sup>, Dinggang Shen<sup>1</sup>, and Guorong Wu<sup>1</sup>(✉)

<sup>1</sup> Department of Radiology and BRIC,  
University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA  
grwu@med.unc.edu

<sup>2</sup> Department of Information Science and Engineering,  
Chongqing Jiaotong University, Chongqing 400074, China

<sup>3</sup> School of Computer Science and OPTIMAL Center,  
Northwestern Polytechnical University, Xi'an 710072, China

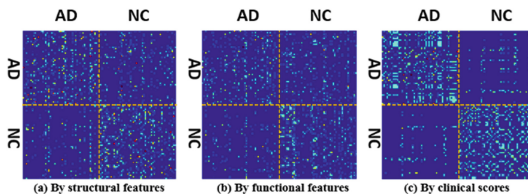
**Abstract.** Graph-based Transductive Learning (GTL) is a powerful tool in computer-assisted diagnosis, especially when the training data is not sufficient to build reliable classifiers. Conventional GTL approaches first construct a fixed subject-wise graph based on the similarities of observed features (i.e., extracted from imaging data) in the feature domain, and then follow the established graph to propagate the existing labels from training to testing data in the label domain. However, such a graph is exclusively learned in the feature domain and may not be necessarily optimal in the label domain. This may eventually undermine the classification accuracy. To address this issue, we propose a progressive GTL (pGTL) method to progressively find an intrinsic data representation. To achieve this, our pGTL method iteratively (1) refines the subject-wise relationships observed in the feature domain using the learned intrinsic data representation in the label domain, (2) updates the intrinsic data representation from the refined subject-wise relationships, and (3) verifies the intrinsic data representation on the training data, in order to guarantee an optimal classification on the new testing data. Furthermore, we extend our pGTL to incorporate multi-modal imaging data, to improve the classification accuracy and robustness as multi-modal imaging data can provide complementary information. Promising classification results in identifying Alzheimer's disease (AD), Mild Cognitive Impairment (MCI), and Normal Control (NC) subjects are achieved using MRI and PET data.

## 1 Introduction

Alzheimer's disease (AD) is the most common neurological disorder in the older population. There is overwhelming evidence in the literature that the morphological patterns are observable by means of *either* structural and diffusion MRI *or* PET [1–3]. However, morphological abnormal patterns are often subtle, compared to high inter-subject variations. Hence, sophisticated pattern recognition methods are of high demand to accurately identify individuals at different stages of AD progression.

Medical imaging applications often deal with high dimensional data and usually less number of samples with ground-truth labels. Thus, it is very challenging to find a general model that can work well for an entire set of data. Hence, GTL method has been investigated with great success in medical imaging area [4, 5], since it can overcome the above difficulties by taking advantage of the data representation on unlabeled testing subjects. In current state-of-the-art methods, graph is used to represent the subject-wise relationship. Specifically, each subject, regardless of being labeled or unlabeled, is treated as a graph node. Two subjects are connected by a graph link (i.e., an edge) if they have similar morphological patterns. Using these connections, the labels can be propagated throughout the graph until all latent labels are determined. Many current label propagation strategies have been proposed to determine the latent labels of testing subjects based on subject-wise relationships encoded in the graph [6].

The 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 morphological patterns are often highly complex and heterogeneous. Figure 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 dot and blue dot denote high and low subject-wise similarities, respectively. Since the clinical data (e.g., MMSE and CDR scores [1]) is more related 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 structural image features and clinical scores are completely *different*. Thus, there is *no* guarantee 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. Many studies take advantage of multi-modal information to improve discrimination power of transductive learning. However, the graphs from different modalities might be *different* too, 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)). Graph diffusion [5] is recently proposed to find the common graph. Unfortunately, as shown in Fig. 1, it is hard to find a combination for the graphs in Fig. 1(a) and (b) that can lead to the graph in Fig. 1(c), which is more related with final classification task.



**Fig. 1.** Affinity matrices using structural image features (a), functional image features (b), and clinical scores (c).

To solve these issues, we propose a pGTL method to learn the intrinsic data representation, which could be eventually optimal for label propagation. Specifically, the intrinsic data representation is required to be (a) close to subject-wise relationships constructed by image features extracted from different modalities, and (b) verified on the training data and guaranteed to be optimal for label classification. To that end, we simultaneously (1) refine the data representation (subject-wise graph) in the feature domain, (2) find the intrinsic data representation based on the constructed graphs on multi-modal imaging data and also the clinical labels of entire subject set (including known labels on training subjects and also 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 classification results have been achieved in classifying 93AD, 202 MCI, and 101NC subjects, each with MR and PET images.

## 2 Methods

Suppose we have  $N$  subjects  $\{I_1, \dots, I_P, I_{P+1}, \dots, I_N\}$ , which sequentially consist of  $P$  training subjects and  $Q (= N - P)$  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 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  $\mathbf{F}^{N \times C} = [\mathbf{F}_P \mathbf{F}_Q]$ .

### 2.1 Progressive Graph-Based Transductive Learning

**Conventional Graph-Based Transductive Learning.** For clarity, we first extract single modality image features from each subject  $I_i$  ( $i = 1, \dots, N$ ), denoted as  $\mathbf{x}_i$ . In conventional GTL methods, the subject-wise relationships are computed based on feature similarity, which is encoded in an  $N \times N$  feature affinity matrix  $\mathbf{S}$ . Each element  $s_{ij}$  ( $0 \leq s_{ij} \leq 1, i, j = 1, \dots, N$ ) represents the feature affinity degree between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . After constructing  $\mathbf{S}$  (based on feature similarity), conventional methods determine the latent label for each testing subject  $I_q$  by solving a classic graph learning problem:

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

As shown in Fig. 1, the affinity matrix  $\mathbf{S}$  might not be strongly related with the intrinsic data representation in the label domain. Therefore, it is necessary to further design a graph based on the labels matrix, rather than solely using the graph constructed by the features. However, the labels on testing subjects are not determined yet. In order to solve this chicken-and-egg dilemma, we propose to construct a dynamic graph which progressively reflects the intrinsic data representation in the label domain.

**Progressive Graph-Based Transductive Learning on Single Modality.** We propose three strategies to remedy the above issue. **(1)** We propose to gradually find an intrinsic data representation  $\mathbf{T} = [t_{ij}]_{i,j=1,\dots,N}$  which is more relevant than  $\mathbf{S}$  to guide the label propagation in Eq. (1). **(2)** Since only the training images have their known clinical labels, exclusively optimizing  $\mathbf{T}$  in the label domain is an ill-posed problem. Thus, we encourage the intrinsic data presentation  $\mathbf{T}$  also respecting the affinity matrix  $\mathbf{S}$ , where image features are complete in the feature domain. **(3)** In order to suppress possible noisy patterns and outlier subjects, we allow the intrinsic data representation  $\mathbf{T}$  to progressively refine the affinity matrix  $\mathbf{S}$  in the feature domain. In this way, the estimations of  $\mathbf{S}$  and  $\mathbf{T}$  are coupled, thus bringing a *dynamic graph learning model* with the following objective function:

$$\arg \min_{\mathbf{S}, \mathbf{T}, \mathbf{F}} \sum_{i,j=1}^N \left\{ \mu \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 t_{ij} + \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij} + \lambda_1 s_{ij}^2 + \lambda_2 \|s_{ij} - t_{ij}\|_2^2 \right\} \quad (2)$$

$$s.t. \quad 0 \leq s_{ij} \leq 1, \mathbf{s}'_i \mathbf{1} = 1, 0 \leq t_{ij} \leq 1, \mathbf{t}'_i \mathbf{1} = 1, \mathbf{F} = [\mathbf{F}_P \mathbf{F}_Q]$$

where  $\mu$  is the scalar balancing the data fitting terms from two different domains (i.e., the first and second terms in Eq. (2)). Suppose  $\mathbf{s}_i \in \mathbb{R}^{N \times 1}$  and  $\mathbf{t}_i \in \mathbb{R}^{N \times 1}$  are vectors with the  $j$ -th element as  $s_{ij}$  and  $t_{ij}$  separately. In order to avoid trivial solution,  $l_2$ -norm is used as the constraint on each element  $s_{ij}$  in affinity matrix  $\mathbf{S}$ .  $\lambda_1$  and  $\lambda_2$  are two scalars to control the strengths of the last two terms in Eq. (2).

**Progressive Graph-Based Transductive Learning on Multiple Modalities.** 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  $m$ -th modality, we optimize the affinity matrix  $\mathbf{S}^m$ . As shown in Fig. 1(a) and (b), the affinity matrices across modalities could be different. Thus, we require the intrinsic data representation  $\mathbf{T}$  to be close to all  $\mathbf{S}^m, m = 1, \dots, M$ . It is straightforward to extend our above pGTL method to the multi-modal scenario:

$$\arg \min_{\mathbf{S}^m, \mathbf{T}, \mathbf{F}} \sum_{i,j=1}^N \left\{ \mu \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 t_{ij} + \sum_{m=1}^M \left[ \|\mathbf{x}_i^m - \mathbf{x}_j^m\|_2^2 s_{ij}^m + \lambda_1 (s_{ij}^m)^2 + \lambda_2 \|s_{ij}^m - t_{ij}\|_2^2 \right] \right\} \quad (3)$$

$$s.t. \quad 0 \leq s_{ij}^m \leq 1, (\mathbf{s}_i^m)' \mathbf{1} = 1, 0 \leq t_{ij} \leq 1, \mathbf{t}'_i \mathbf{1} = 1, \quad \mathbf{F} = [\mathbf{F}_P \mathbf{F}_Q]$$

It is worth noting that, although the multi-modal information leads to multiple affinity matrices in the feature domain, they share the same intrinsic data representation  $\mathbf{T}$ .

## 2.2 Optimization

Since our proposed energy function in Eq. (3) is convex to each variables, i.e.  $\mathbf{S}, \mathbf{T}, \mathbf{F}$ , we present the following divide-and-conquer solution to optimize one set of variables at a time by fixing other sets of variable. We initialize  $S = \exp(-\|x_i - x_j\|_2^2 / 2\sigma^2)$ ,  $\sigma$  is an empirical parameter,  $\mathbf{T} = \sum_{m=1}^M \mathbf{S}^m / M, F_Q = \{0\}^{Q \times C}$ .

**Estimation of Affinity Matrix  $\mathbf{S}^m$  for Each Modality.** Removing the unrelated terms *w.r.t.*  $\mathbf{S}^m$  in Eq. (3), the optimization of  $\mathbf{S}^m$  falls to the following objective function:

$$\arg \min_{\mathbf{S}^m} \sum_{i,j=1}^N \left\| \mathbf{x}_i^m - \mathbf{x}_j^m \right\|_2^2 s_{ij}^m + \lambda_1 \left( s_{ij}^m \right)^2 + \lambda_2 \sum_{i,j=1}^N \left\| s_{ij}^m - t_{ij} \right\|_2^2 \quad (4)$$

where  $(0 \leq s_{ij} \leq 1, (\mathbf{s}_i^m)' \mathbf{1} = 1)$ . Since Eq. (4) is independent of variables  $i$  and  $j$ , we further reformulate Eq. (4) in the vector form as below:

$$\arg \min_{\mathbf{s}_i^m} \left\| \mathbf{s}_i^m + \frac{\mathbf{d}_i}{2r_1} \right\|_2^2 \quad (5)$$

where  $\mathbf{s}_i^m$  is the  $i$ -th column vector of affinity matrix  $\mathbf{S}^m$ ,  $\mathbf{d}_i = [d_{ij}]_{j=1,\dots,N}$  is a vector with each  $d_{ij} = \left\| \mathbf{x}_i^m - \mathbf{x}_j^m \right\|_2^2 - 2\lambda_2 t_{ij}$ , and  $r_1 = \lambda_1 + \lambda_2$ . The problem in Eq. (5) is equivalent to project onto a simplex, which has a closed-form solution in [7]. After we solve each  $\mathbf{s}_i^m$ , we can obtain the affinity matrix  $\mathbf{S}^m$ .

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

$$\arg \min_{\mathbf{T}} \sum_{i,j=1}^N \mu \left\| \mathbf{f}_i - \mathbf{f}_j \right\|_2^2 t_{ij} + \lambda_2 \sum_{m=1}^M \sum_{i,j=1}^N \left( \left\| s_{ij}^m - t_{ij} \right\|_2^2 \right) \quad (6)$$

Similarly, we can reformulate Eq. (6) by solving each  $\mathbf{t}_i$  at a time:

$$\arg \min_{\mathbf{t}_i} \left\| \mathbf{t}_i + \frac{\mathbf{h}_i}{2r_2} \right\|_2^2 \quad (7)$$

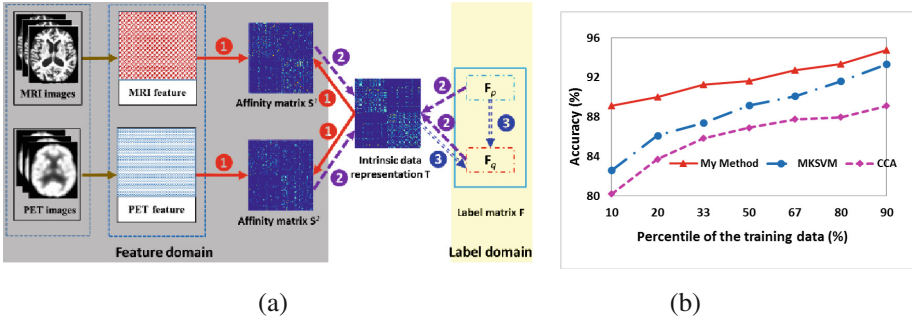
where  $\mathbf{h}_i = [h_{ij}]_{j=1,\dots,N}$  is a vector with each element  $h_{ij} = \mu \left\| \mathbf{f}_i - \mathbf{f}_j \right\|_2^2 - 2\lambda_2 \sum_{m=1}^M s_{ij}^m$ , and  $r_2 = M\lambda_2$  is a scalar.

**Update the Latent Labels  $\mathbf{F}_Q$  on Testing Subjects.** Given both  $\mathbf{S}^w$  and  $\mathbf{T}$ , the objective function for the latent label  $\mathbf{F}_Q$  can be derived from Eq. (3) as below:

$$\arg \min_{\mathbf{F}} \sum_{i,j=1}^N \left\| \mathbf{f}_i - \mathbf{f}_j \right\|_2^2 t_{ij} \Rightarrow \arg \min_{\mathbf{F}} \text{Trace}(\mathbf{F}'\mathbf{L}\mathbf{F}), \quad (8)$$

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  $\mathbf{T}$ . By differentiating Eq. (8) *w.r.t.*  $\mathbf{F}$  and letting the gradient  $\mathbf{L}\mathbf{F} = 0$ , we obtain the following equation:  $\begin{bmatrix} \mathbf{L}_{PP} & \mathbf{L}_{PQ} \\ \mathbf{L}_{QP} & \mathbf{L}_{QQ} \end{bmatrix} \begin{bmatrix} \mathbf{F}_P \\ \mathbf{F}_Q \end{bmatrix} = 0$ , 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  $\hat{\mathbf{F}}_Q = -(\mathbf{L}_{QQ})^{-1} \mathbf{L}_{QP} \mathbf{F}_P$ .

**Discussion.** Taking MRI and PET modalities as example, Fig. 2(a) illustrates the optimization of Eq. (3) by alternating the following three steps. (1) Estimate each



**Fig. 2.** (a) The dynamic procedure of the proposed pGTL method, (b) Classification accuracy as a function of the number of training samples used.

affinity matrix  $S^m$ , which depends on the observed image features  $\mathbf{x}^m$  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  $S^1$  and  $S^2$  and also the subject-wise relationship in the label domain (purple arrows); (3) Update the latent labels  $F_Q$  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

**Subject Information and Image Processing.** In the following experiments, we select 93 AD subjects, 202 MCI subjects, and 101 NC subjects from ADNI dataset. Since MCI is a highly heterogeneous group, we further separate them into 55 progressive MCI subjects (pMCI), who will finally develop into AD patients within the next 24 months, and 63 stable MCI subjects (sMCI), who won't convert to AD after 24 months. The remain MCI subjects included a group not converted in 24 months but converted in 36 months and another group with observation information in baseline but missing information in 24 months. Each subject has both MR and 18-Fluoro-DeoxyGlucose PET (FDG-PET) images.

For each subject, we first align the PET image to MR image. Then we remove the skull and cerebellum from MR image and segment MR image into white matter, gray matter and cerebrospinal fluid. 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. Finally, the gray matter volume and the mean PET intensity image in each ROI are used and form a 186-dimensional feature vector.

**Experiment Settings.** *First*, we evaluate our proposed pGTL method, with comparison to classic classification methods, such as Canonical Correlation Analysis (CCA) [8] based SVM (denoted as CCA in the following context), Multi-Kernel SVM (MKSVM) [9], and a conventional GTL method, since these methods are widely used

in AD studies. In order to demonstrate the overall performance of our method in several classification tasks, i.e. AD vs NC, MCI vs NC, and pMCI vs sMCI, in each experiment, we use 10-fold cross-validation strategy, with 9 folds of data as training dataset and the remaining 1 fold as testing dataset. *Second*, we compare our proposed method with three recently published state-of-the-art classification methods: (1) random-forest based classification method [10], (2) multi-modal graph-fusion method [4], and (3) multi-modal deep learning method [11]. It is worth noting that we only use the classification accuracy reported in their papers, in order for fair comparison.

**Parameter Settings.** In the following experiments, we use the same greedy strategy to select best parameters for CCA, MKSVM and our proposed method. For example, we obtain the optimal values for  $\mu$ ,  $\lambda_1$  and  $\lambda_2$  in our method by exhaustive search in the range from  $10^{-3}$  to  $10^3$  in a small portion of training dataset.

**Comparison with Classic CCA, GTL and Multi-kernel SVM (MKSVM).** The classification accuracies by CCA, MKSVM, GTL and our method are evaluated in three classification tasks (AD vs NC, MCI vs NC, and pMCI vs sMCI), respectively. The averaged classification accuracy (ACC), sensitivity (SEN), and specificity (SPE) with 10-fold cross-validation are summarized in Table 1. It is clear that our proposed method beats other competing classification methods in three classification tasks, with significant improvement under paired t-test ( $p < 0.001$ , designated by ‘\*’ in Table 1).

**Table 1.** Comparison of classification performance by different methods.

Classification Task	Method	ACC (%)	SEN (%)	SPE(%)
AD vs NC	CCA	89.1 ± 1.57	87.6 ± 2.02	90.5 ± <b>1.25</b>
	MKSVM	90.0 ± 1.03	89.1 ± 1.53	90.7 ± 1.28
	GTL	88.0 ± 1.27	90.4 ± 1.59	85.8 ± 1.90
	<b>Our Method</b>	<b>92.6 ± 0.65</b>	<b>92.2 ± 1.34</b>	<b>92.9 ± 1.37</b>
MCI vs NC	CCA	68.3 ± 1.95	78.8 ± 1.92	47.3 ± <b>3.16</b>
	MKSVM	72.6 ± 1.87	84.5 ± <b>1.70</b>	48.9 ± 3.57
	GTL	71.9 ± 0.94	92.8 ± 0.93	29.9 ± 2.31
	<b>Our Method</b>	<b>78.6 ± 1.19</b>	<b>85.1 ± 2.18</b>	<b>65.4 ± 3.35</b>
pMCI vs sMCI	CCA	64.7 ± 1.39	41.5 ± 3.30	84.9 ± 3.20
	MKSVM	67.5 ± 2.79	57.6 ± 5.75	76.2 ± 5.42
	GTL	67.7 ± 1.27	57.8 ± 2.85	76.5 ± 2.33
	<b>Our Method</b>	<b>76.0 ± 1.19</b>	<b>75.6 ± 2.30</b>	<b>76.3 ± 3.00</b>

Furthermore, we evaluate the classification performance *w.r.t.* the number of training samples, as shown in Fig. 2(b). It is clear that (1) our proposed method always has higher classification accuracy than both CCA and MKSVM methods; and (2) all methods can improve the classification accuracy as the number of training samples increases. It is worth nothing that our proposed method achieves large improvement against MKSVM, when only 10 % of data is used as the training dataset. The reason is

**Table 2.** Comparison with the classification accuracies reported in the literatures (%).

Method	Subject information	Modality	AD/NC	MCI/NC
Random forest [10]	37AD + 75MCI + 35NC	MRI + PET + CSF + Genetic	89.0	74.6
Graph fusion [4]	35AD + 75MCI + 77NC	MRI + PET + CSF + Genetic	91.8	79.5
Deep learning [11]	85AD + 169MCI + 77NC	MRI + PET	91.4	<b>82.1</b>
Our method	99AD + 202MCI + 101NC	MRI + PET	<b>92.6</b>	78.6

that supervised methods require a sufficient number of samples to train the reliable classifier. Since the training samples with known labels are expensive to collect in medical imaging area, this experiment indicates that our method has high potential to be deployed in current neuroimaging studies.

**Comparison with Recently Published State-of-the-Art Methods.** Table 2 summarizes the subject information, imaging modality, and average classification accuracy by using state-of-the-art methods. These comparison methods represent four typical machine learning techniques. Since the classification between pMCI and sMCI groups are not reported in [4, 10, 11], we only show the classification results for AD vs NC, and MCI vs NC tasks. 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.

**Discussion.** Deep learning approach in [11] 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 hand-crafted features for classification. It is noteworthy that we can complete the classification on a new dataset (including greedy parameter tuning) within three hours on a regular PC (8 CPU cores and 16 GB memory), which is much more economic than massive training cost in [11]. 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.

## 4 Conclusion

In this paper, we present a novel pGTL method to identify individual subject at different stages of AD progression, using multi-modal imaging data. Compared to conventional methods, our method seeks for the intrinsic data representation, which can be learned from the observed imaging features and simultaneously validated on the existing labels of training data. Since the learned intrinsic data presentation is more relevant to label propagation, our method achieves promising classification performance in AD vs NC, MCI vs NC, and pMCI vs sMCI tasks, after comprehensive comparison with classic and recent state-of-the-art methods.



## References

1. Thompson, P.M., Hayashi, K.M., et al.: Tracking Alzheimer's disease. *Ann. NY Acad. Sci.* **1097**, 198–214 (2007)
2. Zhu, X., Suk, H.-I., et al.: A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. *NeuroImage* **100**, 91–105 (2014)
3. Jin, Y., Shi, Y., et al.: Automated multi-atlas labeling of the fornix and its integrity in Alzheimer's disease. In: 2015 IEEE 12th ISBI, pp. 140–143. IEEE (2015)
4. Tong, T., Gray, K., Gao, Q., Chen, L., Rueckert, D.: Nonlinear graph fusion for multi-modal classification of Alzheimer's disease. In: Zhou, L., Wang, L., Wang, Q., Shi, Y. (eds.) MICCAI 2015. LNCS, vol. 9352, pp. 77–84. Springer, Heidelberg (2015)
5. Wang, B., Mezlini, A.M., et al.: Similarity network fusion for aggregating data types on a genomic scale. *Nat. Methods* **11**, 333–337 (2014)
6. Zhang, Y., Huang, K., et al.: MTC: a fast and robust graph-based transductive learning method. *IEEE Trans. Neural Netw. Learn. Syst.* **26**, 1979–1991 (2015)
7. Huang, H., Yan, J., Nie, F., Huang, J., Cai, W., Saykin, A.J., Shen, L.: A new sparse simplex model for brain anatomical and genetic network analysis. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 625–632. Springer, Heidelberg (2013)
8. Thompson, B.: Canonical correlation analysis. In: *Encyclopedia of Statistics in Behavioral Science* (2005)
9. Gönen, M., Alpaydın, E.: Multiple kernel learning algorithms. *J. Mach. Learn. Res.* **12**, 2211–2268 (2011)
10. Gray, K., Aljabar, P., et al.: Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage* **65**, 167–175 (2013)
11. Liu, S., Liu, S., et al.: Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. *IEEE Trans. Biomed. Eng.* **62**, 1132–1141 (2015)