

Hyper-graph based Adaptive Sparse Multi-view Canonical Correlation Analysis with Application to Neuroimaging Genetics Study of Alzheimer's Disease

Lei Wang¹, Wei Kong^{1,2}, and Shuaiqun Wang³

^{1,2} College of Information Engineering, Shanghai Maritime University, Shanghai, China
Email: 1192196207@qq.com

Abstract. Neuroimaging genetics has gained more and more attention on account of detecting the linkage between the brain imaging phenotypes (i.e., regional volumetric measures) and the genetic variants (i.e., Single Nucleotide Polymorphism (SNP) in Alzheimer's disease (AD)). To overcome the problem of sparse multi-view canonical correlation (SMCCA) 'unfair combination of pairwise covariance', introducing adaptive weights when combining pairwise covariances, a novel formulation of SMCCA, named adaptive SMCCA. In this paper, we integrate multi-modal genomic data from postmortem AD brain and proposed a hyper-graph based sparse multi-view canonical correlation analysis (HGSMCCA) method to extract the most correlated multi-modal biomarkers. Specifically, we utilized the adaptive sparse multi-view canonical correlation analysis (AdsSMCCA) framework, consider the benefit of hyper-graph-based regularization term into consideration that will contribute to the selection of more discriminative biomarkers. We propose a hyper-graph optimization strategy based on the adaptive SMCCA model, and we apply it to neuroimaging genetics data. All these results demonstrate the capability of HGSMCCA in identifying diagnostically genotype-phenotype patterns.

Keywords: neuroimaging genetics, hyper-graph, Alzheimer's disease, sparse multi-view canonical correlation analysis.

1 Introduction

Alzheimer's disease (AD) is the most common form of dementia in the elderly that is characterized by the progressive loss of neurons and their connections. Because of the dramatic increase in the prevalence of AD, the identification of effective biomarkers for the early diagnosis and treatment of AD in individuals at high risk to develop the disease is crucial [1]. Neuroimaging techniques including Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) have been widely used to investigate the neurophysiological characteristics of AD [2]. Neuroimaging genetics is a hot research field for identifying associations between imaging results and genetic variables. Genetic variation affects phenotypes such as brain function and structure. Investigating the relationship between imaging and genetic variables provides a new perspective for a better understanding of mechanisms in the brain [3].

Structured sparse canonical correlation analysis (SCCA) models have been widely used to identify neuroimaging genetic associations. These models either use group lasso or graph-guided fused lasso to conduct feature selection and feature grouping simultaneously [4]. Nevertheless, existing SCCA algorithms use the soft threshold strategy for solving the lasso regularization terms, which assumes the independent structure of data features. Meanwhile, this method only contrite on finding the best linear transformations for imaging and genetic features, respectively.

Recently, several studies combined multiple data types for the AD disease mechanism study. For instance, Bihmeyer et al [5][6] integrated microarray and RNA-seq data to identify important genes in the pathophysiology of AD. Jiang et al [7][8] systematically analyzed AD-related mRNA and miRNA expression profiles to identify active transcription factor and miRNA regulatory pathways in AD. Humphries et al [9][10] proposed an integrative method that combined the transcriptome and DNA methylation data to identify gene networks specific to late-onset Alzheimer's disease. All the above methods can disclosure the hidden mechanism of AD that cannot be fulfilled using a single type of data. But they ignore diagnostic biomarkers identification for disease status and progression stages [11-13].

To analyze multi-modal data sets, feature selection is the key step in biomarker identification. For instance, Jie et al [14] proposed a multi-modal feature selection method that used a multi-task learning framework to preserve the intrinsic relatedness among multiple modalities of data. Zu et al [15] constructed a label aligned multi-task learning framework that can uncover inherent structures among multi-modal data. Besides, Lei [16] proposed a canonical correlation analysis (CCA) based method to fuse multi-modal data for the diagnosis of AD/MCI patients. However, such a pairwise relationship does not reflect the complex connections of different subjects in real applications. Intuitively, the higher-order relationships among subjects can provide useful information to improve the machine learning performance of AD/MCI diagnosis tasks [17].

To bridge this gap, our goal is to apply sparse multi-view CCA (SMCCA) machine learning algorithms on multi-modal AD data to identify biomarkers for highly accurate individual classification of AD patients from normal age-matched people [18]. In this paper, we constructed a hypergraph in each modality and introduced a hyper-graph-based Laplacian regularization term to capture the high-order relationships among different subjects. We adopt a hyper-graph optimization strategy based on the adaptive SMCCA model [19]. To evaluate the effectiveness of the proposed method, we perform the classification on the dataset derived from ADNI. We show that our method outperforms competing SCCA models on both synthetic and real data. In particular, our method identifies stronger canonical correlations and better canonical loading patterns, demonstrating its promising capability in revealing biologically meaningful neuroimaging genetics associations.

2 Methodology

In this paper, we denote scalars as italic letters. The Euclidean norm of a vector u is denoted as $\|u\|$. $X_G \in R^{N \times p}$, $X_D \in R^{N \times q}$, $X_M \in R^{N \times r}$ denote the SNP data, MRI, and PET data, where N denotes the number of samples, p , q , r corresponding to the dimensionalities for each modality of data. Then, the objective function of the adaptive SMCCA model can be formulated as follows:

$$(u_1, u_2, \dots, u_m) = \arg \max_{u_1, u_2, \dots, u_m} \sum_{i < j} \Lambda_{ij} u_i^t X_i^t X_j u_j \quad \text{s.t.} \quad u_i^t u_i \leq 1, P_i(u_i) \leq \lambda_i, \forall i \quad (1)$$

where Λ_{ij} is a weight term used to adaptively fit the varying scales of different pairwise covariances during each iteration step.

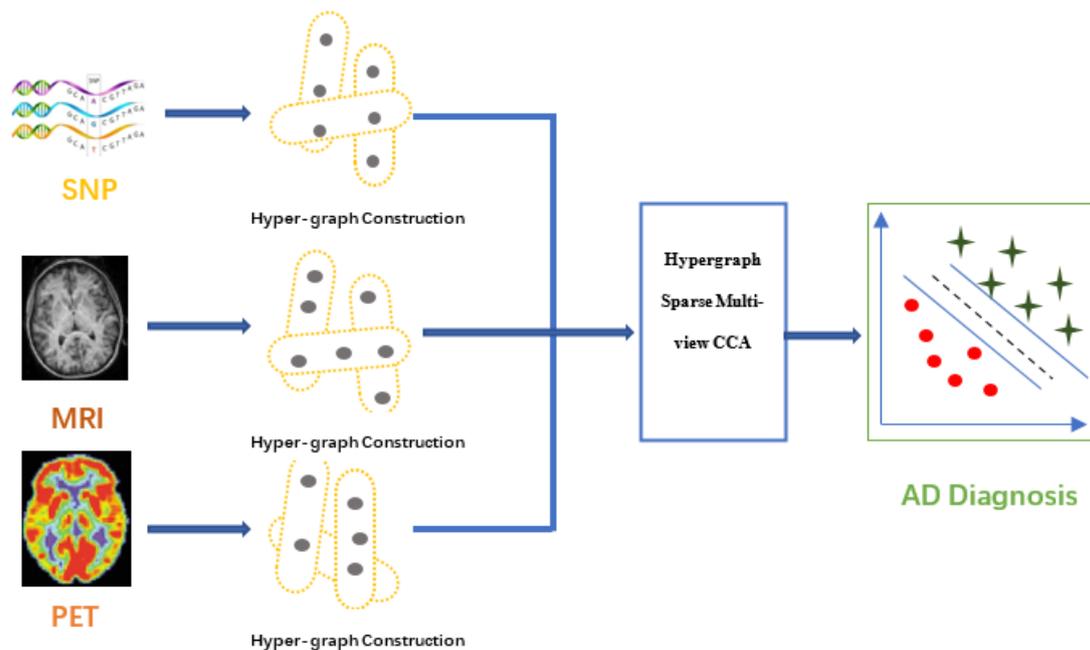


Figure 1. The flowchart of the HGSMCCA method.

2.1 Overview of the HGSMCCA Workflow

Figure 1. illustrates the flowchart of the proposed HGSMCCA method, which includes four steps: 1) for the multi-modal data, filter and preselect the features that are highly correlated to AD status; 2) construct the hyper-graph for each data modality that can capture the high-order correlation among different subjects; 3) based on the derived hypergraphs, a novel HGSMCCA algorithm is used to further select the informative features from the pre-selected features in step 1; and 4) the support vector machine (SVM) classifier is adopted for the diagnosis of AD (or MCI) patients from the normal control subjects.

2.2 Hyper-graph Based Sparse Multi-view Canonical Correlation Analysis (HGSMCCA)

In the AdaSMCCA model, we only consider the inter-correlations among multi-modal data but ignore the mutual dependency among different subjects, which may result in large deviations even for very similar data after projecting to the common representation Q (shown in Eq. (1)). Different from the traditional graph, a hyper-graph can connect more than two vertices by a hyperedge. Hence, we can model complex relationships (i.e., high-order relationships) among samples of our interest using hyper-graph.

Mathematically, a hyper-graph can be represented by $G(V, E, a)$, where $V = \{v_1, v_2, \dots, v_n\}$, $E = \{e_1, e_2, \dots, e_{N_e}\}$ denotes the set of vertices and hyper-edges respectively [20], and $a \in R^{N_e}$ is the weights for the hyper-edges. Each hyper-graph e_i ($i=1, 2, \dots, N_e$) is assigned a weight $a(e_i)$. In hyper-graphs, a hyper-edge can connect more than two vertices, through which high-order relationships can be modeled explicitly. We use the incidence matrix, i.e., $H \in R^{N_e \times N_v}$ to represent the relationships between hyper-edges and vertices, an entry $H(v_i, e_j)$ is set to 1 if the hyperedge e_j contains vertex v_i , and 0 otherwise. Based on this definition of H , the degree of each vertex and hyper-edge are denoted as:

$$d(v_i) = \sum_{e_j \in E} a(e_j)H(v_i, e_j) \tag{2}$$

$$\varepsilon(e_j) = \sum_{v_i \in V} H(v_i, e_j) \tag{3}$$

Suppose D_v and D_E represent the diagonal matrices containing the vertex degrees and the hyperedge degrees, respectively. Denote $A \in R^{N_e \times N_e}$ as the diagonal matrix of hyper-edge weights with the diagonal element $A_{i,i} = a(e_i)$. In order to capture the high-order correlation among different subjects, we follow the method in [42] to define the hyper-graph Laplacian matrix L^H as following:

$$L^H = I - \Phi \tag{4}$$

where I is the identity matrix and $\Phi = D_v^{-\frac{1}{2}}HAD_E^{-1}H^T D_v^{-\frac{1}{2}}$. As can be seen from Eq. (2) and Eq. (3), the most important step for hyper-graph learning is to construct the hyper-graph from the input data. Here, we adopt the K nearest neighbor (KNN) strategy to generate the hyper-graph [19-21].

Based on the hyper-graph Laplacian matrix (shown in Eq. (4)), we define the objective function of the hyper-graph based sparse multi-view canonical correlation analysis method as

$$\begin{aligned} \min_{w_i, Q} \frac{1}{N} \sum_{i \in \{G, D, M\}} X_i w_i - Y_i^2 + \alpha \sum_{i \in \{G, D, M\}} X_i w_i - Q_i^2 + \beta \sum_{i \in \{G, D, M\}} w_i^T X_i^T \\ L_i^H X_i w_i + \sum_{i \in \{G, D, M\}} r_i w_{i1} \end{aligned} \tag{5}$$

The first term corresponds to the linear discriminant function for each data modality, and $Y = [y_1, y_2, \dots, y_N]^T$ is the label vector for the N subjects, and $y_i = 1$ or -1 indicates the disease case or control normal [22]. L_i^H in the third term represents the hyper-graph Laplacian matrix for modality i . In comparison to the adaptive SMCCA model, our proposed HGSMCCA model not only incorporates correlations among multi-modal data, but also preserves the high-order structure information in each modality by adding hyper-graph regularization terms, which can generate more discriminant features for the diagnosis of AD disease.

2.3 Optimization

The objective function in Eq. (5) is not jointly convex with respect to w_i and Q . However, it is convex with respect to w_i when Q is fixed, and also convex with respect to Q given a fixed w_i . In this study, we adopt an alternating optimization algorithm to solve the problem in (5). Specifically, we first fix w_i , $i \in \{G, D, M\}$, then the optimization problem for Q is:

$$\min_Q \sum_{i \in \{G, D, M\}} \|X_i w_i - Q\|_2^2 \quad (6)$$

which has a closed form solution:

$$Q = \frac{1}{3} \sum_{i \in \{G, D, M\}} X_i w_i \quad (7)$$

On the other hand, if other variables are fixed, the optimization problem for each w_i can be formulated as:

$$\min_{w_i} \frac{1}{N} \|X_i w_i - Y\|_2^2 + \alpha \|X_i w_i - Q\|_2^2 + \beta w_i^T X_i^T L_i^H X_i w_i + r_i \|w_i\|_1 \quad (8)$$

Since the L_1 -norm is non-differentiable at zero, a smooth approximation is estimated for L_1 term by including an extremely small value. Then, we can approximate $r_i \|w_i\|_1$ by

$$\|w_i\|_1 \approx \sum_{j=1}^{d_i} \sqrt{(w_i^j)^2 + \epsilon} \quad (9)$$

Here, D_i is a diagonal matrix with the j -th element denoted as $\frac{1}{\sqrt{(w_i^j)^2 + \epsilon}}$. Then, by taking the derivative

of Eq. (8) with respect to w_i and let it be zero, we can obtain:

$$\frac{2}{N} (X_i)^T (X_i w_i - Y) + 2\alpha (X_i)^T (X_i w_i - Q) + 2\beta X_i^T L_i^H X_i w_i + r_i D_i w_i = 0 \quad (10)$$

Then, we can derive

$$w_i = \left[\left(\alpha + \frac{1}{N} \right) X_i^T X_i + \beta X_i^T L_i^H X_i + \frac{r_i}{2} D_i \right]^{-1} \left(\alpha X_i^T Q + \frac{1}{N} X_i^T Y \right) \quad (11)$$

In our experiments, the iterative number is set as 10.

3 Results

3.1 Subjects and Neuroimage Preprocessing

Table 1. Demographic and clinical information for the participants in this study

Subjects	CN	MCI	AD
Num	121	213	87
Gender (M/F, %)	52/48	55/45	53/47
Age (mean \pm std)	77.2 \pm 5.3	73.8 \pm 7.3	74 \pm 7.5

CN = Cognitive Normal, MCI = Mild Cognitive Impairment, AD = Alzheimer's disease

In this study, we select 421 subjects from the ADNI cohort (www.adni-info.org) with complete MRI and PET data at baseline scan, including 121 CN (Control Normal), 213 MCI (Mild Cognitive Impairment), and 87 AD (Alzheimer's Disease). 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 116 pre-defined ROIs, 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 116 ROI-based features from both the MRI and PET data, respectively [23-25].

For the genotyping data, we included 85 SNP markers within the APOE gene, including the APOE e4 SNP rs429358 (*i.e.* the best-know AD genetic risk factor). We aim to evaluate the associations between the amyloid data and the APOE SNP data using the proposed method.

3.2 Identification of Top Selected Neuroimaging Genetic Markers

In our experiment, a 5-fold cross-validation strategy was adopted to evaluate the effectiveness of our proposed method. As for parameters of regularization, we determined their values by nested 5-fold cross-validation on the training set. It was to fine the parameters ($\lambda_1=1$, $\lambda_2=1$ and $\beta= 0.1$ in Eq.(5)).

Table 2 presents the top ten brain imaging ROIs identified by the averaged canonical weights. In this table, the first column exhibits the name of the SNP data, the second column showed the name of the brain region.

Table 2. Top ten brain ROIs and genetics selected by integrated canonical weights

SNP	Brain Region
rs11668327	Occipital_Sup_Right
rs34095326	Parietal_Sup_Left
rs283815	Insula_Right
rs118170342	Cingulum_Ant_Right
rs112019714	Precuneus_Left
rs3925681	Cerebelum_7b_Left
rs75687619	ParaHippocampal_Left
rs79398853	Cerebelum_8_Left
rs115881343	Parietal_Sup_Right
rs76366838	Temporal_Inf_Right

Based on the top ten selected SNPs and brain ROIs, Figure 2 shows the heat map of pairwise correlations of every brain ROI-SNP pair. Interestingly, we notice that rs283815 has a strong correlation with the Cerebellum region, it may provide important information for AD therapy.

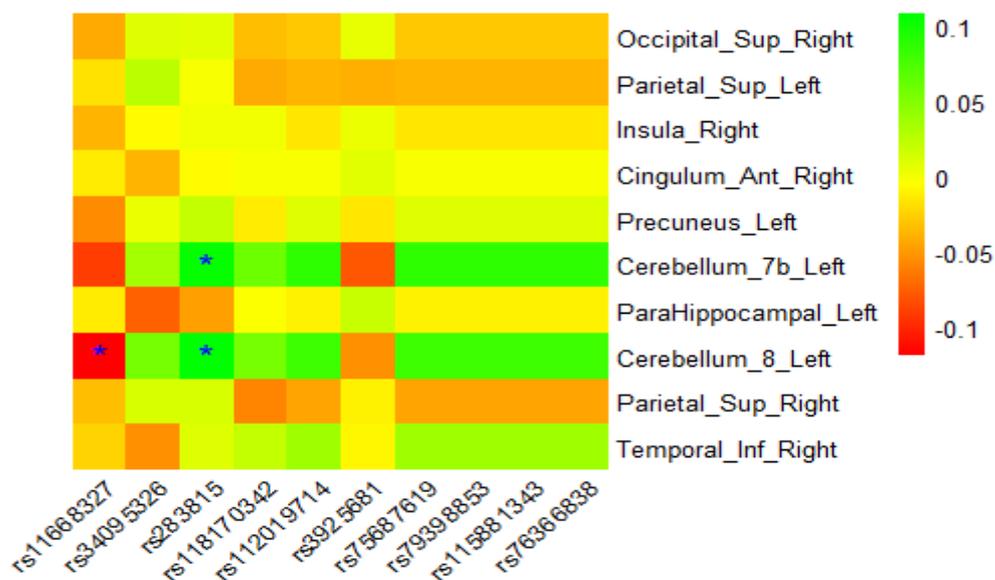


Figure 2. Heat map of brain ROI-SNP associations of top selected markers.

4 Discussion

4.1 Results on Synthetic Data

We generate data sets with different properties in this simulation study to assess the performance of our model in Figure 3.

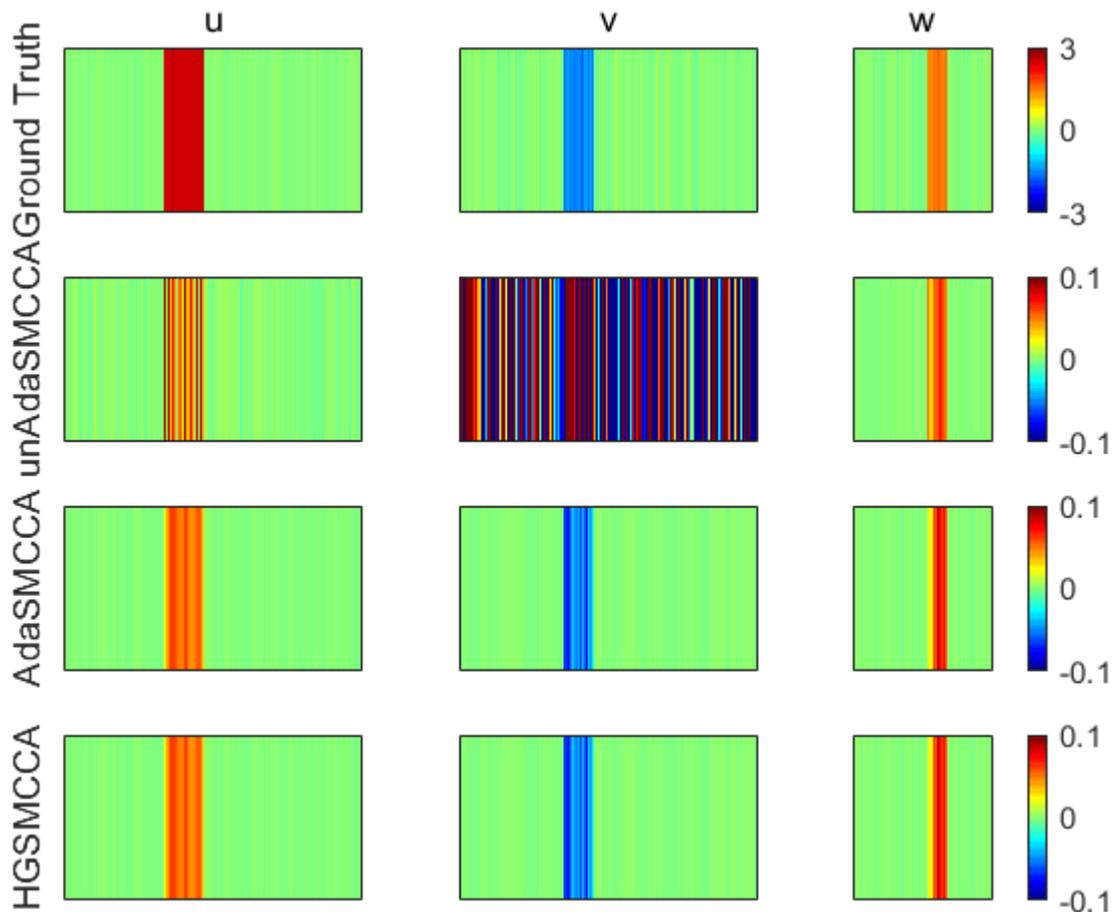


Figure 3. Canonical weights estimated on synthetic data. The first row is the ground truth, and each remaining row corresponds to an SCCA method: (1) unAdaSMCCA, (2) AdaSMCCA, and (3) HGSMCCA. For each method, the estimated weights of u are shown on the left panel, v are shown in the middle, and those of w are shown on the right. In each subfigure, the vertical axis represents the indices of each u (left panel), v (middle panel), or w (right panel).

This demonstrates that HGSMCCA can find the combination of the optimal weight that best reflects the original data importance. The difference of performance is more significant at lower noise levels compared to higher noise levels [26].

4.2 Classification Performance Comparisons among Different Methods

In this section, we compared the diagnosis power of the proposed HGSMCCA with the following algorithms on both AD vs CN and MCI vs CN tasks. 1) unAdaSMCCA: weighted SMCCA model of coefficients are fixed. 2) AdaSMCCA: adaptive adjust the weight coefficients on pairwise covariances. 3) HGSMCCA: Hyper-graphed regularization on the adaptive SMCCA model. For the proposed and all the comparing methods, the SVM classifier with RBF kernel is adopted for the classification task on the selected features [27-29]. The results for different methods are shown in Fig. 4.

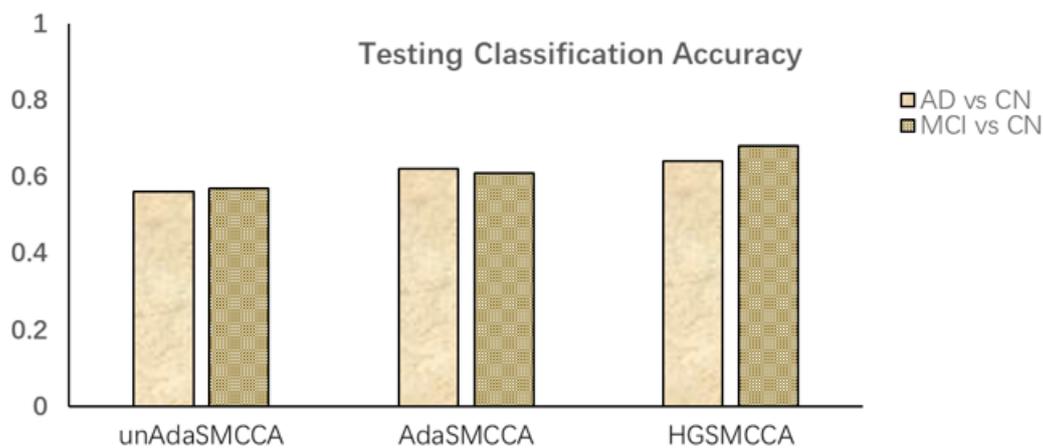


Figure 4. The results of AD vs CN and MCI vs CN tasks with three different feature selection methods.

For MCI vs CN classification, our model achieves 68.9% in terms of testing classification outperformance than adaptive SMCCA (AdaSMCCA) model.

Besides, the performance on each dataset is assessed with the correlation coefficient (CC) between actual and predicted response value, which is widely used in measuring performances of regression and association analysis [30]. The results of CC among the 5-fold training and testing data are shown in Table 2.

Table 3. Performance comparison of real data. Averaged training and testing correlation coefficients by 50 times 5-fold cross-validation

Method	Training Results	Testing Results
unAdaSMCCA	0.45 ± 0.02	0.35 ± 0.08
AdaSMCCA	0.44 ± 0.02	0.42 ± 0.07
HGSMCCA	0.47 ± 0.02	0.48 ± 0.07

As shown in Eq. (5), our HGSMCCA model uses the least square loss to calculate the empirical risk in the training dataset. It is obvious that the objective function (Eq. (5)) of the HGSMCCA is bi-convex concerning w_i and Q . Hence, we use an alternative optimization algorithm to solve the convergence problem, which can theoretically guarantee its convergence to the optimal solution.

5 Conclusion

In this paper, the HGSMCCA framework was successfully applied for neuroimaging genetic study on a candidate gene set for the diagnosis of Alzheimer's disease. This model is within generalized sparse canonical correlation analysis, but we extended it by introducing a hyper-graph-based regularization term to capture the high-order correlation among different subjects. Different from the existing multi-modal learning algorithms, HGSMCCA can incorporate the distribution knowledge of multi-modal data to find more discriminant features with biological meanings. The experimental results demonstrate that it produced much cleaner patterns than competing methods. By discovering a strong association between the APOE SNP data and the amyloid accumulation data in an AD study, for instance, we observe that rs283825 gene loci are bound up with Cerebellum by our experimental result, demonstrating that itself as a promising structured SCCA method in identifying meaningful bi-multivariate associations in brain imaging genetics studies. The classification results on ADNI datasets demonstrated that HGSMCCA can achieve significantly better classification performance for the diagnosis of AD and MCI than the currently available methods that also handle multi-modal genomic data. In the future, we will explore potentially more accurate loss function to further improve the AD classification performance.

Acknowledgments. This work was supported by the Natural Science Foundation of Shanghai (No. 18ZR1417200) and National Natural Science Foundation of China (No. 61803257).

References

1. Witten DM, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*. 2009: kxp008.
2. Du L, Huang H, Yan J, Kim S, Risacher SL, Inlow M, et al. Structured sparse canonical correlation analysis for brain imaging genetics: an improved GraphNet method. *Bioinformatics*. 2016: btw033.
3. Lin D, Calhoun VD, Wang YP. Correspondence between fMRI and SNP data by group sparse canonical correlation analysis. *Medical image analysis*. 2014; 18:891–902. [PubMed: 24247004]
4. J. Saugstad, J. Wiedrick, J. Lapidus, U. Sandau, T. Lusardi, C. Harrington, et al., “Validation of human cerebrospinal fluid microRNAs as biomarkers for Alzheimer's disease,” *Journal of Extracellular Vesicles*, vol. 7, pp. 199-199, 2018.
5. Du L, Huang H, Yan J, Kim S, Risacher SL, Inlow M, Moore JH, Saykin AJ, Shen L. Structured sparse cca for brain imaging genetics via graph oscar. *BMC Systems Biology*. 2016:335–345.
6. Friedman JH, Hastie T, Hofling H, Tibshirani R. Pathwise coordinate optimization. *The Annals of Applied Statistics*. 2007; 1(2):302–332.
7. Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, Foroud T, Pankratz N, Moore JH, Sloan CD, et al. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage*. 2010; 53(3):1051–63. [PubMed: 20100581]
8. Shen,L. and Thompson,P.M. (2020) Brain imaging genomics: integrated analysis and machine learning. *Proc. IEEE*, 108, 125–162.
9. Mukherjee,S. et al. (2018) Genetic data and cognitively defined late-onset Alzheimer's disease subgroups. *Mol. Psychiatr.* doi: 10.1038/s41380-018- 0298-8 .
10. Hao, X., et al., Identifying Multimodal Intermediate Phenotypes between Genetic Risk Factors and Disease Status in Alzheimer's Disease. *Neuroinformatics*, 14(4): p. 439-452, 2016.
11. Zhou T, Liu M, Thung KH, Shen D. Latent Representation Learning for Alzheimer's Disease Diagnosis With Incomplete Multi-Modality Neuroimaging and Genetic Data. *IEEE Trans Med Imaging*. 38(10): 2411–2422, 2019.
12. Wang, Q., et al., Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, Proceedings. Vol. 10541. 2017, Cham: Springer International Publishing AG, 2017.
13. Witten DM, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*. 10(3): 515–34, 2009.
14. Chen X, Liu H. An efficient optimization algorithm for structured sparse cca, with applications to eqtl mapping. *Statistics in Biosciences*. 4(1): 3-26, 2012.
15. Hao X, Li C, Du L, et al. Mining Outcome-relevant Brain Imaging Genetic Associations via Three-way Sparse Canonical Correlation Analysis in Alzheimer's Disease. *Sci Rep*. 7: 44272, 2017.
16. Du L, Yan J, Kim S, Risacher SL, Huang H, et al. A novel structure-aware sparse learning algorithm for brain imaging genetics. *MICCAI*. 329–336, 2014.
17. Du L, Zhang T, Liu K, et al. Sparse Canonical Correlation Analysis via Truncated ℓ_1 -norm with Application to Brain Imaging Genetics. *Proceedings (IEEE Int Conf Bioinformatics Biomed)*. 707–711, 2016.
18. Yan, J., et al., Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm. *Bioinformatics*, 30(17): p. i564-i571, 2014.
19. Viivi Uurtio, Sahely Bhadra, and Juho Rousu. Large-scale sparse kernel canonical correlation analysis. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6383–6391, Long Beach, California, USA, 09–15. PMLR, 2019.
20. Shao, Wei, et al. "Hyper-graph based Sparse Canonical Correlation Analysis for the Diagnosis of Alzheimer's Disease from Multi-dimensional Genomic Data." *Methods* (2020).
21. Zheng W, Zhou X, Zou C, Zhao L. Facial expression recognition using kernel canonical correlation analysis (KCCA). *IEEE Trans Neural Netw*. 17(1):233–238, 2006.
22. Liu Y, Tan L, Wang HF, et al. Multiple Effect of APOE Genotype on Clinical and Neuroimaging Biomarkers across Alzheimer's Disease Spectrum. *Mol Neurobiol*. 53(7): 4539–4547, 2016.

23. Du L, Zhang T, Liu K, et al. Sparse Canonical Correlation Analysis via Truncated ℓ_1 -norm with Application to Brain Imaging Genetics. *Proceedings (IEEE Int Conf Bioinformatics Biomed)*. 707–711, 2016.
24. Du L, Liu K, Zhu L, et al. Identifying progressive imaging genetic patterns via multi-task sparse canonical correlation analysis: a longitudinal study of the ADNI cohort. *Bioinformatics*. 35(14): i474–i483, 2019.
25. Larson NB, Jenkins GD, Larson MC, et al. Kernel canonical correlation analysis for assessing gene-gene interactions and application to ovarian cancer. *Eur J Hum Genet*. 22(1):126–131, 2014.
26. Iusem, A. On the convergence properties of the projected gradient method for convex optimization. *Computational & Applied Mathematics*. 22(1): 37–52, 2003.
27. Voxel-Based Morphometry. In: Schmidt R., Willis W. (eds) *Encyclopedia of Pain*. Springer, Berlin, Heidelberg, 2007.
28. Hu, Wenxing, et al. "Adaptive Sparse Multiple Canonical Correlation Analysis With Application to Imaging (Epi) Genomics Study of Schizophrenia." *IEEE Transactions on Biomedical Engineering* 65.2(2018):390-399.
29. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, et al. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mmimri single-subject brain. *Neuro Image*, 15: 273–289, 2002.
30. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, and M. J. Daly, "PLINK: a tool set for whole-genome association and population-based linkage analyses," *The American Journal of Human Genetics*, vol. 81, pp. 559-575, 2007.