World Wide Web
https://doi.org/10.1007/s11280-018-0637-3

# Group sparse reduced rank regression for neuroimaging genetic study

Xiaofeng Zhu[1,2,4] · Heung-Il Suk[3] · Dinggang Shen[3,4]

Received: 9 March 2018 / Revised: 19 July 2018 / Accepted: 7 September 2018 /
Published online: 17 September 2018

## Abstract

The neuroimaging genetic study usually needs to deal with high dimensionality of both brain imaging data and genetic data, so that often resulting in the issue of curse of dimensionality. In this paper, we propose a group sparse reduced rank regression model to take the relations of both the phenotypes and the genotypes for the neuroimaging genetic study. Specifically, we propose designing a graph sparsity constraint as well as a reduced rank constraint to simultaneously conduct subspace learning and feature selection. The group sparsity constraint conducts feature selection to identify genotypes highly related to neuroimaging data, while the reduced rank constraint considers the relations among neuroimaging data to conduct subspace learning in the feature selection model. Furthermore, an alternative optimization algorithm is proposed to solve the resulting objective function and is proved to achieve fast convergence. Experimental results on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset showed that the proposed method has superiority on predicting the phenotype data by the genotype data, than the alternative methods under comparison.

**Keywords** Reduced rank regression · Subspace learning · Feature selection · Neuroimaging genetic study

This article belongs to the Topical Collection: *Special Issue on Deep vs. Shallow: Learning for Emerging Web-scale Data Computing and Applications*
Guest Editors: Jingkuan Song, Shuqiang Jiang, Elisa Ricci, and Zi Huang

✉ Dinggang Shen
dgshen@med.unc.edu

[1] Guangxi Key Lab of Multi-source Information Mining and Security, Guangxi Normal University, Guilin 541004, Guangxi, People's Republic of China

[2] Institute of Natural and Mathematical Sciences, Massey University, Auckland 0745, New Zealand

[3] Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea

[4] BRIC Center of the University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

## 1 Introduction

In the past decades, the genetic variants have been suspected to associate with development of early and late-onset Alzheimer's Disease (AD). For example, the APOE$\epsilon$4 allele has been observed as the major carrier of cholesterol in the central nervous system compared to other APOE isoform carriers. Specifically, individuals with one or two copies of APOE$\epsilon$4 are more likely to develop late-onset AD [41]. Meanwhile, neuroimaging phenotypes have also been used for studies of genetic variants [25, 31]. Hence, neuroimaging genetic study has become an emergent cross-disciplinary field, where genetic information such as Single Nucleotide Polymorphism (SNP) is combined with neuroimaging data such as structural Magnetic Resonance Imaging (MRI), to analyze both biological and neurobiological systems of the human brain to help with prevention and treatment of AD [31].

Recent interest in neuroimaging genetic study is focused on association studies between phenotypes and genotypes. The motivation of the previous studies is that the genetic variants reflect the variability of phenotypes, while phenotypes may increase the power to detect causal variants of genotypes. For example, Brun et al. [5] proposed to select a subset of neuroimaging features by conducting association studies between neuroimaging features of a whole brain and a small number of genetic information, while the studies in [17, 38, 51] focused on selecting a subset of SNPs to conduct association analysis between a limited number of neuroimaging features and all SNPs.

Vounou et al. categorized the existing association studies between phenotypes and genotypes into four classes [36]: 1) Candidate Phenotype-Candidate Gene Association (CP-CGA), *e.g.,* between a brain surrogate and the MECP2 gene [22]; 2) Candidate Phenotype-Genome-Wide Association (CP-GWA), *e.g.,* between SNPs and a disease-related hippocampal MRI-driven measure [17]; 3) Brain-Wide and Candidate-Gene Association (BW-CGA), *e.g.,* between the gray matter volume in the entire brain and the APOE$\epsilon$4 allele [11]; and 4) Brain-Wide and Genome-Wide Association (BW-GWA), *e.g.,* between voxel-based neuroimaging phenotypes and SNP genotypes [34, 36]. The differences between the four classes lie in the way that they define the number of phenotypes and genotypes under consideration, *i.e.,* which one is available on either predefined candidates or the whole phenotypes and genotypes. The BW-GWA paradigm is a generation version of other three paradigms. More importantly, the BW-GWA paradigm has the potential benefit of helping to discover important associations between neuroimaging based phenotypic markers and genotypes from a different perspective. For example, by identifying high associations between specific SNPs and brain regions related to AD, information of the specific SNPs can be used to predict the risk of AD much earlier, and even before pathological abnormalities onset. This allows clinicians the time to track the course of AD and find solutions to prevent further degeneration of brain regions.

In real applications, a few literature have been designed to conduct the BW-GWA study, *i.e.,* conducting neuroimaging genetic study using all 620, 901 SNPs in the ADNI dataset. Usually, most studies conducted a process of SNP filtering to remove rare genetic variants or variants violating the Hardy-Weinberg Principle, thus resulting in a subset of all the SNPs to conduct neuroimaging genetic study, *i.e.,* the BW-CGA study. For example, the number of the selected SNPs is 437,577 [36], 437,607 [18], 448,293 [34], 501,584 [20], and 15,788 [2] out of 620,901 SNPs in these different studies. The reasons of using a subset of SNPs rather than all SNPs include the high computational cost of the algorithm, the effectiveness of neuroimaging genetic study, *etc.* More specifically, in the Region-Of-Interest (ROI) based neuroimaging genetic study, hundreds of the neuroimaging features can easily result in the lack of the ranks of the coefficient matrix to output ineffective performance of neuroimaging genetic study and expensive computation cost. On the other hand, removing irrelevant/redundant SNPs in the BW-CGA study can always output more stable and effective performance of neuroimaging genetic study, compared to the BW-GWA study [17, 20]. Therefore, in this work, we focus on the BW-CGA study, *i.e.,* the neuroimaging genetic study between ROI-based neuroimaging features and a subset of SNPs.

Inspired by recent advancements in neuroinformatics and bioinformatics, machine-learning techniques have been used for imaging-genetic association studies [51]. However, the high dimensionality of neuroimaging phenotypes and genotypes makes the BW-CGA study challenging. In addition, although the phenotypes and the genotypes have been indicated to have strong correlations, not all are equally informative. The BW-CGA study with all phenotypes and the selected genotypes may result in unreliable association models while no appropriate constraints. In this regard, only a few studies have focused on the BW-CGA problem [17, 34]. For example, pairwise univariate analysis (*e.g.,* Pearson correlation) treats the neuroimaging phenotypes and the genetic information as independent and isolated units without taking into account the interacting relationships among them. The earlier methods (*e.g.,* [1, 4, 18]) used Multi-output Linear Regression (MLR) methods for BW-CGA by estimating the coefficients of the response variables independently. Recent studies in [34, 36] exploited dimensionality reduction techniques for modeling and interpreting associations between phenotypes and genotypes, which limits their power in revealing and interpreting complex imaging-genetic associations. In a nutshell, previous studies mostly consider only inter-relations between genotypes and phenotypes, by ignoring potential informative intra-relations.

In this paper, we formulate the BW-CGA study as a regression problem by regarding the genotypes and the phenotypes as regressors and responses, respectively. By finding optimal weight coefficients in a regularized linear regression model, our proposed method may 1) discover the inherent relations in the phenotypes and the genotypes, which are interpretable with the linear feature selection model; and 2) predict phenotypes (*e.g.,* MRI volumes in our work) from a new genotype sample (*e.g.,* SNPs in our work), based on which potential risk of an incidence of a certain disease, such as AD, may be identified. Specifically, we propose a novel sparse regression model to find matrices that transform variables into subspaces by introducing a reduced rank constraint on a weight coefficient matrix. In the resulting subspaces, it is easier to understand relations among variables. The rationale of the reduced rank constraint is that the high-dimensional data have reduced rank structures due to noise and redundancy inherent in data [13, 19, 36, 50]. We also apply a group sparsity constraint (*i.e.,* an $\ell_{2,1}$-norm regularizer [15, 33, 39, 43, 48, 49]) on each reduced rank matrix, such that highly informative phenotypes and genotypes are selected for the BW-CGA study [38]. The joint use of the reduced rank constraint and the group sparsity in our linear regression model helps select a subset of brain regions and a subset of genotypes, which show high associations in the end. Finally, we conducted experiments on the ADNI cohort to validate our method's effectiveness.

## 2 Method

### 2.1 Sparse reduced rank regression

By denoting $\mathbf{X} \in \mathbb{R}^{n\times d}$ and $\mathbf{Y} \in \mathbb{R}^{n\times c}$, respectively, as the SNP genotype data and MRI phenotype data, where $n$, $d$, and $c$, indicate the number of the samples, the SNP dimensionality, and the MRI dimensionality, respectively, we assume that there is linear relationships between genotypes and phenotypes as well as there is redundancy in the phenotypes, the group sparsity based multi-output linear regression [14, 54] is formulated as follows

$$\min_{\mathbf{W},\mathbf{b},r} \quad \|\mathbf{Y} - \mathbf{X}\mathbf{W} - \mathbf{e}\mathbf{b}^T\|_F^2 + \alpha\|\mathbf{W}\|_{2,1}, \tag{1}$$

where $\mathbf{W} \in \mathbb{R}^{d\times c}$, $\mathbf{b} \in \mathbb{R}^{c\times 1}$, and $\mathbf{e} \in \mathbb{R}^{n\times 1}$, respectively, are the coefficient matrix, the bias term, and the column vector with all ones. $\|\cdot\|_F^2$ and $\|\cdot\|_{2,1}$, respectively, indicate the Frobenius norm and $\ell_{2,1}$ norm [16, 42, 46, 50].

## 2.2 Group sparse reduced rank multi-output linear regression

Equation (1) conducts feature selection on genotypes (*i.e.,* $\mathbf{X}$) and has been used for removing the redundancy of genotypes. However, in multi-output linear regression, both the genotypes (*i.e.,* $\mathbf{X}$) and the phenotypes (*i.e.,* $\mathbf{Y}$) may have noise to add the real ranks of the feature matrix and the response matrix. Moreover, the phenotype $\mathbf{Y}$ may also contain redundancy.

To solve the first issue, we make the hypothesis of reduced rankness of the MRI phenotypes and the SNP genotypes, as shown in [36], to change the sparsity regression model in (1) to a sparse reduced rank regression model. Specifically, we assume $\mathbf{W} = \mathbf{B}\mathbf{A}^T$, where $\mathbf{B} \in \mathbb{R}^{d\times r}$, $\mathbf{A} \in \mathbb{R}^{c\times r}$, $r$ is minimal rank between $\mathbf{X}$ and $\mathbf{Y}$, and $rank(\mathbf{W}) \leq min(n, d, c)$, to have

$$\min_{\mathbf{A},\mathbf{B},\mathbf{b}} \|\mathbf{Y} - \mathbf{X}\mathbf{B}\mathbf{A}^T - \mathbf{e}\mathbf{b}^T\|_F^2 + \alpha\|\mathbf{W}\|_{2,1}, \quad \text{s.t.}, \mathbf{A}^T\mathbf{A} = \mathbf{I}. \tag{2}$$

Further, from the reduced rankness of $\mathbf{W}$ (or $\mathbf{B}\mathbf{A}^T$), we can derive the following inequalities:

$$rank(\mathbf{B}) = r \Rightarrow rank(\mathbf{X}\mathbf{B}) \leq r \tag{3}$$

$$rank(\mathbf{B}\mathbf{A}^T) = r \Rightarrow rank(\mathbf{X}\mathbf{B}\mathbf{A}^T) \leq r. \tag{4}$$

According to (3), we think that the reduced matrix $\mathbf{X}\mathbf{B} \in \mathbb{R}^{n\times r}$, which is then multiplied with $\mathbf{A}^T$ to represent the response variables in (2), has less than $r$ latent factors. The assumption of latent factors in either the phenotypes or the genotypes has been also considered in [8, 9, 50, 54] for achieving well-conditioned estimation. Geometrically, $\mathbf{B}$ has the effect of transforming $\mathbf{X}$ into $r$-dimensional space, and determining $\mathbf{B}$ can be considered as subspace learning by using the correlations among the features, *i.e.,* $d$ SNP genotypes as a group. In the meantime, (4) implies that the rank of the predicted matrix $\hat{\mathbf{Y}} \in \mathbb{R}^{n\times c}$ (*i.e.,* $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{B}\mathbf{A}^T - \mathbf{e}\mathbf{b}^T$) is less than $r$. That is, each $c$ columns of $\mathbf{Y}$ can be represented by a linear combination of at most $r$ latent response variables. This considers the correlations among the response variables to conduct subspace learning on $\mathbf{Y}$, based on which we use to predict $\hat{\mathbf{Y}}$. Therefore, the reduced rank constraint on the coefficient matrix has the effect of subspace learning on both the regressor matrix and the response matrix by considering intra-relations in genotypes and phenotypes, separately.

As a complex system, the brain regions of the human being usually are related to each other [17, 38]. Moreover, a number of literature have shown that redundant brain regions may affect the performance of neuroimaging analysis. By considering the above observations, we add a group sparsity constraint on $\mathbf{Y}$ (*i.e.,* ROIs) to have the our final objective function:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{b},r} \quad \|\mathbf{Y} - \mathbf{X}\mathbf{B}\mathbf{A}^T - \mathbf{e}\mathbf{b}^T\|_F^2 + \alpha\|\mathbf{B}\mathbf{A}^T\|_{2,1} + \beta\|\mathbf{A}\|_{2,1}, \quad \text{s.t.}, \mathbf{A}^T\mathbf{A} = \mathbf{I} \tag{5}$$

where $\alpha$ and $\beta$ are the tuning parameters and $\mathbf{I} \in \mathbb{R}^{r \times r}$ is an identity matrix. The $\ell_{2,1}$-norm regularizers on $\mathbf{BA}^T$ and $\mathbf{A}$, respectively, manifest to conduct regressor/response selection on $\mathbf{X}$ (*i.e.,* SNP genotypes) and $\mathbf{Y}$ (*i.e.,* MRI phenotypes). The orthogonality constraint on $\mathbf{A}$ encourages the finding of un-correlated vectors, which can also be regarded as a transformation matrix of subspace learning on $\mathbf{Y}$. Furthermore, the orthogonality constraint $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ implies that $\|\mathbf{BA}^T\|_{2,1}$ shares the same zero-rows of $\mathbf{X}$ with $\|\mathbf{B}\|_{2,1}$, due to the fact that $\|\mathbf{BA}^T\|_{2,1} = tr(\mathbf{AB}^T\mathbf{DBA}^T) = tr(\mathbf{B}^T\mathbf{DB}) = \|\mathbf{B}\|_{2,1}$ with a diagonal matrix $\mathbf{D}$ whose $i$-th diagonal element defined as $d_{jj} = \frac{1}{2\|\mathbf{B}^j\|_2^2}$, $j = 1, ..., d$. Hence, our final objective function is defined as

$$\min_{\mathbf{A},\mathbf{B},\mathbf{b},r} \|\mathbf{Y} - \mathbf{XBA}^T - \mathbf{eb}^T\|_F^2 + \alpha\|\mathbf{B}\|_{2,1} + \beta\|\mathbf{A}\|_{2,1}, \text{ s.t., } \mathbf{A}^T\mathbf{A} = \mathbf{I}. \tag{6}$$

Clearly, the $\ell_{2,1}$-norm regularizers on $\mathbf{B}$ and $\mathbf{A}$ penalize coefficients of $\mathbf{B}$ and $\mathbf{A}$ in a row-wise manner for joint selection or un-selection of the regressors and the response variables.

After optimizing (6), elaborated in Section 2.3, we conduct feature selection [24, 42–44, 53] by discarding the regressors (or the response variables) whose corresponding coefficients in $\mathbf{B}$ (or $\mathbf{A}$) are zeros in the rows. More specifically, according to (12), the sparse rows on $\mathbf{A}$ imply that their corresponding columns (*i.e.,* ROIs) of $\mathbf{Y}$ will not be selected, while the sparse rows on $\mathbf{B}$ imply that their corresponding features (*i.e.,* SNPs) of $\mathbf{X}$ will be excluded by the proposed model. This results in the selection of a subset of brain ROIs from $\mathbf{Y}$ that is statistically meaningful and associated with a subset of SNPs (*i.e.,* the selected SNPs) from $\mathbf{X}$. By means of our optimization method described below, the reduced rank constraint conducts subspace learning on both $\mathbf{X}$ and $\mathbf{Y}$, so that the sequential feature selection is conducted by avoiding noise in the data thus improving performance. In contrast, the group sparsity constraints ensure the reduced rank constraint to explore the reduced rank representations of data on the 'purified data', *i.e.,* the data after removing uninformative ROIs and SNPs by group sparsity constraints. These two steps alternate until the objective function converges. This iterative learning yields optimal results of both feature selection and subspace learning. That is, the selected ROIs are associated with the selected SNPs.

## 2.3 Optimization

This section describes the optimization process of the parameters $\mathbf{b}$, $\mathbf{B}$, and $\mathbf{A}$. Specifically, we iteratively conduct the following three steps until convergence by means of Iteratively Reweighted Least Square (IRLS) [32, 53]: (i) Update $\mathbf{b}$ with fixed $\mathbf{B}$ and $\mathbf{A}$. (ii) Update $\mathbf{B}$ with fixed $\mathbf{b}$ and $\mathbf{A}$. (iii) Update $\mathbf{A}$ with fixed $\mathbf{b}$ and $\mathbf{B}$.

### 2.3.1 (i) Update b with fixed B and A.

For fixed $\mathbf{B}$ and $\mathbf{A}$, (6) reduces to

$$\min_{\mathbf{b}} \|\mathbf{Y} - \mathbf{XBA}^T - \mathbf{eb}^T\|_F^2, \tag{7}$$

By setting the derivative of (7) with respect to $\mathbf{b}$ to zero, we have:

$$\mathbf{b} = \frac{1}{n}(\mathbf{Y}^T\mathbf{e} - \mathbf{AB}^T\mathbf{X}^T\mathbf{e}) \tag{8}$$

### 2.3.2 (ii) Update B with fixed b and A.

For fixed $\mathbf{b}$, we substitute (8) into (6) by yielding the following:

$$\min_{\mathbf{B},\mathbf{A}} \ \|\mathbf{Y} - \mathbf{X}\mathbf{B}\mathbf{A}^T - \tfrac{1}{n}\mathbf{e}(\mathbf{Y}^T\mathbf{e} - \mathbf{A}\mathbf{B}^T\mathbf{X}^T\mathbf{e})^T\|_F^2 \\ + \alpha\|\mathbf{B}\|_{2,1} + \beta\|\mathbf{A}\|_{2,1}, \ \text{s.t.,}\ \mathbf{A}^T\mathbf{A} = \mathbf{I}. \tag{9}$$

By introducing $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{e}\mathbf{e}^T \in \mathbb{R}^{n\times n}$, we can rewrite (9) as follows:

$$\min_{\mathbf{B},\mathbf{A}} \ \|\mathbf{H}\mathbf{Y} - \mathbf{H}\mathbf{X}\mathbf{B}\mathbf{A}^T\|_F^2 + \beta\|\mathbf{A}\|_{2,1}, +\alpha\|\mathbf{B}\|_{2,1} \ \text{s.t.,}\ \mathbf{A}^T\mathbf{A} = \mathbf{I}. \tag{10}$$

Since $\mathbf{A}$ is subject to having orthogonal columns, there is a matrix $\mathbf{A}^\perp$ with orthogonal columns such that $(\mathbf{A}, \mathbf{A}^\perp)$ is an orthogonal matrix. Thus, we have

$$\begin{aligned} \|\mathbf{H}\mathbf{Y} - \mathbf{H}\mathbf{X}\mathbf{B}\mathbf{A}^T\|_F^2 &= \|(\mathbf{H}\mathbf{Y} - \mathbf{H}\mathbf{X}\mathbf{B}\mathbf{A}^T)(\mathbf{A}, \mathbf{A}^\perp)\|_F^2 \\ &= \|\mathbf{H}\mathbf{Y}\mathbf{A} - \mathbf{H}\mathbf{X}\mathbf{B}\|_F^2 + \|\mathbf{H}\mathbf{Y}\mathbf{A}^\perp\|_F^2 \\ &= \|\mathbf{H}\mathbf{Y}\mathbf{A} - \mathbf{H}\mathbf{X}\mathbf{B}\|_F^2. \end{aligned} \tag{11}$$

The second term in (11) does not involve $\mathbf{B}$. For fixed $\mathbf{b}$ and $\mathbf{A}$, we substitute (11) into (10) and then obtain:

$$\min_{\mathbf{B}} \quad \|\mathbf{H}\mathbf{Y}\mathbf{A} - \mathbf{H}\mathbf{X}\mathbf{B}\|_F^2 + \alpha\|\mathbf{B}\|_{2,1} \tag{12}$$

By employing the framework of IRLS to optimize $\mathbf{B}$, we set the derivative of (12) with respect to $\mathbf{B}$ to zero and have:

$$\mathbf{B} = (\mathbf{X}^T\mathbf{H}\mathbf{X} + \alpha\mathbf{Q})^{-1}\mathbf{X}^T\mathbf{H}\mathbf{Y}\mathbf{A}. \tag{13}$$

where $\mathbf{Q} \in \mathbb{R}^{d\times d}$ is a diagonal matrix and its diagonal element $q_{jj} = \frac{1}{2\|\mathbf{B}^j\|_2^2}$, $j = 1, ..., d$.

### 2.3.3 (iii) Update A with fixed b and B

For fixed $\mathbf{b}$ and $\mathbf{B}$, (6) reduces to

$$\min_{\mathbf{A}} \|\mathbf{H}\mathbf{Y} - \mathbf{H}\mathbf{X}\mathbf{B}\mathbf{A}^T\|_F^2 + \beta\|\mathbf{A}\|_{2,1}, \ \text{s.t.,}\ \mathbf{A}^T\mathbf{A} = \mathbf{I}. \tag{14}$$

Based on the framework of IRLS again, we have:

$$\max_{\mathbf{A}} \ tr(\mathbf{A}^T(\mathbf{Y}^T\mathbf{H}\mathbf{X}(\mathbf{X}^T\mathbf{H}\mathbf{X} + \alpha\mathbf{Q})^{-1}\mathbf{X}^T\mathbf{H}\mathbf{Y} - \beta\mathbf{P})\mathbf{A}), \ \text{s.t.,}\ \mathbf{A}^T\mathbf{A} = \mathbf{I}, \tag{15}$$

where $\mathbf{P} \in \mathbb{R}^{c\times c}$ is a diagonal matrix and its diagonal element $p_{jj} = \frac{1}{2\|\mathbf{A}^j\|_2^2}$, $j = 1, ..., c$. (15) is a generalized eigenvalue problem, and its global optimal solution is obtained from the nonzero eigenvectors of $(\mathbf{Y}^T\mathbf{H}\mathbf{X}(\mathbf{X}^T\mathbf{H}\mathbf{X} + \alpha\mathbf{Q})^{-1}\mathbf{X}^T\mathbf{H}\mathbf{Y} - \beta\mathbf{P})$.

We provide the pseudo algorithm of solving (6) in Algorithm 1. According to [44, 50, 53], the objective function in (6) monotonically decreases after each iteration.

## 3 Experimental analysis

We conducted various experiments on the ADNI dataset ('www.adni-info.org') by comparing our method with the state-of-the-art methods.

### 3.1 Data preprocessing

Both SNP and MRI data used in this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). Please refer to 'www.adni-info.org' for up-to-date information. By following earlier studies [30, 38], in this work, we used samples of 737 non-Hispanic Caucasian participants, including 171 AD, 362 MCI, and 204 healthy Normal Control (NC), who were also genotyped by ADNI.

We downloaded raw Digital Imaging and Communications in Medicine (DICOM) MRI scans from the public ADNI website, and we conducted the image processing of MR images following the same procedures in [45, 52]. Specifically, the MRI scans were processed using a standard protocol, including spatial distortion correction and bias field correction, followed by skull-stripping, cerebellum removal, intensity inhomogeneity correction, segmentation, and registration. Based on the Jacob template [23], we finally obtained gray matter volume measures of 93 cortical and subcortical regions for each MRI scan to characterize its anatomy [45].

**Algorithm 1** Pseudo code of solving (6).

```
Input: $\mathbf{X} \in \mathbb{R}^{n \times d}$, $\mathbf{Y} \in \mathbb{R}^{n \times c}$, $\alpha$, $\beta$;
Output: $\mathbf{b}$, $\mathbf{B}$, $\mathbf{A}$;
Initialize $t = 1$;
Initialize $\mathbf{b}(t)$ and $\mathbf{B}(t)$ as two random vectors;
/* $\mathbf{b}(t)$: the $t$-th iteration result of $\mathbf{b}$. */;
repeat
    Update $\mathbf{b}(t+1)$ via (8);
    Update $\mathbf{B}(t+1)$ via (13);
    Update $\mathbf{Q}(t+1)$ via $q_{jj} = \frac{1}{2\|\mathbf{B}(t+1)^j\|_2^2}$, $j = 1, ..., d$;
    Update $\mathbf{A}(t+1)$ via (15);
    Update $\mathbf{P}(t+1)$ via $p_{jj} = \frac{1}{2\|\mathbf{A}(t+1)^j\|_2^2}$, $j = 1, ..., c$;
    $t = t+1$;
until The difference between the objective function values of (6) within two sequential
    iterations less than $10^{-5}$;
```

We obtained the genotype data of all non-Hispanic Caucasian participants from the ADNI Phase 1 cohort. ADNI genotyping was performed using the Human610-Quad BeadChip, which includes 620,901 SNPs and copy number variations [40]. The SNP of the APOE$\epsilon$4 variant has been separately genotyped by ADNI, but is not included in the original genotyping chip. In this work, the SNP was added to the final genotype dataset. All subjects were unrelated and further detail of genotypes can be found in [34]. Each of the MRI scans had corresponding genetic data obtained from the ADNI Phase 1 cohort, consisting of 620, 901 SNPs. The SNPs were processed by two steps, *i.e.,* the quality control step and the imputation step [3]. The quality control step included 1) call rate check per subject and per SNP marker; 2) gender check; 3) sibling pair identification; 4) the Hardy-Weinberg equilibrium test; 5) marker removal by the minor allele frequency; and 6) population stratification. The imputation step imputed the incomplete SNPs with the modal value. Finally, we obtained 3996 SNPs, within the boundary of 20K base pairs of the 153 Alzheimer's disease (AD) candidate genes listed on the AlzGene database (http://www.alzgene.org/) as of 4/18/2011. Finally, we obtained 2098 SNPs from 153 genes (boundary: 20KB) using the ANNOVAR annotation.[1]

### 3.2 Competing methods

In order to validate the effectiveness of the proposed method, we compared our method to the standard regularized Multi-output Linear Regression (MLR) [21], sparse feature selection with an $\ell_{2,1}$-norm regularizer (L21 for short) [10], Group sparse Feature Selection (GFS) [38], sparse Canonical Correlation Analysis (CCA) [25], and sparse Reduced-Rank Regression (RRR) [36]. The former two are the most widely used methods in both statistical learning and medical image analysis, while the last three are state-of-the-art methods in neuroimaging genetic study. We listed the details of these competing methods as follows:

- MLR consideres the correlations among the features (*i.e.,* SNPs) but independently considers each of the response variables to conduct the BW-CGA study.
- L21 employs a least square loss function in combination with a group sparse regularizer (*i.e.,* an $\ell_{2,1}$-norm regularizer) to consider the correlations among the features.
- GFS considers the inter-linked relationship among the genotypes (*i.e.,* the features) without taking correlations among the response variables into account.
- CCA conducts feature selection on the response matrix as well as the feature matrix, but does not conduct subspace learning.
- RRR conducts subspace learning on both the neuroimaging phenotypes and the genotypes. However, RRR does not explicitly conduct feature selection on the data.
- Baseline is a special case of our proposed method. Specifically, Baseline removes the third term (*i.e.,* $\beta\|\mathbf{A}\|_{2,1}$) of (6) to only conduct SNP selection. In this way, Baseline does not conduct feature selection on genotype data, thus it may be affected by the irrelevant/redundant SNPs for the BW-CGA study.

### 3.3 Experimental setup

We employed the three-fold cross-validation scheme to evaluate all the methods. Specifically, we partitioned the whole dataset into 3 subsets, where one subset was set as the testing dataset and the left two subsets were set as the training set. Given the training set, we conducted five-fold nested cross-validation to conduct the model selection, which outputted the parameters' combination with the best results of RMSE for the testing datasets. We repeated the whole process of every method ten times, and reported the averaged results of all results within ten times. In model selection, we tuned the parameters of all the methods with the range of $\{10^{-5}, ..., 10^{5}\}$, and further varied the rank number $r$ in $\{2, 4, ..., 20\}$ for our proposed method. Furthermore, we followed the literatures [17, 38] to select the top $\{20, 40, 60, ..., 180, 200\}$ genotypes to predict the phenotypes in our experiments.

We used two evaluation metrics, *i.e.,* Root-Mean-Square Error (RMSE) and 'Frequency' defined as the freqency of the genotypes (or the phenotypes) selected in all the experiments. Usually, the range of 'Frequency' is from 0 to 1.

[1] http://www.openbioinformatics.org/annovar/.

### 3.4 Results

Figure 1 presents the RMSE performance of all the methods considered in this work, with the mean and standard deviation obtained from all the experiments. From Figure 1, we have the following conclusions.

- All the methods reduced their RMSE results with the increase of the number of the selected genotypes (*i.e.,* SNPs), indicating that the more the genotypes were used, the better the performance for predicting the phenotypes was, with at most top 200 SNPs to be involved.
- Our method achieved the best RMSE results, followed by Baseline, RRR, GFS, CCA, L21, and MLR. More specifically, our method on average increased 12.75%, compared to all the competing methods. Moreover, paired-sample $t$-test ($p < 0.05$) showed that the $p$-values between our method and each of the competing methods were less than 0.00001. This demonstrated that our method has statistically significant improvements, than all the competing methods. Furthermore, the stability of our method is the best, showing that our method has superiority on combing a reduced rank constraint with a group sparsity constraint in a framework.
- Baseline on average increased 8.26%, than other competing methods. Moreover, paired-sample $t$-test (at 95% confidence level) showed that the $p$-values between our method and each of the competing methods were less than 0.001. Hence, Baseline (*i.e.,* our proposed method without conducting ROIs selection) is still better other competing methods, indicating that simultaneously selecting a subset of genotypes and a subset of phenotypes makes sense for conducting neuroimaging genetic study.

We presented the ‘Frequency’ value of the 10 selected SNPs and the 10 selected ROIs by the competing methods along with our method in Figure 2 and also visualized the coefficients of the 10 selected SNPs and the 10 selected ROIs in Figure 3. The left sub-figures in Figures 2 and 3 indicate that phenotypes can be affected by different degrees based

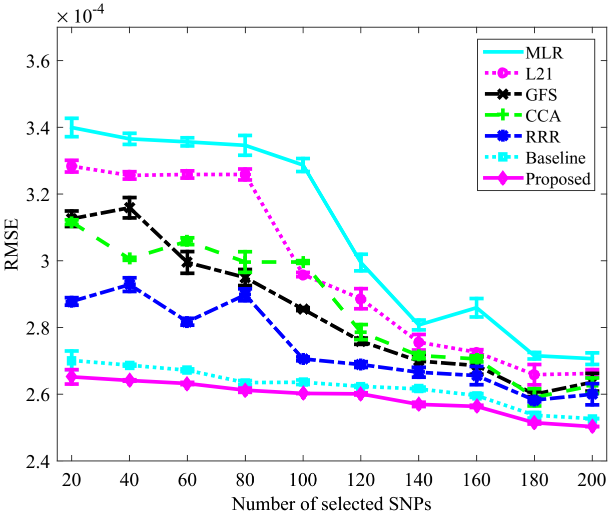

**Figure 1** The RMSE with respect to different number of selected SNPs of all methods

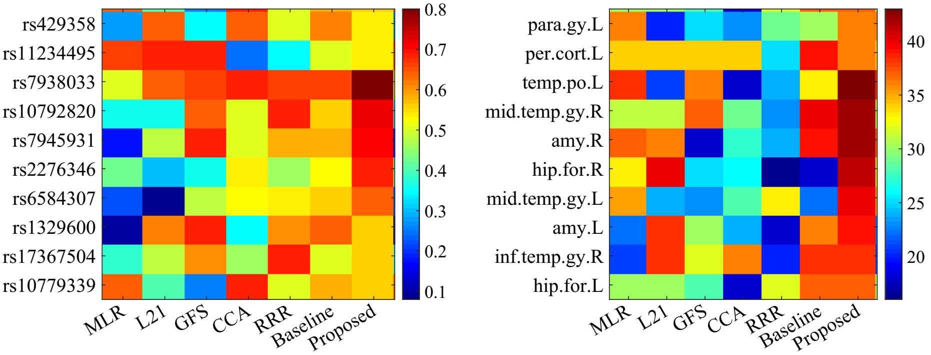

**Figure 2** ‘Frequency’ of the top 10 selected SNPs (left) and ROIs (right) by the proposed method

on genotypes: (i) The selected SNPs, by the proposed method, were the genes PICALM, APOE, SORL1, ENTPD7, DAPK1, MTHFR, and CR1, which have been reported as top AD candidate genes on the AlzGene website.[2] (ii) Although we know little about the underlying mechanisms of genotypes in relation to AD, the left sub-figures in Figures 2 and 3 offer biological insight from the BW-CGA study.

- PICLAM is a new A$\beta$ toxicity modifier of genes and has been demonstrated to be significantly associated with risk of late-onset AD [29]. Here, our experiments verify that the gene PICALM has biological relations to phenotypes. For example, our method selected SNPs from the PICLAM gene that were the top 10 SNPs, such as ‘rs7938033’, ‘rs11234495’, and ‘rs10792820’, which have been reported to be related to inheritable neuro-developmental disorders [41].
- The APOE-$\epsilon$4 variant of the APOE gene has been reported to be responsible for the production of apolipoprotein E [41]. In our experiments, all methods selected the SNP of ‘rs429358’ as one of the top significant SNPs and our method indicated its strongest association with phenotypes.
- SNPs of ‘rs7945931’ and ‘rs2276346’ have been shown to have significant effects on the temporal cortex of the gene SORL1, which influences clinical manifestation of AD and is genetically associated with increased risk for late-onset AD [26, 27].

The right sub-figures in Figures 2 and 3 present that the top 10 selected ROIs by our proposed method are parahippocampal gyrus left (para.gy.L), perirhinal cortex left (per.cort.L), temporal pole left (temp.po.L), middle temporal gyrus right (mid.temp.gy.R), amygdala right (amy.R), hippocampal formation right (hip.for.R), middle temporal gyrus left (mid.temp.gy.L), amygdala left (amy.L), inferior temporal gyrus right (inf.temp.gy.R), and hippocampal formation left (hip.for.L). These selected ROIs were known to be highly related to AD or related dementia (*e.g.,* MCI) in previous studies, such as in neuroimaging genetic study [17, 38], AD classification and regression [35, 45, 47], and clinical diagnosis [6, 7, 12, 28]. Hence, the ROIs selected by our method could be further incorporated for future clinical analysis.

[2] http://www.alzgene.org/

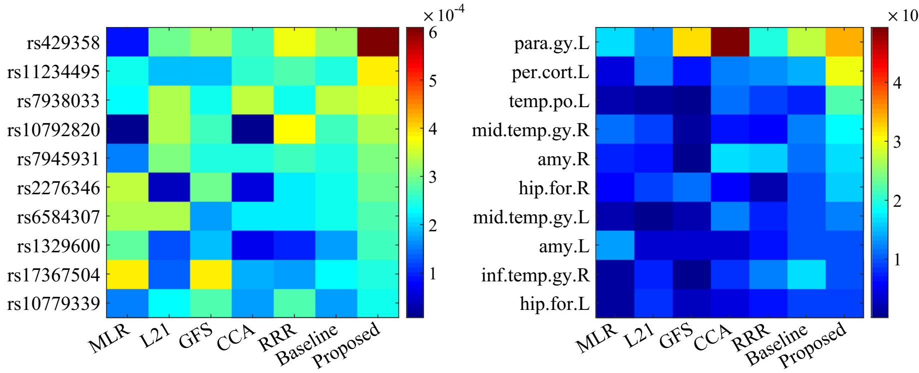

**Figure 3** The coefficients of the top 10 selected SNPs (left) and ROIs (right) by the proposed method

We then interpreted the details on how top selected SNPs (or ROIs) affected the BW-CGA study by reporting a subset of the coefficient matrix of $\mathbf{BA}^T \in \mathbb{R}^{2098\times 93}$ in (6). Specifically, we averaged the absolute value of $\mathbf{BA}^T$ from all 50 experiments to sort the resulting matrix in a descending order along the rows (or the columns) to obtain the top 10 SNPs (or ROIs). The resulting coefficients, whose rows and columns, respectively, correspond to the top 10 SNPs and the top 10 ROIs in the resulting matrix, are illustrated in Figure 4 to explain the association between the selected SNPs and the selected ROIs. Figure 5 illustrates the top 20 selected ROIs associated with each of the selected SNPs obtained by our proposed method. Figures 4 and 5 manifest that the selected set of SNPs and ROIs are related to AD, which is in accordance with previous state-of-the-art methods [17, 37, 38].

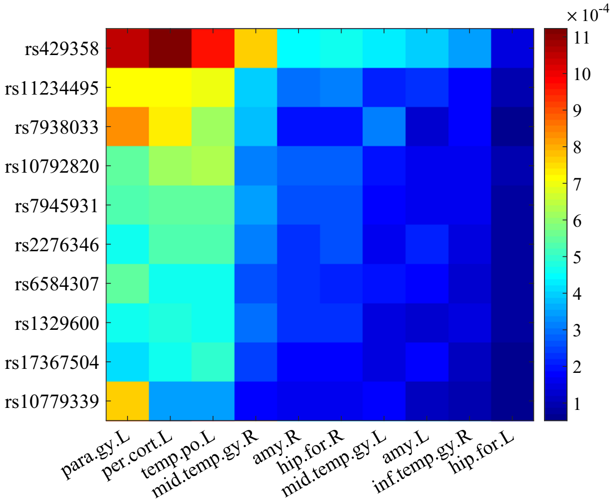

**Figure 4** The relationship between the top 10 ROIs and the top 10 SNPs, selected by the proposed method, in terms of the absolute value of $\mathbf{BA}^T$

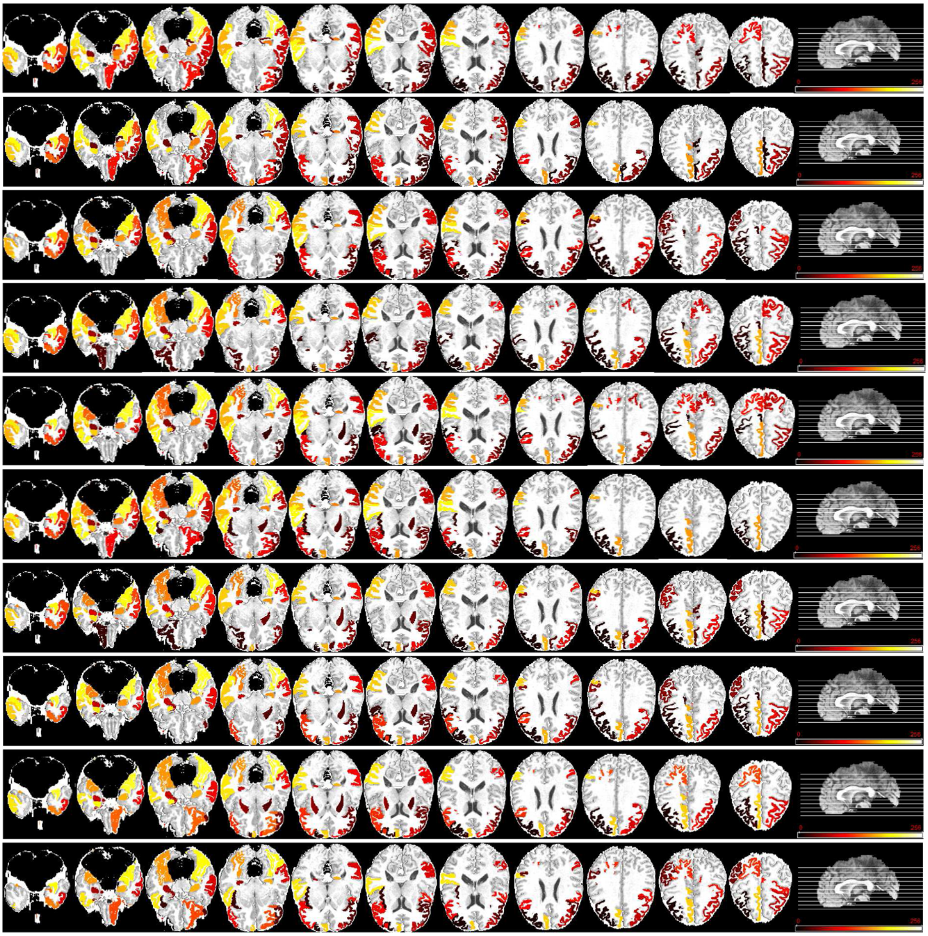

**Figure 5** The top 20 ROIs selected by each of the top 10 SNPs (corresponding to 10 rows) with the proposed method. The names of SNPs and the corresponding name of genes are 'APOE (rs429358)', 'PICALM (rs11234495)', 'PICALM (rs7938033)', 'PICALM (rs10792820)', 'SORL1 (rs7945931)', 'SORL1 (rs2276346)', 'ENTPD7 (rs6584307)', 'DAPK1 (rs1329600)', 'MTHFR (rs17367504)', and 'CR1 (rs10779339)', respectively, from top to bottom

### 3.5 Effects of the matrix rank *r*

We investigated the effect of different numbers of rank $r \in \{2, ..., 20\}$ in our proposed method by reporting the change of RMSE values in Figure 6, where the mean and standard deviation of the RMSE were obtained from all 50 experiments. In the figure, each curve represents the change of RMSE with a fixed number of SNPs in predicting the test data, *e.g.,* 'top-200' represents the change of RMSE using top 200 SNPs in predicting the test ROIs.

From Figure 6, we observed that the best performance of cases with different numbers of SNPs, in predicting test data, was between 8 and 12, which empirically justifies imposing a reduced rank assumption on both neuroimaging phenotypes and genotypes. The reduced rank constraints conducting subspace learning helped find low-dimensional

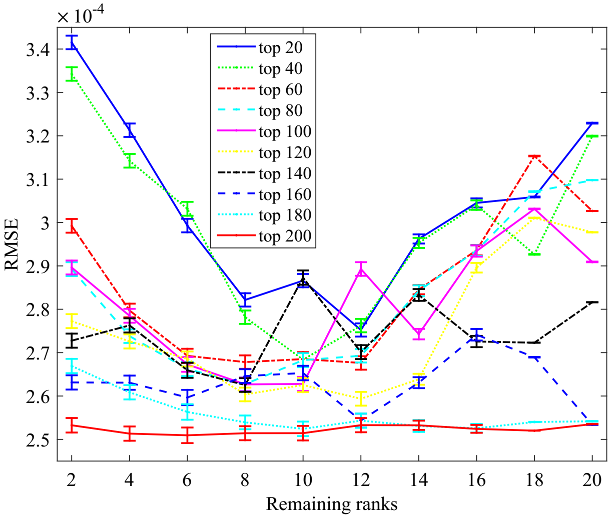

**Figure 6** The RMSE of the proposed method with different numbers of ranks using different numbers of SNPs to predict the test data

structure of high-dimensional neuroimaging data via relational considerations among the response variables.

## 4 Conclusion

We have designed a new group sparse reduced rank regression method to select highly associated phenotypes and genotypes for conducting the BW-CGA study. Experimental results on the ADNI dataset demonstrated that our proposed method outperformed all the competing methods.

Despite superior performance of our proposed method over the competing methods, there are still existing limitations, which inspire us to further improve our method for the BW-CGA study in the future work. First, each of the subjects in ADNI has label information, which offers high-level representations of subjects and should thus be informative for improving performance of the BW-CGA study. Hence, we may use this available information to more effectively explore associations between SNPs and ROIs in the future. Second, our proposed method does not consider any natural structures of SNPs (or ROIs). However, SNPs are naturally connected via different pathways, while ROIs have various functional or structural relations to each other [25, 31]. It would be interesting to extend our method to take inter-linked structures, within both SNPs and ROIs, into account for further improving performance of the BW-CGA study in our future work.

**Acknowledgments** This work was supported in part by NIH grants (EB008374, AG041721, AG049371, AG042599, EB022880). X. Zhu was also supported by the National Natural Science Foundation of China (Grants No: 61573270 and 61876046); the Project of Guangxi Science and Technology (GuiKeAD17195062); and the Guangxi Collaborative Innovation Center of Multi-Source Information Integration and Intelligent Processing. H.I. Suk was also supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (No. 2017-0-00451).

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

1. Ballard, D.H., Cho, J., Zhao, H.: Comparisons of multi-marker association methods to detect association between a candiyear region and disease. Genet. Epidemiol. **34**(3), 201–212 (2010)
2. Batmanghelich, N.K., Dalca, A., Quon, G., Sabuncu, M., Golland, P.: Probabilistic modeling of imaging, genetics and diagnosis. IEEE Trans. Med. Imaging **35**(7), 1765–1779 (2016)
3. Bertram, L., McQueen, M.B., Mullin, K., Blacker, D., Tanzi, R.E.: Systematic meta-analyses of Alzheimer disease genetic association studies: the Alzgene database. Nat. Genet. **39**(1), 17–23 (2007)
4. Bralten, J., Arias-Vásquez, A., Makkinje, R., Veltman, J.A., Brunner, H.G., Fernández, G., Rijpkema, M., Franke, B.: Association of the Alzheimer's gene SORL1 with hippocampal volume in young, healthy adults. American Journal of Psychiatry (2011)
5. Brun, C.C., Leporé, N., Pennec, X., Lee, A.D., Barysheva, M., Madsen, S.K., Avedissian, C., Chou, Y.-Y., Zubicaray, G.I.D., McMahon, K.L., et al.: Mapping the regional influence of genetics on brain structure variabilitya tensor-based morphometry study. Neuroimage **48**(1), 37–49 (2009)
6. Chételat, G., Eustache, F., Viader, F., De La Sayette, V., Pélerin, A., Mézenge, F., Hannequin, D., Dupuy, B., Baron, J.-C., Desgranges, B.: FDG-PET Measurement is more accurate than neuropsychological assessments to predict global cognitive deterioration in patients with mild cognitive impairment. Neurocase **11**(1), 14–25 (2005)
7. Convit, A., De Asis, J., De Leon, M.J., Tarshish, C.Y., De Santi, S., Rusinek, H.: Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to alzheimers disease. Neurobiol. Aging **21**(1), 19–26 (2000)
8. Deng, X., Li, Y., Weng, J., Zhang, J.: Feature selection for text classification: a review. Multimedia Tools Appl., pp. 1–20 (2018)
9. Du, L., Yan, J., Kim, S., et al.: A novel structure-aware sparse learning algorithm for brain imaging genetics. In: MICCAI, pp. 329–336 (2014)
10. Evgeniou, A., Pontil, M.: Multi-task feature learning. NIPS **19**, 41–48 (2007)
11. Filippini, N., Rao, A., Wetten, S., Gibson, R.A., Borrie, M., Guzman, D., Kertesz, A., Loy-English, I., Williams, J., Nichols, T., et al.: Anatomically-distinct genetic associations of APOE$\epsilon$4 allele load with regional cortical atrophy in Alzheimer's disease. Neuroimage **44**(3), 724–728 (2009)
12. Fox, N.C., Schott, J.M.: Imaging cerebral atrophy: normal ageing to Alzheimer's disease. The Lancet **363**(9406), 392–394 (2004)
13. Gao, L., Guo, Z., Zhang, H., Xu, X., Shen, H.T.: Video captioning with attention-based lstm and semantic consistency. IEEE Trans. Multimedia **19**(9), 2045–2055 (2017)
14. Gao, L., Song, J., Liu, X., Shao, J., Liu, J., Shao, J.: Learning in high-dimensional multimedia data: the state of the art. Multimedia Syst. **23**(3), 303–313 (2017)
15. Guo, Y., Wu, G., Jiang, J., Shen, D.: Robust anatomical correspondence detection by hierarchical sparse graph matching. IEEE Trans. Med. Imaging **32**(2), 268–277 (2013)
16. Guo, Y., Gao, Y., Shen, D.: Deformable mr prostate segmentation via deep feature learning and sparse patch matching. In: Deep Learning for Medical Image Analysis, 197–222 (2017)
17. Hao, X., Yu, J., Zhang, D.: Identifying genetic associations with MRI-derived measures via tree-guided sparse learning. In: MICCAI 2014, pp. 757–764 (2014)
18. Hibar, D.P., Stein, J.L., Kohannim, O., Jahanshad, N., Saykin, A.J., Shen, L., Kim, S., Pankratz, N., Foroud, T., Huentelman, M.J., et al.: Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. Neuroimage **56**(4), 1875–1891 (2011)
19. Hu, R., Zhu, X., Cheng, D., He, W., Yan, Y., Song, J., Zhang, S.: Graph self-representation method for unsupervised feature selection. Neurocomputing **220**, 130–137 (2017)
20. Huang, M., Nichols, T., Huang, C., Yu, Y., Lu, Z., Knickmeyer, R.C., Feng, Q., Zhu, H.: Alzheimer's Disease Neuroimaging Initiative, et al. Fvgwas: fast voxelwise genome wide association analysis of large-scale imaging genetic data. Neuroimage **118**, 613–627 (2015)
21. Izenman, A.J.: Reduced-rank regression for the multivariate linear model. J. Multivar. Anal. **5**(2), 248–264 (1975)
22. Joyner, A.H., Bloss, C.S., Bakken, T.E., Rimol, L.M., Melle, I., Agartz, I., Djurovic, S., Topol, E.J., Schork, N.J., Andreassen, O.A., et al.: A common mecp2 haplotype associates with reduced cortical surface area in humans in two independent populations. Proc. Natl. Acad. Sci. **106**(36), 15483–15488 (2009)

23. Kabani, N.J.: 3D anatomical atlas of the human brain. In: Human Brain Mapping (1998)
24. Lei, C., Zhu, X.: Unsupervised feature selection via local structure learning and sparse learning. pages. https://doi.org/10.1007/s11042–017–5381–7, 11 (2017)
25. Lin, D., Cao, H., Calhoun, V.D., Wang, Y.-P.: Sparse models for correlative and integrative analysis of imaging and genetic data. J. Neurosci. Methods **237**, 69–78 (2014)
26. Louwersheimer, E., Ramirez, A., Cruchaga, C., Becker, T., Kornhuber, J., Peters, O., Heilmann, S., Wiltfang, J., Jessen, F., Visser, P.J., et al.: The influence of genetic variants in SORL1 gene on the manifestation of Alzheimer's disease. Neurobiol. Aging **36**(3), 1605–e13 (2015)
27. McCarthy, J.J., Saith, S., Linnertz, C., Burke, J.R., Hulette, C.M., Welsh-Bohmer, K.A., Chiba-Falek, O.: The Alzheimer's associated 5' region of the SORL1 gene cis regulates SORL1 transcripts expression. Neurobiol. Aging **33**(7), 1485–e1 (2012)
28. Misra, C., Fan, Y., Davatzikos, C.: Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI. Neuroimage **44**(4), 1415–1422 (2009)
29. Rosenthal, S.L., Wang, X., et al.: Beta-amyloid toxicity modifier genes and the risk of alzheimers disease. Am. J. Neurodegener. Dis. **1**(2), 191–198 (2012)
30. Shen, L., Kim, S., Risacher, S.L., Nho, K., Swaminathan, S., West, J.D., Foroud, T., Pankratz, N., Moore, J.H., Sloan, C.D., 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 **53**(3), 1051–1063 (2010)
31. Shen, L., Thompson, P.M., Potkin, S.G., et al.: Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. Brain Imaging Behav. **8**(2), 183–207 (2014)
32. Shen, F., Xu, Y., Liu, L., Yang, Y., Huang, Z., Shen, H.T.: Unsupervised deep hashing with similarity-adaptive and discrete optimization (2018)
33. Song, J., Gao, L., Li, L., Zhu, X., Sebe, N.: Quantization-based hashing: a general framework for scalable image and video retrieval. Pattern Recogn. **75**, 175–187 (2018)
34. Stein, J.L., Hua, X., Lee, S., Ho, A.J., Leow, A.D., Toga, A.W., Saykin, A.J., Shen, L., Foroud, T., Pankratz, N., et al.: Voxelwise genome-wide association study (vGWAS). Neuroimage **53**(3), 1160–1174 (2010)
35. Thung, K.-H., Wee, C.-Y., Yap, P.-T., Shen, D.: Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion. Neuroimage **91**, 386–400 (2014)
36. Vounou, M., Nichols, T.E., Montana, G.: ADNI discovering genetic associations with high-dimensional neuroimaging phenotypes: a sparse reduced-rank regression approach. Neuroimage **53**(3), 1147–1159 (2010)
37. Wang, H., Nie, F., Huang, H., et al.: From phenotype to genotype: an association study of longitudinal phenotypic markers to Alzheimer's disease relevant snps. Bioinformatics **28**(18), i619–i625 (2012)
38. Wang, H., Nie, F., Huang, H., et al.: Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. Bioinformatics **28**(2), 229–237 (2012)
39. Wang, X., Gao, L., Wang, P., Sun, X., Liu, X.: Two-stream 3d convnet fusion for action recognition in videos with arbitrary size and length. IEEE Transactions on Multimedia (2017)
40. Weiner, M.W., Aisen, P.S., Jack, C.R., Jagust, W.J., Trojanowski, J.Q., Shaw, L., Saykin, A.J., Morris, J.C., Cairns, N., Beckett, L.A., et al.: The Alzheimer's disease neuroimaging initiative: progress report and future plans. Alzheimer's & Dementia **6**(3), 202–211 (2010)
41. Xia, K., Guo, H., Hu, Z., et al.: Common genetic variants on 1p13. 2 associate with risk of autism. Mol. Psychiatry **19**(11), 1212–1219 (2014)
42. Zhang, S., Li, X., Zong, M., Zhu, X., Wang, R.: Efficient knn classification with different numbers of nearest neighbors. IEEE Trans. Neural Netw. Learn. Syst. **29**(5), 1774–1785 (2018)
43. Zheng, W., Zhu, X., Zhu, Y., Hu, R., Lei, C.: Dynamic graph learning for spectral feature selection. Multimedia Tools and Applications, pages. https://doi.org/10.1007/s11042–017–5272–y (2017)
44. Zheng, W., Zhu, X., Wen, G., Zhu, Y., Yu, H., Gan, J.: Unsupervised feature selection by self-paced learning regularization. Pattern Recognition Letters, page. https://doi.org/10.1016/j.patrec.2018.06.029 (2018)
45. Zhu, X., Suk, H.-I., Shen, D.: A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. Neuroimage **100**, 91–105 (2014)
46. Zhu, X., Zhang, L., Zi, H.: A sparse embedding and least variance encoding approach to hashing. IEEE Trans. Image Process. **23**(9), 3737–3750 (2014)
47. Zhu, X., Suk, H.-I., Lee, S.-W., Shen, D.: Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification. IEEE Trans. Biomed. Eng. **63**(3), 607–618 (2016)
48. Zhu, X., Li, X., Zhang, S.: Block-row sparse multiview multilabel learning for image classification. IEEE Trans. Cybernet. **46**(2), 450–461 (2016)

49. Zhu, X., Li, X., Zhang, S., Ju, C., Wu, X.: Robust joint graph sparse coding for unsupervised spectral feature selection. IEEE Trans. Neural Netw. Learn. Syst. **28**(6), 1263–1275 (2017)
50. Zhu, X., Li, X., Zhang, S., Xu, Z., Yu, L., Wang, C.: Graph pca hashing for similarity search. IEEE Trans. Multimedia **19**(9), 2033–2044 (2017)
51. Zhu, X., Suk, H.-I., Huang, H., Shen, D.: Low-rank graph-regularized structured sparse regression for identifying genetic biomarkers. IEEE Trans. Big Data **3**(4), 405–414 (2017)
52. Zhu, X., Suk, H.-I., Wang, L., Lee, S.-W., Shen, D.: A novel relational regularization feature selection method for joint regression and classification in AD diagnosis. Med. Image Anal. **38**, 205–214 (2017)
53. Zhu, X., Zhang, S., Li, Y., Zhang, J., Yang, L., Fang, Y.: Low-rank sparse subspace for spectral clustering. IEEE Transactions on Knowledge and Data Engineering (2018). https://doi.org/10.1109/TKDE.2018.2858782
54. Zhu, X., Zhang, S., Hu, R., Zhu, Y., et al.: Local and global structure preservation for robust unsupervised spectral feature selection. IEEE Trans. Knowl. Data Eng. **30**(3), 517–529 (2018)