Improved Prediction of Cognitive Outcomes via Globally Aligned Imaging Biomarker Enrichments over Progressions

Lyujian Lu[1], Saad Elbeleidy[1], Lauren Baker[1], Hua Wang[1(✉)], Heng Huang[2,3], Li Shen[4], and for the ADNI

[1] Department of Computer Science, Colorado School of Mines, Golden, CO, USA
{lyujianlu,selbeleidy,laurenzoebaker}@mymail.mines.edu,
huawangcs@gmail.com
[2] Department of Electrical and Computer Engineering, University of Pittsburgh,
Pittsburgh, PA, USA
[3] JD Finance America Corporation, Mountain View, CA, USA
heng.huang@pitt.edu
[4] Department of Biostatistics, Epidemiology and Informatics, University
of Pennsylvania, Philadelphia, PA, USA
Li.Shen@pennmedicine.upenn.edu

Abstract. Incomplete or inconsistent temporal neuroimaging records of patients over time pose a major challenge to accurately predict clinical scores for diagnosing Alzheimer's Disease (AD). In this paper, we present an unsupervised method to learn enriched imaging biomarker representations that can simultaneously capture the information conveyed by all the baseline neuroimaging measures and the progressive variations of the available follow-up measurements of every participant. Our experiments on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset show improved performance in predicting cognitive outcomes thereby demonstrating the effectiveness of our proposed method.

Keywords: Alzheimer's Disease · Longitudinal representations · Representation enrichment · Imaging biomarker

1 Introduction

Alzheimer's Disease (AD) is a chronic neurodegenerative disease that severely impacts patients' thinking, memory, and behavior. AD is listed as the sixth

leading cause of death in the United States of America, threatening 5.7 million American and 44 millions individuals worldwide [1]. Over the past decade, neuroimaging measures have been widely studied to predict disease status and/or cognitive performance [14]. However, there are some critical limitations in many predictive models, *i.e.*, they routinely perform learning at each time points *separately*, ignoring the longitudinal variations among temporal brain phenotypes. First, since AD is a progressive neurodegenerative disorder, multiple consecutive neuroimaging records are usually required to monitor the disease progressions. It would be beneficial to explore the temporal relations among the longitudinal records of the biomarkers. Second, the records of neuroimaging biomarkers are often missing at some time points for some participants over the entire courses of the AD progressions, because it is difficult to conduct medical scans consistently across a large group of participants. To be more specific, higher mortality risk and cognitive impairment hinder older adults from staying in studies requiring multiple visits and thus result in incomplete data.

To overcome the first limitation of longitudinal data, many studies [3,9,17,18] explore the temporal data structures of brain phenotypes over time. However, these models often formulate the temporal neuroimaging records as a tensor, which inevitably complicates the prediction problems. To handle the second limitation of data inconsistency, most longitudinal studies of AD only utilize the samples with complete temporal records for model analysis and ignore the samples with fewer time points. But discarding of observed samples may potentially neglects substantially valuable information in the data. To solve this problem, data imputation methods [4,10] have been used to generate missing records over AD progressions, after which temporal regression studies can then be conducted. However, missing data imputation methods can introduce undesirable artifacts, which in turn can worsen the predictive power of the longitudinal models.

To deal with the longitudinal prediction problem with incomplete temporal inputs, we propose a novel method to learn an enriched biomarker representation to integrate the baseline records of the neuroimaging biomarkers of the entire cohort and the dynamic measures across all following time points of the individuals. First, it learns a global projection from the baseline records of the biomarkers from all the participants to preserve the global structure of a given neuroimaging dataset. Then, for each participant we learn a local projection from available follow-up neuroimaging records in the later couple of years to maintain the local data structure for the participant. Finally, a soft constraint is used to build consistency between the global and local projections. A schematic illustration of this formulation is shown in Fig. 1. Armed with the learned projection, we can transform the records with inconsistent sizes in a neuroimaging dataset into fixed-length vectors. With these fixed-length biomarker representations, we can make use of conventional regression methods to predict clinical scores.

2   Problem Formalization and the Solution Algorithm

In the task of representation learning for predicting cognitive outcomes, our goal is to learn a fixed-length biomarker representation vector for a participant from
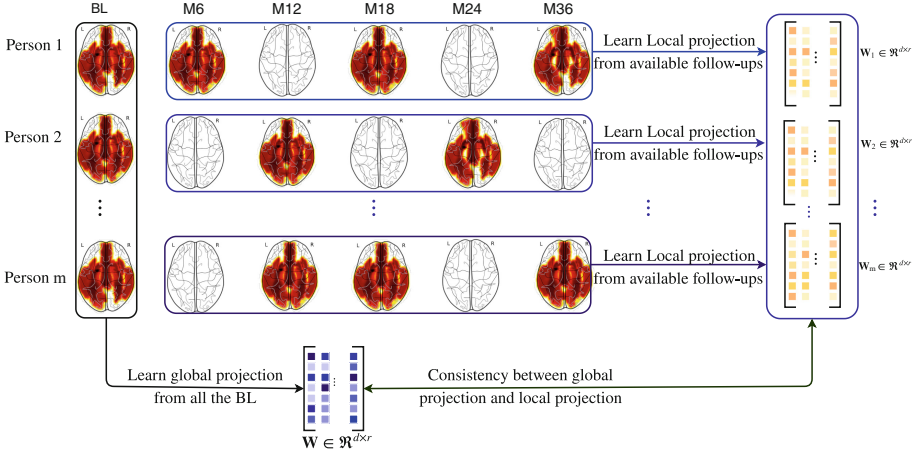
Fig. 1. Illustration of the proposed framework to learn the global and local projections. The blank plots denote the absence of scans for the participants of an AD dataset.

her or his baseline neuroimaging measures and all available follow-up measures. For the observed neuroimaging measures of a participant, we denote this information as: $\mathcal{X}_i = \{\mathbf{x}_i, \mathbf{X}_i\}$, where $i = 1, 2, \cdots, m$ is the index of a participant in a neuroimaging dataset. In every $\mathcal{X}_i$, $\mathbf{x}_i \in \Re^d$ is the baseline neuroimaging measure of the i-th participant where d denotes the dimensions of the neuroimaging feature set, and $\mathbf{X}_i = \left[\mathbf{x}_1^i, \ldots, \mathbf{x}_{n_i}^i\right] \in \Re^{d \times n_i}$ collects all available follow-up biomarker records of the i-th participant after the baseline where n_i denotes the number of available follow-up neuroimaging records of the same participant. Here we note that n_i varies over the dataset, because many participants of a cohort may have some missing records at certain time points.

Because the neuroimaging measures usually reside in a high-dimensional space, they can be redundant and noisy. Thus, we learn the compact representations of these measures via a global projection to only keep the most useful information of an input dataset. To achieve this goal, Principal Component Analysis (PCA) is the right tool, because it learns from the input data a linear projection to preserve as much information as possible in a low-dimensional projected subspace. Mathematically, PCA minimizes the reconstruction error via the projection $\mathbf{W} \in \Re^{d \times r}$ (usually $r \ll d$) that minimizes the following objective:

$$\mathcal{J}_{\text{Global}}(\mathbf{W}) = \sum_{i=1}^{m} \left\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T\mathbf{x}_i\right\|_2^2, \quad s.t. \ \mathbf{W}^T\mathbf{W} = \mathbf{I}. \tag{1}$$

Usually the neuroimaging measures of each participant do not experience drastic changes in a short time, thus we still want to preserve the local consistency in the projected space for each participant. To uncover the local consistency among the neuroimaging records of each participant, we keep the local pairwise patterns

of measures in the projected subspace. To achieve this, locality preserving projections (LPP) is the right tool to leverage. Given a pairwise similarity matrix $\mathbf{S}_i \in \Re^{n_i \times n_i}$ of the i-th participant of the neuroimaging dataset, LPP preserves the local relationships and maximizes the smoothness of the manifold of the data in the embedding space by minimizing the following objective:

$$\mathcal{J}_{\text{Local}}(\mathbf{W}_i) = \sum_{\mathbf{x}_j^i, \mathbf{x}_k^i \in \mathbf{X}_i} s_{jk}^i \left\| \mathbf{W}_i^T \mathbf{x}_j^i - \mathbf{W}_i^T \mathbf{x}_k^i \right\|_2^2, \quad s.t. \ \mathbf{W}_i^T \mathbf{W}_i = \mathbf{I}, \quad (2)$$

where s_{jk}^i is the pairwise relationship coefficient between the available records of the i-th participant at j-th and k-th time points.

To integrate the global and local consistencies of the neuroimaging measures at the same time, following the same idea in our previous works [6,7,15,16] we can develop a combined objective using the above two objectives to minimize:

$$\mathcal{J}_{\ell_2^2}(\mathcal{W}) = \sum_{i=1}^m \left\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \right\|_2^2 + \alpha \sum_{i=1}^m \sum_{\mathbf{x}_j^i, \mathbf{x}_k^i \in \mathbf{X}_i} s_{jk}^i \left\| \mathbf{W}_i^T \mathbf{x}_j^i - \mathbf{W}_i^T \mathbf{x}_k^i \right\|_2^2$$

$$+ \beta \sum_{i=1}^m \left\| \mathbf{W} - \mathbf{W}_i \right\|_2^2, \qquad s.t. \quad \mathbf{W}^T \mathbf{W} = \mathbf{I}, \ \mathbf{W}_i^T \mathbf{W}_i = \mathbf{I}, \qquad (3)$$

where we denote $\mathcal{W} = \{\mathbf{W}, \mathbf{W}_1, \cdots, \mathbf{W}_m\}$. Through the third term of Eq. (3), the projections $\mathbf{W}_i$ learned from each participant are aligned with the projection $\mathbf{W}$ learned globally from the baseline measures of the entire dataset, such that the information conveyed by the global projection $\mathbf{W}$ learned from all the participants as a whole can be transferred to each individual participant.

Taking into account that the squared ℓ_2-norm is notorious for being very sensitive to the outliers in a dataset which usually results in the failure of traditional regression models, to improve the robustness of the learned enriched representations by our new method, we substitute the squared ℓ_2-norm terms in Eq. (3) by their *not-squared* counterparts as follows:

$$\mathcal{J}_{\ell_2}(\mathcal{W}) = \sum_{i=1}^m \left\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \right\|_2 + \alpha \sum_{i=1}^m \sum_{\mathbf{x}_j^i, \mathbf{x}_k^i \in \mathbf{X}_i} s_{jk}^i \left\| \mathbf{W}_i^T \mathbf{x}_j^i - \mathbf{W}_i^T \mathbf{x}_k^i \right\|_2$$

$$+ \beta \sum_{i=1}^m \left\| \mathbf{W} - \mathbf{W}_i \right\|_2^2, \qquad s.t. \quad \mathbf{W}^T \mathbf{W} = \mathbf{I}, \ \mathbf{W}_i^T \mathbf{W}_i = \mathbf{I}. \qquad (4)$$

Upon solving the optimization problem in Eq. (4), we learn a fixed-length representation for each participant by computing $\{\mathbf{y}_i = \mathbf{W}_i^T \mathbf{x}_i\}_{i=1}^m$, which can be readily fed into traditional machine learning models.

Although the motivations of the formulation of our new method in Eq. (4) is clear and justifiable, it is a non-smooth objective, which is difficult to efficiently solve in general. Using the optimization framework presented in our earlier work [8] that solves non-smooth objectives using not-squared ℓ_2-norm terms, we can

solve Eq. (4) by an iterative procedure [8, Algorithm 1] in which the key step is to minimize the following objective:

$$\mathcal{J}_{\ell_2}^{\mathrm{R}}(\mathcal{W}) = \mathrm{tr}\left(\mathbf{X} - \mathbf{W}\mathbf{W}^T\mathbf{X}\right)\mathbf{\Gamma}\left(\mathbf{X} - \mathbf{W}\mathbf{W}^T\mathbf{X}\right)^T + \alpha\sum_{i=1}^{m}\mathrm{tr}\left(\mathbf{W}_i^T\mathbf{X}_i\mathbf{L}_i\mathbf{X}_i^T\mathbf{W}_i\right)$$

$$+ \beta\sum_{i=1}^{m}\|\mathbf{W} - \mathbf{W}_i\|_2^2, \quad s.t. \quad \mathbf{W}^T\mathbf{W} = \mathbf{I},\ \mathbf{W}_i^T\mathbf{W}_i = \mathbf{I}, \tag{5}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m] \in \Re^{d\times m}$ summarizes all the baseline measurements, and $\mathbf{\Gamma} \in \Re^{m\times m}$ is a diagonal matrix whose i-th element is $\gamma^i = \frac{1}{2\sqrt{\|\mathbf{x}_i - \mathbf{W}\mathbf{W}^T\mathbf{x}_i\|_2^2 + \delta}}$. Defining $\theta_{jk}^i = \frac{1}{2\sqrt{\|\mathbf{W}_i^T\mathbf{x}_j^i - \mathbf{W}_i^T\mathbf{x}_k^i\|_2^2 + \delta}}$ and $\tilde{\mathbf{S}}_i \in \Re^{n_i\times n_i}$ whose element value is $\tilde{s}_{jk}^i = \theta_{jk}^i s_{jk}^i$, in Eq. (5) we compute $\mathbf{L}^i = \mathbf{D}^i - \tilde{\mathbf{S}}^i$ where $\mathbf{D}^i$ is a diagonal matrix whose diagonal entries are the column (or row) sums of $\tilde{\mathbf{S}}_i$, i.e., $d_{jj} = \sum_j \tilde{s}_{jk}$.

The objective in Eq. (5) can be solved by the Alternating Direction Method of Multipliers (ADMM) [2] that minimizes the following equivalent objective:

$$\mathcal{J}_{\ell_2}^{\mathrm{ADMM}}(\mathcal{W}, \mathcal{P}) = \mathrm{tr}\left(\mathbf{X} - \mathbf{W}\mathbf{W}^T\mathbf{X}\right)\mathbf{\Gamma}\left(\mathbf{X} - \mathbf{W}\mathbf{W}^T\mathbf{X}\right)^T$$

$$+ \alpha\sum_{i=1}^{m}\mathrm{tr}\left(\mathbf{W}_i^T\mathbf{X}_i\mathbf{L}_i\mathbf{X}_i^T\mathbf{W}_i\right) + \beta\sum_{i=1}^{m}\|\mathbf{P} - \mathbf{P}_i\|_F^2 + \frac{\mu}{2}\left\|\mathbf{W} - \mathbf{P} + \frac{1}{\mu}\mathbf{\Lambda}\right\|_2^2$$

$$+ \sum_{i=1}^{m}\frac{\mu}{2}\left\|\mathbf{W}_i - \mathbf{P}_i + \frac{1}{\mu}\mathbf{\Lambda}_i\right\|_2^2, \quad s.t. \quad \mathbf{P}^T\mathbf{P} = \mathbf{I},\ \mathbf{P}_i^T\mathbf{P}_i = \mathbf{I}, \tag{6}$$

where $\mathcal{P} = \{\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_m\}$, $\mathbf{\Lambda} \in \Re^{d\times r}$ is the Lagrangian multiplier for the constraint of $\mathbf{W} = \mathbf{P}$, and $\mathbf{\Lambda}_i \in \Re^{d\times r}$ is the Lagrangian multiplier for the constraint of $\mathbf{W}_i = \mathbf{P}_i$. Algorithm (1) summaries the solution to Eq. (6).

3   Experiments

Data used in our experiments were obtained from the ADNI database (adni.loni.usc.edu). We downloaded 1.5T MRI scans and the demographic information for 821 ADNI-1 participants. We performed voxel-based morphometry (VBM) on the MRI data by following [12] and extracted mean modulated gray matter (GM) measures for 90 target regions of interest (ROI). These measures were adjusted for the baseline intracranial volume (ICV) using regression weights derived from the Healthy Control (HC) participants at the baseline. We also downloaded the longitudinal scores of the participants in the cognitive assessment of Mini-Mental State Examination (MMSE). The time points examined in this study for both imaging biomarkers and cognitive assessments include baseline (BL), Month 6 (M6), Month 12 (M12), Month 18 (M18) Month 24 (M24) and Month 36 (M36). All the participants' data used in our enriched biomarker representation study are required to have a BL MRI measurement, BL cognitive score and at least two available measures from M6/M12/M18/M24/36. A total

---

Algorithm 1: Solve the optimization problem in Eq. (6).

Initialization: $\mathbf{W}$, $\mathbf{W}_i$, $\mathbf{P}$, $\mathbf{P}_i$, $\mathbf{\Lambda}$, $\mathbf{\Lambda}_i$, $1 < \rho < 2$, $\mu, \alpha, \beta > 0$;

while *not converge* **do**

    1. Update $\mathbf{W}_i$ by $\mathbf{W}_i = \left(2\alpha \mathbf{X}_i \mathbf{L}_i \mathbf{X}_i^T + \mu \mathbf{I}\right)^{-1} (\mu \mathbf{P}_i - \mathbf{\Lambda}_i)$;

    2. Update $\mathbf{P}_i$ by $\mathbf{P}_i = \mathbf{U}_i \mathbf{V}_i^T$, where $\mathbf{N}_i = 2\beta \mathbf{P} + \mu \mathbf{W}_i + \mathbf{\Lambda}_i$ and svd $(\mathbf{N}_i) = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{V}_i^T$;

    3. Update $\mathbf{W}$ by $\mathbf{W} = \left(\mu \mathbf{I} - 2\mathbf{X} \mathbf{\Gamma} \mathbf{X}^T\right)^{-1} (\mu \mathbf{P} - \mathbf{\Lambda})$;

    4. Update $\mathbf{P}$ by $\mathbf{P} = \mathbf{U}\mathbf{V}^T$, where $\mathbf{N} = 2\beta \sum_{i=1}^{m} \mathbf{P}_i + \mu \mathbf{W} + \mathbf{\Lambda}$ and svd $(\mathbf{N}) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$;

    5. Update $\mathbf{\Lambda}_i$ by $\mathbf{\Lambda}_i = \mathbf{\Lambda}_i + \mu (\mathbf{W}_i - \mathbf{P}_i)$;

    6. Update $\mathbf{\Lambda}$ by $\mathbf{\Lambda} = \mathbf{\Lambda} + \mu (\mathbf{W} - \mathbf{P})$;

    7. Update μ by $\mu = \rho\mu$;

Output: $\mathbf{W}$, $\mathbf{W}_i$.

---

of 544 sample subjects are selected to perform in our study, among which we have 92 AD samples, and 205 Mild Cognitive Impairment (MCI) samples and 247 HC samples.

Experiment Settings. To validate the usefulness of our proposed method, we compare the performance to predict cognitive outcomes using two types of the neuroimaging inputs—the learned enriched representation and BL biomarker measurement. In our experiments, several methods proven to generalize well, such as ridge regression (RR), Lasso, support vector regression (SVR), and convolutional neural networks (CNN), are leveraged. For RR, Lasso and SVR models, we conduct a standard 5-fold cross-validation to fine tune the parameters and compute the root mean square error (RMSE) between the predicted values and ground truth values of the cognitive scores on the testing data. For the SVR model, the linear kernel is leveraged, for which the box constraint parameters are fine tuned by a grid search in standard 5-fold cross-validation as well. For the CNN regression model, we randomly select 70% of the neuroimaging measurements as the training set, 20% of the neuroimaging measurements as the validation set and the remaining 10% of the neuroimaging measurements as the testing set. The evaluation metrics are reported based on the results on the testing dataset. We construct a two-layer convolution architecture for the cognitive outcomes prediction and the dropout technique [13] is also leveraged to reduce overfitting in CNN models and prevent complex co-adaptations on training data. For the model parameters, α, β are fine tuned by searching the grid of $\{10^{-5}, \ldots, 10^{-1}, 1, 10, \cdots, 10^5\}$.

Experiment Results. From Table 1, we can see that the proposed enriched neuroimaging representation is consistently better than baseline representations when we use the four different methods, LR, RR Lasso, SVR and CNN. It can be attributed to the following reasons. First, the original baseline neuroimaging biomarker representation only contains the cognitive information of a participant at one single time point, which does not benefit from the correlation across

different cognitive measures over the time. Instead, our proposed enriched neuroimaging biomarker representation could capture not only the baseline cognitive measurement, but also the temporal information conveyed by the longitudinal biomarkers over AD progressions. Our enriched neuroimaging representation could integrate the neuroimaging measurement at the fixed time point and the dynamic temporal changes. As AD is a progressively degenerative disease, incorporation of future information about subjects benefits the prediction model. Second, the original baseline neuroimaging measurements exhibits high dimensionality, which could be redundant and noisy. Thus the traditional methods easily suffer from "the curse of dimensionality". Via the projection, we map the baseline cognitive measurement into a low dimensional space thereby mitigating the issue of high dimensionality.

Table 1. RMSE values (smaller is better) when predicting MMSE score using the VBM biomarkers.

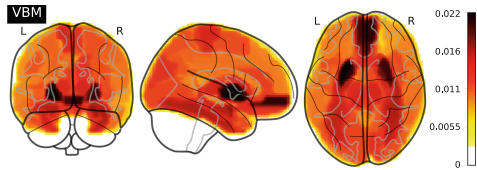|	RR	Lasso	SVR	CNN
Baseline	0.4398	0.3845	0.5035	0.1572
Enriched (our method)	0.3192	0.3509	0.3898	0.1546

Fig. 2. Frequency map of regions of interest of brain.

Apart from the cognitive outcomes prediction task, another primary goal of our regression analysis is to identify a subset of imaging biomarkers which are highly correlated to AD progressions. Therefore, we examine the imaging biomarkers of each participant identified by the proposed methods encoded by the cognitive scores. A brief summary of all the patients' projections can be conducted to determine the most important biomarkers across patients. We select the 10 most frequently appearing regions of interest from each patient's projection coefficients to reconstruct a global frequency map, shown in Fig. 2. The bilateral hippocampus, amygdala, fusiform, putamen regions are found to be in the top selected biomarkers by our model, which are in nice accordance with the known clinical knowledge [5,11].

4   Conclusion

Missing data is a critical challenge in longitudinal AD studies. In this paper, we proposed a formulation to learn a consistent-length representation for all the participants in the ADNI dataset. The enriched fixed length biomarker representation could capture the global consistency from baseline measurements and local pairwise pattern from available follow-up measurements of each participant at the same time. Our results show that the learned enriched representation beat the performance of the baseline measurement in predicting the clinical scores. Furthermore, the identified biomarkers are highly suggestive and strongly agree with existing research findings, which warrants the correctness of our approach.

References

1. Association, A., et al.: 2018 Alzheimer's disease facts and figures. Alzheimer's Dement. **14**(3), 367–429 (2018)
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends® Mach. Learn. **3**(1), 1–122 (2011)
3. Brand, L., et al.: Joint high-order multi-task feature learning to predict the progression of Alzheimer's disease. In: The Twenty-First International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2018), pp. 555–562 (2018)
4. Campos, S., Pizarro, L., Valle, C., Gray, K.R., Rueckert, D., Allende, H.: Evaluating imputation techniques for missing data in ADNI: a patient classification study. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. LNCS, vol. 9423, pp. 3–10. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25751-8_1
5. De Jong, L.: Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. Brain **131**(12), 3277–3285 (2008)
6. Liu, K., Wang, H., Han, F., Zhang, H.: Visual place recognition via robust ℓ_2-norm distance based holism and landmark integration. In: The thirty-Third AAAI Conference on Artificial Intelligence (AAAI 2019) (2019)
7. Liu, K., Wang, H., Nie, F., Zhang, H.: Learning multi-instance enriched image representations via non-greedy ratio maximization of the ℓ_1-norm distances. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018 (CVPR 2018), pp. 7727–7735 (2018)
8. Liu, Y., Guo, Y., Wang, H., Nie, F., Huang, H.: Semi-supervised classifications via elastic and robust embedding. In: The thirty-First AAAI Conference on Artificial Intelligence (AAAI 2017) (2017)

9. Lu, L., Wang, H., Yao, X., Risacher, S., Saykin, A., Shen, L.: Predicting progressions of cognitive outcomes via high-order multi-modal multi-task feature learning. In: IEEE ISBI 2018, pp. 545–548. IEEE (2018)

10. Minhas, S., Khanum, A., Riaz, F., Alvi, A., Khan, S.A.: Early Alzheimer's disease prediction in machine learning setup: empirical analysis with missing value computation. In: Jackowski, K., Burduk, R., Walkowiak, K., Woźniak, M., Yin, H. (eds.) IDEAL 2015. LNCS, vol. 9375, pp. 424–432. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24834-9_49

11. Poulin, S.P., Dautoff, R., Morris, J.C., Barrett, L.F., Dickerson, B.C., Initiative, A.D.N., et al.: Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. Psychiatry Res. Neuroimaging **194**(1), 7–13 (2011)

12. Risacher, S.L., et al.: Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. Neurobiol. Aging **31**(8), 1401–1418 (2010)

13. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**, 1929–1958 (2014)

14. Stonnington, C.M., et al.: Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. Neuroimage **51**(4), 1405–1413 (2010)

15. Wang, H., Huang, H., Ding, C.: Discriminant laplacian embedding. In: The twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI 2010) (2010)

16. Wang, H., Nie, F., Huang, H.: Globally and locally consistent unsupervised projection. In: The twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI 2014), pp. 1328–1333 (2014)

17. Wang, H., et al.: High-order multi-task feature learning to identify longitudinal phenotypic markers for Alzheimer's disease progression prediction. In: NIPS, pp. 1277–1285 (2012)

18. Wang, X., et al.: Longitudinal genotype-phenotype association study via temporal structure auto-learning predictive model. In: Sahinalp, S.C. (ed.) RECOMB 2017. LNCS, vol. 10229, pp. 287–302. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56970-3_18