Multi-modal Neuroimaging Data Fusion via Latent Space Learning for Alzheimer's Disease Diagnosis

Tao Zhou[1], Kim-Han Thung[1], Mingxia Liu[1], Feng Shi[2], Changqing Zhang[1,3],
and Dinggang Shen[1](✉)

[1] Department of Radiology and BRIC, University of North Carolina,
Chapel Hill, USA
dgshen@med.unc.edu
[2] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China
[3] School of Computer Science and Technology, Tianjin University, Tianjin, China

Abstract. Recent studies have shown that fusing multi-modal neuroimaging data can improve the performance of Alzheimer's Disease (AD) diagnosis. However, most existing methods simply concatenate features from each modality without appropriate consideration of the correlations among multi-modalities. Besides, existing methods often employ feature selection (or fusion) and classifier training in two independent steps without consideration of the fact that the two pipelined steps are highly related to each other. Furthermore, existing methods that make prediction based on a single classifier may not be able to address the heterogeneity of the AD progression. To address these issues, we propose a novel AD diagnosis framework based on latent space learning with ensemble classifiers, by integrating the latent representation learning and ensemble of multiple diversified classifiers learning into a unified framework. To this end, we first project the neuroimaging data from different modalities into a common latent space, and impose a joint sparsity constraint on the concatenated projection matrices. Then, we map the learned latent representations into the label space to learn multiple diversified classifiers and aggregate their predictions to obtain the final classification result. Experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset show that our method outperforms other state-of-the-art methods.

1 Introduction

Alzheimer's disease (AD) impairs patients' memory and other cognitive functions and is often found in people over 65 years old [1]. As there is no cure for AD, timely and accurate diagnosis of AD and its prodromal stage (i.e., Mild Cognitive Impairment (MCI)) is highly desirable in clinical practices.

Neuroimaging techniques including Magnetic Resonance Imaging (MRI) and Positr-on Emission Topography (PET) have been widely used to investigate the neurophysiological characteristics of AD [15,18]. As neuroimaging data are very

I. Rekik et al. (Eds.): PRIME 2018, LNCS 11121, pp. 76–84, 2018.
https://doi.org/10.1007/978-3-030-00320-3_10

high-dimensional, existing methods often use Region-Of-Interest (ROI) based features instead of the original voxel based features for analysis [4,17]. Recently, many studies have been proposed to fuse the complementary information from multi-modality data for accurate AD diagnosis [10,14,19]. For example, Zhu et al. [19] use Canonical Correlation Analysis (CCA) to first transform multi-modality data into a common CCA space, and then use the transformed features for classification. Hinrichs et al. [8] use Multiple Kernel Learning (MKL) to fuse multi-modality data by learning an optimal linearly combined kernels for classification.

Most of the multi-modality data based AD studies in the literature are based on the 2-step strategy, where feature selection or fusion is first performed, and then a classifier (e.g., Support Vector Machine (SVM)) is learned [10,19]. Because the features selected in the first step may not be best to the classifier in the second step, which will degrade the final classification performance. Further, most methods [10,19] also focus on learning a single classifier for AD diagnosis, which is difficult to address the heterogeneity of complex brain disorder. To deal with this disease heterogeneity issue, it is more reasonable to train a set of diversified classifiers and ensemble them (instead of training a single classifier), which has been shown effective in previous studies [2,5].

To this end, we propose a novel multi-modal neuroimaging data fusion via latent space learning with ensemble classifier for AD diagnosis framework, which can seamlessly perform latent space learning and ensemble of diversified classifiers learning in a unified framework. Specifically, we first project neuroimaging data from different modalities (i.e., MRI and PET) into a common latent space, to exploit the correlation between MRI and PET features, and learn the latent representations. Concurrently, we also select a subset of discriminative ROI-based features from both modalities jointly, by imposing a cross-modality joint sparsity constraint on the concatenated projection matrices (as shown in Fig. 1). This is based on the assumption that, both the structure and function of an ROI could be affected by the disease progression, hence it is intuitive to select the same ROI based features for MRI and PET data in the latent space. Further, we learn multiple diversified classifiers by mapping the latent representations into the label space, and use an ensemble strategy to obtain the final result. Note that we integrate all the above learning tasks into a unified framework, so that all components can work together to achieve a better AD diagnostic model. We have verified the effectiveness of our method on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset.

2 Methodology

Latent Space Learning with Cross-Modality Joint Sparsity. Given a multi-modality data set $\mathbf{X} = \{\mathbf{X}_1, \cdots, \mathbf{X}_M\}$, where $\mathbf{X}_m \in \mathbb{R}^{d_m \times n}$ denotes the feature matrix for the m-th modality with d_m features and n subjects, and M is

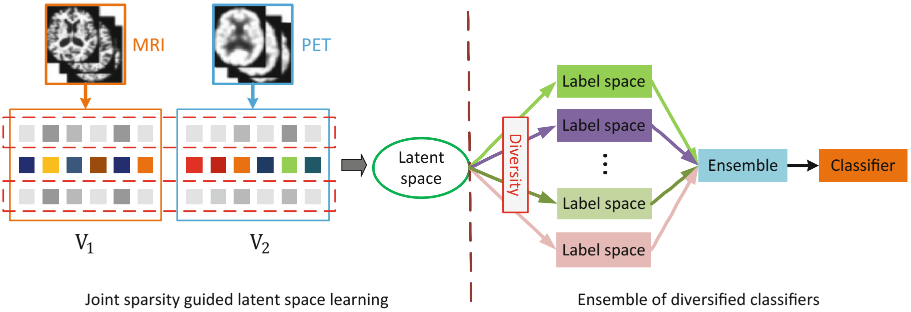


Fig. 1. A flow diagram of our proposed AD diagnosis framework. We project multi-modality data (i.e., MRI and PET in our case) into a common latent space to exploit the correlation among multi-modal neuroimaging data. Besides, a joint sparsity constraint (denoted by the dashed red rectangles) is imposed on different modalities, to encourage the selection of same ROIs from MRI and PET data. Furthermore, multiple classifiers with diversity constraint are trained and an ensemble strategy is used to obtain the final classification result.

the number of modalities. To exploit the correlations among different modalities, we project different modalities into a common latent space as follows:

$$\min_{\mathbf{V}_m,\mathbf{H}} \sum\nolimits_{m=1}^{M} \left(\|\mathbf{V}_m^T\mathbf{X}_m - \mathbf{H}\|_F^2 + \gamma\|\mathbf{V}_m\|_{2,1}\right), \tag{1}$$

where $\mathbf{V}_m \in \mathbb{R}^{d_m \times h}$ is a projection matrix, $\mathbf{H} \in \mathbb{R}^{h \times n}$ is a matrix of latent representation, γ is the regularization parameter, and h is the dimension of the latent space. We use $\ell_{2,1}$-norm regularizer (i.e., $\|\mathbf{V}\|_{2,1} = \sum\nolimits_{i=1}^{d} \sqrt{\sum\nolimits_{j=1}^{h} \mathbf{v}_{ij}^2}$, where $\mathbf{V} \in \mathbb{R}^{d \times h}$) in Eq. (1) to enforce row-wise sparsity in $\mathbf{V}_m$, by penalizing the coefficients in each row of $\mathbf{V}_m$ together. In other words, the $\ell_{2,1}$-norm encourages the selection of useful (ROI-based) features from $\mathbf{X}_m$ during the latent space learning. To encourage cross-modality joint sparsity, assuming the features from different modalities are related, the objective function in Eq. (1) is extended to the following formulation:

$$\min_{\mathbf{V}_m,\mathbf{H}} \sum\nolimits_{m=1}^{M} \|\mathbf{V}_m^T\mathbf{X}_m - \mathbf{H}\|_F^2 + \gamma\|[\mathbf{V}_1 \cdots \mathbf{V}_M]\|_{2,1}, \tag{2}$$

where a joint sparsity constraint is imposed on the concatenated projection matrices. In our case, ROI-based features are used for both the MRI and PET data, thus Eq. (2) will enforce the features from the same ROI to be selected for multi-modalities. This is based on the assumption that both the brain structure (quantified by MRI features) and function (quantified by the PET features) will be degraded for the same AD-affected ROIs. In this way, the correlations among multi-modality data can effectively be exploited.

It is worth noting that the Frobenius norm in Eq. (2) is sensitive to sample outliers. To address this issue, we reformulate Eq. (2) as:

$$\begin{aligned} \min_{\mathbf{V}_m,\mathbf{E}_m,\mathbf{H}} \sum\nolimits_{m=1}^{M} \|\mathbf{E}_m\|_1 + \gamma \|[\mathbf{V}_1 \cdots \mathbf{V}_M]\|_{2,1}, \\ s.t. \ \ \mathbf{H} = \mathbf{V}_m^T \mathbf{X}_m + \mathbf{E}_m, m = 1, \ldots, M, \end{aligned} \tag{3}$$

where an error term $\mathbf{E}_m \in \mathbb{R}^{h \times n}$ is introduce to model the reconstruction error (i.e., the first term in Eq. (2)), and use ℓ_1-norm to penalize $\mathbf{E}_m$.

Ensemble of Diversified Classifiers Learning. After obtaining the latent representations from multi-modality data, we regard the new representations in the latent space as input to train a classifier. As SVM is a widely used classifier due to its promising performance in many applications [13], we incorporate the latent space learning and classifier learning into a unified framework as follows:

$$\begin{aligned} \min_{\mathbf{V}_m,\mathbf{E}_m,\mathbf{w},\mathbf{H},b} \sum\nolimits_{i=1}^{n} f(y_i, \mathbf{h}_i^T \mathbf{w} + b) + \lambda \Psi(\mathbf{w}) \\ + \beta \sum\nolimits_{m=1}^{M} \|\mathbf{E}_m\|_1 + \gamma \|[\mathbf{V}_1 \cdots \mathbf{V}_M]\|_{2,1} \\ s.t. \ \ \mathbf{H} = \mathbf{V}_m^T \mathbf{X}_m + \mathbf{E}_m, m = 1, \ldots, M, \end{aligned} \tag{4}$$

where $\mathbf{h}_i \in \mathbb{R}^{h \times 1}$ is the latent representation of the i-th sample (i.e., i-th column of $\mathbf{H}$), $y_i \in \{-1, 1\}$ is the corresponding label, and $\mathbf{w}$ and b denote the weight vector and bias of the classifier, respectively. Besides, $f(\cdot)$ in Eq. (4) is the classifier loss function, while the second term in Eq. (4) is the regularizer for $\mathbf{w}$ (e.g., ℓ_1 or ℓ_2-norm of $\mathbf{w}$). If we use hinge loss function for $f(\cdot)$, the first term in Eq. (4) can be given as:

$$\sum\nolimits_{i=1}^{n} f(y_i, \mathbf{h}_i^T \mathbf{w} + b) = \sum\nolimits_{i=1}^{n} (1 - (\mathbf{h}_i^T \mathbf{w} + b) y_i)_+^p, \tag{5}$$

where the operation $(x)_+ := max(x, 0)$ keeps x unchanged if it is non-negative, and returns zero otherwise, and p is a constant with either value 1 or 2 to have physical meaning. In Eq. (4), only a classifier is trained, which may not be able to address the heterogeneity of AD progression. In addition, some studies have also indicated that ensembling multiple classifiers may result in a more robust and accurate classifier. Thus, following the work in [6], we replace the loss function in Eq. (4) with the loss functions from multiple classifiers, as follows:

$$\begin{aligned} \min_{\mathbf{V}_m,\mathbf{E}_m,\mathbf{W},\mathbf{H},\mathbf{b}} \sum\nolimits_{c=1}^{C} \sum\nolimits_{i=1}^{n} (1 - (\mathbf{h}_i^T \mathbf{w}_c + b_c) y_i)_+^p + \lambda \Psi(\mathbf{W}) \\ + \beta \sum\nolimits_{m=1}^{M} \|\mathbf{E}_m\|_1 + \gamma \|[\mathbf{V}_1 \cdots \mathbf{V}_M]\|_{2,1}, \\ s.t. \ \ \mathbf{H} = \mathbf{V}_m^T \mathbf{X}_m + \mathbf{E}_m, m = 1, \ldots, M, \end{aligned} \tag{6}$$

where $\mathbf{W} = [\mathbf{w}_1 \cdots \mathbf{w}_C] \in \mathbb{R}^{h \times C}$ is a matrix with each of its column denoting the weight vector for one classifier, $\mathbf{b} = [b_1 \cdots b_C] \in \mathbb{R}^{C \times 1}$ is the corresponding bias vector, and C is the number of classifiers. To ensure that we have a diversity of classifiers rather than redundant classifiers, we minimize the exclusivity function between each pair of classifier weight vectors, given as $\{\min \|\mathbf{w}_i \circ \mathbf{w}_j\|_0, i \neq j\}$ [6], where $\circ$ denotes Hadamard product, and $\|\cdot\|_0$ denotes ℓ_0-norm. This constraint will ensure the column weight vectors in $\mathbf{W}$ be exclusive and orthogonal to each other, thus giving us diversified classifiers.

However, as this constraint is too strong and difficult to optimize, we choose to minimize the relaxed exclusivity function instead, i.e., given by $\{\min \|\mathbf{w}_i \circ \mathbf{w}_j\|_1 = \min \sum_k |\mathbf{w}_i(k)| \cdot |\mathbf{w}_j(k)|, i \neq j\}$, where $\mathbf{w}_i(k)$ denotes the k-th element in $\mathbf{w}_i$, and $|\cdot|$ denotes the absolute operator. Following the work in [6], we use the following regularizer as a diversity constraint to encourage the learning of diversified classifiers. The regularizer is given as:

$$\begin{aligned}\Psi(\mathbf{W}) =& \frac{1}{2}\|\mathbf{W}\|_F^2 + \sum\nolimits_{i,j\neq i} \|\mathbf{w}_i \circ \mathbf{w}_j\|_1 \\ =& \frac{1}{2}\sum\nolimits_{k=1}^{K}(\sum\nolimits_{c=1}^{C} |\mathbf{w}_c(k)|)^2 = \frac{1}{2}\|\mathbf{W}^T\|_{1,2}^2.\end{aligned} \tag{7}$$

The derivation details for the above equation can be found in [6].

Unified AD Diagnosis Framework. By integrating the latent space learning and ensemble learning of diversified classifiers into a unified framework, the final objective function of our proposed model is given as:

$$\begin{aligned}\min_{\mathbf{V}_m,\mathbf{E}_m,\mathbf{W},\mathbf{H},\mathbf{b}} & \sum\nolimits_{c=1}^{C}\sum\nolimits_{i=1}^{n}(1-(\mathbf{h}_i^T\mathbf{w}_c+b_c)y_i)_+^p + \frac{\lambda}{2}\|\mathbf{W}^T\|_{1,2}^2 \\ & + \beta\sum\nolimits_{m=1}^{M}\|\mathbf{E}_m\|_1 + \gamma\|[\mathbf{V}_1\cdots\mathbf{V}_M]\|_{2,1}, \\ & s.t.\ \ \mathbf{H} = \mathbf{V}_m^T\mathbf{X}_m + \mathbf{E}_m, m = 1,\ldots,M.\end{aligned} \tag{8}$$

2.1 Optimization and Prediction

The objective function in Eq. (8) is not jointly convex with respect to all variables. Therefore, we utilize the Augmented Lagrange Multiplier (ALM) [11] algorithm to solve Eq. (8) efficiently and effectively. After we train our model and obtain $\mathbf{W}$ and $\mathbf{b}$, we can obtain the ensemble classifier weight and bias via $\mathbf{w} = \frac{1}{C}\sum_{c=1}^{C}\mathbf{w}_c$, and $b = \frac{1}{C}\sum_{c=1}^{C} b_c$. Then, for a testing sample $\mathbf{X}^{test} = \{\mathbf{X}_1^{test},\ldots,\mathbf{X}_M^{test}\}$, the corresponding testing label $\mathbf{y}_{test}$ is estimated by using $\mathbf{y}_{test} = \text{sign}(\mathbf{h}_{test}^T\mathbf{w} + b)$, where $\mathbf{h}_{test} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{V}_m^T\mathbf{X}_m^{test}$, denoting the latent representation of the testing sample.

3 Experiments

3.1 Subjects and Neuroimage Preprocessing

In this study, we select 379 subjects from the ADNI cohort (www.adni-info.org) with complete MRI and PET data at baseline scan, including 101 Normal Control (NC), 185 MCI, and 93 AD. In our experiments, we used ROI-based features from both MRI and PET images (i.e., $M{=}2$ in our study). Then, we further processed the MR images using a standard pipeline including the following steps: (1) intensity inhomogeneity correction, (2) brain extraction, (3) cerebellum removal, (4) tissues segmentation, and (5) template registration. After that, the processed MR images were divided into 93 pre-defined ROIs [9], and the gray matter volumes in these ROIs were computed as MRI features. For PET data, we aligned the PET images to their corresponding MR images by using affine registration, and calculated the average intensity value of each ROI as PET features. Thus, we have 93 ROI-based features from both the MRI and PET data, respectively.

3.2 Experimental Setup

We evaluated the effectiveness of the proposed model by conducting the following two binary classification experiments: i.e., AD vs. NC and MCI vs. NC classifications. We used classification accuracy (ACC) and Area Under Curve (AUC) as evaluation metrics.

We compared our proposed framework with the following comparison methods: (1) baseline method ("ORI"), which concatenates MRI and PET ROI-based features into a long vector for SVM classification, (2) Lasso based feature selection method [16], which selects features from both modalities using ℓ_1-norm, (3) CCA [7] and MKL [8] based multi-modality fusion methods; and (4) Deep learning based feature representation method, i.e., Stacked Auto-encoder (SAE) [12].

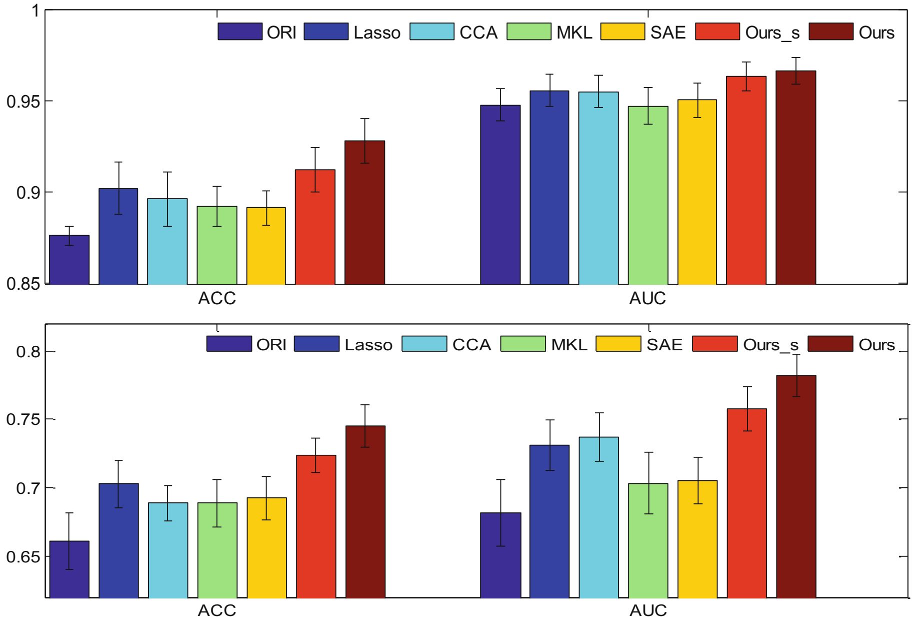
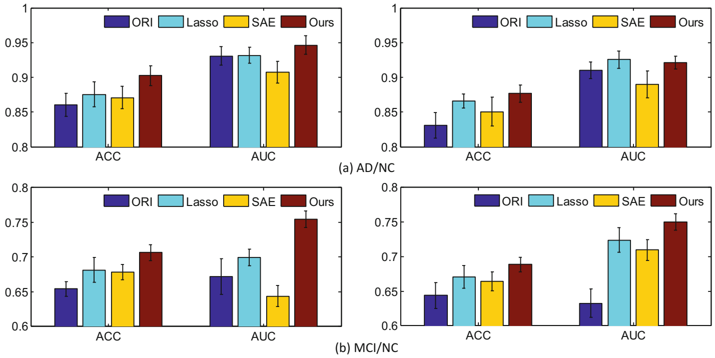


Fig. 2. Comparison of classification results using two evaluation metrics (i.e., ACC and AUC) for two classification tasks: AD/NC (top) and MCI/NC (bottom).

Note that all the above comparison methods are based on 2-step strategy, where feature selection and feature fusion (or feature learning) are first performed, before using SVM (from LIBSVM toolbox [3]) for classification. We performed 10-fold cross validation for all the methods under comparison, and reported the means and standard deviations of the experimental results. For parameter setting of our method, we determined the regularization parameter values (i.e., $\{\lambda, \beta, \gamma\} \in \{10^{-5}, \ldots, 10^{2}\}$) and the dimension of the latent space (i.e., $h \in \{10, \ldots, 60\}$) via an inner cross-validation search on the training data, and searched the number of classifiers C in the range $\{5, 10, 15, 20\}$. We also used inner cross-validation to select hyper-parameter values for all the comparison methods. Besides, we further determined the soft margin parameter of SVM classifier via grid search in the range of $\{10^{-4}, \ldots, 10^{4}\}$.

Figure 2 shows the classification performance of all the competing methods. From Fig. 2, it can be clearly seen that our proposed method performs consistently better than all the comparison methods in terms of ACC and AUC. Compared with the Lasso based feature selection method, which fuses multi-modality data without effective consideration of the correlation between MRI and PET, our method performs significantly better. In addition, our method also outperforms SAE, which uses high-level features learned from auto encoder for classification. This is probably due to the fact the SAE is an unsupervised feature learning method that does not consider label information. In addition, to verify the effectiveness of ensemble of diversified classifiers, we also compare the performance of our proposed method for the cases of using single classifier and multi-classifiers, with "Ours_s" denoting the results of our proposed method using a single classifier. From the results shown in Fig. 2, our proposed method using the ensemble of diversified classifiers performs better than the case of using only a single classifier.

To analyze the benefit of multi-modalities fusion, Fig. 3 illustrates the performance comparison of different methods using independent modality (i.e., MRI or PET). Note that, multi-modality fusion methods (i.e., CCA and MKL) are



(a) AD/NC

(b) MCI/NC

Fig. 3. Comparison of results for two classification tasks (i.e., (a) AD/NC and (b) MCI/NC) using two different modalities: MRI (left) and PET (right).

excluded in this comparison. From Fig. 3, it can be seen that our method still outperforms other comparison methods. Besides, comparing Figs. 2 and 3, we can see that all the methods perform better when using multi-modality data, compared to the use of just the single modality data.

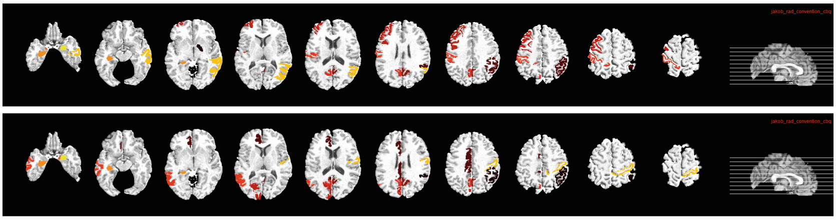


Fig. 4. Top selected regions for two classification tasks: AD/NC (top) and MCI/NC (bottom).

Furthermore, we also identified the potential brain regions that can be used as AD biomarkers. We ranked the ROIs based on their average weight values. Figure 4 shows the top ranked ROIs by using our proposed method for different tasks. Specifically, for AD/NC task, the top selected ROIs (common to both MRI and PET data) are globus palladus right, precuneus right, precuneus left, entorhinal cortex left, hippocampal formation left, middle temporal gyrus right, and amygdala right. For MCI/NC task, the top selected ROIs are angular gyrus right, precuneus right, precuneus left, middle temporal gyrus left, hippocampal formation left, postcentral gyrus right, and amygdala right. These regions are consistent with some previous studies [10,19] and can be used as potential biomarkers for AD diagnosis.

4 Conclusion

In this paper, we have proposed an AD diagnosis model based on latent space learning with diversified classifiers. This is different from the conventional AD diagnosis models that often perform feature selection (or fusion) and classifier training separately. Specifically, we project the original ROIs-based features into a latent space to effectively exploit the correlations among multi-modality data. Besides, we impose a cross-modality joint sparsity constraint to encourage the selection of same ROIs for MRI and PET data, based on the assumption that the degenerated brain regions would affect both brain structure and function. Then, using the learned latent representations as input, we learn multiple diversified classifiers and further use an ensemble strategy to obtain the final result, so that the ensemble classifier is more robust to disease heterogeneity. Experimental results on ADNI dataset have demonstrated the effectiveness of the proposed method against other methods.

Acknowledgment. This research was supported in part by NIH grants EB006733, EB008374, MH100217, AG041721 and AG042599.

References

1. Alzheimer's Association: 2013 Alzheimer's disease facts and figures. Alzheimer's Dement. **9**(2), 208–245 (2013)
2. Brown, G., Wyatt, J., Harris, R., Yao, X.: Diversity creation methods: a survey and categorisation. Inf. Fusion **6**(1), 5–20 (2005)
3. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Tech. (TIST) **2**(3), 27 (2011)
4. Chaves, R., Ramírez, J.: SVM-based computer-aided diagnosis of the Alzheimer's disease using t-test nmse feature selection with feature correlation weighting. Neurosci. Lett. **461**(3), 293–297 (2009)
5. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. **55**(1), 119–139 (1997)
6. Guo, X., Wang, X., Ling, H.: Exclusivity regularized machine: a new ensemble SVM classifier. In: IJCAI, pp. 1739–1745 (2017)
7. Hardoon, D.R., Szedmak, S., Shawe-Taylor, J.: Canonical correlation analysis: an overview with application to learning methods. Neural Comput. **16**(12), 2639–2664 (2004)
8. Hinrichs, C., Singh, V., Xu, G., Johnson, S.: MKL for robust multi-modality AD classification. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 786–794. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04271-3_95
9. Kabani, N.J.: 3D anatomical atlas of the human brain. NeuroImage **7**, P-0717 (1998)
10. Lei, B., Yang, P.: Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis. IEEE Trans. Cybern. **47**(4), 1102–1113 (2017)
11. Lin, Z., Liu, R., Su, Z.: Linearized alternating direction method with adaptive penalty for low-rank representation. In: NIPS, pp. 612–620 (2011)
12. Suk, H.: Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. Brain Struct. Funct. **220**(2), 841–859 (2015)
13. Suykens, J.A., Vandewalle, J.: Least squares support vector machine classifiers. Neural Process. Lett. **9**(3), 293–300 (1999)
14. Thung, K.H.: Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion. Med. Image Anal. **45**, 68–82 (2018)
15. Thung, K.H., Wee, C.Y., Yap, P.T., Shen, D., Initiative, A.D.N., et al.: Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion. NeuroImage **91**, 386–400 (2014)
16. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B, **58**, 267–288 (1996)
17. Zhou, T., Thung, K.H., Liu, M., Shen, D.: Brain-wide genome-wide association study for Alzheimer's disease via joint projection learning and sparse regression model. IEEE Trans. Biomed. Eng. (2018, in press)
18. Zhou, T., Thung, K.-H., Zhu, X., Shen, D.: Feature learning and fusion of multimodality neuroimaging and genetic data for multi-status dementia diagnosis. In: Wang, Q., Shi, Y., Suk, H.-I., Suzuki, K. (eds.) MLMI 2017. LNCS, vol. 10541, pp. 132–140. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67389-9_16
19. Zhu, X.: Canonical feature selection for joint regression and multi-class identification in Alzheimer's disease diagnosis. Brain Imaging Behav. **10**(3), 818–828 (2016)