

# Inherent Structure-Guided Multi-view Learning for Alzheimer's Disease and Mild Cognitive Impairment Classification

Mingxia Liu<sup>1,2</sup>, Daoqiang Zhang<sup>2</sup>, and Dinggang Shen<sup>1</sup>(✉)

<sup>1</sup> Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

dgshen@med.unc.edu

<sup>2</sup> School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

**Abstract.** Multi-atlas based morphometric pattern analysis has been recently proposed for the automatic diagnosis of Alzheimer's disease (AD) and its early stage, i.e., mild cognitive impairment (MCI), where multi-view feature representations for subjects are generated by using multiple atlases. However, existing multi-atlas based methods usually assume that each class is represented by a specific type of data distribution (i.e., a single cluster), while the underlying distribution of data is actually a priori unknown. In this paper, we propose an inherent structure-guided multi-view learning (ISML) method for AD/MCI classification. Specifically, we first extract multi-view features for subjects using multiple selected atlases, and then cluster subjects in the original classes into several sub-classes (i.e., clusters) in each atlas space. Then, we encode each subject with a new label vector, by considering both the original class labels and the coding vectors for those sub-classes, followed by a multi-task feature selection model in each of multi-atlas spaces. Finally, we learn multiple SVM classifiers based on the selected features, and fuse them together by an ensemble classification method. Experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) database demonstrate that our method achieves better performance than several state-of-the-art methods in AD/MCI classification.

## 1 Introduction

Multi-atlas based morphometric pattern analysis using magnetic resonance imaging (MRI) data are recently proposed for automatic diagnosis of Alzheimer's disease (AD) and its early stage, i.e., mild cognitive impairment (MCI) [1,2,3,4]. Generally, multi-atlas based methods mainly focus on the direct morphometric measurement of spatial brain atrophy of subjects, by non-linearly registering a brain image onto multiple atlases. Thus, multi-view feature representations can be generated from those multi-atlas spaces for each subject, where each atlas is regarded as a specific view. Compared with single-atlas based methods, multi-atlas based methods can reduce registration errors by using multiple atlases, which is helpful in improving subsequent learning performance [1,2,5].

In the literature, most of existing multi-atlas based methods simply assume that each class is represented by a specific type of data distribution (i.e., a single cluster). Although such assumption may simplify the problem at hand, it will definitely degrade the learning performance because the underlying distribution structure of data is actually a prior unknown. In practice, the potentially complicated distribution structure of neuroimaging data within a specific class could result from several facts [6], e.g., 1) different sub-types of a specific disease, and 2) an inaccurate clinical diagnosis. Intuitively, modeling such inherent structure of data distribution can bring more prior information to the learning process. However, no previous methods employ such information in their learning models.

In this paper, we propose an inherent structure-guided multi-view learning (ISML) method for AD/MCI classification. Specifically, we first non-linearly register each brain image onto multiple selected atlases, through which multi-view feature representations for each subject can be obtained from different atlases. To uncover the inherent distribution structure of data, we partition subjects in original classes into several sub-classes (i.e., clusters) by using a clustering algorithm. Then, we encode each of sub-classes with a unique coding vector, and regard these coding vectors as new class labels for corresponding subjects. Next, we adopt a multi-task feature selection method to select the most informative features in each atlas space. Based on these selected features, we then learn multiple SVM classifiers, with each SVM corresponding to a specific atlas space. Finally, we fuse these classifiers by an ensemble classification method. Experiments on the ADNI database demonstrate that our method outperforms several state-of-the-art methods in AD/MCI classification.

## 2 Proposed Method

Figure 1 illustrates the overview of our inherent structure-guided multi-atlas learning (ISML) method, which includes three main steps, i.e., 1) feature extraction, 2) inherent structure-guided sparse feature selection and 3) ensemble classification. Specifically, we first non-linearly register the brain images of those subjects onto multiple selected atlases, and then extract volumetric features from the gray matter (GM) tissue density map within each of multi-atlas spaces. Afterwards, we perform feature selection using the proposed inherent structure-guided sparse feature selection method, where we cluster the original classes into several sub-classes and perform sparse feature selection using a multi-task feature selection method. With the selected features, we then learn a support vector machine (SVM) classifier in each of multi-atlas spaces, followed by an ensemble classification approach to combine those SVMs for making a final decision. In what follows, we will introduce each step in detail.

### 2.1 Feature Extraction

In this study, we apply a standard image pre-processing procedure to the T1-weighted MR brain images for all studied subjects. Specifically, we first perform

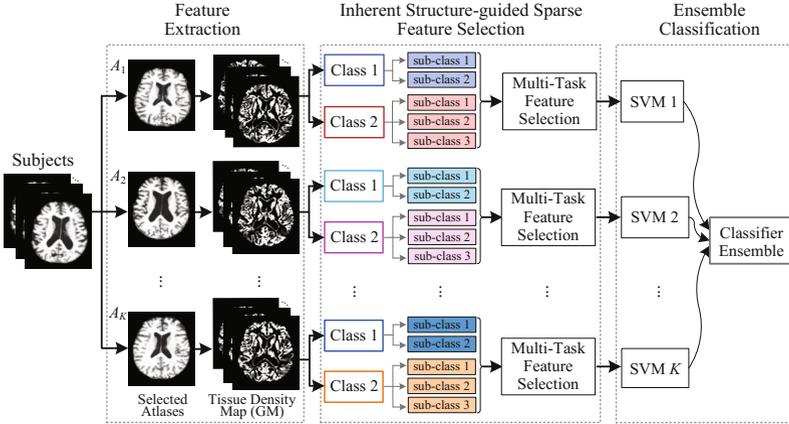


Fig. 1. The overview of our proposed ISML method.

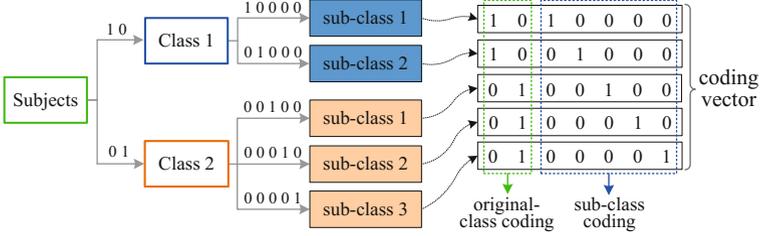
a non-parametric non-uniform bias correction [7] on the MR images to correct intensity in-homogeneity. Then, we perform a skull stripping [8] procedure, followed by a manual review or correction to ensure that the skull and the dura have been removed cleanly. Next, we remove the cerebellum by warping a labeled atlas to each skull-stripping image. Afterwards, we use the FAST method proposed in [9] to segment each brain image into three tissues, i.e., gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). Finally, all brain images are affine-aligned by the FLIRT method [10].

To select appropriate atlases, we adopt a clustering method to select multiple atlases from all studied subjects. Specifically, we first partition the whole population of AD and NC brain images into  $K$  ( $K = 10$  in this study) non-overlapping clusters by using the Affinity Propagation (AP) clustering algorithm [11], through which an exemplar image can be automatically determined for each cluster. We then regard these exemplar images as selected atlases (i.e.,  $A_1, \dots, A_K$ ), which is shown in Fig. 1. By performing feature extraction as described in [12] in each atlas space, we can obtain  $D$ -dimensional ( $D = 1500$  in this study) features from the GM tissue density map for each subject. Given  $K$  atlases, we now have  $K$  sets of feature representations for each subject.

## 2.2 Inherent Structure-Guided Sparse Feature Selection

To utilize the structure information of data, we first divide subjects in original classes into several sub-classes, and then encode these sub-classes with a popular one-versus-all (OVA) encoding strategy [13], followed by a multi-task feature selection model to select the most informative features in each atlas space.

**Sub-class Clustering and Encoding.** For a specific class, we exploit the affinity propagation (AP) algorithm [11] to automatically partition the subjects within this class into several sub-classes (i.e., clusters) in each atlas space, where



**Fig. 2.** Illustration of the sub-class clustering and encoding method.

the cluster number can be determined by cross validation on training data. As shown in Fig. 2, subjects belong to two original classes, i.e., Class 1 and Class 2. Using the AP algorithm, we partition the subjects in Class 1 into two sub-classes, and divide subjects in Class 2 into three sub-classes (see Fig. 2).

Then, we label all subjects with new class labels by encoding the original classes and those sub-classes using the OVA encoding strategy. For the original classes (see Fig. 2), each class is represented by a unique OVA coding vector, i.e.,  $[1\ 0]$  for Class 1 and  $[0\ 1]$  for Class 2. For those sub-classes in Class 1 and Class 2, they are encoded in a combined encoding manner, by concatenating the coding vectors of their original classes and their own OVA coding vectors. For example, the coding vector for sub-class 1 in Class 1 is set as  $[1\ 0\ 1\ 0\ 0\ 0\ 0]$ , where the first two bits denote the OVA coding for Class 1, and the last five bits represent its unique OVA coding among five sub-classes (2 sub-classes in Class 1 and 3 sub-classes in Class 2). Afterwards, we regard these 7-bit coding vectors as the new class labels for corresponding subjects. In this way, the original binary classification problem is transformed into a multi-task (e.g., 7 tasks in Fig. 2) learning problem, through which the structure information of the original classes can be incorporated into the learning process.

**Multi-task Feature Selection.** As high-dimensional features could be redundant or noisy that has attracted research attention in many fields (e.g., image retrieval [14] and classification [15]), feature selection remains a popular research topic for dealing with high-dimensional features. Here, we exploit a multi-task feature selection model to perform feature selection in each atlas space. We denote matrices as boldface uppercase letters, vectors as boldface lowercase letters, and scalars as normal italic letters, respectively. Denote  $\mathbf{a}_i$  and  $\mathbf{a}^j$  as the  $i^{\text{th}}$  row and the  $j^{\text{th}}$  column of a matrix  $\mathbf{A}$ , respectively. We further denote the Frobenius norm and the  $\ell_{2,1}$  norm of  $\mathbf{A}$  as  $\|\mathbf{A}\|_F = \sqrt{\sum_i \|\mathbf{a}_i\|^2}$  and  $\|\mathbf{A}\|_{2,1} = \sum_i \|\mathbf{a}_i\|$ , respectively. Denote  $\mathbf{X} \in \mathbb{R}^{N \times D}$  as the data matrix with  $N$  subjects of  $D$ -dimensional features. Let  $\mathbf{Y} \in \mathbb{R}^{N \times C}$  denote the new class label matrix for  $N$  subjects, where each subject is labeled with a  $C$ -bit row vector. Here,  $C$  is the sum of the number of original classes and the number of sub-classes of all original classes. Denote  $\mathbf{W} = [\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^c, \dots, \mathbf{w}^C] \in \mathbb{R}^{D \times C}$  as the weight matrix for  $C$  learning tasks, where  $\mathbf{w}^c$  is the column weight vector for the  $c^{\text{th}}$  learning task. To jointly select common features among different tasks, the multi-task

feature selection (MTFS) model is defined as follows:

$$\min_{\mathbf{W}} \|\mathbf{Y} - \mathbf{X}\mathbf{W}\|_F^2 + \lambda \|\mathbf{W}\|_{2,1} \quad (1)$$

where  $\|\mathbf{W}\|_{2,1}$  is a group sparsity regularization, and  $\lambda$  is a parameter to trade off the balance between the empirical loss on the training data and the regularization term. Due to the group sparsity nature of  $\ell_{2,1}$  norm, the estimated optimal coefficient matrix  $\widehat{\mathbf{W}}$  will have some zero-value row vectors, implying that the corresponding features are not useful in predicting the new class labels of training data. With the MTFS model defined in (1), we can select the most informative features in each atlas space, where the distribution structure of the original classes are used to guide the feature selection process.

### 2.3 Ensemble Classification

Then, we further propose an ensemble classification approach. To be specific, after feature selection in each atlas space, we can obtain  $K$  feature subsets using  $K$  atlases. Based on those selected features, we then learn  $K$  SVM classifiers, with each SVM corresponding to a specific atlas space. Finally, we combine these classifiers together by a majority voting strategy that is a simple and effective classifier fusion method. Given a new test sample, its class label can be determined by majority voting for the outputs of those SVMs.

## 3 Experiments

### 3.1 Data and Experimental Settings

To demonstrate the efficacy of our proposed method, we perform experiments on part of subjects in the ADNI database with T1-weighted MR images. There are totally 459 subjects randomly selected from those scanned with 1.5T scanners, including 97 AD, 117 progressive MCI (pMCI), 117 stable MCI (sMCI) and 128 normal controls (NC). In this study, we perform four groups of experiments, including 1) AD vs. NC classification, 2) pMCI vs. sMCI classification, 3) pMCI vs. NC classification, and 4) sMCI vs. NC classification. We employ a 10-fold cross-validation strategy to evaluate the performance of different methods. Specifically, all samples are partitioned into 10 subsets (each subset with a roughly equal size), and each time samples in one sub-set are successively selected as the test data, while those in the other nine subsets are used as the training data to perform feature selection and classifier construction. Finally, we report the average values of classification results among those 10 folds.

We demonstrate the advantages of our proposed ISML method from two aspects. First, we compare ISML with three feature selection algorithms using the proposed ensemble classification strategy, including 1) Pearson Correlation (PC), 2) COMPARE proposed in [12], and 3) LASSO. Second, we compare our ensemble classification method with conventional feature concatenation methods (i.e., PC\_con, COMPARE\_con, LASSO\_con, and ISML\_con). That is, we

**Table 1.** Comparison of ISML with different methods in four classification tasks

Method	AD vs. NC			pMCI vs. sMCI			pMCI vs. NC			sMCI vs. NC		
	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)
PC_con	84.01	81.56	89.23	72.78	74.62	70.91	75.42	65.17	88.41	61.53	62.35	66.32
PC	85.59	82.44	89.93	73.92	73.38	72.32	78.09	70.15	89.17	65.93	64.92	72.63
COMPARE_con	84.93	80.11	87.03	73.35	75.76	70.83	76.71	68.22	85.26	62.75	55.68	69.68
COMPARE	86.61	85.44	89.23	75.56	75.75	73.48	80.63	73.18	89.55	67.81	65.90	<b>76.97</b>
LASSO_con	86.62	84.78	89.80	71.49	76.06	66.67	84.96	79.54	89.74	62.02	68.86	56.54
LASSO	87.27	84.78	89.23	75.32	81.36	69.17	85.35	82.48	87.15	68.38	70.69	54.80
ISML_con(ours)	88.06	88.78	87.50	76.13	78.63	73.33	86.99	85.38	88.02	69.29	71.17	71.03
ISML(ours)	<b>93.83</b>	<b>92.78</b>	<b>95.69</b>	<b>80.90</b>	<b>85.95</b>	<b>78.41</b>	<b>89.09</b>	<b>87.66</b>	<b>90.25</b>	<b>71.20</b>	<b>72.42</b>	70.91

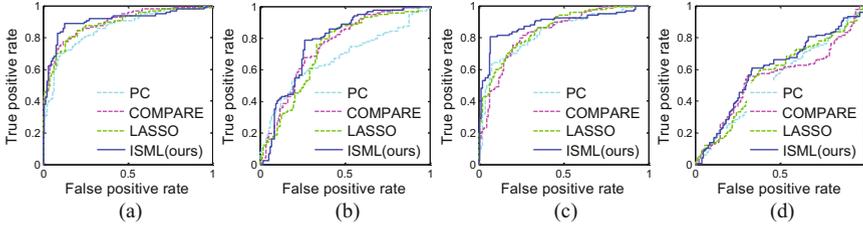
first concatenate those multi-view feature representations as a long vector, and then use a specific feature selection method (e.g., PC, COMPARE, LASSO, and ISML) to select the most informative features, followed by a SVM classifier. For fair comparison, all compared methods share the same multi-view feature representations generated from  $K$  atlases.

The regularization parameter (i.e.,  $\lambda$ ) in equation 1 and that in LASSO are chosen from the range  $\{2^{-10}, 2^{-9}, \dots, 2^0\}$  through inner cross-validation on the training data. There are  $K$  ( $K = 10$  in this study) atlases selected from AD and NC subjects. Using these selected atlases, we extract  $D$ -dimensional ( $D = 1500$  in this study) features from each of  $K$  atlas spaces. Here, we use the linear SVM with default parameters as the benchmark classifier. In addition, we evaluate the performance of different methods via three criteria, including classification accuracy (ACC), sensitivity (SEN), and specificity (SPE).

### 3.2 Results and Discussion

We first record the averaged number of sub-classes (i.e., clusters) identified by our ISML method in the cross-validation process, i.e., 2.1/2.9 for AD/NC, 3.2/3.8 for pMCI/sMCI, 3.1/2.5 for pMCI/NC, and 3.2/2.3 for sMCI/NC. These results validate our intuition that there exists heterogeneity within a specific class.

Then, we report the results achieved by different methods in four classification tasks in Table 1, where the best results are shown in boldface. We further plot the ROC curves achieved by four ensemble-based methods in Fig. 3. From Table 1 and Fig. 3, one can observe two main points. First, in four classification tasks, our proposed ISML method generally outperforms the compared methods in terms of accuracy, sensitivity and specificity. For example, the accuracy achieved by ISML in AD vs. NC classification is 93.83%, which is much higher than the second best accuracy achieved by ISML\_con (i.e., 88.06%). Although the COMPARE method obtains the best specificity in sMCI vs. NC classification, its accuracy and sensitivity are much lower than those achieved by ISML. Second, methods using the ensemble classification strategy (i.e., PC, COMPARE, LASSO, and ISML) generally outperform their conventional counterparts that adopt the feature concatenation strategy (i.e., PC\_con, COMPARE\_con, LASSO\_con, and ISML\_con).



**Fig. 3.** ROC curves achieved by four ensemble-based methods in (a) AD vs. NC, (b) pMCI vs. sMCI, (c) pMCI vs. NC, and (d) sMCI vs. NC classification.

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

Method	AD vs. NC			pMCI vs. sMCI			pMCI vs. NC		
	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)	ACC (%)	SEN (%)	SPE (%)
Cuingnet et al. [16]	88.58	81.00	95.00	70.40	57.00	78.00	81.17	73.00	85.00
Wolz et al. [1]	89.00	85.00	93.00	68.00	67.00	69.00	84.00	82.00	86.00
Koikkalainen et al. [2]	86.00	81.00	91.00	72.10	77.00	71.00	-	-	-
Min et al. [3]	91.64	88.56	93.85	72.41	72.12	72.58	-	-	-
Liu et al. [5]	92.51	<b>92.89</b>	88.33	78.88	85.45	76.06	88.26	86.43	89.94
ISML(ours)	<b>93.83</b>	92.78	<b>95.69</b>	<b>80.90</b>	<b>85.95</b>	<b>78.41</b>	<b>89.09</b>	<b>87.66</b>	<b>90.25</b>

Furthermore, we compare our ISML method with several state-of-the-art methods using MRI data of ADNI, with results given in Table 2. Since very few works reports the results of sMCI vs. NC classification, we only show the results of three classification tasks (i.e., AD vs. NC, pMCI vs. sMCI, and pMCI vs. NC classification) in Table 2. It is worth noting that the method proposed in [16] employs single-atlas feature representation extracted from one single atlas, while the others use multi-atlas based feature representations. It can be seen from Table 2 that, in three classification tasks, our ISML method yields the best classification results in terms of both accuracy and specificity, and comparable sensitivity with the work in [5].

## 4 Conclusion

In this paper, we propose an inherent structure-guided multi-view learning (ISML) method for AD/MCI classification. Specifically, we first extract multi-view features for subjects, and then cluster subjects in original classes into several sub-classes, followed by a multi-task feature selection algorithm. Finally, we develop an ensemble classification method by fusing multiple classifiers constructed in multi-atlas spaces. Experimental results on the ADNI database demonstrate the efficacy of our ISML method.

**Acknowledgments.** 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).

## References

1. Wolz, R., Julkunen, V., Koikkalainen, J., Niskanen, E., Zhang, D.P., Rueckert, D., Soininen, H., Lötjönen, J., Initiative, A.D.N., et al.: Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. *PLoS One* **6**(10), e25446 (2011)
2. Koikkalainen, J., Lötjönen, J., Thurfjell, L., Rueckert, D., Waldemar, G., Soininen, H., Initiative, A.D.N., et al.: Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease. *NeuroImage* **56**(3), 1134–1144 (2011)
3. Min, R., Wu, G., Cheng, J., Wang, Q., Shen, D.: Multi-atlas based representations for alzheimer's disease diagnosis. *Human Brain Mapping* **35**(10), 5052–5070 (2014)
4. Jin, Y., Shi, Y., Zhan, L., Gutman, B.A., de Zubicaray, G.L., 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)
5. Liu, M., Zhang, D., Shen, D.: View-centralized multi-atlas classification for alzheimer's disease diagnosis. *Human Brain Mapping* **36**(5), 1847–1865 (2015)
6. Noppeney, U., Penny, W.D., Price, C.J., Flandin, G., Friston, K.J.: Identification of degenerate neuronal systems based on intersubject variability. *NeuroImage* **30**(3), 885–890 (2006)
7. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical 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., et al.: 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 Transactions on Medical Imaging* **20**(1), 45–57 (2001)
10. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* **5**(2), 143–156 (2001)
11. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. *Science* **315**(5814), 972–976 (2007)
12. Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C.: COMPARE: Classification of morphological patterns using adaptive regional elements. *IEEE Transactions on Medical Imaging* **26**(1), 93–105 (2007)
13. Liu, M., Zhang, D., Chen, S., Xue, H.: Joint binary classifier learning for ecoc-based multi-class classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2015)
14. 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)
15. Ji, R., Gao, Y., Hong, R., Liu, Q., Tao, D., Li, X.: Spectral-spatial constraint hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **52**(3), 1811–1824 (2014)
16. Cuingnet, R., Gerardin, E., Tessieras, J., Auzias, G., Lehéricy, S., Habert, M.O., Chupin, M., Benali, H., Colliot, O., Initiative, A.D.N., et al.: 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)