

# MCI Identification by Joint Learning on Multiple MRI Data

Yue Gao<sup>1</sup>, Chong-Yaw Wee<sup>1</sup>, Minjeong Kim<sup>1</sup>, Panteleimon Giannakopoulos<sup>2</sup>,  
Marie-Louise Montandon<sup>3</sup>, Sven Haller<sup>3</sup>, and Dinggang Shen<sup>1</sup>

<sup>1</sup> Department of Radiology and BRIC,

University of North Carolina at Chapel Hill, NC 27599, USA

<sup>2</sup> Division of Psychiatry, Geneva University Hospitals, Switzerland

<sup>3</sup> Department of Neuroradiology, University Hospitals of Geneva  
and Faculty of Medicine of the University of Geneva, Switzerland

**Abstract.** The identification of subtle brain changes that are associated with mild cognitive impairment (MCI), the at-risk stage of Alzheimer’s disease, is still a challenging task. Different from existing works, which employ multimodal data (e.g., MRI, PET or CSF) to identify MCI subjects from normal elderly controls, we use four MRI sequences, including T1-weighted MRI (T1), Diffusion Tensor Imaging (DTI), Resting-State functional MRI (RS-fMRI) and Arterial Spin Labeling (ASL) perfusion imaging. Since these MRI sequences simultaneously capture various aspects of brain structure and function during clinical routine scan, it simplifies finding the relationship between subjects by incorporating the mutual information among them. To this end, we devise a hypergraph-based semi-supervised learning algorithm. In particular, we first construct a hypergraph for each of MRI sequences separately using a star expansion method with both the training and testing data. A centralized learning is then performed to model the optimal relevance between subjects by incorporating mutual information between different MRI sequences. We then combine all centralized hypergraphs by learning the optimal weight of each hypergraph based on the minimum Laplacian. We apply our proposed method on a cohort of 41 consecutive MCI subjects and 63 age-and-gender matched controls with four MRI sequences. Our method achieves at least a 7.61% improvement in classification accuracy compared to state-of-the-art methods using multiple MRI data.

## 1 Introduction

Alzheimer’s disease (AD) is the most common form of dementia in elderly over 65 years of age. The number of AD patients has reached 26.6 million in nowadays and is expected to double within the next 20 years, leading to 1 in every 85 people worldwide being affected by AD by 2050. Therefore, the diagnosis of AD at its at-risk stage of mild cognitive impairment (MCI) [7] becomes extremely essential and has attracted extensive research efforts in recent years [11, 9]. Previous studies [10] have shown that structural and functional brain changes may start before clinically converted to AD and can be used as potential biomarkers for MCI identification.

Recent studies [4, 11] show great promises for integrating multiple modalities, e.g., MRI, PET and CSF, for improving AD/MCI diagnosis accuracy, and semi-supervised

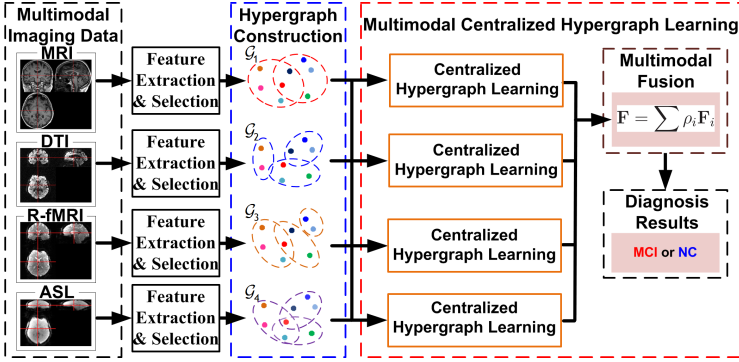


Fig. 1. An overview of the proposed centralized hypergraph learning for MCI diagnosis.

learning for multimodal data has also been investigated [2]. However, in most previous works, modeling the relationship among subjects is often performed separately for each modality, ignoring the crucial mutual information between different modalities. In practice, integrating the information acquired from different modalities is a challenging task, since the relationship among subjects may differ for different modalities.

On the other hand, multiple MR sequences, e.g., T1-weighted (T1), Diffusion Tensor Imaging (DTI) and Resting-State functional MRI (RS-fMRI), can be used in clinical routine scans to capture different aspects of the brain structures and functions. For instance, T1 provides the tissue type information of the brain, DTI measures macroscopic axonal organization in nervous system tissues, and RS-fMRI provides the regional interactions that take place when the subject is in the absence of an explicit task. As a relatively new technique, Arterial Spin Labeling (ASL) [1] perfusion imaging is introduced to measure brain perfusion without any injection of a contrast agent and demonstrated consistent reduction in basal perfusion notably in the posterior cingulate cortex in MCI and AD [1]. More recently, ASL was even able to predict very early cognitive decline in healthy elderly controls, *i.e.*, the earliest stage of neurodegeneration. Multiple MRI sequences can be easily and simultaneously captured during clinical routine scans.

In this work, we propose a centralized hypergraph learning method (CHL) to better model relationship among subjects with multiple MRIs for the purpose of MCI diagnosis. The basic idea of the proposed method is to estimate the relevance between different subjects that reflects how likely two subjects belong to the same category by integrating multiple imaging data in a semi-supervised manner. Then, the relationship information among subjects is represented by a hypergraph structure that connects all subjects in both training and testing sets. Compared to the simple graph, in which an edge can only link two vertices, hypergraph [12] conveys more information through a set of hyperedges that connects more than two vertices at the same occasion. It has been successfully applied to various applications, such as image and object retrieval [3]. In this way, a hypergraph structure is able to capture the higher-order relationship among different subjects, *i.e.*, whether a group of subjects share similar content. Then, MCI diagnosis is formulated as a binary classification task in the hypergraph structure to classify each subject as MCI patient or normal control (NC). Figure 1 presents the schematic diagram

of the proposed framework. We first construct a hypergraph using both the training and testing data for each of multiple MRIs separately to reflect the higher-order relationship among subjects. In hypergraph construction, each time one subject is selected as a centroid. It is then connected to its  $K$  nearest subjects in the feature space via a hyperedge. We then conduct a centralized hypergraph learning to explore the underneath relationship of a set of samples, where the relevance among subjects and the hyperedge weights are optimized simultaneously via an alternating optimization approach. Specifically, each time one hypergraph is first selected as the core and the rest as auxiliary information in the learning process. This procedure is repeated for each hypergraph, producing a set of relevance scores for each subject for classification. To obtain the final decision, we assemble the relevance scores based on the optimal weights learned by minimizing the overall hypergraph Laplacian. Note that, for the training subjects, we just use their imaging features to construct hypergraphs. Therefore, the relevance scores are conveyed globally, leading to a semi-supervised learning model, and better avoiding over-fitting to the training set.

## 2 Method

**Data and Preprocessing** - A dataset containing T1, DTI, RS-fMRI and ASL from 41 MCI patients and 63 normal controls was collected through the University Hospital of Geneva in Switzerland. The T1 images were preprocessed by skull stripping, cerebellum removal, and tissue segmentation [6]. The anatomical automatic labeling (AAL) atlas, parcellated with 90 predefined regions-of-interest (ROIs) in cerebrum area, are aligned to the native space of each subject using a deformable registration algorithm. For T1 data, WM and GM tissue volumes in each region are computed and further normalized to generate a 180-dimensional feature vector, i.e., 90 WM and 90 GM features.

The DTI were first parcellated into 90 regions by propagating the AAL ROIs to each image using a deformable DTI registration algorithm. Whole-brain streamline fiber tractography was performed on each DTI image. The number of fibers passing through each pair of regions was counted and the averages of on-fiber fractional anisotropy (FA) for each ROI pair were computed to form a structural connectivity matrix.

For RS-fMRI, a 9min ON-OFF CO<sub>2</sub> challenge was employed, i.e., 1 min OFF, 2 min ON, 2 min OFF, 2 min ON, 2 min OFF, during data acquisition. Slice timing correction and head-motion correction were performed using the Statistical Parametric Mapping software package. To ensure magnetization equilibrium, the first 10 acquired RS-fMRI images of each subject were discarded. The remaining 170 images were first corrected for acquisition time delays among different slices before they were realigned to the first volume of the remaining images for head-motion corrections. A Pearson correlation-based connectivity matrix of dimension  $90 \times 90$  based on AAL atlas was constructed for each subject.

Preprocessing of ASL images was performed using ASLtbx. Similar to RS-fMRI, a symmetric connectivity matrix of dimension  $90 \times 90$  was constructed for each subject.

For DTI, RS-fMRI and ASL, the local clustering coefficients, which quantify the cliquishness of the nodes [8], are computed for the connectivity networks. For these three MRI sequences, both the raw feature (8100-D) and the clustering coefficients-based feature (90-D) are employed.

**Hypergraph Construction** - For each type of multiple imaging data, a hypergraph is constructed. For the  $i$ -th imaging with  $N$  subjects, a hypergraph  $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i, \mathbf{W}_i)$  with  $N$  vertices is constructed where each subject is represented by a vertex. Here,  $\mathcal{V}_i$  is the vertex set,  $\mathcal{E}_i$  is the hyperedge set, and  $\mathbf{W}_i$  is the corresponding weight for the hyperedges. A star expansion method is employed to generate a set of hyperedges among vertices. Specifically, in each feature space, a vertex is selected as the centroid vertex and then a hyperedge is developed by connecting to its nearest neighbors within  $\varphi \bar{d}$  distance, where  $\bar{d}$  is the average distance between subjects in the feature space and  $\varphi$  is set as 1 in our experiment. Incidence matrix of the hypergraph is then generated to represent the relationship among different vertices. The  $(v, e)$ -th entry of the incidence matrix indicates whether the vertex  $v$  is connected via the hyperedge  $e$  to other vertices. The incidence matrix  $\mathbf{H}_i$  of hypergraph  $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i, \mathbf{W}_i)$  is generated as

$$\mathbf{H}_i(v, e) = \begin{cases} \exp\left(-\frac{d_i(v, v_c)}{0.1\hat{d}_i}\right) & \text{if } v \in e \\ 0 & \text{if } v \notin e \end{cases}, \quad (1)$$

where  $d_i(v, v_c)$  is the distance for the  $i$ -th imaging data between a vertex  $v$  and its corresponding centroid vertex  $v_c$  in hyperedge  $e$ , and  $\hat{d}_i$  is the average pairwise subject distance for the  $i$ -th imaging data. Accordingly, the vertex degree of the vertex  $v \in \mathcal{V}_i$  and hyperedge degree of the hyperedge  $e \in \mathcal{E}_i$  are calculated, respectively, as  $d_i(v) = \sum_{e \in \mathcal{E}_i} \mathbf{W}_i(e) \mathbf{H}_i(v, e)$  and  $\delta_i(e) = \sum_{v \in \mathcal{V}_i} \mathbf{H}_i(v, e)$ . Thus, we can define  $\mathbf{D}_i^v$  and  $\mathbf{D}_i^e$  as diagonal matrices representing vertex degrees and hyperedge degrees, respectively. Note that all the hyperedges are initialized with an equal weight, *e.g.*, 1.

**Centralized Hypergraph Learning** - To model the relationship of subjects with multiple imaging data, we propose a centralized hypergraph learning method. MCI diagnosis is formulated as a binary classification task in the hypergraph structure. Given four hypergraphs, at each stage one hypergraph is selected as the core hypergraph, while the rest are used to provide extra guidance for updating the hypergraphs.

During centralized learning, hypergraphs are assigned with different weights according to their influences on the structure of the core hypergraph. The core hypergraph is assigned with a weight of  $\alpha_1$ , while the other hypergraphs are assigned with an equal weight,  $\alpha_2$ . Assume the  $j$ -th hypergraph is selected as the core hypergraph, we can have the  $j$ -th centralized hypergraph. To learn the relevance among vertices and improve the structure of the  $j$ -th centralized hypergraph, we optimize the following objective function including the weights of hyperedges by imposing an  $\ell_2$ -norm regularizer on  $\mathbf{W}_i$  ( $i=1,2,3,4$ ) as

$$\begin{aligned} \arg \min_{\mathbf{F}_j, \mathbf{W}_i (i=1,2,3,4)} & \left\{ \Omega_j^c(\mathbf{F}_j) + \lambda \mathcal{R}_{emp}(\mathbf{F}_j) + \mu \sum_i \sum_{e \in \mathcal{E}_i} \mathbf{W}_i(e)^2 \right\}, \\ \text{s.t.} & \quad \mathbf{H}_i \text{diag}(\mathbf{W}_i) = \text{diag}(\mathbf{D}_i^v), \text{diag}(\mathbf{W}_i) \geq 0 \end{aligned} \quad (2)$$

where  $\Omega_j^c(\mathbf{F}_j)$  is the regularizer for the  $j$ -th centralized hypergraph, aiming to smooth the relationship among vertices on the hypergraph structure,  $\mathbf{W}_i(e)$  is the weight of the hyperedge  $e \in \mathcal{E}_i$ ,  $\mathbf{F}_j$  is the to-be-learned relevance matrix, and  $\mathcal{R}_{emp}(\mathbf{F}_j)$  is the empirical loss, and  $\sum_i \sum_{e \in \mathcal{E}_i} \mathbf{W}_i(e)^2$  is the  $\ell_2$  regularizer term to learn the weights

of hyperedges. The two constraints guarantee that the vertex degree is not changed during the learning process and all the weights are non-negative, respectively.  $\Omega_j^c(\mathbf{F}_j)$  is defined as

$$\Omega_j^c(\mathbf{F}_j) = \alpha_1 \Omega_j(\mathbf{F}_j) + \sum_{i \neq j} \alpha_2 \Omega_i(\mathbf{F}_j), \quad (3)$$

where the first term is the regularizer for the core hypergraph and the second term is the regularizer for the other hypergraphs. We define the regularizer term  $\Omega_i(\mathbf{F}_j)$  as

$$\begin{aligned} \Omega_i(\mathbf{F}_j) &= \frac{1}{2} \sum_{k=1}^2 \sum_{e \in \mathcal{E}_i} \sum_{u, v \in \mathcal{V}_i} \frac{\mathbf{W}_i(e) \mathbf{H}_i(u, e) \mathbf{H}_i(v, e)}{\delta_i(e)} \left( \frac{\mathbf{F}_j(u, k)}{\sqrt{d_i(u)}} - \frac{\mathbf{F}_j(v, k)}{\sqrt{d_i(v)}} \right)^2 \\ &= \mathbf{F}_j^T (\mathbf{I} - \Theta_i) \mathbf{F}_j, \end{aligned} \quad (4)$$

where  $\Theta_i = \mathbf{D}_i^{e-\frac{1}{2}} \mathbf{H}_i \mathbf{W}_i \mathbf{D}_i^{e-1} \mathbf{H}_i^T \mathbf{D}_i^{v-\frac{1}{2}}$ .  $\Omega_j(\mathbf{F}_j)$  can be accordingly obtained by  $\Omega_j(\mathbf{F}_j) = \mathbf{F}_j^T (\mathbf{I} - \Theta_j) \mathbf{F}_j$ . Then,  $\Omega_j^c(\mathbf{F}_j)$  can be simplified as

$$\Omega_j^c(\mathbf{F}_j) = \mathbf{F}_j^T \left( \mathbf{I} - (\alpha_1 \Theta_j + \alpha_2 \sum_{i \neq j} \Theta_i) \right) \mathbf{F}_j = \mathbf{F}_j^T \Delta_j^c \mathbf{F}_j, \quad (5)$$

where  $\Delta_j^c = \mathbf{I} - (\alpha_1 \Theta_j + \alpha_2 \sum_{i \neq j} \Theta_i)$  is the  $j$ -th centralized hypergraph Laplacian.

The empirical loss  $\mathcal{R}_{emp}(\mathbf{F}_j)$  is defined as

$$\mathcal{R}_{emp}(\mathbf{F}_j) = \sum_{k=1}^2 \|\mathbf{F}_j(:, k) - \mathbf{Y}(:, k)\|^2, \quad (6)$$

where  $\mathbf{Y} \in \mathbb{R}^{n \times 2}$  is the label matrix and each of its entry denotes the category a subject belongs to.  $\mathbf{Y}(:, k)$  is the  $k$ -th column of  $\mathbf{Y}$ . If the vertex  $v$  belongs to the first category, then the  $(v, 1)$ -th and  $(v, 2)$ -th entries in  $\mathbf{Y}$  are set to 1 and 0, respectively.

**Solution:** To solve the optimization problem in Eq. (2), we employ an alternating optimization approach. We first fix  $\mathbf{W}_i$  and optimize  $\mathbf{F}_j$  as follows

$$\arg \min_{\mathbf{F}_j} \{ \Omega_j^c(\mathbf{F}_j) + \lambda \mathcal{R}_{emp}(\mathbf{F}_j) \}, \quad (7)$$

which leads to the following closed-form solution for

$$\mathbf{F}_j = \frac{\lambda}{1 + \lambda} \left( \mathbf{I} - \frac{1}{1 + \lambda} (\alpha_1 \Theta_j + \alpha_2 \sum_{i \neq j} \Theta_i) \right)^{-1} \mathbf{Y}. \quad (8)$$

We then fix  $\mathbf{F}_j$  and optimize  $\mathbf{W}_i$  as follows

$$\begin{aligned} \arg \min_{\mathbf{W}_i (i=1,2,3,4)} & \left\{ \Omega_j^c(\mathbf{F}_j) + \mu \sum_i \sum_{e \in \mathcal{E}_i} \mathbf{W}_i(e)^2 \right\} \\ \text{s.t.} & \quad \mathbf{H}_i \text{diag}(\mathbf{W}_i) = \text{diag}(\mathbf{D}_i^v), \quad \text{diag}(\mathbf{W}_i) \geq 0. \end{aligned} \quad (9)$$

The above optimization task is convex on  $\mathbf{W}_i$  and can be solved via quadratic programming. The learning process is repeated several times, once for each of the hypergraphs as the core hypergraph, to generate a centralized relevance matrix for each imaging modality.

## 2.1 Relevance Matrices Fusion

To optimally integrate information from multiple MRIs, we learn the weights of each hypergraph by minimizing the overall hypergraph Laplacian. Let  $\rho_i$  be the weight for the  $i$ -th centralized hypergraph, we impose an  $\ell_2$ -norm penalization on the weights of all centralized hypergraphs as follows

$$\arg \min_{\rho_i} \left\{ \sum \rho_i \Omega_i^c(\mathbf{F}_i) + \eta \sum \rho_i^2 \right\} \quad \text{s.t.} \quad \sum \rho_i = 1, \quad (10)$$

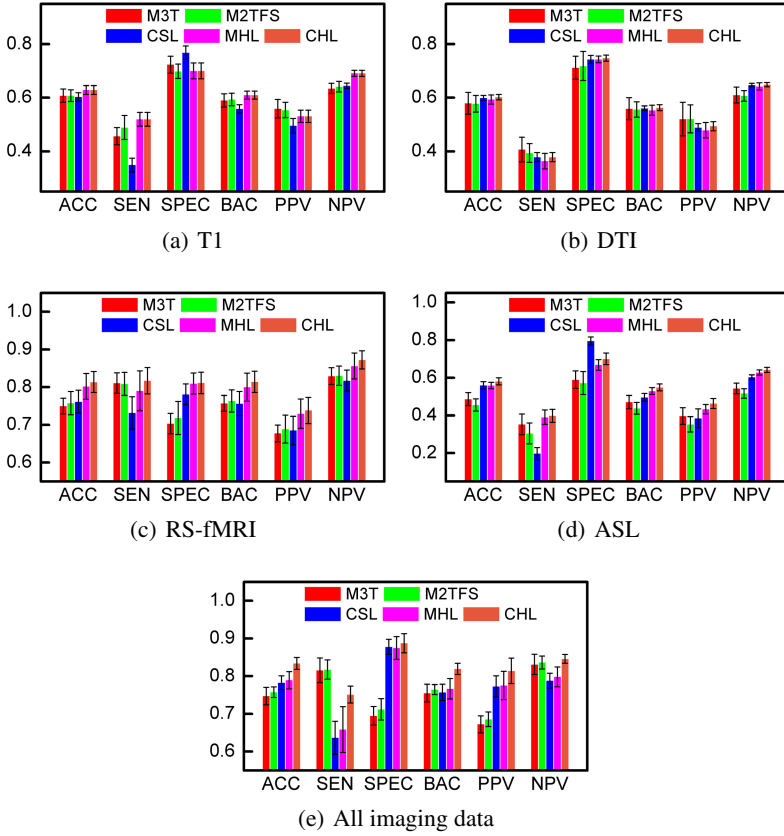
where  $\eta$  is to balance the Laplacian and the weight regularizer. Eq. (10) can be solved using the Lagrangian method. The final relevance matrix is then computed as  $\mathbf{F} = \sum \rho_i \mathbf{F}_i$ . Classification of a subject could therefore be determined by its corresponding value in the final relevance matrix. If the relevance score to MCI is larger than that to NC, a subject is classified as MCI, and vice versa.

## 3 Experiments

**Experimental Settings** - To evaluate the performance of the proposed CHL method, a 10-fold cross-validation strategy is used in our experiments. Specifically, we randomly partition the subjects in the MCI and NC groups into 10 non-overlapping approximately equal sets. One set is first left out as testing set and the rest sets are used as the training set. The training set is further divided into 5 subsets for a 5-fold inner cross-validation to learn the optimal parameters. This procedure is repeated 10 times, once for each of the 10 sets, to compute the overall cross-validation classification performance. Six statistical measures are used for performance evaluation, which include accuracy (ACC), sensitivity (SEN), specificity (SPEC), balance accuracy (BAC), positive predictive value (PPV), and negative predictive value (NPV).

**Results** - We compare the proposed method with two state-of-the-art methods: 1) Multimodal multitask learning (M3T) [11], which employs multimodal imaging data for MCI classification, and 2) Manifold regularized multitask feature learning (M2TFS) [5], which is another semi-supervised method on MCI classification. We further compare with two simplified versions of the proposed method: 1) centralized simple graph learning (CSL), in which only simple graphs are employed instead of hypergraphs in our method, and 2) Multi-hypergraph learning (MHL) [3] without centralized learning.

Figure 2 shows the MCI classification results of all compared methods using different imaging data individually and together. The proposed method (CHL) outperforms all competing state-of-the-art methods. Specifically, by using four types of imaging data, CHL achieves an improvement of  $8.65 \pm 2.50\%$  and  $7.61 \pm 1.92\%$ , in terms of ACC, compared with M3T and M2TFS, respectively. Better performance of our method



**Fig. 2.** MCI identification performance using four types of imaging data individually and together.

is attributed to the following aspects. First, different from existing methods that employ the labeled samples to train classifiers, the proposed method employs both the labeled and unlabeled samples to explore the underlying relationships among different subjects in a semi-supervised learning way, leading to a more general model that is not over-fitted to the training data. Second, the employed hypergraph structure [12] is superior in formulating the joint relationship among multiple vertices compared with simple graph. As shown in the results, CHL achieves an improvement of  $5.10 \pm 2.13\%$  in terms of ACC compared with CSL when all four imaging data are employed. Third, the proposed method employs a centralized learning approach to model the subject relationship by using one of multiple imaging data with the guidance from others, enabling more accurate modeling of underlying relationships among samples.

Compared with MHL, the proposed CHL demonstrates better performance, especially when multiple types of imaging data are utilized. More specifically, CHL achieves an improvement of  $4.42 \pm 1.57\%$  in terms of ACC compared with MHL by using four types of imaging data. This result demonstrates that the proposed centralized method can better explore the relationship underlying the multiple imaging data.

## 4 Conclusion

In this paper, we proposed a centralized hypergraph learning method to model the relationship among subjects with multiple MRIs for MCI identification. In our method, this relationship is encoded by the structure and the weights of hyperedges in a hypergraph. Multiple hypergraphs are constructed using multiple MRIs, respectively. In the learning process, each time one hypergraph is selected as the core and the relationships among subjects from other hypergraphs help to provide extra guidance to meliorate the structure of the core hypergraph. Integrating hypergraphs with different weights enables optimal utilization of supplementary information conveyed by different imaging data. Our findings demonstrate the effectiveness of the CHL method on MCI diagnosis.

## References

- [1] Binnewijzend, M.A., Kuijter, J.P., Benedictus, M.R., van der Flier, W.M., Wink, A.M., Wattjes, M.P., van Berckel, B.N., Scheltens, P., Barkhof, F.: Cerebral blood flow measured with 3D pseudocontinuous arterial spin-labeling MR imaging in alzheimer disease and mild cognitive impairment: a marker for disease severity. *Radiology* 267(1), 221–230 (2013)
- [2] Cai, X., Nie, F., Cai, W., Huang, H.: Heterogeneous image features integration via multi-modal semi-supervised learning model. In: *IEEE International Conference on Computer Vision*, pp. 1737–1744 (2013)
- [3] Gao, Y., Wang, M., Tao, D., Ji, R., Dai, Q.: 3-D object retrieval and recognition with hypergraph analysis. *IEEE Transactions on Image Processing* 21(9), 4290–4303 (2012)
- [4] Hinrichs, C., Singh, V., Xu, G., Johnson, S.C., Initiative, A.D.N., et al.: Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. *NeuroImage* 55(2), 574–589 (2011)
- [5] Jie, B., Zhang, D., Cheng, B., Shen, D.: Manifold regularized multi-task feature selection for multi-modality classification in alzheimer’s disease. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *MICCAI 2013, Part I. LNCS*, vol. 8149, pp. 275–283. Springer, Heidelberg (2013)
- [6] Lim, K.O., Pfefferbaum, A.: Segmentation of mr brain images into cerebrospinal fluid spaces, white and gray matter. *Journal of Computer Assisted Tomography* 13(4), 588–593 (1989)
- [7] Richiardi, J., Monsch, A.U., Haas, T., Barkhof, F., Van de Ville, D., Radü, E.W., Kressig, R.W., Haller, S.: Altered cerebrovascular reactivity velocity in mild cognitive impairment and Alzheimer’s disease. *Neurobiology of Aging* (2014)
- [8] Rubinov, M., Sporns, O.: Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* 52(3), 1059–1069 (2010)
- [9] Wang, H., Nie, F., Huang, H., Risacher, S.L., Saykin, A.J., Shen, L., et al.: Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. *Bioinformatics* 28(12), i127–i136 (2012)
- [10] Ye, J., Wu, T., Li, J., Chen, K.: Machine learning approaches for the neuroimaging study of Alzheimer’s disease. *Computer* 44(4), 99–101 (2011)
- [11] Zhang, D., Shen, D.: Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer’s disease. *NeuroImage* 59(2), 895–907 (2012)
- [12] Zhou, D., Huang, J., Schokopf, B.: Learning with hypergraphs: Clustering, classification, and embedding. In: *Proceedings of Advances in Neural Information Processing Systems*. pp. 1601–1608 (2006)