Relationship Induced Multi-atlas Learning for Alzheimer's Disease Diagnosis

Mingxia Liu[1,2], Daoqiang Zhang[2], Ehsan Adeli-Mosabbeb[1], and Dinggang Shen[1(✉)]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
dgshen@med.unc.edu

[2] School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Abstract. Multi-atlas based methods using magnetic resonance imaging (MRI) have been recently proposed for automatic diagnosis of Alzheimer's disease (AD) and its prodromal stage, i.e., mild cognitive impairment (MCI). However, most existing multi-atlas based methods simply average or concatenate features generated from multiple atlases, which ignores the important underlying structure information of multi-atlas data. In this paper, we propose a novel relationship induced multi-atlas learning (RIML) method for AD/MCI classification. Specifically, we first register each brain image onto multiple selected atlases separately, through which multiple sets of feature representations can be extracted. To exploit the structure information of data, we develop a relationship induced sparse feature selection method, by employing two regularization terms to model the relationships among atlases and among subjects. Finally, we learn a classifier based on selected features in each atlas space, followed by an ensemble classification strategy to combine multiple classifiers for making a final decision. Experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) database demonstrate that our method achieves significant performance improvement for AD/MCI classification, compared with several state-of-the-art methods.

1 Introduction

Brain morphometric pattern analysis using magnetic resonance imaging (MRI) is one of the most popular approaches for automatic diagnosis of Alzheimer's disease (AD) and its early stage, i.e., mild cognitive impairment (MCI). In these methods, all subjects are spatially normalized onto a common space (i.e., a predefined atlas), through which the same brain region across different subjects can be compared [1]. However, due to the potential bias associated with the use of a specific atlas, feature representations extracted from a single atlas may not be sufficient to reveal the underlying complicated differences between populations of disease-affected patients and normal controls (NC).

Recently, several studies [2–4] have shown that multi-atlas based methods usually achieve more accurate diagnosis results than single-atlas based ones.

B. Menze et al. (Eds.): MCV Workshop 2015, LNCS 9601, pp. 24–33, 2016.
DOI: 10.1007/978-3-319-42016-5_3

In multi-atlas based methods, one brain image is non-linearly registered onto multiple atlases, and thus multiple feature representations can be generated for this image. Using multiple atlases could reduce errors due to misregistration, which is helpful for improving subsequent learning performance. However, most of existing multi-atlas based methods simply average or concatenate multiple sets of features generated from multiple atlases, which do not take advantage of the underlying structure information [5,6] of multi-atlas data. In fact, there exists some important structure information, e.g., the relationships among atlases and among subjects. Intuitively, modeling such relationships can bring more prior information into the learning process, which can further boost the learning performance. However, to the best of our knowledge, previous multi-atlas based methods seldome utilize such relationship information in their models.

In this paper, we propose a relationship induced multi-atlas learning (RIML) method for AD/MCI classification. We first non-linearly register each brain image onto multiple selected atlases, and then extract multiple sets of feature representations for each subject from those atlas spaces. Next, we develop a novel relationship induced sparse feature selection model, by considering the relationships among multiple atlases and among different subjects. Finally, we develop an ensemble classification method to better make use of feature representations generated from multiple atlases. Experimental results on the ADNI database demonstrate the efficacy of our method.

2 Proposed Method

Figure 1 illustrates the overview of our proposed method, which includes three major steps: (1) feature extraction, (2) relationship induced sparse feature selection, and (3) ensemble classification. In the first step, brain images are non-linearly registered onto multiple selected atlases separately, and then multiple sets of volumetric features are extracted for each subject in each atlas space. Afterwards, our proposed relationship induced sparse feature selection method is used to select the most discriminative features by considering the underlying structure information in multi-atlas data. Finally, multiple SVM classifiers are constructed based on multiple sets of selected features, followed by an ensemble classification strategy to combine the outputs of multiple classifiers.

2.1 Feature Extraction

For all studied subjects, we first perform a standard pre-processing procedure on the T1-weighted MR brain images. Specifically, we first use the non-parametric non-uniform bias correction [7] method to correct intensity in-homogeneity. Next, we perform skull stripping [8], and double check it to ensure the clean removal of skull and dura. Then, we remove the cerebellum by warping a labeled atlas to each skull-stripped image. Afterwards, we apply the FAST method [9] to segment each brain image into three tissues: gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Here, we only use the GM density map

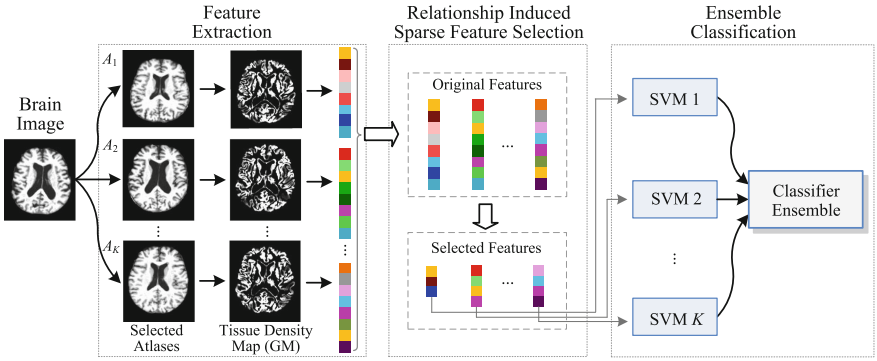

Fig. 1. The overview of our proposed RIML method.

in our feature set, because GM is mostly affected by AD and is widely used in the literature [3,10]. Finally, all brain images are affine-aligned by FLIRT [11].

To obtain multiple atlases, we adopt the affinity propagation (AP) clustering algorithm [12] to partition the whole population of AD and NC images into K non-overlapping groups. The exemplar image of each group is then selected as an atlas, and a total of $K = 10$ atlases (i.e., $A_1, \cdots, A_{10}$) are obtained (see Fig. 2) empirically in this study. We then employ these atlases to capture multiple sets of feature representations for each subject by performing feature extraction as described in [10]. Specifically, for a given subject with three segmented tissues (i.e., GM, WM and CSF), its brain image is first non-linearly registered onto K atlases separately by using a high-dimensional elastic warping tool, i.e., HAMMER [13]. Then, based on these K estimated deformation fields, for each tissue we quantify its voxel-wise tissue density map [14] in each of K atlas spaces, to reflect the unique deformation behavior of a given subject with respect to each specific atlas. In this study, we only use the gray matter (GM) density map for feature extraction and classification, since GM is mostly affected by AD and is widely used in the literature [4,15]. After registration and quantification, we group voxel-wise morphometric features into regional features by using the clustering method proposed in [10] for adaptive feature grouping, followed by a Watershed segmentation [16] process for obtaining the region of interest (ROI) partitions for each of multiple atlases. Here, each atlas will yield its unique ROI partition, because different tissue density maps of the same subject are generated from different atlases. To improve the discriminative power as well as the robustness of volumetric features computed from each ROI, we further refine ROI by choosing the voxels with reasonable representation power. To be specific, we first select the most relevant voxel according to the Pearson correlation between this voxels tissue density values and class labels among all training subjects. Then, we iteratively include the neighboring voxels until no increase in Pearson correlation when adding new voxels. Such voxel selection process will lead to a voxel set for a specific region, and then the mean of tissue density values of those selected voxels can be computed as the feature representation for this

region. Such voxel selection process is important in helping eliminate irrelevant and noisy features, confirmed by several previous studies [4,15,17]. Finally, the top 1500 most discriminative ROI features are selected in each atlas space in this study. By using K atlases, one subject is represented by K sets of feature vectors, where each feature vector is of 1500 dimensions.

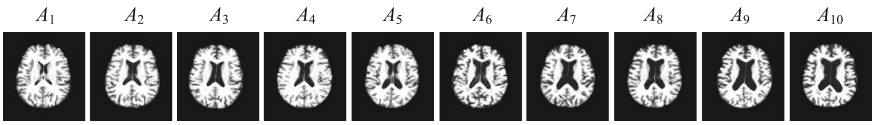

Fig. 2. Selected atlases achieved by the AP clustering algorithm.

2.2 Relationship Induced Sparse Feature Selection

Since multiple atlases are used in this study, feature representations for each subject are high-dimensional, while the number of subjects is usually very limited. In such a case, features could be noisy or redundant, which could degrate the performances of subsequent classifiers [5,18–20]. To this end, we propose a relationship induced sparse feature selection algorithm to find the most informative features in multi-atlas data. Assume we have K learning tasks (corresponding to K atlases). Denote $\mathbf{X}^k = [\mathbf{x}_1^k, \mathbf{x}_2^k, \cdots, \mathbf{x}_n^k, \cdots, \mathbf{x}_N^k]^\top \in \mathbb{R}^{N \times D}$ as the training data for the k^{th} learning task with N training subjects, where $\mathbf{x}_n^k$ represents the column feature vector for the n^{th} training subject in the k^{th} atlas space. Let $\mathbf{y} = [y_1, y_2, \cdots, y_n, \cdots, y_N]^\top \in \mathbb{R}^N$ represent the column response vector for the training data, where $y_n \in \{-1, 1\}$ is the class label for the n^{th} subject. Denote $\mathbf{W} = [\mathbf{w}^1, \mathbf{w}^2, \cdots, \mathbf{w}^k, \cdots, \mathbf{w}^K] \in \mathbb{R}^{D \times K}$ as the weight matrix for K tasks, where $\mathbf{w}^k \in \mathbb{R}^D$ is a column weight vector for the k^{th} task, and $\mathbf{w}_d \in \mathbb{R}^K$ that will be used below as the d^{th} row of $\mathbf{W}$. To encourage the sparsity of $\mathbf{W}$, and to select the most informative features in each atlas space, we propose the following multi-task sparse feature learning model:

$$\min_{\mathbf{W}} \sum_{k=1}^{K} \|\mathbf{y} - \mathbf{X}^k \mathbf{w}^k\|^2 + \lambda_1 \|\mathbf{W}\|_{1,1} \tag{1}$$

where the first term is the empirical loss on the training data, and $\|\mathbf{W}\|_{1,1} = \sum_{d=1}^{D} |\mathbf{w}_d|$ is the sum of ℓ_1-norm of the rows of $\mathbf{W}$ to ensure that only a small subset of features will be selected in each task.

In (1), a linear mapping function (i.e., $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{w}$) is learned to transform the data in original feature space to a one-dimensional label space, which only considers the relationship between samples and class labels. Nevertheless, there exists some other important structure information when we use multiple atlases for extracting feature representations, e.g., (1) the relationship among multiple atlases, and (2) the relationship among subjects. As illustrated in the left panel of Fig. 3, one subject $\mathbf{x}_n$ is represented as $\mathbf{x}_n^{k_1}$ in the $k_1{}^{\text{th}}$ atlas space, and as

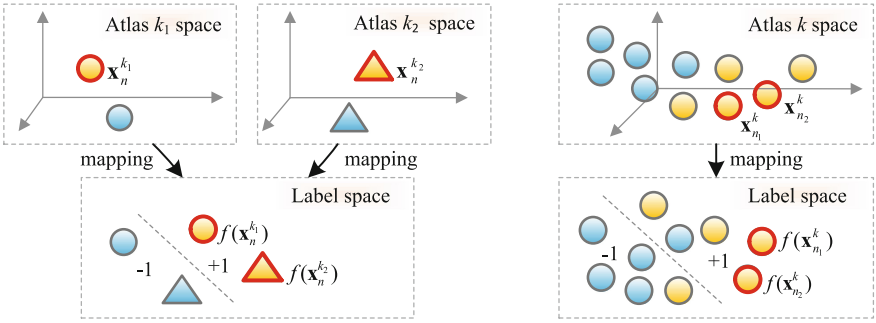

Fig. 3. Illustration of the relationship between two atlases (left panel), and the relationship between two subjects in the same atlas space (right panel).

$\mathbf{x}_n^{k_2}$ in the $k_2{}^{\text{th}}$ atlas space, respectively. After being mapped to the label space, they should be close to each other (i.e., $f(\mathbf{x}_n^{k_1})$ should be similar to $f(\mathbf{x}_n^{k_2})$), since they represent the same subject. Similarly, as shown in the right panel of Fig. 3, if two subjects $\mathbf{x}_{n_1}^k$ and $\mathbf{x}_{n_2}^k$ in the k^{th} atlas space are very similar, their estimated label information should be also similar, i.e., the distance between $f(\mathbf{x}_{n_1}^k)$ and $f(\mathbf{x}_{n_2}^k)$ should be small. To achieve these goals, we first introduce a novel atlas-relationship induced regularization term P as follows:

$$P = \sum_{n=1}^{N} \sum_{k_1=1}^{K} \sum_{k_2=1}^{K} \left(f(\mathbf{x}_n^{k_1}) - f(\mathbf{x}_n^{k_2}) \right)^2 = \sum_{n=1}^{N} tr\left((\mathbf{B}_n \mathbf{W})^\top \mathbf{L}_n (\mathbf{B}_n \mathbf{W}) \right) \quad (2)$$

where $tr(\cdot)$ denotes the trace of a square matrix, $\mathbf{B}_n = [\mathbf{x}_n^1, \cdots, \mathbf{x}_n^K]^\top \in \mathbb{R}^{K \times D}$ represents the n^{th} subject with multiple sets of features generated from K atlas spaces, and $\mathbf{L}_n \in \mathbb{R}^{K \times K}$ is a diagonal matrix with diagonal elements equal to $K-1$ and all the other elements as -1. By using (2), we can model the relationships among multiple atlases explicitly.

We then also propose a subject-relationship induced regularizer Q as follows:

$$Q = \sum_{k=1}^{K} \sum_{n_1=1}^{N} \sum_{n_2=1}^{N} S_{n_1 n_2}^k \left(f(\mathbf{x}_{n_1}^k) - f(\mathbf{x}_{n_2}^k) \right)^2 = \sum_{k=1}^{K} \left(\mathbf{X}^k \mathbf{w}^k \right)^\top \mathbf{L}^k \left(\mathbf{X}^k \mathbf{w}^k \right) \quad (3)$$

where $\mathbf{X}^k$ denotes the data matrix in the k^{th} learning task, and $S_{n_1 n_2}^k$ represents the similarity between the n_1^{th} subject and the n_2^{th} subject in the k^{th} atlas space. Here, $S_{n_1 n_2}^k$ is defined as $e^{-\frac{\|\mathbf{x}_{n_1}^k - \mathbf{x}_{n_2}^k\|^2}{\sigma}}$ if $\mathbf{x}_{n_1}^k$ and $\mathbf{x}_{n_2}^k$ are neighbors, and 0 otherwise. Therefore, $\mathbf{S}^k = \{S_{n_1 n_2}^k\}_{n_1, n_2 = 1}^N \in \mathbb{R}^{N \times N}$ could be a similarity matrix with elements defining the similarities between subjects. Then, $\mathbf{L}^k = \mathbf{D}^k - \mathbf{S}^k$ represents the Laplacian matrix, where $\mathbf{D}^k$ is a diagonal matrix with the element $D_{n_1 n_1}^k = \sum_{n_2=1}^{N} S_{n_1 n_2}^k$. It is easy to see that (3) aims to preserve the local neighboring structure of original data during the mapping, through which the relationships among subjects can be captured explicitly.

By incorporating two relationship induced regularization terms defined in (2) and (3) into (1), our proposed relationship induced sparse feature selection model can be finally formulated as follows:

$$\min_{\mathbf{W}} \sum_{k=1}^{K} \|\mathbf{y} - \mathbf{X}^k \mathbf{w}^k\|^2 + \lambda_1 \|\mathbf{W}\|_{1,1} + \lambda_2 P + \lambda_3 Q \tag{4}$$

where λ_1, λ_2 and λ_3 are positive constants to balance the relative contributions of those four terms in (4), and their values can be determined via inner cross-validation on training data. In (4), the second term is used to find the most discriminative features, and the last two terms are to capture the relationships among atlases and among subjects. Since the objective function in (4) is convex but non-smooth because of the non-smooth term $l_{1,1}$-norm, we adopt a smooth approximation algorithm [21] to solve the proposed problem.

2.3 Ensemble Classification

To better make use of feature representations generated from multiple atlases, we further propose using an ensemble classification approach. Particularly, after feature selection using our relationship induced sparse feature selection algorithm, we obtain K feature subsets from the K different atlases. Based on these selected features, we then construct K SVM classifiers separately, with each classifier corresponding to a specific atlas space. Next, we adopt the majority voting strategy, which is a simple and effective classifier fusion method, to combine the outputs of K SVM classifiers for making a final decision. In this way, the class label of a new test subject can be determined by majority voting for the outputs of those K classifiers.

3 Experiments

3.1 Subjects and Experimental Settings

We evaluate our method on T1-weighted MRI data in ADNI-1 for AD/MCI classification. In the experiments, there are totally 459 subjects randomly selected from those scanned with 1.5 T scanners, including 97 AD, 128 NC, 117 progressive MCI (pMCI), and 117 stable MCI (sMCI) subjects. We perform two groups of experiments, including AD vs. NC classification and pMCI vs. sMCI classification. We compare our RIML method with four widely used feature selection methods, including Pearson correlation (PC), COMPARE [10], t-test, and LASSO [22]. We first use single-atlas (sa) based methods to perform classification, denoted as PC_sa, COMPARE_sa, t-test_sa, and LASSO_sa. Then, we adopt two strategies to deal with features from multiple atlases, i.e., feature concatenation and ensemble. In feature concatenation methods (including PC_con, COMPARE_con, t-test_con, and LASSO_con), we first concatenate features extracted from K atlases ($K = 10$ in this study), and then use a specific feature selection method to select features, followed by a SVM classifier. In ensemble methods (including PC_ens, COMPARE_ens, t-test_ens, and LASSO_ens), we first select features in each atlas space

by using a specific feature selection algorithm, and then construct K SVM classifiers, followed by an ensemble classification process (see Fig. 1).

In the experiments, we use a 10-fold cross-validation strategy to evaluate the performance of different methods, and record the average results among those 10 folds. The regularization parameters in (4) and that for LASSO are chosen from $\{10^{-10}, 10^{-9}, \cdots, 10^{0}\}$, and the p-value in t-test method is chosen from $\{0.05, 0.08, 0.10, 0.12, 0.15\}$ via inner cross-validation on the training data. We use the linear SVM with default parameters as classifier, and evaluate the performance of different methods via four criteria, including classification accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under the receiver operating characteristic curve (AUC).

3.2 Results and Analysis

First, we report the results achieved by four single-atlas based methods and nine multi-atlas based methods in Table 1. For the single-atlas based methods, the averaged results using K individual atlases are provided. From Table 1, we can observe three main points: First, multi-atlas based methods generally achieve better performances than single-atlas based methods (i.e., PC_sa, COMPARE_sa, t-test_sa, and LASSO_sa). For instance, in AD vs. NC classification, the best accuracy achieved by single-atlas based methods is only 84.32 % (by LASSO_sa), which is usually lower than those of multi-atlas based methods. Second, when using multiple atlases, our proposed ensemble strategy (i.e., PC_en, COMPARE_ens, t-test_ens, and LASSO_ens) usually outperforms their counterparts using feature concatenation strategy (i.e., PC_con, COMPARE_con, t-test_con, and LASSO_con). Third, in most cases, our RIML method achieves better results than the compared methods. Also, our method

Then, we compare the results achieved by RIML with several state-of-the-art multi-atlas based methods using MRI data from ADNI, with results shown in Table 2. From Table 2, it is obvious that our RIML method generally outperforms the other three methods in both AD vs. NC classification and pMCI vs. sMCI

Table 1. Comparison of RIML with different methods in two classification tasks.

Method		AD vs. NC				pMCI vs. sMCI			
		ACC (%)	SEN (%)	SPE (%)	AUC (%)	ACC (%)	SEN (%)	SPE (%)	AUC (%)
Single-atlas	PC_sa	84.00	79.53	87.45	76.92	68.49	67.80	69.10	62.85
	COMPARE_sa	84.18	75.33	89.17	78.70	70.06	68.08	72.02	63.56
	t-test_sa	76.27	68.50	83.01	74.96	61.99	64.93	73.11	65.16
	LASSO_sa	84.32	81.66	86.36	84.02	72.06	72.04	72.02	72.03
	PC_con	84.01	81.56	89.23	81.91	72.78	74.62	70.91	72.45
	COMPARE_con	84.93	80.11	87.03	79.07	73.35	75.76	70.83	74.05
	t-test_con	81.87	70.77	**90.71**	81.78	61.60	64.32	75.01	71.63
	LASSO_con	86.62	84.78	89.80	87.29	71.49	76.06	66.67	71.36
Multi-atlas	PC_ens	85.59	82.44	89.93	91.51	73.92	73.38	72.32	76.29
	COMPARE_ens	86.61	85.44	89.23	90.85	75.56	75.75	73.48	76.58
	t-test_ens	84.31	74.56	89.70	88.78	63.36	60.60	71.74	63.33
	LASSO_ens	87.27	84.78	89.23	92.79	75.32	81.36	69.17	76.02
	RIML(ours)	**93.06**	**94.85**	90.49	**95.79**	**79.25**	**87.92**	**75.54**	**83.44**

Table 2. Comparison with the state-of-the-art methods using MRI data of ADNI.

Method	AD vs. NC			pMCI vs. sMCI		
	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)
Wolz et al. [2]	89.00	85.00	93.00	68.00	67.00	69.00
Koikkalainen et al. [3]	86.00	81.00	91.00	72.10	77.00	71.00
Min et al. [15]	91.64	88.56	**93.85**	72.41	72.12	72.58
RIML(ours)	**93.06**	**94.85**	90.49	**79.25**	**87.92**	**75.54**

classification. Although the study in [15] reported the best specificity in AD vs. NC classification, its accuracy and sensitivity are lower than those of the proposed RIML method.

3.3 Diversity Analysis

In order to understand how our ensemble classification method works, we further plot a diversity-error diagram [23] to evaluate the level of agreement between the outputs of two classifiers. In Fig. 4, we show diagrams achieved by five ensemble-based methods. For each method, the corresponding ensemble contains K ($K = 10$ in this study) individual SVM classifiers. In a diversity-error diagram, the value on the x-axis denotes the kappa diversity of a pair of classifiers in the ensemble, and that on the y-axis represents the averaged individual error of a pair of classifiers. The most desirable pairs of classifiers will be close to the bottom left corner of the diagram, since a small value of kappa diversity indicates better diversity and a small value of averaged error indicates a better accuracy. For visual evaluation of relative positions of kappa-error point clouds, we also plot the centroids of clouds for different methods in Fig. 4 (denoted as squares). It can be seen from Fig. 4 that our RIML method usually yields better diversity and lower classification error than the other four methods, implying that RIML can achieve a better trade-off between the classification error and the classifier

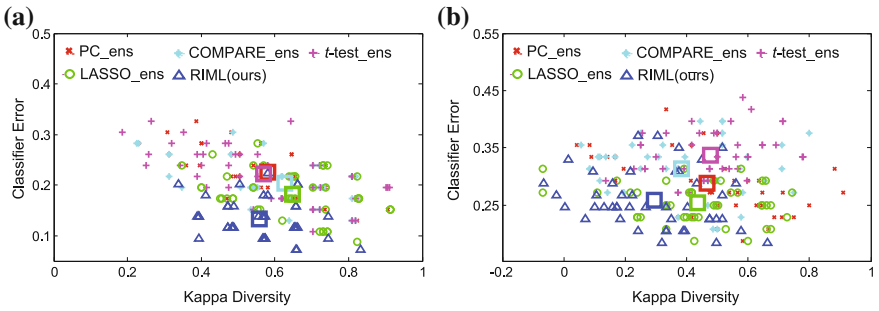

Fig. 4. Diversity-error diagrams of classifiers achieved by five ensemble-based methods in (a) AD vs. NC classification, and (b) pMCI vs. sMCI classification.

diversity than the compared methods. That is, it builds a classifier ensemble based on the reasonably diverse and accurate individual components.

4 Conclusion

In this paper, we have proposed a relationship induced multi-atlas learning (RIML) method for AD/MCI classification. Specifically, we first extracted feature representations from multiple selected atlases, and then proposed a relationship induced sparse feature selection method, followed by an ensemble classification method. Experimental results on the ADNI database demonstrate that our RIML method achieves significant performance improvement in both AD vs. NC classification and pMCI vs. sMCI classification, compared with several state-of-the-art methods.

Acknowledgements. This study was supported by NIH grants (EB006733, EB008374, EB009634, MH100217, AG041721, and AG042599), the National Natural Science Foundation of China (Nos. 61422204, 61473149), the Jiangsu Natural Science Foundation for Distinguished Young Scholar (No. BK20130034), and the NUAA Fundamental Research Fund under grant number NE2013105.

References

1. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., Initiative, A.D.N.: Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. NeuroImage **56**(2), 766–781 (2011)
2. Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D.P., Rueckert, D., Soininen, H., Lötjönen, J., Alzheimer's Disease Neuroimaging Initiative: Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. PLOS ONE **6**(10), e25446 (2011)
3. Koikkalainen, J., Lötjönen, J., Thurfjell, L., Rueckert, D., Waldemar, G., Soininen, H., Alzheimer's Disease Neuroimaging Initiative: Multi-template tensor-based morphometry: application to analysis of Alzheimer's disease. NeuroImage **56**(3), 1134–1144 (2011)
4. Liu, M., Zhang, D., Shen, D.: View-centralized multi-atlas classification for Alzheimer's disease diagnosis. Hum. Brain Mapp. **36**, 1847–1865 (2015)
5. Liu, M., Zhang, D.: Pairwise constraint-guided sparse learning for feature selection. IEEE Transactions on Cybernetics **46**, 298–310 (2015)
6. Liu, M., Zhang, D., Chen, S., Xue, H.: Joint binary classifier learning for ECOC-based multi-class classification. IEEE Trans. Pattern Anal. Mach. Intell. (2015)
7. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging **17**(1), 87–97 (1998)
8. Wang, Y., Nie, J., Yap, P.T., Li, G., Shi, F., Geng, X., Guo, L., Shen, D., Initiative, A.D.N.: Knowledge-guided robust MRI brain extraction for diverse large-scale neuroimaging studies on humans and non-human primates. PLOS ONE **9**(1), e77810 (2014)

9. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging **20**(1), 45–57 (2001)
10. Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C.: COMPARE: classification of morphological patterns using adaptive regional elements. IEEE Trans. Med. Imaging **26**(1), 93–105 (2007)
11. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. Med. Image Anal. **5**(2), 143–156 (2001)
12. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science **315**(5814), 972–976 (2007)
13. Shen, D., Davatzikos, C.: Hammer: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imaging **21**(11), 1421–1439 (2002)
14. Goldszal, A.F., Davatzikos, C., Pham, D.L., Yan, M.X., Bryan, R.N., Resnick, S.M.: An image-processing system for qualitative and quantitative volumetric analysis of brain images. J. Comput. Assist. Tomogr. **22**(5), 827–837 (1998)
15. Min, R., Wu, G., Cheng, J., Wang, Q., Shen, D.: Multi-atlas based representations for Alzheimer's disease diagnosis. Hum. Brain Mapp. **35**(10), 5052–5070 (2014)
16. Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 583–598 (1991)
17. Jin, Y., Shi, Y., Zhan, L., Gutman, B.A., de Zubicaray, G.I., McMahon, K.L., Wright, M.J., Toga, A.W., Thompson, P.M.: Automatic clustering of white matter fibers in brain diffusion mri with an application to genetics. Neuroimage **100**, 75–90 (2014)
18. Gao, Y., Wang, M., Tao, D., Ji, R., Dai, Q.: 3-d object retrieval and recognition with hypergraph analysis. IEEE Trans. Image Process. **21**(9), 4290–4303 (2012)
19. Ji, R., Gao, Y., Hong, R., Liu, Q., Tao, D., Li, X.: Spectral-spatial constraint hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **52**(3), 1811–1824 (2014)
20. Liu, M., Miao, L., Zhang, D.: Two-stage cost-sensitive learning for software defect prediction. IEEE Trans. Reliab. **63**(2), 676–686 (2014)
21. Nesterov, Y.: Smooth minimization of non-smooth functions. Math. Program. **103**(1), 127–152 (2005)
22. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B (Methodological) **58**, 267–288 (1996)
23. Rodriguez, J.J., Kuncheva, L.I., Alonso, C.J.: Rotation forest: a new classifier ensemble method. IEEE Trans. Pattern Anal. Mach. Intell. **28**(10), 1619–1630 (2006)