Nonlinear Feature Space Transformation to Improve the Prediction of MCI to AD Conversion

Pin Zhang[1], Bibo Shi[2], Charles D. Smith[3], and Jundong Liu[1(✉)]

[1] School of Electrical Engineering and Computer Science,
Ohio University, Athens, OH, USA
liuj1@ohio.edu
[2] Department of Radiology, Duke University, Durham, NC, USA
[3] Department of Neurology, University of Kentucky, Lexington, KY, USA

Abstract. Accurate identification of patients with Mild Cognitive Impairment (MCI) at high risk for conversion to Alzheimer's Disease (AD) offers an opportunity to target the disease process early. In this paper, we present a novel nonlinear feature transformation scheme to improve the prediction of MCI-AD conversion through semi-supervised learning. Utilizing Laplacian SVM (LapSVM) as a host classifier, the proposed method learns a smooth spatially varying transformation that makes the input data more linearly separable. Our approach has a broad applicability to boost the classification performance of many other semi-supervised learning solutions. Using baseline MR images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, we evaluate the effectiveness of the proposed semi-supervised framework and demonstrate the improvements over the state-of-the-art solutions within the same category.

1 Introduction

Alzheimer's Disease (AD), the most common form of dementia, affects more than 34 million people in 2016. Amnestic mild cognitive impairment (MCI) is often regarded as a prodromal stage of AD, where some patients convert to AD over time, and the others remain stable for many years. Identifying the differences between "converters" and "stable" groups of subjects can offer an opportunity to target the disease early.

A number of solutions have been proposed in recent years to tackle AD/MCI early diagnosis problem. Common practices include the utilization of multi-modality [1,8,15] and longitudinal [7,21] data to exploit complementary information, dimension reduction to diminish data redundancy and feature selection to extract the most discriminative feature set. Ensembles of classifiers from multiple domains or/and levels have also been employed to improve the overall classification performance [4,7,8,21].

As accurate diagnosis of MCI-AD conversion is often not available until a later time, semi-supervised learning (SSL), utilizing unlabeled data in conjunction with labeled samples (the valuable gold standard confirmed cases) to

improve classification performance, is uniquely suitable to predict patients' clinical trajectories. In [19], MCI subjects were used as unlabeled data to boost the classification accuracy in discriminating AD vs. normal control (NC) subjects. Compared with using AD/NC subjects only, a significant improvement was achieved. Similar approaches were proposed in [5,17] to predict disease labels for MCI subjects. Moradi *et al.* [9] developed a semi-supervised classifier for AD conversion prediction in MCI patients based on low-density separation (LDS). All these studies demonstrate that label augmentation through unlabeled data samples equips SSL with better predictive power over supervised learning.

Despite all the strides made in recent years, insufficient attention has been given to rationally selecting appropriate metrics from the training data that could maximize the power of various SSL solutions. Learning a metric from the training input is equivalent to learning a feature transformation [11,13,16], and such transformations can often significantly boost the performance of many metric-based algorithms, such as kNN, k-means, and even SVMs in various tasks [3,12,20].

In this paper, we propose to enhance the prediction of MCI-AD conversion via a novel nonlinear feature transformation scheme. We take Laplacian SVM (LapSVM), a classic graph-based SSL model, as the host classifier, and generalize it through the application of deformable geometric models to transform the feature space. The *Coherent Point Drifting* (CPD) method from [10] is chosen as the geometric model in this study, due to its remarkable versatility and representation power in accounting for high-order deformations. In this work, we focus on the classification of progressive MCI (pMCI) vs. stable MCI (sMCI) subjects. The connection to MCI-AD conversion is that: if our solution can differentiate pMCI/sMCI subjects at their baseline time, very accurately and robustly, it should be able to make reliable MCI-to-AD predictions for unseen subjects. The data we used are baseline MR images, obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.usc.edu/ADNI).

2   Method

Formulated under the standard support vector machine (SVM) framework, LapSVM solves the classification problem by employing two regularization terms: one for SVM maximal margin classification, and the other for label smoothness across the data graph – neighboring nodes should have identical or similar labels.

Let $\mathcal{X} = \{\mathbf{x}_i | \ \mathbf{x}_i \in \mathbb{R}^d, i = 1, \cdots, l + u\}$ be the training set, where $\{\mathbf{x}_i, y_i\}_{i=1}^{l}$ are labeled samples with labels $y_i \in \{-1, +1\}$, and the remaining $\{\mathbf{x}_i\}_{i=l+1}^{l+u}$ have no labels. A graph needs to be established to specify the adjacency relationships among samples, and then LapSVM estimates a membership function $f(\mathbf{x})$ on the graph, by solving the following optimization problem:

$$\min_{f \in \mathcal{H}_\mathcal{K}} \quad J = \frac{1}{l} \sum_{i=1}^{l} \xi_i + C_1 \|f\|_K^2 + C_2 \sum_{i,j=1}^{l+u} W_{ij}(f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 \tag{1}$$

$$\text{s.t.} \quad y_i f(\mathbf{x}_i) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i = 1 \ldots l;$$

where W_{ij} is the weight of the edge that connects $\mathbf{x}_i$ and $\mathbf{x}_j$ in the data adjacency graph. ξ_i are slack variables to penalize misclassifications. $\|f\|_K^2$ is the squared norm of f in the Reproducing Kernel Hilbert Space (RKHS). C_1 and C_2 are hyperparameters controlling the contributions of the two regularization terms.

2.1 Feature Transformation Through Coherent Point Drifting (CPD)

For distance or dissimilarity based classification algorithms, the application of a smooth nonlinear transformation across the space is equivalent to assigning spatially varying metrics at different locations. Our goal is to learn such a transformation so that the displaced samples would better conform to the data distribution assumed by the ensuing classifier, LapSVM.

In this work, the CPD model is chosen as the geometric model to drive the deformations. CPD was originally designed for landmark matching. For two sets $\mathcal{X}$ and $\mathcal{U}$, each with n points of dimension d, CPD seeks a continuous velocity function $v(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^d$ that moves the source dataset $\mathcal{X}$ towards the target $\mathcal{U}$. Estimation of an optimal $v(\cdot)$ is formulated under Tikhonov regularization framework: $\mathcal{R}[v] = \frac{1}{2} \sum_{i=1}^n [\mathbf{u}_i - (\mathbf{x}_i + v(\mathbf{x}_i))]^2 + \frac{1}{2}\lambda\|\mathbf{D}v\|^2$, where $\mathbf{D}$ is a linear differentiation operator, $\|\cdot\|$ is the norm operation, and λ controls the strength of the regularization term. CPD chooses a particular regularization term whose kernel function is a Gaussian low-pass filter. According to [10], the optimal solution $v(\mathbf{x})$ in CPD can be written in the matrix format as:

$$v(\mathbf{x}_i) = \boldsymbol{\Psi} \begin{pmatrix} G(\mathbf{x}_i, \mathbf{x}_1) \\ \cdots \\ G(\mathbf{x}_i, \mathbf{x}_n) \end{pmatrix} = \boldsymbol{\Psi} G(\mathbf{x}_i, \mathcal{X}), \qquad (2)$$

where $\boldsymbol{\Psi}$ (size $d \times n$) is the weight matrix for the Gaussian kernel functions. $G(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2\sigma^2}}$, where σ is the width of the Gaussian filter.

2.2 CPD Based LapSVM (CPD-LapSVM)

Allowing samples to be moved, our proposed CPD-LapSVM is designed to learn a spatial transformation and a classifier at the same time. Let $\mathbf{x}_i^0$ be the original position of a sample $\mathbf{x}_i$. Through the motion regulated by CPD, $\mathbf{x}_i$ will be moved to a new location $\mathbf{x}_i^1$:

$$\mathbf{x}_i^1 = \mathbf{x}_i^0 + v(\mathbf{x}_i^0) = \mathbf{x}_i^0 + \boldsymbol{\Psi} G(\mathbf{x}_i^0, \mathcal{X}^0) \qquad (3)$$

The CPD transformation can be applied in both the original input space and the feature space after kernelization. The classifier to be learned is a LapSVM defined on the CPD-transformed samples: $f(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i^1 + b$.

Under the input space, our linear version CPD-LapSVM (note: "linear" refer to decision boundary; the transformation is nonlinear) is built on the LapSVM objective function in Eq. (1). Two modifications are made. First, quadratically

smoothed hinge loss functions, $\xi_i = \max[0, 1 - y_i f(\mathbf{x}_i^1)]^2$, are included as slack variables to convert the constrained optimization in Eq. (1) to an unconstrained optimization problem. The choice of quadratic form is to make the computation of derivatives more convenient. Second, to reduce the chance of overfitting, we add the squared Frobenius norm $\|\mathbf{\Psi}\|_F^2$, to penalize non-smoothness in the estimated transformations. With these two added terms, our linear CPD-LapSVM minimizes the following updated objective function to find the optimal transformation and classifier, specified by $\mathbf{\Psi}$ and $\{\mathbf{w}, b\}$, respectively:

$$
\min_{\mathbf{\Psi}, \mathbf{w}, b} J = \frac{1}{l} \sum_{i=1}^{l} \max[0, 1 - y_i(\mathbf{w}^T \mathbf{x}_i^1 + b)]^2 + C_1 \|\mathbf{w}\|_K^2 + C_2 \|\mathbf{\Psi}\|_F^2
$$
$$
+ C_3 \sum_{j,k=1}^{l+u} W_{jk}(\mathbf{w}^T \mathbf{x}_j^1 - \mathbf{w}^T \mathbf{x}_k^1)^2
$$
(4)

where C_1, C_2 and C_3 are trade-off hyperparameters. In this paper, we choose full graphs as the neighborhood adjacency graphs, where every sample pair is assumed to be connected. The edge weight of each connection is calculated as $W_{jk} = \exp(-\frac{1}{2\alpha^2}(\|\mathbf{x}_j^1 - \mathbf{x}_k^1\|^2))$, where α is the width of heat kernel function.

   To solve Eq. (4), an EM-like iterative strategy is applied to update $\mathbf{\Psi}$ and $\{\mathbf{w}, b\}$ alternatingly. First, when $\mathbf{\Psi}$ is fixed, Eq. (4) is reduced to the original LapSVM, performing on deformed training samples $\mathcal{X}^1$. A standard LapSVM solver, as in [2], can be used to search for the optimal classifier. Second, with $\{\mathbf{w}, b\}$ fixed, the classification decision boundary becomes explicit. The updated objective function now only depends on parameter $\mathbf{\Psi}$ (note: $\mathbf{x}_i^1$ is also parameterized with $\mathbf{\Psi}$):

$$
\min_{\mathbf{\Psi}} \quad J = \frac{1}{l} \sum_{i=1}^{l} \max[0, 1 - \mathbf{y}_i(\mathbf{w}^T \mathbf{x}_i^1 + b)]^2
$$
$$
+ C_2 \|\mathbf{\Psi}\|_F^2 + C_3 \sum_{j,k=1}^{l+u} W_{jk}(\mathbf{w}^T \mathbf{x}_j^1 - \mathbf{w}^T \mathbf{x}_k^1)^2
$$
(5)

   The objective Eq. (5) is differentiable w.r.t. $\mathbf{\Psi}$. In this work, we used the "fmincon" function in Matlab, a gradient-based nonlinear programming solver, to search for the optimal $\mathbf{\Psi}$.

   Our SSL framework is a general paradigm. While the above derivations are based on a particular classifier, LapSVM, integrating CPD with many other SSL solutions is often straightforward. We can commonly use CPD to parameterize data samples at new locations, and apply the same two-stage EM procedure to estimate an optimal pair of transformation and classifier jointly. For example, CPD can be integrated with another SSL solution, Transductive SVM (CPD-TSVM), in a similar fashion as CPD-LapSVM. In addition, it should be noted that our model is different from kernel learning methods, where the kernel bases need to be pre-defined and only their weights are learned from the training

samples. Our CPD model realizes a fully deformable nonlinear feature transformation, and it is directly estimated from the training samples.

Kernelization of CPD-LapSVM. The proposed CPD-LapSVM we introduced so far is developed and therefore applicable under the input space. It can be further kernelized to deal with more complicated data. In this paper, we adopt a kernel principal component analysis (KPCA) based framework in [18]. By first projecting the entire input samples into a kernel feature space introduced by KPCA, we can train CPD-LapSVM under the kernel space to learn both $\boldsymbol{\Psi}$ and $\{\mathbf{w}, b\}$, in the same way as it is carried out under the original input space. No derivation of any new mathematical formula is needed. The detailed procedures can be found in [18].

3   Experimental Results

In this section, we evaluate the proposed CPD-LapSVM through two binary classification problems: AD vs. NC with MCI subjects as unlabeled samples, and progressive MCI (pMCI) vs. stable MCI (sMCI) with unknown MCI (uMCI) as unlabeled samples (the definition of "unknown MCI" will be given later). The data used in the experiment were obtained from ADNI. We focus on the features extracted from baseline T1 weighted MRIs. Overall, 185 patients with AD, 242 with MCI and 227 with NC (654 subjects in total) were used in our experiments.

The features utilized in this study are 113 cortical and sub-cortical regional volumes under "Cross-Sectional Processing aseg" files, available under ADNI. The anatomical structures include left/right Hippocampi, left/right Caudates, etc. The whole list of the structure names can be found in one of our previous studies [12]. All features have been normalized by the corresponding whole brain volumes obtained from "Intracranial Volume Brain Mask" files.

The performance of various classification solutions is compared based on four measures: classification accuracy (ACC), sensitivity (SEN), specificity (SPE), and area under the receiver operating characteristic curve (AUC). Three semi-supervised methods: Laplacian regularized least squares (LapRLS), Laplacian SVM (LapSVM) and Optimized LapSVM (OLapSVM) [6] are utilized in all experiments as the competing solutions. For each solution, both linear and RBF Gaussian kernel versions are evaluated. In the end, we also compare our method with five state-of-the-art pMCI/sMCI classification solutions [4,5,9,14,17], which also used baseline T1-weighted MRIs from the ADNI database.

3.1   AD vs. NC with MCI as Unknown

The first set of experiments classify AD and NC subjects, with MCI subjects as unlabeled samples. The AD and NC groups, which are used as labeled subjects, were randomly divided into four folds for cross validation (2 for training, 1 for validation and 1 for testing). All MCI subjects were shared as unlabeled data across folds.

Table 1. Performance comparison of CPD-LapSVM with other methods for AD vs. NC classifications. Boldface denotes the best ACC & AUC performance.

AD vs. NC								
Methods	Linear kernel				RBF kernel			
	ACC (%)	SEN (%)	SPE (%)	AUC (%)	ACC (%)	SEN (%)	SPE (%)	AUC (%)
LapRLSC	83.65	83.11	84.02	89.56	84.91	79.05	89.72	90.34
LapSVM	84.11	83.21	84.78	89.98	84.77	79.80	88.76	90.40
OLapSVM	85.36	84.26	86.38	90.89	86.38	81.75	90.03	91.69
CPD-LapSVM	**86.33**	85.77	86.82	**91.44**	**87.47**	81.73	91.97	**92.89**

The involved hyper-parameters C_1, C_2 and C_3 are all chosen from $\{2^{-5} \sim 2^{10}\}$ over the cross validations. C_1 and C_2 are the slackness tradeoff and graph regularization parameters used in all models. C_3 is the regularization parameter, only used in our model. All the RBF kernel versions of the methods have an additional parameter to tune: the RBF Gaussian kernel width σ, which is also chosen from $\{2^{-5} \sim 2^{10}\}$ in our experiments.

Table 1 summarizes the AD vs. NC classification results for all methods, averaged from 50 random repeats. It is evident that our CPD-LapSVM achieves the highest ACC and AUC scores among the competing methods, for both the linear and kernel versions. It is also noteworthy that the linear version CPD-LapSVM obtains comparable results with the kernel versions of other competing methods. The standard deviations of the measures were also computed; however, they are not included in the table mainly due the page limit.

3.2 PMCI vs. sMCI with uMCI as Unknown

The second set of experiments are designed to predict AD conversion from MCI patients through classification of pMCI vs sMCI, with uMCI as unlabeled samples. If the initial diagnosis was MCI at baseline, but the follow-up diagnosis are missing or not stable, the patient is categorized as "unknown MCI". Overall, 110 patients with pMCI, 38 with sMCI and 227 with uMCI (242 MCI subjects in total) were used in our experiments.

The same experimental setting and hyperparameter selection approach as in the previous AD/NC classifications are adopted here. uMCI subjects were shared as unlabeled data during 4-fold cross validation. The results are reported in Table 2. Similar results as that in AD vs. NC classification can be observed. The ACC score of our linear-version CPD-LapSVM is significantly higher than all other methods. Comparing with LapSVM, which is the host solution of CPD-LapSVM, the ACC improvement on LapSVM is from 67.09% to 69.12%. For RBF Gaussian kernel versions, the highest ACC and AUC scores were both achieved by our model.

In order to investigate the effect of the number of labeled data on our method, a test was performed by decreasing the number of revealed labels. The ratio of revealed labels are decreased as: $\{100\%, 80\%, 60\%, 40\%$ and $20\%\}$. The ACC scores with different labeled sample ratios are shown in Fig. 1. The Solid lines

Table 2. Performance comparison of CPD-LapSVM with other methods for pMCI vs. sMCI classifications. Boldface denotes the best ACC & AUC performance.

pMCI vs. sMCI								
Methods	Linear kernel				RBF kernel			
	ACC (%)	SEN (%)	SPE (%)	AUC (%)	ACC (%)	SEN (%)	SPE (%)	AUC (%)
LapRLSC	67.09	62.66	80.42	79.11	76.09	92.82	29.61	76.49
LapSVM	66.37	61.86	79.61	79.03	76.14	88.70	40.17	76.46
OLapSVM	67.40	63.69	78.26	**79.60**	76.54	83.28	55.34	76.79
CPD-LapSVM	**69.12**	67.92	72.03	78.07	**78.27**	86.38	52.32	**78.58**
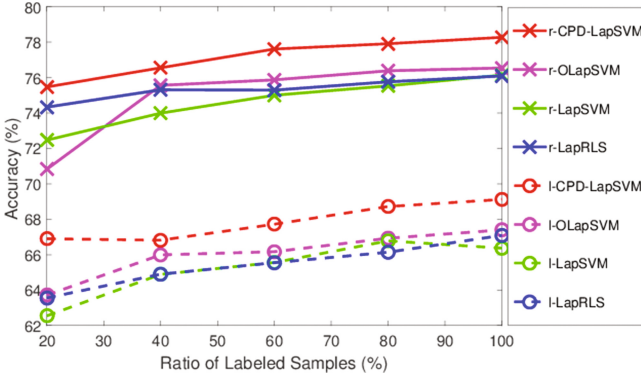
Fig. 1. ACC score w.r.t different ratio of revealed labels in pMCI vs. sMCI classifications

and prefix "*r*" denote the results from kernelized classifiers, while dashed lines and "*l*" are for linear classifiers. It is clear that the ACC values of our CPD-LapSVM with both linear and RBF Gaussian kernel are always performing the best.

Finally, we summarize several recent works in pMCI vs. sMCI classification as a comparison in Table 3. To best of our knowledge, the best result so far was achieved by [14]. Among the methods using baseline MRIs only, our work achieved the best performance in ACC score. However, it should be noted that direct comparisons of the published neuroimaging algorithms are often not feasible. When different datasets and experimental setups are utilized, higher accuracy or better results over a competing solution ought to be interpreted as more of a side evidence of the model efficacy, rather than the proof of superiority for head-to-head competitions.

As for the experiment and data setup, there could be different approaches to include the unlabeled samples. For example, MCI can be used as unknown for AD/NC [5,19], AD/NC as unknown for pMCI/sMCI [17], and uMCI as unknown for pMCI/sMCI [5,9]. We chose the last scheme, but other different settings can be certainly carried out and tested.

Table 3. Comparisons of pMCI vs. sMCI classification solutions using ADNI database. Boldface denotes the best performance for the measure of ACC & AUC.

pMCI vs. sMCI				
Methods	ACC (%)	SEN (%)	SPE (%)	AUC (%)
Ye *et al.* [17]	56.10	94.10	40.80	73.00
Filipovych *et al.* [5]	–	79.40	51.7	69.00
Moradi *et al.* [9]	74.74	88.85	51.46	76.61
Cheng *et al.* [4]	73.80	69.00	77.40	**79.60**
Suk *et al.* [14]	74.82	70.93	78.82	75.89
CPD-LapSVM	**78.27**	86.38	52.32	78.58

4   Conclusions

In this paper, we have proposed a nonlinear feature transformation based semi-supervised learning strategy to enhance the prediction of MCI-AD conversion through MR images. The proposed CPD-LapSVM model takes advantage of the space deformations regulated by CPD to push the data samples towards better linear separability, which leads to improved LapSVM classification performance. Exploring more transformation models is the directions of our future efforts. We are also interested in applying the proposed strategy to other neuroimage analysis problems.

References

1. An, L., Adeli, E., Liu, M., Zhang, J., Shen, D.: Semi-supervised hierarchical multi-modal feature and sample selection for alzheimer's disease diagnosis. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 79–87. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_10
2. Belkin, M., Niyogi, P.: Semi-supervised learning on riemannian manifolds. Mach. Learn. **56**(1–3), 209–239 (2004)
3. Bellet, A., Habrard, A., Sebban, M.: A survey on metric learning for feature vectors and structured data. arXiv preprint (2013). arXiv:1306.6709
4. Cheng, B., Liu, M., Shen, D., Li, Z., Zhang, D.: Multi-domain transfer learning for early diagnosis of alzheimer's disease. Neuroinformatics **15**(2), 115–132 (2017). https://doi.org/10.1007/s12021-016-9318-5
5. Filipovych, R., Davatzikos, C., Initiative, A.D.N., et al.: Semi-supervised pattern classification of medical images: application to mild cognitive impairment (MCI). Neuroimage **55**(3), 1109–1119 (2011)
6. Gu, Y., Feng, K.: Optimized laplacian svm with distance metric learning for hyperspectral image classification. AEORS **6**(3), 1109–1117 (2013)
7. Huang, M., Yang, W., Feng, Q., Chen, W.: Longitudinal measurement and hierarchical classification framework for the prediction of alzheimer's disease. Nature **7**, 39880 (2017)

8. Liu, M., Zhang, J., Yap, P.T., Shen, D.: View-aligned hypergraph learning for alzheimer's disease diagnosis with incomplete multi-modality data. Med. Image Anal. **36**, 123–134 (2017)

9. Moradi, E., Pepe, A., Gaser, C., Huttunen, H., Tohka, J., Initiative, A.D.N., et al.: Machine learning framework for early mri-based alzheimer's conversion prediction in MCI subjects. Neuroimage **104**, 398–412 (2015)

10. Myronenko, A., Song, X.: Point set registration: coherent point drift. TPAMI **32**(12), 2262–2275 (2010)

11. Shi, B., Chen, Y., Hobbs, K., Liu, J., Smith, C.D.: Nonlinear metric learning for alzheimer's disease with integration of longitudinal neuroimaging features. In: BMVC (2015)

12. Shi, B., Chen, Y., Zhang, P., Smith, C.D., Liu, J.: Nonlinear feature transformation and deep fusion for alzheimer's disease staging analysis. Pattern Recognit. **63**, 487–498 (2017)

13. Shi, B., Wang, Z., Liu, J.: Distance-informed metric learning for alzheimer's disease staging. In: EMBC, pp. 934–937. IEEE (2014)

14. Suk, H.I., Lee, S.W., Shen, D.: Deep ensemble learning of sparse regression models for brain disease diagnosis. Med. Image Anal. **37**, 101–113 (2017)

15. Tong, T., Gray, K., Gao, Q., Chen, L., Rueckert, D.: Multi-modal classification of alzheimer's disease using nonlinear graph fusion. Pattern Recognit. **63**, 171–181 (2017)

16. Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning with application to clustering with side-information. In: NIPS, vol. 15, p. 12 (2002)

17. Ye, D.H., Pohl, K.M., Davatzikos, C.: Semi-supervised pattern classification: application to structural MRI of alzheimer's disease. In: PRNI, pp. 1–4. IEEE (2011)

18. Zhang, C., Nie, F., Xiang, S.: A general kernelization framework for learning algorithms based on kernel PCA. Neurocomputing **73**(4), 959–967 (2010)

19. Zhang, D., Shen, D.: Semi-supervised multimodal classification of alzheimer's disease. In: ISBI, pp. 1628–1631. IEEE (2011)

20. Zhang, P., Shi, B., Smith, C.D., Liu, J.: Nonlinear metric learning for semi-supervised learning via coherent point drifting. In: ICMLA, pp. 314–319. IEEE (2016)

21. Zhu, Y., Zhu, X., Kim, M., Shen, D., Wu, G.: Early diagnosis of alzheimer's disease by joint feature selection and classification on temporally structured support vector machine. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9900, pp. 264–272. Springer, Cham (2016). doi:10.1007/978-3-319-46720-7_31