

Early Diagnosis of Alzheimer's Disease by Joint Feature Selection and Classification on Temporally Structured Support Vector Machine

Yingying Zhu, Xiaofeng Zhu, Minjeong Kim,
Dinggang Shen, and Guorong Wu^(✉)

Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, Chapel Hill, USA
grwu@med.unc.edu

Abstract. The diagnosis of Alzheimer's disease (AD) from neuroimaging data at the pre-clinical stage has been intensively investigated because of the immense social and economic cost. In the past decade, computational approaches on longitudinal image sequences have been actively investigated with special attention to Mild Cognitive Impairment (MCI), which is an intermediate stage between normal control (NC) and AD. However, current state-of-the-art diagnosis methods have limited power in clinical practice, due to the excessive requirements such as equal and immoderate number of scans in longitudinal imaging data. More critically, very few methods are specifically designed for the early alarm of AD uptake. To address these limitations, we propose a flexible spatial-temporal solution for early detection of AD by recognizing abnormal structure changes from longitudinal MR image sequence. Specifically, our method is leveraged by the non-reversible nature of AD progression. We employ temporally structured SVM to accurately alarm AD at early stage by enforcing the monotony on classification result to avoid unrealistic and inconsistent diagnosis result along time. Furthermore, in order to select best features which can well collaborate with the classifier, we present as joint feature selection and classification framework. The evaluation on more than 150 longitudinal subjects from ADNI dataset shows that our method is able to alarm the conversion of AD 12 months prior to the clinical diagnosis with at least 82.5 % accuracy. It is worth noting that our proposed method works on widely used MR images and does not have restriction on the number of scans in the longitudinal sequence, which is very attractive to real clinical practice.

1 Introduction

Alzheimer's disease is an incurable neurodegenerative disease. Typical clinical symptoms include memory loss, disorientation, language and behavioral issues. The progression of AD is not reversible, however, there are treatment available to modify disease effect in the early stage of AD [1]. Thus, early diagnosis or prognosis of AD is of high value in clinical practice since it can save more time for treatment and then improve the life quality for not only patients but also their caregivers.

AD introduces both structural and functional loss that is known to have dynamically evolving morphological patterns [1–3, 12–15]. In the past decade, longitudinal studies have been actively investigated for AD diagnosis with special attention to MCI [1, 4], which is an intermediate stage between NC and AD. For example, tensor-based morphometry is used in [4] to reveal brain atrophy patterns from 91 probable AD patients and 189 MCI subjects scanned at baseline, and after 6, 12, 18, and 24 months. Moreover, the trend of longitudinal cortical thickness is used as the morphological patterns in [5] to identify subjects which eventually convert to AD. However, current longitudinal AD diagnosis methods have very strong restriction on the longitudinal image sequence. For example, each subject recruited in [5] should at least 5 time-points in every six months, and should develop AD after at least 12 months after the baseline scan. For convenience, many longitudinal approaches assume the number of scans is equal, albeit implicitly. In real clinical setting, however, not all patients have a large or an equal number of imaging scans.

In order to accurately measure the tiny structural changes along time, current state-of-the-art computer assisted diagnosis methods have to wait until the patient has enough number of longitudinal scans. More critically, the prediction is short term, e.g., only 6 months before real onset of AD in [5]. Although promising results have been achieved in predicting whether the subject has progressed to AD or stays in MCI stage, the limitation of short-term prediction substantially hamper the deployment in clinical practice.

In light of this, we propose a flexible solution for early detection of AD by sequentially and consistently recognizing abnormal patterns of structure change from longitudinal MR image sequence. *First*, we present a novel temporally structured SVM (TS-SVM) which is trained based on a set of partial image sequences cut from the complete longitudinal data. Compared to conventional SVM, our TS-SVM has two major improvements to achieve early alarm and high accuracy in detecting AD progression: **(1) Temporal consistency.** We enforce monotonic constraint to avoid inconsistent detection results along time. Since convergent evidence suggests that AD progression is non-reversible [6, 7], we require the risk of AD progression should monotonically increase within each subject as more and more time-points are inspected. **(2) Early detection.** We employ sequential recognition to achieve best balance of early alarm and detection accuracy. In the training stage, we specifically train the classifiers by making the classification margin adaptive to the length of partial image sequence. Given the longitudinal image sequence of new subject with arbitrary number of scans, we sequentially examine the longitudinal imaging patterns from baseline and alarm the AD conversion as long as the detection of abnormal change is of high confidence. Thus, our proposed AD early detection method does not have requirement on the number of scans. *Second*, we further present a joint feature selection and classification framework, in order to make the selected best features are eventually optimal to work with the learned support vector machine. We have evaluated the performance of AD early detection on more than 150 longitudinal subjects from ADNI dataset. Our method achieved promising results by alarming AD onset 12 months prior to the clinical diagnosis with at least 82.5 % accuracy.

2 Methods

2.1 Temporally Structured SVM for Early Detection of AD

The goal of our method is to accurately predict AD converting as early as possible by longitudinally tracking the structure changes. Since magnetic resonance (MR) image is non-invasive and widely used in clinic practice, we present a novel temporally structured SVM on longitudinal MR image sequences.

Morphological Features. Suppose we have N training subjects, each subject S_n has a MR image sequences $I^n = \{I_t^n | t = 1, \dots, T_n\}$ ($n = 1, \dots, N$) with T_n longitudinal scans. For each volumetric image I_t^n , we first register the template image (<http://qnl.bu.edu/obart/explore/AAL/>) with 90 manually labeled ROIs (regions of interest) using hammer registration tool to the underlying image I_t^n and extract seven morphological features in each ROI which include tissue percentiles (volumetric percentiles of the ROI volume) of white matter (WM), gray matter (GM), cerebral-spinal fluid (CSF), and background, and the averaged voxel-wise Jacobian determinant in WM, and GM and CSF regions. Therefore, the image feature \mathbf{f}_t^n for each volumetric image I_t^n is a $90 \times 7 = 630$ dimension feature vector.

Decomposition to Partial Image Sequences. We can decompose the complete longitudinal image sequence I^n into $(T_n - 1)$ partial image sequences $P^n = \{P^n(b) | b = 2, \dots, T_n\}$, where each $P^n(b) = \{I_t^n | t = 1, \dots, b\}$ is the partial image sequence with b time points from baseline to $(b - 1)$ -th follow-up. For each $P^n(b)$, we further extract longitudinal feature representations and form a column vector $\mathbf{h}(b, n) = \left[\sum_{t=1}^b \mathbf{f}_t^n / b, (\mathbf{f}_1^n - \mathbf{f}_b^n) \right]^T$, where the first half elements are the average of morphological features from baseline to last time point and the second half elements measure the longitudinal difference of morphological features from baseline to the last follow-up. It is apparent that each feature representation $\mathbf{h}(b, n)$ describes both the spatial and temporal morphological patterns. As we will explain in Sect. 2.2, feature selection is of necessity to remove data redundancy from such high dimension ($d = 1, 260$).

Naive Way to Achieve Early Detection by Classic SVM. In our application, the goal of classification is to determine (1) whether we can detect the conversion of AD on the new testing subject based on its MR image sequence $Z = \{Z_t | t = 1, \dots, T_z\}$ up to the current time point T_z ; and (2) whether we could detect the AD onset as early as possible, i.e., push T_z as close to baseline as possible. Thus, we regard the early detection of AD as a binary classification problem between MCI non-converter (MCI-NC for short) and MCI converter (MCI-C for short). Without loss of generality, we assume the first M subjects belong to MCI-NC group and the remaining subjects belong to MCI-C group. Therefore, we divide all partial image sequences for training purpose into two groups: MCI-NC group $\mathbf{X} = \{\mathbf{x}_{b,p} | \mathbf{x}_{b,p} = \mathbf{h}(b, p), p = 1, \dots, M, b =$

$1, \dots, T_p\}$ and MCI-C group $\mathbf{Y} = \{\mathbf{y}_{b,p} | \mathbf{y}_{b,p} = \mathbf{h}(b, q) | q = M + 1, \dots, N, b = 1, \dots, T_q\}$. To achieve above goal, the naïve way is to train a SVM by:

$$\arg \min_{\mathbf{w}} \|\mathbf{W}\|_F^2 + \lambda \varepsilon^2, s.t. \begin{cases} \delta_x - (\mathbf{w}_x^T - \mathbf{w}_y^T) \mathbf{x}_{b,p} < \varepsilon, \varepsilon > 0, \forall \mathbf{x}_{b,p} \in \mathbf{X} \\ \delta_y - (\mathbf{w}_y^T - \mathbf{w}_x^T) \mathbf{y}_{b,q} < \varepsilon, \varepsilon > 0, \forall \mathbf{y}_{b,q} \in \mathbf{Y} \end{cases}, \quad (1)$$

where $\mathbf{W} = [\mathbf{w}_x \mathbf{w}_y]$ is a matrix consisting of classifier $\mathbf{w}_x \in \mathcal{R}^{d \times 1}$ for MCI-NC group and $\mathbf{w}_y \in \mathcal{R}^{d \times 1}$ for MCI-C group. The intuition behind the constraint is that the probability score $(\mathbf{w}_x^T \mathbf{x}_{b,p})$ for each MCI-NC sample $\mathbf{x}_{b,p}$ staying the MCI-NC group should be greater than the score $(\mathbf{w}_y^T \mathbf{x}_{b,p})$ for jumping to MCI-C group by an inter-class margin δ_x . Similar principle also applies to the sample $\mathbf{y}_{b,q}$ from MCI-C group. ε is the slack variable which compensates for the mis-classification errors.

It is clear that there is strong structural correlations along partial image sequences in each subject. However, the naïve SVM solution shown in Eq. (1) treats each partial sequence separately. As shown in the left of Fig. 1, the probability scores of AD conversion and staying in MCI stage are not stable along time, which is not realistic since the structural change and AD progression are normally regarded as non-reversible.

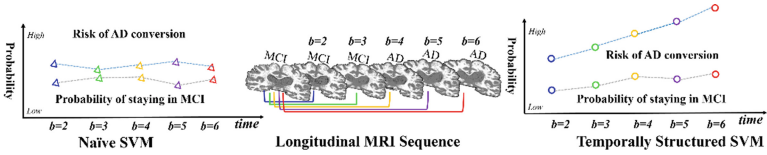


Fig. 1. Advantages of our TS-SVM (right) over the naïve SVM solution (left). In our TS-SVM method, we enforce the temporal monotony and consistency constraints on the extracted partial image sequences (shown in the middle))

Temporally Structured SVM on Longitudinal MR Image Sequences. To improve the accuracy of early AD detection, we propose the temporally structured SVM as:

$$\arg \min_{\mathbf{w}} \|\mathbf{W}\|_F^2 + \lambda \varepsilon^2, s.t. \begin{cases} \delta_x(b) - (\mathbf{w}_x^T - \mathbf{w}_y^T) \mathbf{x}_{b,p} < \varepsilon, \varepsilon > 0, \forall \mathbf{x}_{b,p} \in \mathbf{X} \\ \delta_y(b) - (\mathbf{w}_y^T - \mathbf{w}_x^T) \mathbf{x}_{b,q} < \varepsilon, \varepsilon > 0, \forall \mathbf{x}_{b,q} \in \mathbf{Y} \end{cases}, \text{ and} \quad (2)$$

$$C_2 : \tau_y(l) - \mathbf{w}_y^T(\mathbf{y}_{b,q} - \mathbf{y}_{a,q}) < \varepsilon, l = b - a, \varepsilon > 0, 2 \leq a < b, \forall \mathbf{y}_{a,q}, \mathbf{y}_{b,q} \in \mathbf{Y}.$$

Compared to the objective function of naïve SVM in Eq. (1), two new constraints (C_1 and C_2) are used. (1) we first turn the inter-class margins δ_x and δ_y in Eq. (1) from scalar values into the monotonically increasing functions of b (the length of partial

image sequence). The constraint C_1 is mainly used to achieve early detection, i.e., we require the probability of making accurate classification should increase as more time points are available. (2) The second constraint C_2 takes advantage of the non-reversible nature of AD progression. Suppose $\mathbf{y}_{a,q}$ and $\mathbf{y}_{b,q}$ are the morphological features from the same MCI-C subject but $\mathbf{y}_{b,q}$ is extracted at the later time points after $\mathbf{y}_{a,q}$ (i.e., $a < b$). Then we require the probability of the underlying MCI-C subject being converted to AD should higher at later time point b than at earlier time point a , i.e., $\mathbf{w}_y^T \mathbf{y}_{b,q} > \mathbf{w}_y^T \mathbf{y}_{a,q}$ since AD conversion is irreversible. Furthermore, the intra-class margin τ_y is a monotonically increasing function of l ($l = b - a$ is the length difference between two partial image sequences). Intuitively, the bigger the gap between two time points is, the larger the increase of AD conversion risk becomes. It is worth noting that the constraint C_2 is *not* applicable to MCI-NC subjects since the MCI-NC subject might convert to AD as more and more follow-ups will be scanned in future. Thus it is unreasonable to assume the MCI-NC subject can keep staying at MCI stage. As shown in the right of Fig. 1, for particular MCI-C subject, not only the probability score of AD conversion but also the difference between the probability scores of converting to AD and staying MCI monotonically increase as the partial image sequence becomes longer and longer. Thus, our TS-SVM can detect AD onset at early stage with high confidence. It is worth noting that we set $\delta_x(b) = b$, $\delta_y(b) = b$, $\tau_y(l) = l$ in all experiments.

2.2 Joint Feature Selection and Classification on TS-SVM

Since the morphological features are in high dimension, feature selection is a standard procedure to remove the data redundancy. Usually feature selection is independently applied prior to train the classifiers. In order to make the selected best features are also optimal for using TS-SVM, we proposed to jointly select best features and train the classifiers by introducing $L_{2,1}$ norm on the classification matrix \mathbf{W} :

$$\arg \min_{\mathbf{w}} \|\mathbf{W}\|_{2,1} + \lambda \mathbf{e}^2, s.t. C_1 \text{ and } C_2. \quad (3)$$

The intuitions behind using $\|\mathbf{W}\|_{2,1}$ are that (1) sparsity constraint on each column of \mathbf{W} : only a small number of features are selected which is useful to suppress the noisy and redundant patterns, and (2) group-wise constraint on each row of \mathbf{W} : both MCI-NC and MCI-C classifiers select/discard the same morphological features. In this way, \mathbf{W} can be simultaneously regarded as a coefficient matrix for feature selection and a classifier for classification.

2.3 Optimization

Although Eq. (3) is a convex problem, it is hard to optimize it directly due to a large number of linear inequality constraints. To solve this problem efficiently, we reformulate it as an unconstrained problem following the framework of Alternating Direction Method of Multipliers (ADMM) [8, 9, 16]. Specifically, we rewrite Eq. (3) as

an unconstrained convex optimization problem by introducing a dummy variable \mathbf{Z} to break the group sparse constraint with other inequality constraints:

$$\arg \min_{\mathbf{W}, \mathbf{Z}} \|\mathbf{Z}\|_{2,1} + \lambda \left[\sum_{\mathbf{x}_{b,p} \in \mathbf{X}} \left\| \delta_{\mathbf{x}}(b) - (\mathbf{w}_{\mathbf{x}}^T - \mathbf{w}_{\mathbf{y}}^T) \mathbf{x}_{b,p} \right\|_h + \sum_{\mathbf{y}_{b,q} \in \mathbf{Y}} \left\| \delta_{\mathbf{y}}(b) - (\mathbf{w}_{\mathbf{y}}^T - \mathbf{w}_{\mathbf{x}}^T) \mathbf{y}_{b,p} \right\|_h + \sum_{\mathbf{y}_{a,q} \in \mathbf{Y}, \mathbf{y}_{b,q} \in \mathbf{Y}} \left\| \tau_{\mathbf{y}}(b-a) - \mathbf{w}_{\mathbf{y}}^T (\mathbf{y}_{b,q} - \mathbf{y}_{a,q}) \right\|_h \right] + \mu \|\mathbf{W} - \mathbf{Z}\|_F^2 + \text{Tr}(\mathbf{\Lambda}^T (\mathbf{W} - \mathbf{Z})) \quad (4)$$

where $\|\cdot\|_h$ is a hinge loss function which measures the mis-classification error with the quadratic loss: $\|\mathbf{x}\|_h = \|\max(0, \mathbf{x})\|_2^2$, μ is the penalty parameters for the constraint $\mathbf{W} = \mathbf{Z}$, $\mathbf{\Lambda} \in \mathfrak{R}^{d \times 2}$ is the Lagrange multiplier matrix for the equality constraint $\mathbf{W} = \mathbf{Z}$, $\text{Tr}(\cdot)$ represents the trace operator, and λ is the penalty parameter for the constraints C_1 and C_2 , respectively. Equation (4) can be optimized by alternatively solving \mathbf{W} , \mathbf{Z} until the overall energy function converges.

3 Experiments

In the following experiments, we select 70 MCI-C subjects from ADNI dataset which have AD onset in the middle of longitudinal image sequence and 81 MCI-NC subject which stay in MCI stage until the last scan in the latest ADNI dataset. For all subjects, 95.3 % have 4 follow-ups every 6 months, and the remaining 4.7 % having more than 4 follow-ups. Specifically, for 70 MCI-C subjects, 11.1 % are diagnosed AD at 6 months, 31.8 % at 12 months, 25.3 % are diagnosed AD at 18 months after baseline scan, while the remaining 31.8 % are diagnosed AD more than 24 months after baseline scan. We compare our proposed TS-SVM based early detection method with standard SVM based method. Furthermore, we evaluate the importance of feature selection in both TS-SVM and standard SVM method. Thus, we compare the classification performance for four method in total, denoted by SVM, SVM+FS, TS-SVM, and TS-SVM+FS, respectively. In all experiments, we split the data into 10 non-overlap folders and report the averaged classification accuracy after 10-fold cross validation. The parameters are tuned using grid search strategy only in the training dataset.

Performance of AD Early Detection. In each cross validation case, we train our TS-SVM on the training data and sequentially apply the trained classifier to the testing subject image sequence from the first follow-up. Since the month of converting to AD after baseline scans varies across MCI-C subjects, we show the detection accuracy for MCI-C subjects converting to AD 12 months, 18 months, and 24 months after the baseline scan in Tables 1, 2 and 3, separately. It is clear our TS-SVM beat the standard SVM with more than 10 % improvement in terms of classification accuracy, which shows the advantage of temporal consistency and monotony constraints in our proposed method. Also, feature selection is very important to improve the detection accuracy, where SVM+FS and TS-SVM+FS can obtain average 3.8 % and 2.9 %

increase over SVM and TS-SVM, respectively. In brief, our full method (TS-SVM+FS) can detect AD 6 months prior to AD onset with 86.8 % accuracy, 12 months prior to AD onset with 82.5 % accuracy, and 18 months prior to AD onset with 76.5 % accuracy. Note, the early detection performance in Table 3 is worse than Tables 1 and 2 at corresponding pre-diagnosis windows. The reason is that the subjects in Table 3 mostly have 5 time points and have AD onset exactly at the last time point. Thus, the unbalanced partial image sequences before and after AD onset challenge the learning of robust classifiers.

Critical Brain Regions Related with AD Progression. Since our method jointly select morphological features in training TS-SVM, it is interesting to examine the critical brain regions where the morphological features extracted from these region contribute significantly to detect AD progression via longitudinal tracking. Figure 2 show the top 20 regions selected by our TS-SVM+FS method. It is apparent that the selected brain regions are located at AD related sub-cortical regions (such as putamen, thalamus, and hippocampus) and cortical areas (such as orbitofrontal cortex, medial/lateral temporal lobe, and medial/lateral parietal lobe), which is in consensus with the neuroimaging observations in the literatures [10, 11]. We also compared the top selected ROI for short term and long term detection and found that the cortical regions contribute more for short term detection and the sub-cortical regions, such as such as putamen, thalamus, and hippocampus, contribute more for long term converters detection. This may indicate that the sub-cortical regions changes are more significant compared with the cortical regions at the earlier AD progression stage. We did not visualize this result due to the page limitation.

Table 1. Accuracy of AD early detection at 6 months, and 0 month before AD onset for the MCI-C subjects converting to AD 12 months after baseline scan.

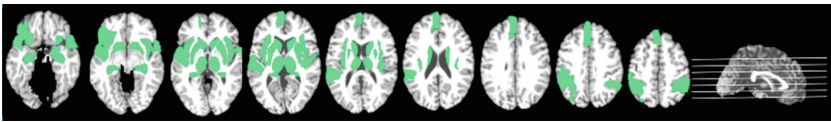
Method	18 months		12 months		6 months		0 month	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
SVM	-	-	-	-	0.7110	0.7612	0.7345	0.7937
SVM+FS	-	-	-	-	0.7557	0.7862	0.7735	0.8237
TS-SVM	-	-	-	-	0.8816	0.9327	0.8975	0.9431
TS-SVM+FS	-	-	-	-	0.9025	0.9649	0.9075	0.9776

Table 2. Accuracy of AD early detection at 12 months, 6 months, and 0 month before AD onset for the MCI-C subjects converting to AD 18 months after baseline scan.

Method	18 months		12 months		6 months		0 month	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
SVM	-	-	0.7325	0.7822	0.7455	0.7917	0.7535	0.8223
SVM+FS	-	-	0.7537	0.7912	0.7685	0.8123	0.7725	0.8314
TS-SVM	-	-	0.8425	0.8851	0.8593	0.9042	0.8635	0.9128
TS-SVM+FS	-	-	0.8475	0.8932	0.8720	0.9277	0.8812	0.9216

Table 3. Accuracy of AD early detection at 18 months, 12 months, 6 months, and 0 month before AD onset for the MCI-C subjects converting to AD 24 months after baseline scan.

Method	18 months		12 months		6 months		0 month	
	ACC	AUC	ACC	AUC	ACC	AUC	ACC	AUC
SVM	0.6016	0.6542	0.6025	0.6677	0.6325	0.6712	0.6515	0.6931
SVM+FS	0.6557	0.6862	0.6735	0.6127	0.6675	0.6983	0.6464	0.6926
TS-SVM	0.7345	0.7734	0.7675	0.8116	0.7805	0.8334	0.7875	0.8503
TS-SVM+FS	0.7653	0.7983	0.8125	0.8434	0.8345	0.8672	0.8431	0.8894

**Fig. 2.** The top 20 critical brain regions which contributed in AD early detection.

4 Conclusion

In this paper, we present a novel early AD diagnosis method using temporally structural SVM. In order to avoid inconsistent and unrealistic classification results, we propose the monotony on the output of SVM since the AD progression is generally non-reversible. In order to achieve early alarm of AD onset, we propose to adjust the classification margin such that the confidence of detecting AD progression becomes high as more and more follow-up scans are examined. Furthermore, we jointly perform feature selection and training of TS-SVM, in order to make the selected features can work well with the trained classifiers.

References

1. Thompson, P.M., Hayashi, K.M., Dutton, R.A., Chiang, M.C., Leow, A.D., Sowell, E.R., De Zubicaray, G., Becker, J.T., Lopez, O.L., Aizenstein, H.J., Toga, A.W.: Tracking Alzheimer's disease. *Ann. NY Acad. Sci.* **1097**, 198–214 (2007)
2. aël Chetelat, G., Baron, J.-C.: Early diagnosis of Alzheimer's disease: contribution of structural neuroimaging. *NeuroImage* **18**, 525–541 (2003)
3. Reisberg, B., Ferris, S.H., Kluger, A., Franssen, E., Wegiel, J., de Leon, M.J.: Mild cognitive impairment (MCI): a historical perspective. *Int. Psychogeriatr.* **20**, 18–31 (2008)
4. Hua, X., Lee, S., Hibar, D.P., Yanovsky, I., Leow, A.D., Toga Jr., A.W., Jack, C.R., Bernstein, M.A., Reiman, E.M., Harvey, D.J., Kornak, J., Schuff, N., Alexander, G.E., Weiner, M.W., Thompson, P.M.: Mapping Alzheimer's disease progression in 1309 MRI scans: power estimates for different inter-scan intervals. *NeuroImage* **51**, 63–75 (2010)
5. Li, Y., Wang, Y., Wu, G., Shi, F., Zhou, L., Lin, W., Shen, D.: Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features. *Neurobiol. Aging* **33**, 427.e415–430 (2012)

6. Hua, X., Gutman, B., Boyle, C.P., Rajagopalan, P., Leow, A.D., Yanovsky, I., Kumar, A.R., Toga Jr., A.W., Jack, C.R., Schuff, N., Alexander, G.E., Chen, K., Reiman, E.M., Weiner, M.W., Thompson, P.M.: Accurate measurement of brain changes in longitudinal MRI scans using tensor-based morphometry. *NeuroImage* **57**, 5–14 (2011)
7. Filley, C.: Alzheimer's disease: it's irreversible but not untreatable. *Geriatrics* **50**, 18–23 (1995)
8. Boyd, S., et al.: Distributed optimization and statistical learning via the ADMM. *Found. Trends Mach. Learn.* **3**, 1–122 (2011)
9. Nie, F., Huang, Y., Wang, X., Huang, H.: New primal SVM solver with linear computational cost for big data classifications. In: *ICML* (2014)
10. Hoesen, G.W.V., Parvizi, J., Chu, C.-C.: Orbitofrontal cortex pathology in Alzheimer's disease. *Cereb. Cortex* **10**, 243–251 (2000)
11. Risacher, S., Saykin, A.: Neuroimaging biomarkers of neurodegenerative diseases and dementia. *Semin. Neurol.* **33**, 386–416 (2013)
12. Antila, K., Lötjönen, J., Thurfjell, L., et al.: The PredictAD project: development of novel biomarkers and analysis software for early diagnosis of the Alzheimer's disease. *Interface Focus* **3**(2012)
13. Lorenzi, M., Ziegler, G., Alexander, D.C., Ourselin, S.: Efficient Gaussian process-based modelling and prediction of image time series. In: Ourselin, S., Alexander, D.C., Westin, C.-F., Cardoso, M. (eds.) *IPMI 2015. LNCS*, vol. 9123, pp. 626–637. Springer, Heidelberg (2015)
14. Young, A.L., et al.: A data-driven model of bio-marker changes in sporadic Alzheimer's disease. *Brain* **25**, 64–77 (2014)
15. Fonteijn, H.M., et al.: An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage* **60**, 1880–1889 (2012)
16. Zhu, Y., Lucey, S.: Convolutional sparse coding for trajectory reconstruction. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(3), 529–540 (2015)