Brain-wide Genome-wide Association Study for Alzheimer’s Disease via Joint Projection Learning and Sparse Regression Model

Tao Zhou, Member, IEEE, Kim-Han Thung, Mingxia Liu, and Dinggang Shen*, Fellow, IEEE

Abstract—Brain-Wide and Genome-Wide Association (BW-GWA) study is presented in this paper to identify the associations 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). Brain-wide genome-wide association (GW A) study focuses on identifying the genetic variations that affect the brain structures and functions. Most of the GW A studies are based on the mass-univariate linear model, which fixes one variable (e.g., imaging phenotypes) while performing regression analysis using the other variables (e.g., genetic variations). However, as one genotype is probably associated with multiple imaging phenotypes (e.g., one gene may affect the cortical thickness of several brain regions) and one imaging phenotype is also probably associated with multiple genotypes (e.g., the cortical thickness of one brain region could be influenced by multiple genetic variations), the association between genotypes and phenotypes is actually a complex many-to-many relationship. Therefore, univariate analysis of conventional GW A studies is not sufficient to address the complex phenotype-genotype relationship. Consequently, more and more attentions have been given to multivariate analysis of the imaging genetic, i.e., Brain-Wide Genome-Wide Association (BW-GWA) study [24], [25], [26], which can simultaneously identify the genetic variations and imaging phenotypes that are associated with each other in AD study.

The aim of BW-GWA study is to evaluate the influence of genetic variations on brain structures and functions, so that we can better understand the complex neuropathology of the disease (e.g., AD), from genetic variants, to cellular processes, to brain structure and function, and to the symptoms of memory and cognitive impairment. This kind of study is highly desirable in clinical applications, as it enables the researchers to monitor the progression of AD starting from its origin. That Alzheimer’s disease (AD) is the most common form of dementia that often affect individuals over 65 years old, which strongly impacts human’s thinking, memory and behavior [1], [2], [3], [4], [5], [6], [7], [8]. This disease always progresses along a temporal continuum, initially from a normal control (NC) stage, subsequently to mild cognitive impairment (MCI) and eventually deteriorating to AD. According to a recent report from Alzheimer’s Association [9], the total estimated prevalence of AD is expected to be 60 million worldwide over the next 50 years, thus gaining growing attention in clinical research.

In recent years, a large number of neuroimaging studies [10], [11], [12], [13] have been done to investigate the associations between the AD and the brain structural or functional changes [14], [15], [16], [17], [18], [19]. Besides, some genome studies have also been done to understand the genetic causes of AD [20], while some genome-wide association (GWA) studies focus on identifying the genetic variations that affect the brain structures and functions [21], [22], [23]. Most of the GWA studies are based on the mass-univariate linear model, which fixes one variable (e.g., imaging phenotypes) while performing regression analysis using the other variables (e.g., genetic variations). However, as one genotype is probably associated with multiple imaging phenotypes (e.g., one gene may affect the cortical thickness of several brain regions) and one imaging phenotype is also probably associated with multiple genotypes (e.g., the cortical thickness of one brain region could be influenced by multiple genetic variations), the association between genotypes and phenotypes is actually a complex many-to-many relationship. Therefore, univariate analysis of conventional GWA studies is not sufficient to address the complex phenotype-genotype relationship. Consequently, more and more attentions have been given to multivariate analysis of the imaging genetic, i.e., Brain-Wide Genome-Wide Association (BW-GWA) study [24], [25], [26], which can simultaneously identify the genetic variations and imaging phenotypes that are associated with each other in AD study.

I. INTRODUCTION

T. Zhou, K-H. Thung and M. Liu are with the Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA (e-mails: taizhou.dreams@gmail.com, henrythung@gmail.com, mxliu@med.unc.edu).

D. Shen* is with the Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, and also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 136-713, Republic of Korea (e-mail: dgshen@med.unc.edu)

* Corresponding author.

This research was supported in part by NIH grants EB006733, EB008374, EB009634, MH100217, AG041721 and AG042599.
is, such study may provide vital information on how the AD-related genes interact with other non-genetic factors, which is important to possibly slow down the progression or even prevent the occurrence of AD.

However, the BW-GWA study is challenging, primarily due to the high-dimensional genotypes (e.g., Single Nucleotide Polymorphisms: SNPs) and the heterogeneity between the genotypes and the neuroimaging phenotypes (e.g., Region-Of-Interest (ROI)-based features in Magnetic Resonance Imaging (MRI) data). To address the high-dimensional data issue, some studies regarded this imaging genetic study as a feature selection problem [27], [28], [29], [30], e.g., to select SNPs that are most correlated with the imaging phenotypes. For instance, Wang et al. [30] proposed a group sparse multi-task regression model to predict the neuroimaging measures using the most related SNPs. This method focused on selecting the most relevant SNPs, while the neuroimaging measures were preselected based on prior knowledge. Vounou et al. [31], [32] proposed to use a sparse reduced-rank regression model and a sub-sampling strategy to select a subset of SNPs that are most associated with the multivariate phenotypes. Zhu et al. [24] developed a structured sparse low-rank regression model to investigate the associations between the imaging phenotypes and the genotypes. In addition, considering that the neuroimaging phenotypes are often measured over time, the longitudinal profile should provide increased power for identifying the associated genotypes. Following this line of research, Wang et al. [33] presented a temporal structure learning model that used the longitudinal genotype-phenotype interrelations to predict the phenotypes, while Wang et al. [29] presented a task-correlated longitudinal sparse regression model to identify longitudinal imaging markers that are associated with the AD-relevant SNPs. In summary, to address the high-dimensionality issue in imaging genetic study, the current approaches project the SNP data into a lower dimensional space via different sparsity constraint and/or low-rank constraint methods.

However, there are still several limitations in the current BW-GWA studies. First, the existing methods seldom consider the data heterogeneity between the genotypes and the imaging phenotypes. For instance, the conventional methods [24], [29], [33], [34] often simply adopt a linear regression model using the imaging phenotypes and the genotypes as the responses and the input variables, respectively, without considering the heterogeneity between these two data modalities. Second, the current methods also lack of consideration of complex associations between the genotypes and the imaging phenotypes. For example, most of the current methods [30], [33], [32], [24] only focus on finding the associated SNPs for a subset of neuroimaging phenotypes, or finding the associated neuroimaging phenotypes for a subset of SNPs, without simultaneously exploit the many-to-many associations between the SNPs and the imaging phenotypes. Third, the issue of phenotypes outliers is seldom considered in the existing studies. Many current BW-GWA methods [30], [33], [32], [24], [35] use squared Frobenius-norm loss function in their linear regression models, i.e., $\|W_2^T X_2 - X_1\|_F^2$, where $X_2$ and $X_1$ denote the SNPs and the MRI phenotype data, respectively, and $W_2$ is the learned weight matrix that reflects the relations between the SNPs and the imaging phenotypes (as shown in Fig. 1). However, since not all the brain regions (phenotypes) are AD-related, using squared Frobenius-norm loss function to measure the reconstruction error of all the phenotypes (i.e., $W_2^T X_2 - X_1$) is vulnerable to those non-AD-related phenotypes (outliers) [24].

Fourth, the valuable diagnostic label information is also often ignored in the previous studies [30], [32], [36]. As the main aim of BW-GWA study for AD is to find the associations between genotypes and phenotypes for AD study, the efficient usage of diagnosis label information may improve the learning performance in BW-GWA study [37].

In this paper, we address the above limitations by proposing a BW-GWA method that simultaneously selects the associated genotypes and imaging phenotypes, while considering the data heterogeneity, phenotype outliers, and diagnostic label information. We achieve this by proposing a novel joint-projection and sparse feature learning algorithm, which projects the genotypes and the imaging phenotypes into a diagnostic-label-guided joint latent feature space, while at the same time employs a sparse regression model to find the associations between these two data modalities. In Fig. 1 we show the overview of our framework as well as how it is different from typical framework of the conventional BW-GWA studies. Specifically, to alleviate the negative influence of the data heterogeneity, we first map the genotypes (i.e., SNPs) to the neuroimaging-phenotype-like space. Then, we project the mapped genotypes (which is now imaging-phenotype-like) and the imaging phenotypes into a joint latent feature space. We also introduce the diagnostic label information to constrain the projected data to have small intra-class distance in this joint feature space, which will consequently enhance the disease discriminability of the projected data in terms of their disease labels. In addition, to select only a subset of genotypes and phenotypes that are associated with each other, we impose a $\ell_{2,1}$-norm regularizer on the transformation matrix that maps the phenotypes to the joint latent space (i.e., $W_1$ in Fig. 1), and also on the transformation matrix that maps the genotypes to the intermediate phenotypes space (i.e., $W_2$ in Fig. 1). The $\ell_{2,1}$-norm regularizer has been widely used to encourage group sparsity (e.g., row-wise sparsity, where each row is corresponding to one genotype or phenotype index), which

![Fig. 1. Comparison of the basic framework and our proposed framework in BW-GWA study. (Left) The basic framework: It projects the genotypes (e.g., SNPs, denoted as $X_2$) into brain imaging phenotypes (e.g., MRI data, denoted as $X_1$). A sparse weight matrix $W_2$ (e.g., with $\ell_2$, $\ell_1$, or $\ell_{2,1}$-norm regularization) can be learned to reflect the relationships between SNPs and MRI. (Right) Our proposed framework: We first learn a weight matrix $W_2$ to intermediate transform the raw SNP data to the imaging phenotypes space. Then, we project the transformed SNP data (which is imaging-phenotype-like) and the MRI phenotype data into a disease label guided joint latent feature space via the orthogonal projection matrix $P$ and the transformation matrix $W_1$, respectively. The feature selection for SNP and MRI phenotypes is achieved by imposing sparsity constraints on $W_1$ and $W_2$, respectively.]

Transactions on Biomedical Engineering
in our case, is used to select the associated phenotypes and genotypes. Besides, in order to address the issue of phenotype outliers, we use a \( \ell_{2,1} \)-norm (instead of the conventional squared Frobenius-norm) based loss function in our proposed method.

We evaluate our model using the genetic and MRI data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI). More specifically, we learn our model by using the SNPs as genotypes, and the ROI-based volumetric measures of the MRI data as phenotypes in this BW-GWA study. We use the learned model to predict the imaging phenotypes by using the selected associated SNPs, and the experimental results show that our proposed model outperforms the state-of-the-art methods in terms of the average Root-Mean-Square-Error (RMSE) of genotype-to-phenotype predictions.

The rest of this paper is organized as follows. We describe materials and data preprocessing steps in Sec. II introduce our model in Sec. III present the experimental results in Sec. IV and conclude our study in Sec. V.

II. MATERIALS AND IMAGE DATA PREPROCESSING

A. Subjects

We use the public ADNI dataset \(^1\) to verify the performance of our proposed framework. The ADNI dataset was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and nonprofit organizations with a five-year public private partnership. In this study, we used 737 ADNI-1 subjects, including 171 AD, 362 MCI, and 204 NC subjects, as described in \(^2\).

B. Neuroimaging Data Preprocessing

We used the ROI-based MRI features as phenotypes in this study. We first downloaded ADNI preprocessed 1.5T MR images from the ADNI website \(^3\). The structural MR images were collected by using a variety of scanners with protocols individualized for each scanner. To ensure the quality of all images, ADNI had reviewed these MR images and corrected them for spatial distortion caused by B1 field inhomogeneity and gradient nonlinearity. After that, following the previous studies \(^40\), \(^41\), \(^42\), we processed these neuroimaging data and extracted the ROI-based features. Specifically, for each MR image, we processed it by using the following steps: anterior commissure-posterior commissure (AC-PC) correction by using MIPAV software \(^4\), intensity inhomogeneity correction by using N3 algorithm \(^43\), brain extraction by using a robust skull-stripping algorithm \(^44\), cerebellum removal, tissues segmentation by using FAST algorithm in FSL package \(^45\) to obtain three main tissues (i.e., white matter (WM), gray matter (GM), and cerebrospinal fluid), and registration to a template \(^46\) by using HAMMER algorithm \(^47\). We then dissected the images into 93 regions-of-interest (ROIs) by using the labeling information from the template \(^46\), and used the gray matter tissue volumes, one from each ROI, as the MRI features of a subject. Finally, we have 93 ROI-based MRI features, or phenotypes, for each subject.

C. Genetic Data Preprocessing

We used SNPs as genotypes in this study. The SNP data were genotyped using the Human 610-Quad BeadChip \(^48\). According to the AlzGene dataset \(^4\) only SNPs that belong to the top AD gene candidates were selected after the standard quality control and imputation steps. The Illumination annotation information was used to select a subset of SNPs \(^49\), and the dimensionality of the processed SNP data is 3123 in this study.

III. METHODOLOGY

A. Basic Notations

In this paper, we denote matrices, vectors, and scalars as boldface uppercase letters, boldface lowercase letters, and normal italic letters, respectively. For a matrix \( X \in \mathbb{R}^{d \times n} \), where \( d \) and \( n \) denote the feature dimension and the number of samples, respectively, we use \( \| X \|_F = \sqrt{\sum_{i} \| X_i \|_2^2} \) and \( \| X \|_{2,1} = \sum_{i=1}^{d} (\sum_{j=1}^{n} x_{ij}^2)^{\frac{1}{2}} \) to denote its Frobenius-norm and \( \ell_{2,1} \)-norm, respectively, where \( x_{ij} \) is the \((i,j)\)-th element of the matrix \( X \). In addition, we use \( X^T \) and \( X_j \) to denote the \(i\)-th row and the \(j\)-th column of \( X \), respectively, and use \( X^T \) and \( Tr(X) \) to denote the transpose and the trace operator of the matrix \( X \), respectively.

B. Joint Projection Learning and Sparse Regression Model

Let \( X_1 \in \mathbb{R}^{d_1 \times n} \) and \( X_2 \in \mathbb{R}^{d_2 \times n} \) denote the feature matrices for MRI (ROI-based features) and SNP data, respectively, where \( n \) is the number of samples, and \( d_1 \) and \( d_2 \) denote their corresponding numbers of feature variables. We can use linear regression to model the relationship between the MRI phenotypes and the SNP genotypes, i.e.,

\[
\min_{W_2} \| W_2^T X_2 - X_1 \|_F^2,
\]

where \( W_2 \in \mathbb{R}^{d_2 \times d_1} \) is a coefficient matrix. With Eq. (1), we can project the SNP data into the MRI feature space. Furthermore, assuming that only certain SNPs are associated with the brain imaging features, we impose a group sparsity constraint (i.e., \( \ell_{2,1} \)-norm) on \( W_2 \) in Eq. (1), and obtain the following objective function

\[
\min_{W_2} \| W_2^T X_2 - X_1 \|_F^2 + \lambda \| W_2 \|_{2,1},
\]

where \( \lambda \) is the regularization parameter to balance the contributions of the reconstruction error (i.e., the first term in Eq. (2)) and the regularizer (i.e., the second term). The \( \ell_{2,1} \)-norm regularizer has been widely applied in multi-task feature learning \(^50\), \(^51\), \(^52\), \(^53\) to impose row sparsity (i.e., many all-zero rows) on the matrix of interest. Given that \( \| W_2 \|_{2,1} = \sum_{i=1}^{d_2} \| W_{2i} \|_2 \) where \( W_{2i} \) is the \(i\)-th row of \( W_2 \), the minimization in Eq. (2) will enforce only certain rows of \( W_2 \)}
to be non-zero. Thus, by ranking $\|W_1^i\|_2$ ($i = 1, 2, \ldots, d_2$), the first $k$ ($k \ll d_2$) highest magnitude of $\|W_2^i\|_2$ would be corresponding to the top $k$ SNPs that are associated with the neuroimaging phenotypes.

In addition, some studies have shown that the first term in Eq. (2) is more vulnerable to the row-wise outliers of $X_1$ [54], which, in our case, are the phenotype outliers. This is because the “square” of the squared Frobenius-norm will magnify the reconstruction errors of the phenotype outliers, and cause them to dominate the minimization of the objective function. One solution to address this issue is to replace the squared Frobenius-norm of the fidelity term with the $\ell_{2,1}$-norm, which has been shown to be robust to row-wise outliers [53]. Thus, by adopting $\ell_{2,1}$-norm to measure the reconstruction error, we obtain

$$\min_{W_2} \|W_2^T X_2 - X_1\|_{2,1} + \lambda \|W_2\|_{2,1}. \quad (3)$$

In summary, the $\ell_{2,1}$-norm is used for both the data fidelity term and the regularization term in Eq. (3), where its usage in fidelity term is to increase its robustness to phenotype outliers, while its usage in regularization term is to sparsely select SNPs that are associated with the ROI-based MRI phenotypes.

However, as shown in many works [22, 33, 16, 17, 20], not all the ROI-based MRI features are related to AD, since AD mainly affects brain regions that are related to memory and cognitive functions. Similarly, we would also expect that the SNPs that are associated with AD would only affect certain brain regions only. Thus, features in $X_1$ that are not useful for the AD diagnosis should be removed, and only certain features in $X_1$ should be used as targets in Eq. (3) to find the associated SNPs. But how should we select features in $X_1$ for BW-GWA study? The simplest selection method perhaps is using the disease label as target and $X_1$ as regressors in sparse regression, to select the most discriminate features in $X_1$. Then, the proposed Eq. (3) can be subsequently used to find the associated SNPs. However, such 2-step approach is sub-optimal in BW-GWA study, as it ignores the complex pathogenesis of the disease and the inter-correlated relationship among the genetic variations (SNPs), imaging phenotypes and the disease labels. In this study, we would like to jointly select the associated SNPs and imaging phenotypes by exploiting the disease label information and its relationship with SNPs and imaging phenotypes. To achieve this, we first assume there exists a joint latent feature space $H$ for $X_1$ and $X_2$, i.e., a space which is guided by the disease label and thus closer to the disease labels. We further assume that, under different transformation matrices $W_1 \in \mathbb{R}^{d_1 \times h}$ and $W_2 \in \mathbb{R}^{d_2 \times h}$, the feature matrices $X_1$ and $X_2$ are respectively mapped to the same latent feature space $H \in \mathbb{R}^{h \times n} \subset H$, i.e.,

$$\min_{W_1, W_2, H} \|W_1^T X_1 - H\|_F^2 + \lambda \|W_2^T X_2 - H\|_F^2 + R_1(W_1, W_2), \quad (4)$$

where $R_1(W_1, W_2)$ denotes the regularizer that involves $W_1$ and $W_2$, e.g., $\ell_{2,1}$-norm regularizer that encourages group sparsity for these two matrices, such as $\|W_1\|_{2,1} + \|W_2\|_{2,1}$. However, there is an issue with this formulation, i.e., both the genotypes and phenotypes are assumed to be directly correlated with $H$, which could be untrue. Let us consider a special case of $H$, which is the corresponding disease label, and then Eq. (4) would end up selecting the discriminate SNPs and ROI-based neuroimaging features. But SNPs as genotype features are much less sensitive to the disease labels (or diagnostic performance) than the neuroimaging features [55], and thus should have taken a longer path to reach the joint latent feature space, as shown in Fig. 1. Accordingly, we redefine the objective function in Eq. (4) as follows,

$$\min_{W_1, W_2, H, P} \|W_1^T X_1 - H\|_F^2 + \lambda \|W_2^T X_2 - H\|_F^2 + R_1(W_1, W_2), \quad (5)$$

$$s.t. \quad PP^T = I,$$

where $P \in \mathbb{R}^{h \times d_1}$ is a projection matrix, $I \in \mathbb{R}^{d_1 \times d_1}$ is an identity matrix, and the orthogonal constraint $P^T P = I$ ensures $P$ to be a basis transformation matrix. In Eq. (5), we compute the latent feature space $H$ to help us find the genotype-phenotype associations, but this $H$ is not really used in analysis. Thus, by omitting the explicit computation of the joint feature matrix $H$, we propose to use a simplified alternative of Eq. (5), as follows,

$$\min_{W_1, W_2, P} \|PW_2^T X_2 - W_1^T X_1\|_F^2 + R_1(W_1, W_2), \quad (6)$$

$$s.t. \quad PP^T = I.$$

**Remarks.** The projection matrix $P$ and the coefficient matrix $W_2$ in Eq. (6) are used to map the genomic features to the joint latent space, to assist us in finding the associations between the SNPs and the neuroimaging phenotypes. Also note that, although we do not compute the latent feature representation $H$ explicitly, the two modality data ($X_1$ and $X_2$) are still implicitly transformed into the joint feature space through the learning of $W_1$, $W_2$, and $P$. Furthermore, as shown in Fig. 1 and Eq. (6), the proposed model takes several measures to address the heterogeneity between the high-dimensional SNPs and the low-dimensional ROI-based MRI phenotypes. Specifically, it first maps the SNPs to an intermediate imaging-phenotype-like space via $W_2$, and then projects both the imaging-phenotype-like features (from genotypes) and the original imaging phenotypes into a joint latent feature space.

So far, the label information is still not exploited in Eq. (6). Inspired by [37], to learn a more meaningful joint latent feature space, in this study, we use linear discriminant analysis.
strategy to encourage compactness of intra-class samples in the joint latent feature space, as shown in Fig. 1. For example, for MRI data, if the $i$-th sample $X_{1i}$ and the $j$-th sample $X_{1j}$ are belonged to the same class (i.e., disease label), we would like to have their projections $W_{1i}^T X_{1i}$ and $W_{1j}^T X_{1j}$ to be close to each other. Similar criterion can also be applied to the samples from the SNP data. Based on these criteria, the regularizer in Eq. (6) is given as,

$$R(X_1, X_2, W_1, W_2, P) = \text{Tr}((W_{1i}^T X_{1i})L_1((W_{1i}^T X_{1i})^T) + \text{Tr}((PW_{2i}^T X_{2i})L_2((PW_{2i}^T X_{2i})^T),$$

where $L_1 \in \mathbb{R}^{n \times n}$ and $L_2 \in \mathbb{R}^{n \times n}$ are the corresponding Laplacian matrices for $X_1$ and $X_2$, respectively. The Laplacian matrix is defined as $L_1 = D_1 - S_1$, where $D_1 \in \mathbb{R}^{n \times n}$ is the degree matrix, in which its $i$-th diagonal element is given as $D_{1i} = \sum_j S_{1ij}$, and $S_1 \in \mathbb{R}^{n \times n}$ is the similarity matrix defined as

$$S_{1ij} = \begin{cases} 1, & \text{if } y((X_1)_i) = y((X_1)_j), \\ 0, & \text{else} \end{cases},$$

where $y((X_1)_i)$ denotes the corresponding label value for the sample $X_{1i}$. Similar formulation is also used to obtain the Laplacian matrix $L_2$.

Finally, by combining Eqs. (3), (6), and (7), we have the following final objective function,

$$\min_{W_1, W_2, P} ||W_{2i}^T X_{2i} - X_{1i}||_{2,1} + \beta ||PW_{2i}^T X_{2i} - W_{1i}^T X_{1i}||_F^2 + \gamma R(X_1, X_2, W_1, W_2, P) + \lambda_1 ||W_1||_{2,1} + \lambda_2 ||W_2||_{2,1},$$

s.t. $PP^T = I,$

where $R(X_1, X_2, W_1, W_2, P)$ is given by Eq. (7), and $\beta, \gamma, \lambda_1$ and $\lambda_2$ are the regularization parameters. To ease the comprehension of the proposed formulation in Eq. (9), we have summarized the main notations in Table I. The properties of our proposed formulation are also summarized below.

- The first term, i.e., $||W_{2i}^T X_{2i} - X_{1i}||_{2,1}$, is a reconstruction term (or data fidelity term) which is mainly used to find the association between the SNP and MRI data. All the other terms are used to improve the reconstruction error of this term for the testing samples, for AD study. Looking from another angle, it can also be regarded as an intermediate-level feature learning for the SNP data, i.e., to first map the SNP data to a neuroimaging-feature-like space (i.e., $W_{2i}^T X_{2i}$), before using the orthogonal projection matrix $P$ to further project the mapped data into the label-guided joint latent feature space. This intermediate feature learning for the SNP data is also reasonable as it is high dimensional and less discriminative to the disease label than the neuroimaging data. Besides, we use $\ell_{2,1}$-norm regularizer for this term (rather than the popularly used Frobenius norm) to alleviate the issue about feature outliers.

- The second term, i.e., $||PW_{2i}^T X_{2i} - W_{1i}^T X_{1i}||_F^2$, is a joint feature projection term, to project the SNPs and the MRI phenotypes into a joint latent feature space, and assist the association study. We further utilize the diagnostic label information when learning the joint feature projection, by adding the Laplacian regularizer in the third term, so that the intra-class samples (i.e., diagnostic cohort) will be close to each other in the joint feature space.

- We use $\ell_{2,1}$-norm regularizer to impose row-wise sparsity (i.e., all zero-value rows) on $W_1$ and $W_2$. As each row in $W_1$ and each row in $W_2$ corresponds to one ROI-based neuroimaging feature and one SNP, respectively, the row sparsity implies the selection of associated ROIs and SNPs. In addition, we do not impose $\ell_{2,1}$-norm on $P$, but rather use the orthogonal constraint $PP^T = I$ to avoid the trivial solution, which has been used widely in many computer vision and machine learning applications.

C. Optimization

The objective function in Eq. (9) simultaneously learns $W_1$, $W_2$, and $P$. Since it is not jointly convex with respect to all the variables (i.e., $W_1$, $W_2$, and $P$), we can update each of these variables by fixing other variables iteratively. More specifically, we adopt the Augmented Lagrange Multiplier (ALM) [50] to solve the optimization problem specified in Eq. (9). We have provided the detailed optimization steps in Supplementary Materials of this manuscript.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we used ADNI dataset to evaluate our method. We evaluated our proposed method by using the associated SNPs to predict the MRI imaging phenotypes. Besides, using our proposed method, we also identified the subset of SNPs that are associated with AD-related imaging phenotypes and vice versa.

A. Experimental Setup

In order to verify the effectiveness of our proposed method, we compared it with some state-of-art methods, including Ridge Regression (RR) [61], Group-sparse Feature Selection (GFS) [30], sparse feature selection with an $\ell_{2,1}$-norm regularizer (L21) [62], Structured-sparse Low-rank Regression model (SLR) [24], and Robust Feature Selection via joint $\ell_{2,1}$-norm (RFS) [63]. For clarity, we also describe the details of these comparison methods below:
Table II

<table>
<thead>
<tr>
<th>Method</th>
<th>Objective function</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>(\min |W^T X_1 - X_2|_F^2 + \lambda_1 |W|_F^2)</td>
</tr>
<tr>
<td>GFS</td>
<td>(\min |W^T X_1 - X_2|_F^2 + \gamma |X_2|_G + \lambda_1 |W|_2)</td>
</tr>
<tr>
<td>L21</td>
<td>(\min |W^T X_1 - X_2|_F^2 + \lambda_1 |W|_2)</td>
</tr>
<tr>
<td>SLR</td>
<td>(\min {X_1^T - X_2^T \beta - \epsilon}_+^2 + \lambda_1 |\beta|_2 + \lambda_2 |W|_2)</td>
</tr>
<tr>
<td>RFS</td>
<td>(\min |W^T X_1 - X_2|_2 + \lambda_1 |W|_2)</td>
</tr>
</tbody>
</table>

- **Ridge Regression (RR)**: RR employs a least square loss function to exploit the association between SNPs and MRI data (e.g., ROIs).
- **Group-sparse Feature Selection (GFS)**: GFS employs two types of regularizers, i.e., one that considers the group level structural information within the SNP data, and another that considers the sparsity among the SNP groups.
- **Sparse feature selection with a \(\ell_{2,1}\)-norm regularizer (L21)**: L21 employs a least square loss function in combination with a \(\ell_{2,1}\) regularizer to exploit the association among the features.
- **Structured-sparse Low-rank Regression model (SLR)**: SLR projects SNP data to MRI data (e.g., ROIs) via \(BA^T\), and the \(\ell_{2,1}\)-norm is respectively imposed on both matrices \(A\) and \(B\). SLR does not consider the phenotype outliers and label information in its formulation.
- **Robust Feature Selection via joint \(\ell_{2,1}\)-norm (RFS)**: RFS imposes the \(\ell_{2,1}\)-norm on both the loss function and the regularization term. But, it does not consider label information.

We also summarize the objective functions of all the comparison methods in Table II where \(X_1\) and \(X_2\) denote the MRI phenotypes and the SNP data in our study, respectively. In summary, the objective functions of all the comparison methods consist of two parts, i.e., the reconstruction term and the regularization term. It is worth noting that the RFS method is the degraded version of our method, i.e., when we only adopt the genotype-to-phenotype linear regression model without using the joint latent space, as in Eq. (3).

We used a 5-fold nested cross-validation to select the optimal regularization parameters for our method, i.e., \(\beta, \gamma, \lambda_1\), and \(\lambda_2\). Similarly, we also used the same method to select the optimal parameters for all the competing methods in this study. The best parameter values are determined by searching in the range of \(\{10^{-3}, 10^{-2}, \ldots, 10^2, 10^3\}\). In addition, we empirically searched the value for the parameter \(h\) in the range of \(\{10, 20, \ldots, 80\}\). Based on our empirical results, we found that the RMSE performance of our proposed model is stable for the value of parameter \(h\) in the range of \(\{20, \ldots, 60\}\). Thus, we empirically set the parameter \(h\) to 30 as its default parameter value. Finally, we used the Root-Mean-Square-Error (RMSE) of the genotype-to-phenotype predictions as our evaluation method.

After training our method using the training data, we obtained the coefficient matrix for the SNPs (\(W_2\)), where each of its row corresponds to a SNP. We sorted the \(l_2\)-norm of each row of \(W_2\), and selected the top \(\{10, 20, \ldots, 250\}\) SNPs (that correspond to the largest row-wise \(l_2\)-norm of \(W_2\)) to predict the imaging phenotypes of the testing SNP data. For each method, we conducted five random repetitions of the 5-fold cross-validation experiments, and reported the average results as the final result. In addition, we also investigated how the selections of SNPs and ROIs are affected by the disease labels, as different disease labels would affect the implicitly learned joint latent feature space. Thus, we performed experiments using different combinations of disease label cohorts (called as tasks, for simplicity) as guidance, including AD/NC, MCI/NC, and AD/MCI/NC.

B. Performance Comparison

Fig. 3 shows the performances of all the comparison methods, using 3 different disease cohort combinations (i.e., tasks). The mean (bar height) and standard deviation (error bar) of the RMSE values of the comparison methods are shown. It can be easily observed that our proposed method consistently outperforms other state-of-the-art methods in all tasks, both in terms of mean and standard deviation. This can be explained by a few advantages of our proposed method as follows.

- Our method adopts the \(\ell_{2,1}\)-norm to model the reconstruction loss, which can effectively reduce the effects of phenotype outliers. The advantage of \(\ell_{2,1}\)-norm over Frobenius-norm in modeling the reconstruction error can be seen by comparing the results given by RFS and L21 in Fig. 3. Note that, RFS and L21 methods only differ in their reconstruction terms, where RFS uses the \(\ell_{2,1}\)-norm while L21 uses the Frobenius-norm, as shown in Table II. The results in Fig. 3 demonstrate that RFS is better than L21. However, as RFS only simply projects the SNP data to MRI data without considering label information, it performs inferior to our proposed method.
- Our method projects the SNP and MRI data into a joint feature space, with the guidance from diagnostic label information. Note, diagnostic labels provide extra information for learning the joint feature space, which in turn helps the selection of associated SNP and MRI phenotypes.
C. Identification of Top Selected SNPs

Fig. 4 shows the normalized selection frequency of the top 10 selected SNPs by using our method and all the comparison methods. From Fig. 4 we have the following observations:

- Most methods include APOE-rs429358 as one of the top 10 associated SNPs, indicating its strong association with MRI phenotypes. In the literature, APOE-rs429358 has been shown to be highly correlated with the atrophy patterns of the entire cortex as well as medial temporal regions (including amygdala, hippocampus, and parahippocampal gyrus) in AD [64, 65, 33]. It is worth noting that, although APOE-rs429358 is one of the well-known major AD risk factors, it is not always selected by other comparison methods (i.e., with the normalized selection frequency less than 1), whereas our proposed BW-GWAS method always selects this SNP (i.e., with the normalized selection frequency being one) in all the three tasks. This is probably due to the fact that we have utilized disease label information in our association study, which make the results more related with AD and useful for AD study.

- For AD/NC task, the top selected SNPs are from ApoE gene (e.g., rs429358), PICLAM gene (e.g., rs11234495), CR1 gene (e.g., rs3818361, rs6701713, rs33737002), SERPINA13 gene (e.g., rs6575450), SORCS1 gene (e.g., rs7073924), etc. For MCI/NC task, the top selected SNPs are from ApoE gene (e.g., rs429358), CTNNA3 gene (e.g., rs997225, rs10997232, rs2082415), VEGF gene (e.g., rs3025035), etc. For AD/MCI/NC task, the top selected SNPs are from ApoE gene (e.g., rs429358), PICLAM gene (e.g., rs11234495), VEGF gene (e.g., rs3025035), SORL1 gene (e.g., rs2276364), etc. These findings are consistent with the existing GWA studies in [24, 66, 53, 67, 23, 68, 69]. Specifically, the ApoE gene and SORCS1 gene have been shown to have significant relationship with AD and MCI [24]. The VEGF gene has been shown to have greater risk for sporadic AD [67], and the PICALM gene affects AD risk primarily by modulating production, transportation, and clearance of \( \beta \)-amyloid peptide [70].

- We also found some new associated SNPs that are worthy for further investigation, including TFAM gene (e.g., rs11006132), EBF3 gene (e.g., rs4751153), ESR1 gene (e.g., rs2504070), BCR gene (e.g., rs140500). For example, some studies have investigated that the TFAM gene encodes the mitochondrial transcription factor A to control the transcription, replication, damage sensing, and repair of mitochondrial DNA (mtDNA), and lots of mutations in mtDNA have been found in the brains of patients with late-onset Alzheimer’s disease [71]. In other words, these SNPs are associated with the ROI-based MRI phenotypes, which could be helpful in AD pathogenesis.

In brief, that our main findings are consistent with the findings in the literature, has not only verified the effectiveness of our proposed method, but also justified our results. Besides, our method has also found other new associated SNPs that are worthy for further investigation in AD pathogenesis study.

D. Identification of Top Related ROIs

We are also interested in identifying the brain regions (i.e., ROIs) that are most associated with the top selected SNPs in this BW-GWA study. In our experiments, the most frequently selected ROI-based features in cross-validation experiments are regarded as the most AD-associated brain regions. These top ROIs and the associated SNPs are important as they can become potential biomarkers and genotypes for pathogenesis study and AD diagnosis. In our model, we can select the ROIs that are most associated with the selected SNPs by ranking \( \| W_i \|_2 \) (\( i = 1, \ldots, 93 \)) for MRI data, where \( \| W_i \|_2 \) denotes the \( l_2 \)-norm of the \( i \)-th row in \( W_i \), and 93 is the total number of ROI features in MRI data. Fig. 5 shows the top ten ROIs by using our proposed method for different disease cohort combinations. Specifically, for AD/NC task, the top selected ROIs are frontal lobe WM left, hippocampal formation right, middle temporal gyrus left, temporal lobe WM left, hippocampal formation left, corpus callosum, and amygdala right. For MCI/NC task, the top selected ROIs are globus pallidus left, hippocampal formation right, parietal lobe WM left, middle temporal gyrus left, hippocampal formation left, postcentral gyrus right, and amygdala right. For AD/MCI/NC task, the top selected ROIs are globus pallidus right, frontal lobe WM right, hippocampal formation right, parietal lobe WM left, temporal lobe WM left, hippocampal formation left, and amygdala right. These findings are consistent with those reported in many AD diagnostic studies [72, 41, 16, 17, 73] and also genome-wide association studies [24, 33, 37]. Thus, the ROIs selected by our proposed method are considered trustworthy.
E. Validation of Key Components

Experimental results so far have clearly demonstrated the effectiveness of our proposed method, as it can obtain lower RMSE by using a subset of SNPs to predict the MRI data. There are three key components in our proposed method, including 1) phenotype outliers robustness via the $\ell_{2,1}$-norm of the data fidelity term, 2) the joint projection into a common feature space, and 3) the label guiding feature projection/selection. In the following text, we will verify the effectiveness of these three key components in our method.

First, we use $F$-norm to replace the $\ell_{2,1}$-norm in Eq. (9), so that the first term is changed to $\|W_2^T X_2 - X_1\|_F^2$, while the other terms remain unchanged; this model is denoted as "Degraded 1". Second, if we do not project the two modalities into a joint latent feature space, the proposed model can be further degraded to

$$\min_{W_2} \|W_2^T X_2 - X_1\|_{2,1} + \lambda_2 \|W_2\|_{2,1}.$$  \hspace{1cm} (10)

Remarks. Without projecting the two modalities into a joint feature space, it will remove the joint projection term in Eq. (11). Besides, without $P$ and $W_1$, the term for label-guided feature selection can be automatically removed. In this case, our proposed method is degraded into RFS method, and it is denoted as "Degraded 2".

Third, if we do not use label information in our method, i.e., removing symmetric normalized Laplacian operation for guiding feature selection, we can obtain another degraded model denoted as "Degraded 3", with its objective function described as

$$\min_{W_1, W_2, \beta} \|W_2^T X_2 - X_1\|_{2,1} + \beta \|PW_2^T X_2 - W_1^T X_1\|_F^2$$

$$+ \lambda_1 \|W_1\|_{2,1} + \lambda_2 \|W_2\|_{2,1}, \ s.t. \ PP^T = I.$$  \hspace{1cm} (11)

Fig. 6 shows the RMSE performance comparison of our proposed method and the three degraded models. From Fig. 6 it can be clearly seen that our proposed method performs better than its degraded counterparts (without one of the key components).

In addition, in order to evaluate the effectiveness of Laplacian regularizer (i.e., the third term in Eq. 9) in encouraging the intra-class sample compactness and class separation when learning the label-guided joint feature space, we define an intra-class distance ratio as separation measure. More specifically, we use the ratio of mean intra-class squared distance to mean inter-class squared distance. Intuitively, if our Laplacian regularization is effective, we will have smaller distance among intra-class samples and larger distance between two class centers after the joint feature projection, resulting in smaller intra-class distance ratios. We show the detailed computation steps for this separation measure as Algorithm 1 in Table III. Table IV shows the comparison results, i.e., intra-class distance ratios before and after the joint feature projection, using MRI and SNP data from different disease cohort combinations. As expected, all the ratios have decreased after the joint feature projection. Furthermore, as a qualitative measure, we also visualize the sample distributions before and after the joint feature projection in Fig. 7 and Fig. 8 by using t-
Distributed Stochastic Neighbor Embedding (t-SNE) algorithm [74]. From Fig. 7 and Fig. 8, it can be seen that the intra-class samples become closer after the joint feature projection, for both the SNP and MRI data. In conclusion, we have quantitatively and qualitatively showed that our label-guided joint feature space has closer intra-class samples and better label separation, compared with the original feature space.

Besides, from Fig. 7 and Fig. 8, it seems that SNP data are more separable in the joint feature space, compared with the MRI data. This is interesting, as genotype features (of SNP data) should be much less sensitive to the disease labels than the neuroimaging features (of MRI data). This is probably due to the fact that the points in Fig. 7 consist of samples from all the data in the training and testing sets, and the more separable samples in the joint feature space are mostly coming from the training set. In other words, SNP data are kind of “overfitting” to the label in the joint feature space. In contrast, although MRI data seem less separable than the SNP data in the joint feature space, they have less “overfitting” issue, as we can still observe a drop in intra-inter class distance ratio for the testing samples, when we examine further. Nevertheless, this “overfitting” issue did not affect much on the performance of our proposed method, as shown in Fig. 6 since we are focusing on BW-GWA study, not classification.

Further, we have computed the intra-inter class distance ratio in the neuroimaging-feature-like space for SNP data. The results for three disease cohort combinations (i.e., AD/NC, MCI/NC, and AD/MCI/NC) are 0.3218, 0.3655, and 0.7819, respectively. This means that the disease cohort separation has also been significantly improved for the neuroimaging-feature-like space of SNP data. However, when compared these results with the results in Table IV, the intra-inter class distance ratios are only slightly improved after joint feature projection. The small improvement of intra-inter class distance ratio from the neuroimaging-feature-like space to the joint feature space could be explained by the fact that the learning of \( W_2 \) is intrinsically guided by the label information. As the label-guided joint feature space and the SNP to MRI projections are learned together, the label information could guide the learning of \( W_2 \), which consequently improves the disease cohort separation in the neuroimaging-feature-like space for SNP data.

### V. CONCLUSION

In this paper, we propose a novel joint projection learning and sparse regression model to study the associations between the genetic variations (characterized by SNPs) and neuroimaging phenotypes (characterized by ROI-based GM volumetric features). Guided by label information of each subject, our proposed method can uncover the interrelation between genetic data and brain image phenotypes, and find the most associated SNPs and ROIs. We achieve these by projecting the genetic variations and neuroimaging phenotypes into a joint latent feature space, where the samples with same labels are constrained to be close. In this joint latent feature space, the heterogeneity between the SNP and MRI data is reduced, as they are much closer to each other in the disease label space. Thus, their correlation can be exploited more effectively. Furthermore, our main objective term to find the associations between the SNPs and neuroimaging phenotypes, if looks from another angle, is equivalent to first intermediated map the SNP data into a neuroimaging-like feature space, before introducing an orthogonal projection matrix to project the mapped SNP data to the joint feature space. The intermediate mapping of SNP data is necessary to alleviate the high dimensionality and less discriminative issues of the SNP data. The orthogonal property of the projection matrix, on the other hand, enables a more efficient solution for our objective function. The experimental results have validated the effectiveness of our proposed method when compared with some state-of-the-art methods, in terms of RMSE value of the reconstruction errors between the mapped SNP and MRI data. Besides, we have also verified the effectiveness of the key components in our proposed method, i.e., 1) the use of \( \ell_2,1 \)-norm to overcome the phenotype outliers, 2) the mapping to the joint feature space, and 3) the use of label information in guiding feature projection and feature selection, by com-
paring our proposed method with its degraded counterparts (without one of the key components in our proposed method). Furthermore, the selected associated SNPs and ROIs are also consistent with the findings in the literature, confirming the effectiveness of our proposed method. In conclusion, our proposed method is able to identify important associated SNPs and ROIs for future pathogenesis of AD and possibly also other neuropsychological diseases.

REFERENCES


