

Longitudinal Genotype–Phenotype Association Study through Temporal Structure Auto-Learning Predictive Model

XIAOQIAN WANG,¹ JINGWEN YAN,^{2,3} XIAOHUI YAO,^{2,3} SUNGEUN KIM,²
KWANGSIK NHO,² SHANNON L. RISACHER,² ANDREW J. SAYKIN,²
LI SHEN,² and HENG HUANG¹; for the ADNI

ABSTRACT

With the rapid development of high-throughput genotyping and neuroimaging techniques, imaging genetics has drawn significant attention in the study of complex brain diseases such as Alzheimer’s disease (AD). Research on the associations between genotype and phenotype improves the understanding of the genetic basis and biological mechanisms of brain structure and function. AD is a progressive neurodegenerative disease; therefore, the study on the relationship between single nucleotide polymorphism (SNP) and longitudinal variations of neuroimaging phenotype is crucial. Although some machine learning models have recently been proposed to capture longitudinal patterns in genotype–phenotype association studies, most machine-learning models base the learning on fixed structure among longitudinal prediction tasks rather than automatically learning the interrelationships. In response to this challenge, we propose a new automated time structure learning model to automatically reveal the longitudinal genotype–phenotype interactions and exploits such learned structure to enhance the phenotypic predictions. We proposed an efficient optimization algorithm for our model and provided rigorous theoretical convergence proof. We performed experiments on the Alzheimer’s Disease Neuroimaging Initiative (ADNI) cohort for longitudinal phenotype prediction, including 3123 SNPs and 2 biomarkers (Voxel-Based Morphometry and FreeSurfer). The empirical results validate that our proposed model is superior to the counterparts. In addition, the best SNPs identified by our model have been replicated in the literature, which justifies our prediction.

Keywords: Alzheimer’s disease, genotype–phenotype association prediction, longitudinal study, temporal structure auto-learning, low-rank model.

¹Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania.

²Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, Indiana.

³Department of Biostatistics, Epidemiology and Bioinformatics, Perelman School of Medicine, University of Pennsylvania.

Data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data, but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: [http://adni.loni.usc.edu/wp-content/uploads/how to apply/ADNI Acknowledgement List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

1. INTRODUCTION

ALZHEIMER'S DISEASE (AD) IS THE MOST PREVALENT and severe type of neurodegenerative disorder, which strongly affects human's memory, thinking, and behavior (Khachaturian, 1985). This disease is characterized by progressive memory impairment and other cognitive disabilities triggered by neuronal damage (Burns and Iliffe, 2009). AD usually progresses continuously, initially starting from the preclinical stage, then into mild cognitive impairment (MCI) and eventually into AD (Wenk, 2003). According to Alzheimer's Association (2016), AD is the sixth leading cause of death in the United States. Every 66 seconds, one person in the United States is developing AD. Until 2016, an estimate of 5.4 million people is living with AD in the United States, compared with 44 million worldwide. Worse, the world will see even more dramatic growth in these numbers in the short term if no breakthroughs are found.

With these facts, AD is gaining growing attention in this day and age. The current consensus emphasizes the need to understand the genetic causes of AD to prevent or slow down the progression in worsening the disease (Petrella et al., 2003). Recent advances in neuroimaging and microbiology have improved understanding of the links between genes, brain structure, and behavior (Hariri et al., 2006). At the same time, the rapid development of high-throughput genotyping enables the simultaneous measurement of hundreds of thousands or even millions of single nucleotide polymorphisms (SNPs) (Potkin et al., 2009a). These progresses have facilitated the pullulation of imaging genetics, which holds great promise for better understanding complex neurobiological systems.

In imaging genetics, an emerging strategy to facilitate the identification of susceptible disease-related genes is to use outcome-related biomarkers as quantitative traits (QTs) for genetic variation evaluation. Correlation studies between genetic variations and imaging studies often have significant advantages over case-control studies due to the following reason: QT is a quantitative measure with four to eight times more statistical power and significantly reduces the required sample size (Potkin et al., 2009b). A number of works have been reported to identify genetic factors affecting imaging phenotypes of biomedical importance (Harold et al., 2009; Shen et al., 2010; Bis et al., 2012).

In the genotype-phenotype association study, we can denote our inputs in the matrix format as follows: the SNP matrix $X \in \mathbb{R}^{d \times n}$ (n is the number of samples, d is the number of SNPs) and the imaging phenotype matrices $Y = [Y_1, Y_2, \dots, Y_T] \in \mathbb{R}^{n \times cT}$ (c is the number of QTs, T is the number of time points, and $Y_t \in \mathbb{R}^{n \times c}$ is the phenotype matrix at time t). The goal is to find the weight matrix $W = [W_1, W_2, \dots, W_T] \in \mathbb{R}^{d \times cT}$, which properly reflect the relations between SNPs and QTs and capture a subset of SNPs responsible for phenotype prediction at the same time. If we treat the prediction of one phenotype at one time point as a task, then the association study between genotypes with multiple longitudinal phenotypes can be seen as a multitask learning problem.

Conventional strategies (Ashford and Schmitt, 2001; Sabatti et al., 2008; Kim et al., 2009), which perform standard regression at all time points, are equivalent to carrying out regression at individual time point separately, thus ignore the longitudinal variations in brain phenotypes. Because AD is a progressive disorder, the imaging phenotype is a quantitative reflection of its neurodegenerative status, and predictive tasks at different points in time are reasonably assumed to be related. For a certain QT, its value at different times may be related, and different QTs at a certain time may also maintain a certain mutual influence. To explore the correlation between longitudinal prediction tasks, several multitasking models have been proposed based on sparse learning (Wang et al., 2012a,b). The main idea of these models is to apply the low-rank constraint to the entire parameter matrix such that the common subspace globally shared by different forecasting tasks can be extracted.

However, longitudinal forecasting tasks are interrelated rather than holistic. Existing methods cannot learn the correct relationship between tasks. Finding such a task group structure is tricky. The straightforward idea is to conduct clustering analysis on the prediction tasks and extract the group structure as a preprocessing step. However, such heuristic steps are not related to the overall longitudinal learning model, so the detected group structure is not optimal for longitudinal study. To bridge this gap, we propose a novel self-learned low-rank structured multitask learning model to simultaneously uncover the interrelations among different imaging phenotype predictions at different time points and utilize such interrelated structures to enhance the prediction tasks. We utilize the Schatten p -norm to identify the shared low-rank common subspace and impose the structure sparsity constraint through $\ell_{2,0+}$ -norm. The new optimization algorithm is derived for the proposed objective. Our new model is applied to the Alzheimer's Disease

Neuroimaging Initiative (ADNI) cohort for imaging phenotype prediction. All empirical results show that the proposed longitudinal feature learning model can more accurately predict phenotypes than related approaches.

Notations: In this article, matrices are all written as uppercase letters whereas vectors as bold lower case letters. For a matrix $M \in \mathbb{R}^{d \times n}$, its i -th row and j -th column are denoted by \mathbf{m}^i , \mathbf{m}_j , respectively, whereas its ij -th element is written as m_{ij} or $M(i,j)$. For a positive value p , the $\ell_{2,p}$ -norm of M is defined

$$\|M\|_{2,p} = \left(\sum_{i=1}^d \left(\sum_{j=1}^n m_{ij}^2 \right)^{\frac{p}{2}} \right)^{\frac{1}{p}} = \left(\sum_{i=1}^d (\|\mathbf{m}^i\|_2)^p \right)^{\frac{1}{p}}.$$

2. TEMPORAL STRUCTURE AUTO-LEARNING PREDICTIVE MODEL

2.1. Illustration of our idea

Our main purpose is to construct a model to simultaneously detect the latent group structure of longitudinal phenotype prediction tasks and study SNP associations across all endophenotypes. As is shown in Figure 1, the four phenotypes marked by the green rectangles have similar distributions and are very likely to be correlated. However, previous models are not capable of capturing such interlinked structures in different task groups. In our model, we expect to uncover such latent subspaces in different groups through low-rank constraints. Meanwhile, the SNP loci marked by the red rectangle, rs429358, appear to be predominant for most response variables. Correspondingly, we impose a sparsity constraint to pick it out. In consequence, our model should be able to capture the group structure within the prediction tasks and utilize this information to select prominent SNPs across the relevant phenotypes. In the next subsection, we will elaborate how to translate these ideas into the new learning model.

2.2. New objective function

To discover the group structure of phenotype prediction tasks, we introduce and optimize a group index matrix set Q . Suppose the tasks come from g groups, then we have $Q = \{Q_1, Q_2, \dots, Q_g\}$, where Q_i is a diagonal matrix and $Q_i \in \{0, 1\}^{cT \times cT}$. For each Q_i , $(Q_i)_{(k,k)} = 1$ means the k -th feature belongs to the i -th group, whereas $(Q_i)_{(k,k)} = 0$ means not. To avoid overlap of subspaces, we maintain the constraint that $\sum_{i=1}^g Q_i = I$.

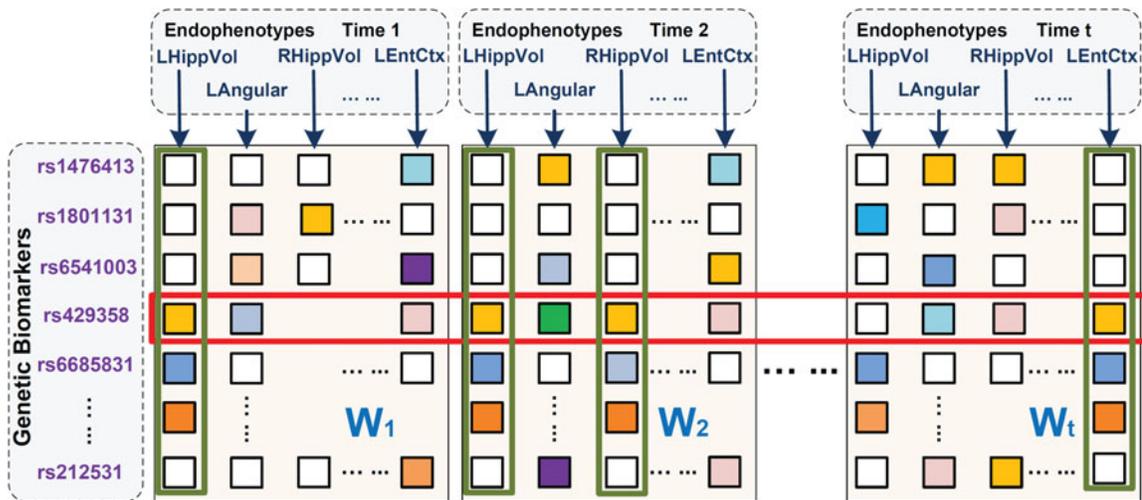


FIG. 1. Illustration of our temporal structure auto-learning regression model. In this figure, parameter matrices at different time points are arranged in the column order. Four tasks marked out by green rectangles obey similar patterns thus have high interrelations while one SNP loci rs429358 enclosed by a red rectangle appears to be correlated with most phenotypes. Our model is meant to uncover the group information among all prediction tasks along the time continuum, that is, cluster the phenotypes in the same latent subspace (phenotypes marked out by green rectangles) into the same group and meanwhile discover genetic biomarkers responsible for most prediction tasks (SNPs marked out by red rectangles). SNPs, single nucleotide polymorphisms.

On the other hand, since SNPs are often correlated and have an overlap in impacting phenotypes, we impose low-rank constraint to uncover the common subspaces shared by prediction tasks. The traditional method to impose low-rank constraint is minimizing trace norm, which is a convex relaxation of rank minimization problem. However, trace norm is not an ideal approximation of the rank minimization. In this study, we use the Schatten p -norm regularization term instead, which approximates the rank minimization better than trace norm (Nie et al., 2012). The definition of Schatten p -norm of a matrix $M \in \mathbb{R}^{m \times n}$ is:

$$\| M \|_{S_p} = (Tr((M^T M)^{\frac{p}{2}}))^{\frac{1}{p}} = \left(\sum_{i=1}^{\min\{m, n\}} \sigma_i^p \right)^{\frac{1}{p}}, \quad (1)$$

where σ_i is the i -th singular value of M . Specially, when $p=1$, the Schatten p -norm of M is exactly the trace norm since

$$\| M \|_* = \| M \|_{S_1} = Tr((M^T M)^{\frac{1}{2}}) = \sum_{i=1}^{\min\{m, n\}} \sigma_i. \quad (2)$$

As we recall, the rank of M can be denoted as $\text{rank}(M) = \sum_{i=1}^{\min\{m, n\}} \sigma_i^0$, where $0^0=0$. Thus, when $0 < p < 1$, Schatten p -norm is a better low-rank regularization than trace norm.

Moreover, since we intend to integrate the SNP selection procedure across multiple learning tasks, here we impose a sparsity constraint. One possible approach is $\ell_{2,1}$ -norm regularization (Obozinski et al., 2006), which is popularly utilized owing to its convex property. However, the real data usually do not satisfy the restricted isometry property (RIP), thus the solution of $\ell_{2,1}$ -norm may not be sparse enough. See (Candès, 2008; Cai and Zhang, 2013) for details. To solve this problem, in our model, we resort to a stricter sparsity constraint, $\ell_{2,0+}$ -norm, which is defined as follows:

$$\| M \|_{2,q} = \left(\sum_{i=1}^d \left(\sum_{j=1}^n m_{ij}^2 \right)^{\frac{q}{2}} \right)^{\frac{1}{q}} = \left(\sum_{i=1}^d (\| \mathbf{m}^i \|_2)^q \right)^{\frac{1}{q}},$$

where q is a positive value. Similarly to the previous discussion, when $0 < q < 1$, $\ell_{2,0+}$ -norm can achieve a more sparse solution than $\ell_{2,1}$ norm.

All in all, we propose a new temporal structure auto-learning model:

$$\begin{aligned} \min_{\| W, Q_i \|_{i=1}^g} & \| X^T W - Y \|_F^2 + \gamma_1 \sum_{i=1}^g (\| W Q_i \|_{S_p}^k) + \gamma_2 \| W \|_{2,0+}, \\ \text{s.t. } & \| Q_i \|_{i=1}^g \in \{0, 1\}^{cT \times cT}, \sum_{i=1}^g Q_i = I. \end{aligned} \quad (3)$$

In Equation (3), we adopt the k -power of Schatten p -norm to make our model more robust. The use of parameter k will be articulated in Section 4. Since it is difficult to solve the proposed new nonconvex and nonsmooth objective, in the next section we propose a novel alternating optimization method for Problem (3).

3. OPTIMIZATION ALGORITHM

In this section, we first introduce an optimization algorithm for general problems with Problem (3) being a special case, and then discuss the detailed optimization steps of Problem (3).

Lemma 1 *Let $g_i(x)$ denote a general function over x , where x can be a scalar, vector or matrix, then we can claim:*

When $\delta \rightarrow 0$, The optimization problem

$$\min_{x \in \mathcal{C}} f(x) + \sum_i Tr((g_i^T(x) g_i(x))^{\frac{\delta}{2}})$$

is equivalent to

$$\min_{x \in \mathcal{C}} f(x) + \sum_i \text{Tr}(g_i^T(x)g_i(x)D_i), \quad \text{where } D_i = \frac{p}{2}(g_i^T(x)g_i(x) + \delta I)^{\frac{p-2}{2}}. \quad (4)$$

Proof. When $\delta \rightarrow 0$, it is apparent that the optimization problem

$$\min_{x \in \mathcal{C}} f(x) + \sum_i \text{Tr}((g_i^T(x)g_i(x) + \delta I)^{\frac{p}{2}}) \quad (5)$$

will reduce to

$$\min_{x \in \mathcal{C}} f(x) + \sum_i \text{Tr}((g_i^T(x)g_i(x))^{\frac{p}{2}}). \quad (6)$$

So we turn the nonsmooth Problem (6) to the smooth Problem (5) where δ is fairly small. The Lagrangian function of Problem (5) is:

$$\mathcal{L}(x, \lambda) = f(x) + \sum_i \text{Tr}((g_i^T(x)g_i(x) + \delta I)^{\frac{p}{2}}) - r(x, \lambda), \quad (7)$$

where $r(x, \lambda)$ is a Lagrangian term for the domain constraint $x \in \mathcal{C}$. Take derivative w.r.t. x and set it to 0, we have:

$$f'(x) + \sum_i \frac{\partial \text{Tr}((g_i^T(x)g_i(x) + \delta I)^{\frac{p}{2}})}{\partial x} - \frac{\partial r(x, \lambda)}{\partial x} = 0. \quad (8)$$

Based on the chain rule (Bentler and Lee, 1978), Equation (8) can be rewritten as:

$$f'(x) + \sum_i \frac{\text{tr}\left(2\frac{p}{2}(g_i^T(x)g_i(x) + \delta I)^{\frac{p-2}{2}}g_i^T(x)\partial g_i(x)\right)}{\partial x} - \frac{\partial r(x, \lambda)}{\partial x} = 0. \quad (9)$$

According to the Karush–Kuhn–Tucker conditions (Rockafellar, 1970), if we can find a solution x that satisfies Equation (9), then we usually find a local/global optimal solution to Problem (5). However, it is intractable to directly find the solution x that satisfies Equation (9). Here we come up with a strategy as follows:

If we define $D_i = \frac{p}{2}(g_i^T(x)g_i(x) + \delta I)^{\frac{p-2}{2}}$ as a given constant, then Equation (9) can be reduced to

$$f'(x) + \sum_i \frac{\text{tr}(2D_i g_i^T(x)\partial g_i(x))}{\partial x} - \frac{\partial r(x, \lambda)}{\partial x} = 0. \quad (10)$$

Based on the chain rule (Bentler and Lee, 1978), the optimal solution x^* of Equation (10) is also an optimal solution to the following problem:

$$\min_{x \in \mathcal{C}} f(x) + \sum_i \text{Tr}(g_i^T(x)g_i(x)D_i). \quad (11)$$

Based on this observation, we can first guess a solution x , next calculate D_i based on the current solution x , and then update the current solution x by the optimal solution of Problem (11) on the basis of the calculated D_i . We can iteratively perform this procedure until it converges. ■

According to Lemma 1 and the property of Q_i that $Q_i Q_i^T = Q_i$, Problem (3) is equivalent to:

$$\begin{aligned} \min_{W, Q_i} \quad & \|X^T W - Y\|_F^2 + \gamma_1 \sum_{i=1}^g \text{Tr}(W Q_i W^T D_i) + \gamma_2 \text{Tr}(W W^T B) \\ \text{s.t.} \quad & Q_i \in \{0, 1\}^{cT \times cT}, \sum_{i=1}^g Q_i = I, \end{aligned} \quad (12)$$

where D_i is defined as:

$$D_i = \frac{kp}{2} (\|WQ_i\|_{S_p}^p)^{k-1} (WQ_i W^T + \delta_1 I)^{\frac{p-2}{2}}, \quad (13)$$

and B is defined as a diagonal matrix with the l -th diagonal element to be:

$$b_{ll} = \frac{q}{2} (\mathbf{w}^l (\mathbf{w}^l)^T + \delta_2 I)^{\frac{q-2}{2}}, \quad (14)$$

and δ_1 and δ_2 are two fairly small parameters close to 0.

We can solve Problem (12) through alternating optimization.

The first step is fixing W and solving Q , then Problem (12) becomes:

$$\min_{\|Q_i\|_{i=1}^g} \sum_{i=1}^g \text{Tr}((W^T D_i W) Q_i) \quad \text{s.t.} \quad \|Q_i\|_{i=1}^g \in \{0, 1\}^{cT \times cT}, \sum_{i=1}^g Q_i = I.$$

Let $A_i = W^T D_i W$, then the solution of each Q_i is as follows:

$$Q_i(k, k) = \begin{cases} 1, & i = \arg \min_j A_j(k, k) \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

The second step is fixing Q and solving W . Problem (12) becomes:

$$\min_W \|X^T W - Y\|_F^2 + \gamma_1 \sum_{i=1}^g \text{Tr}(W Q_i W^T D_i) + \gamma_2 \text{Tr}(W W^T B),$$

which can be further decoupled for each column of W as follows:

$$\min_{\mathbf{w}_k} \|X^T \mathbf{w}_k - \mathbf{y}_k\|_2^2 + \gamma_1 \sum_{i=1}^g (Q_i(k, k) \mathbf{w}_k^T D_i \mathbf{w}_k) + \gamma_2 \mathbf{w}_k^T B \mathbf{w}_k.$$

Taking derivative w.r.t. \mathbf{w}_k in the above problem and set it to 0, we get:

$$\mathbf{w}_k = (X X^T + \gamma_1 (\sum_{i=1}^g Q_i(k, k) D_i) + \gamma_2 B)^{-1} X \mathbf{y}_k. \quad (16)$$

We can iteratively update D , Q , B , and W with the alternative steps mentioned above and the algorithm for Problem (12) is summarized in Algorithm 1.

Algorithm 1 Algorithm to solve Problem (12).

Input:

SNP matrix $X \in \mathbb{R}^{d \times n}$, longitudinal phenotype matrix $Y = [Y_1 Y_2 \dots Y_T]$ where $Y_t|_{t=1}^T \in \mathbb{R}^{n \times c}$, parameter $\delta_1 = 10^{-12}$, and $\delta_2 = 10^{-12}$, number of groups g .

Output:

Weight matrix $W = [W_1 W_2 \dots W_T]$ where $W_t|_{t=1}^T \in \mathbb{R}^{d \times c}$ and g different group matrices $Q_i|_{i=1}^g \in \mathbb{R}^{cT \times cT}$, which groups the tasks into exactly g subspaces.

Initialize W by the optimal solution to the ridge regression problem:

$$\min_W \|W^T X - Y\|_F^2 + \|W\|_F^2$$

Initialize Q matrices randomly

while not converge **do**

1. Update D according to the definition in Equation (13).
2. Update Q according to the solution in Equation (15).
3. Update B according to the definition in Equation (14).
4. Update W , where the solution to the k -th column of W is displayed in Equation (16).

end while

3.1. Convergence analysis of the reweighted algorithm

Our algorithm uses the alternating optimization method to update variables, whose convergence has already been proved in Bezdek and Hathaway (2003). As for the newly proposed reweighted algorithm, we will provide its convergence proof as follows. Before proving convergence of the reweighted algorithm shown in Lemma 1, we first introduce several lemmas:

Lemma 2 For any $\sigma > 0$, the following inequality holds when $0 < p \leq 2$:

$$\frac{p}{2}\sigma - \sigma^{\frac{p}{2}} + \frac{2-p}{2} \geq 0. \tag{17}$$

Proof. Denote $f(\sigma) = p\sigma - 2\sigma^{\frac{p}{2}} + 2 - p$, we have the following derivatives:

$$f'(\sigma) = p(1 - \sigma^{\frac{p-2}{2}}), \quad \text{and} \quad f''(\sigma) = \frac{p(2-p)}{2} \sigma_i^{\frac{p-4}{2}}.$$

Obviously, when $\sigma > 0$ and $0 < p \leq 2$, then $f''(\sigma) \geq 0$ and $\sigma = 1$ is the only point that $f'(\sigma) = 0$. Note that $f(1) = 0$, thus when $\sigma > 0$ and $0 < p \leq 2$, then $f(\sigma) \geq 0$, which indicates Equation (17). ■

Lemma 3 (Ruhe, 1970) For any positive definite matrices \tilde{M}, M with the same size, suppose the eigendecomposition $\tilde{M} = U\Sigma U^T$, $M = V\Lambda V^T$, where the eigenvalues in Σ is in increasing order and the eigenvalues in Λ is in decreasing order. Then the following inequality holds:

$$\text{Tr}(\tilde{M}M) \geq \text{Tr}(\Sigma\Lambda). \tag{18}$$

Lemma 4 For any positive definite matrices \tilde{M}, M with the same size, the following inequality holds when $0 < p \leq 2$:

$$\text{Tr}(\tilde{M}^{\frac{p}{2}}) - \frac{p}{2}\text{Tr}(\tilde{M}M^{\frac{p-2}{2}}) \leq \text{Tr}(M^{\frac{p}{2}}) - \frac{p}{2}\text{Tr}(MM^{\frac{p-2}{2}}). \tag{19}$$

Proof. For any $\sigma > 0$, $\lambda > 0$ and $0 < p \leq 2$, according to Lemma 2 we have $\frac{p}{2}(\frac{\sigma}{\lambda}) - (\frac{\sigma}{\lambda})^{\frac{p}{2}} + \frac{2-p}{2} \geq 0$, which indicates

$$\frac{p}{2}\sigma\lambda^{\frac{p-2}{2}} - \sigma^{\frac{p}{2}} + \frac{2-p}{2}\lambda^{\frac{p}{2}} \geq 0. \tag{20}$$

Suppose the eigendecomposition $\tilde{M} = U\Sigma U^T$, $M = V\Lambda V^T$, where the eigenvalues in Σ is in increasing order and the eigenvalues in Λ is in decreasing order. According to Equation (20), we have

$$\frac{p}{2}\text{Tr}(\Sigma\Lambda^{\frac{p-2}{2}}) - \text{Tr}(\Sigma^{\frac{p}{2}}) + \frac{2-p}{2}\text{Tr}(\Lambda^{\frac{p}{2}}) \geq 0, \tag{21}$$

and according to Lemma 3 we have

$$\begin{aligned} \frac{p}{2}\text{Tr}(\tilde{M}M^{\frac{p-2}{2}}) - \frac{p}{2}\text{Tr}(\Sigma\Lambda^{\frac{p-2}{2}}) &\geq 0, \\ \frac{p}{2}\text{Tr}(\tilde{M}M^{\frac{p-2}{2}}) - \text{Tr}(\Sigma^{\frac{p}{2}}) + \frac{2-p}{2}\text{Tr}(\Lambda^{\frac{p}{2}}) &\geq 0. \end{aligned} \tag{22}$$

Note that $\text{Tr}(\tilde{M}^{\frac{p}{2}}) = \text{Tr}(\Sigma^{\frac{p}{2}})$ and $\text{Tr}(M^{\frac{p}{2}}) = \text{Tr}(\Lambda^{\frac{p}{2}})$, so we have

$$\begin{aligned}
& \frac{p}{2} \text{Tr}(\tilde{M}M^{\frac{p-2}{2}}) - \text{Tr}(\tilde{M}^{\frac{p}{2}}) + \frac{2-p}{2} \text{Tr}(M^{\frac{p}{2}}) \geq 0 \\
& \Rightarrow \text{Tr}(\tilde{M}^{\frac{p}{2}}) - \frac{p}{2} \text{Tr}(\tilde{M}M^{\frac{p-2}{2}}) \leq \frac{2-p}{2} \text{Tr}(M^{\frac{p}{2}}) \\
& \Rightarrow \text{Tr}(\tilde{M}^{\frac{p}{2}}) - \frac{p}{2} \text{Tr}(\tilde{M}M^{\frac{p-2}{2}}) \leq \text{Tr}(M^{\frac{p}{2}}) - \frac{p}{2} \text{Tr}(MM^{\frac{p-2}{2}}),
\end{aligned}$$

which completes the proof. \blacksquare

As a result, we have the following theorem:

Theorem 1 *The reweighted algorithm shown in Lemma 1, which optimizes Problem (4) instead of Problem (5), will monotonically decrease the objective of Problem (5) in each iteration, until the algorithm converges.*

Proof: In Problem (4), suppose the updated x is \tilde{x} . Thus we know

$$f(\tilde{x}) + \sum_i \text{Tr}(g_i^T(\tilde{x})g_i(\tilde{x})D_i) \leq f(x) + \sum_i \text{Tr}(g_i^T(x)g_i(x)D_i), \quad (23)$$

where the equality holds when and only when the algorithm converges.

For each i , according to Lemma 4, we have

$$\begin{aligned}
& \text{Tr}((g_i^T(\tilde{x})g_i(\tilde{x}) + \delta I)^{\frac{p}{2}}) - \frac{p}{2} \text{Tr}(g_i^T(\tilde{x})g_i(\tilde{x})(g_i^T(x)g_i(x) + \delta I)^{\frac{p-2}{2}}) \\
& \leq \text{Tr}((g_i^T(x)g_i(x) + \delta I)^{\frac{p}{2}}) - \frac{p}{2} \text{Tr}(g_i^T(x)g_i(x)(g_i^T(x)g_i(x) + \delta I)^{\frac{p-2}{2}}).
\end{aligned} \quad (24)$$

Note that $D_i = \frac{p}{2}(g_i^T(x)g_i(x) + \delta I)^{\frac{p-2}{2}}$, so for each i we have

$$\begin{aligned}
& \text{Tr}((g_i^T(\tilde{x})g_i(\tilde{x}) + \delta I)^{\frac{p}{2}}) - \text{Tr}(g_i^T(\tilde{x})g_i(\tilde{x})D_i) \\
& \leq \text{Tr}((g_i^T(x)g_i(x) + \delta I)^{\frac{p}{2}}) - \text{Tr}(g_i^T(x)g_i(x)D_i).
\end{aligned} \quad (25)$$

Then we have

$$\begin{aligned}
& \sum_i \text{Tr}((g_i^T(\tilde{x})g_i(\tilde{x}) + \delta I)^{\frac{p}{2}}) - \sum_i \text{Tr}(g_i^T(\tilde{x})g_i(\tilde{x})D_i) \\
& \leq \sum_i \text{Tr}((g_i^T(x)g_i(x) + \delta I)^{\frac{p}{2}}) - \sum_i \text{Tr}(g_i^T(x)g_i(x)D_i).
\end{aligned} \quad (26)$$

Summing Equations (23) and (26) in the two sides, we arrive at

$$\begin{aligned}
& f(\tilde{x}) + \sum_i \text{Tr}((g_i^T(\tilde{x})g_i(\tilde{x}) + \delta I)^{\frac{p}{2}}) \\
& \leq f(x) + \sum_i \text{Tr}((g_i^T(x)g_i(x) + \delta I)^{\frac{p}{2}}).
\end{aligned} \quad (27)$$

Note that the equality in Equation (27) holds only when the algorithm converges. Thus the reweighted algorithm shown in Lemma 1 will monotonically decrease the objective of Problem (5) in each iteration until the algorithm converges.

4. DISCUSSION OF PARAMETERS

In our model, we introduced several parameters to make it more general and adaptive to various circumstances. In this study, we analyze the functionality of each parameter in detail.

In Problem (3), parameter p and q are norm parameters proposed for the two regularization terms. For p , it adjusts the stringency of the low-rank constraint. As analyzed in Section 3, Schatten p -norm makes a

stricter low-rank constraint than trace norm when $0 < p < 1$. The closer p is to 0, the more rigorous low-rank constraint the regularization term $\|M\|_{S_p}^p$ imposes. The rationale is similar for parameter q . The $\ell_{2,0+}$ -norm is a better approximation of $\ell_{2,0}$ -norm than $\ell_{2,1}$ -norm when q lies in the range of $(0, 1)$, thus makes our learned parameter matrix more sparse.

In the low-rank regularization term $\|M\|_{S_p}^p$, when p is small, the number of local solutions becomes more, thus leading our Model (3) to be more sensitive to outliers. Under this condition, we use k -power of this term as $(\|M\|_{S_p}^p)^k$ to make our model robust. According to experimental experience, the value of k can be determined in the range of 2–3.

The parameters γ_1 and γ_2 are proposed to balance the importance of two regularization terms. Larger γ_1 leads to more attention on the low-rank constraint, whereas larger γ_2 lays more emphasis on the sparse structure. These two parameters can be adjusted to accommodate different cases.

In our empirical section, we did not make too much effort on tuning these parameters. Instead, in fairness to other comparing methods, we just simply set each parameters to a reasonable value. Although these parameters brought about great challenges in solving our optimization problem, they make our model more flexible and adaptive to different settings and conditions.

5. EXPERIMENTAL RESULTS

In this section, we evaluate the prediction performance of our proposed method using both synthetic and real data. Our goal is to uncover the latent subspace structure of the prediction tasks and meanwhile select a subset of SNPs responsible for their variation. An executable program of our model is provided (Wang, 2017).

5.1. Experiments on synthetic data

First of all, we utilize synthetic data to illustrate the effectiveness of our model. Our synthetic data are composed of 30 features and 3 groups of tasks from 4 different time points. Each group includes four tasks, which are identical to each other up to a scaling. After we generated weight matrix W in this way, we constructed a random X , including 10,000 samples and get Y according to $Y = X^T W$.

Figure 2a shows the original W_o matrix, where weight matrices in different time points are arranged in a column order. According to the construction process of W_o , tasks in W_o should form three low-rank subspaces. For easily visualizing this low-rank structure, we reshuffled tasks in W_o by putting tasks in the same group together and formed Figure 2b. Now the low-rank structure within the synthetic data can be easily detected, where every four columns in Figure 2b make a low-rank subspace. We applied our method to this synthetic data and plotted our learned W matrix in Figure 2c. To evaluate the structure of W , we did the rearrangement likewise. By comparing Figure 2b and d, we note that our method has successfully uncovered the group structure of the synthetic data and recovered the parameter matrix W in an accurate way.

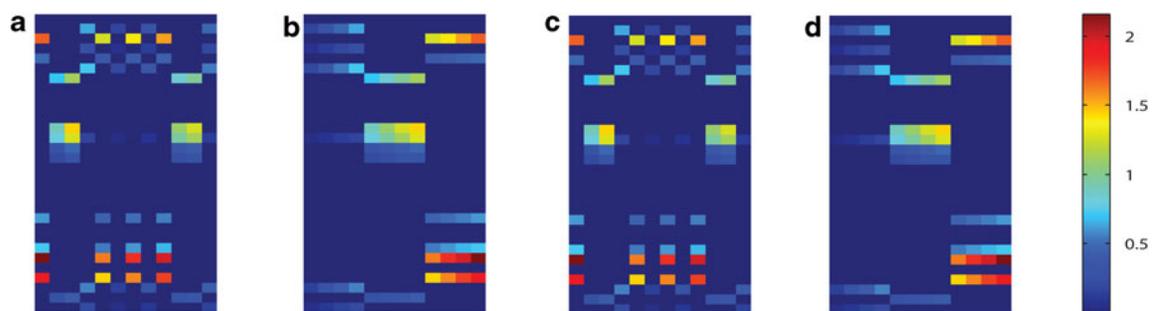


FIG. 2. Visualization of the synthetic parameter matrix W . Columns of W denote 12 prediction tasks from 4 different time points, whereas rows of W correspond to 30 features. These 12 tasks are equally divided into 3 groups, where tasks in the same group are identical to each other up to a scaling factor. (a) The original weight matrix W_o . (b) Rearrangement of columns in W_o by putting tasks in the same group together, such that the low-rank structure of W_o is more clear. (c) The learned W matrix by our model. (d) Rearrangement of W learned by our model.

5.2. Experimental settings on real benchmark data

In the following we evaluate our method on real benchmark datasets. We compare our method with all the counterparts discussed in the introduction section, which are: multivariate linear regression (LR), multivariate ridge regression (RR), multitask trace-norm regression (MTT) (Ji and Ye, 2009), Multitask $\ell_{2,1}$ -norm Regression (MTL21) (Evgeniou and Pontil, 2007), and their combination (MTT+L21) (Wang et al., 2012a,b).

In our preexperiments, we found the performance of our method to be relatively stable with parameters in the reasonable range (data not shown). For simplicity, we set $\gamma_1 = 1$, $\gamma_2 = 1$, $p = 0.8$, $q = 0.1$, and $k = 2.5$ without tuning. The definition and reasonable range of these parameters has been discussed in Section 4.

As the evaluation metric, we reported the root mean square error (RMSE) and correlation coefficient (CorCoe) between the predicted and actual scores in out-of-sample prediction. In our experiment, the RMSE was normalized by the Frobenius norm of the ground truth matrix. Better performance relates with lower RMSE or higher CorCoe value. We utilized the fivefold cross validation technique and reported the average RMSE and CorCoe on these five trials for each method.

5.3. Description of ADNI data

Data used in this work were obtained from the ADNI database (adni.loni.usc.edu). One goal of ADNI is to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org.

The genotype data (Saykin et al., 2010) of all non-Hispanic Caucasian participants from the ADNI Phase 1 cohort were used in this study. They were genotyped using the Human 610-Quad BeadChip. Among all the SNPs, only SNPs, within the boundary of ± 20 kbp base pairs of the 153 AD candidate genes listed on the AlzGene database (www.alzgene.org) as of April 18, 2011 (Bertram et al., 2007), were selected after the standard quality control (QC) and imputation steps. The QC criteria for the SNP data include (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. As the second preprocessing step, the QC'ed SNPs were imputed using the MaCH software (Li et al., 2010) to estimate the missing genotypes. As a result, our analyses included 3576 SNPs extracted from 153 genes (boundary: ± 20 kbp) using the ANNOVAR annotation (www.openbioinformatics.org/annovar/).

Two widely employed automated MRI analysis techniques were used to process and extract imaging phenotypes from scans of ADNI participants as previously described (Shen et al., 2010). First, Voxel-Based Morphometry (VBM) (Ashburner and Friston, 2000) was performed to define global gray matter (GM) density maps and extract local GM density values for 90 target regions. Second, automated parcellation through FreeSurfer V4 (Fischl et al., 2002) was conducted to define volumetric and cortical thickness values for 90 regions of interest (ROIs) and to extract total intracranial volume (ICV). Further details are available in Shen et al. (2010). All these measures were adjusted for the baseline ICV using the regression weights derived from the healthy control (HC) participants. The time points examined in this study for imaging markers included baseline (BL), Month 6 (M6), Month 12 (M12), and Month 24 (M24). All the participants with no missing BL/M6/M12/M24 MRI measurements were included in this study, including 96 AD samples, and 219 MCI samples and 174 HC samples.

5.4. Performance comparison on ADNI cohort

In this study, we assessed the ability of our method to predict a set of imaging biomarkers through genetic variations. We tracked the process along the time axis and intended to uncover the latent subspace structure maintained by phenotypes and meanwhile capture a subset of SNPs influencing the phenotypes in a certain subspace.

We examined the cases where the number of selected SNPs were $\{30, 40, \dots, 80\}$. The experimental results are summarized in Tables 1 and 2. We observe that our method consistently outperforms other methods in most cases. The reasons go as follows: multivariate regression and ridge regression assumed

TABLE 1. BIOMARKER “VBM” AND “FREESURFER” (UPPER TABLE FOR “VBM,” LOWER TABLE FOR “FREESURFER”) PREDICTION COMPARISON THROUGH ROOT MEAN SQUARE ERROR MEASUREMENT WITH DIFFERENT NUMBER OF SELECTED SNPs

<i>No. of SNPs</i>	<i>LR</i>	<i>RR</i>	<i>MTT</i>	<i>MTL21</i>	<i>MTT+L21</i>	<i>OURS</i>
30	0.9277±0.0122	0.9278±0.0122	0.9270±0.0135	0.9278±0.0122	0.9276±0.0126	0.9266±0.0124
40	0.9106±0.0147	0.9100±0.0146	0.9099±0.0141	0.9100±0.0146	0.9099±0.0141	0.9061±0.0087
50	0.8914±0.0177	0.8916±0.0176	0.8934±0.0159	0.8916±0.0176	0.8919±0.0176	0.8913±0.0084
60	0.8843±0.0216	0.8846±0.0216	0.8831±0.0192	0.8846±0.0216	0.8848±0.0215	0.8726±0.0051
70	0.8661±0.0204	0.8663±0.0203	0.8675±0.0211	0.8663±0.0203	0.8666±0.0202	0.8615±0.0045
80	0.8509±0.0250	0.8503±0.0244	0.8500±0.0242	0.8503±0.0244	0.8500±0.0242	0.8482±0.0042
30	0.9667±0.0145	0.9664±0.0145	0.9667±0.0145	0.9664±0.0145	0.9664±0.0146	0.9558±0.0147
40	0.9569±0.0113	0.9569±0.0112	0.9569±0.0113	0.9569±0.0112	0.9569±0.0114	0.9436±0.0113
50	0.9502±0.0141	0.9503±0.0141	0.9502±0.0141	0.9503±0.0141	0.9503±0.0142	0.9350±0.0143
60	0.9416±0.0106	0.9417±0.0106	0.9416±0.0106	0.9417±0.0106	0.9415±0.0107	0.9234±0.0106
70	0.9319±0.0105	0.9316±0.0105	0.9319±0.0105	0.9316±0.0105	0.9321±0.0106	0.9096±0.0106
80	0.9246±0.0093	0.9247±0.0093	0.9246±0.0093	0.9247±0.0093	0.9247±0.0094	0.9012±0.0094

Better performance corresponds to lower root mean square error.

LR, linear regression; MTT, multitask trace-norm regression; RR, ridge regression; OURS, the proposed temporal structure auto-learning predictive model; SNPs, single nucleotide polymorphisms; VBM, Voxel-Based Morphometry.

the imaging features at different time points to be independent, thus did not consider the correlations within. Their neglects of the interrelations among the data were detrimental to their prediction performance.

As for MTTTrace, MTL21 and their combination MTTTrace+MTL21, even though they take into account the inner connection information of imaging phenotypes, they simply constrain all phenotypes to a global space thus cannot handle the possible group structure therein. That is why they may outweigh the standard methods in some cases, but cannot outperform our proposed method. As for our proposed method, not only did we capture the latent structure among the longitudinal phenotypes, but we also selected a set of responsible SNPs at the same time. All in all, our model can capture SNPs responsible for some but not necessarily all imaging phenotypes along the time continuum, which save more effective information in the prediction.

TABLE 2. BIOMARKER “VBM” AND “FREESURFER” (UPPER TABLE FOR “VBM,” LOWER TABLE FOR “FREESURFER”) PREDICTION COMPARISON VIA CORRELATION COEFFICIENT MEASUREMENT WITH DIFFERENT NUMBER OF SELECTED SNPs

<i>No. of SNPs</i>	<i>LR</i>	<i>RR</i>	<i>MTT</i>	<i>MTL21</i>	<i>MTT+L21</i>	<i>OURS</i>
30	0.4186±0.0092	0.4180±0.0098	0.4232±0.0088	0.4180±0.0098	0.4216±0.0077	0.6193±0.0177
40	0.4284±0.0163	0.4282±0.0166	0.4333±0.0181	0.4282±0.0166	0.4312±0.0157	0.6294±0.0071
50	0.4476±0.0395	0.4470±0.0400	0.4526±0.0384	0.4470±0.0400	0.4508±0.0385	0.6362±0.0094
60	0.4477±0.0391	0.4471±0.0397	0.4513±0.0388	0.4471±0.0397	0.4510±0.0380	0.6435±0.0196
70	0.4502±0.0345	0.4490±0.0356	0.4547±0.0335	0.4490±0.0356	0.4529±0.0339	0.6460±0.0129
80	0.4467±0.0287	0.4470±0.0286	0.4514±0.0271	0.4470±0.0286	0.4508±0.0268	0.6521±0.0083
30	0.9007±0.0145	0.9008±0.0145	0.9007±0.0145	0.9008±0.0145	0.9010±0.0146	0.9019±0.0147
40	0.9049±0.0113	0.9051±0.0112	0.9049±0.0113	0.9051±0.0112	0.9052±0.0114	0.9060±0.0113
50	0.9045±0.0141	0.9046±0.0141	0.9045±0.0141	0.9046±0.0141	0.9048±0.0142	0.9057±0.0143
60	0.9079±0.0106	0.9080±0.0106	0.9079±0.0106	0.9080±0.0106	0.9081±0.0107	0.9089±0.0106
70	0.9094±0.0105	0.9096±0.0105	0.9094±0.0105	0.9096±0.0105	0.9097±0.0106	0.9106±0.0106
80	0.9114±0.0093	0.9115±0.0093	0.9114±0.0093	0.9115±0.0093	0.9117±0.0094	0.9124±0.0094

Better performance corresponds to higher correlation coefficient.

5.5. Experiments on convergence analysis

We applied our method to the ADNI cohort with two p -values (i.e., 0.1 and 0.8), and recorded the objective values of our proposed model in Equation (12) in each iteration. As shown in Figure 3, no matter what the p -value is, our model always converges within 10 iterations.

5.6. Identification of top selected SNPs

Shown in Figure 4 are the heat maps of top regression weights between imaging QTs and SNPs. APOE-rs429358, the well-known major AD risk factor, demonstrated relatively strong predictive power in both analyses: (1) In FreeSurfer analysis, it predicts mainly the cerebral cortex volume at M06, M12, and M24. (2) In VBM analysis, it predicts the GM densities of amygdala, hippocampus, and parahippocampal gyrus at multiple time points (Fig. 5). Both patterns are reassuring, since APOE-rs429358 and atrophy patterns of the entire cortex as well as medial temporal regions (including amygdala, hippocampus, and parahippocampal gyrus) are all highly associated with AD.

In addition, APOE-rs429358 has been shown to be related to medial temporal lobe atrophy (Pereira et al., 2014), hippocampal atrophy (Andrawis et al., 2012), and cortical atrophy (Risacher et al., 2013). Variants within membrane-spanning 4-domains subfamily A (MS4A) gene cluster are some other recently discovered AD risk factors (Ma et al., 2014). Our analysis also demonstrated their associations with imaging QTs: (1) MS4A4E-rs670139 predicts cerebral white matter volume at BL, M06, and M12; and

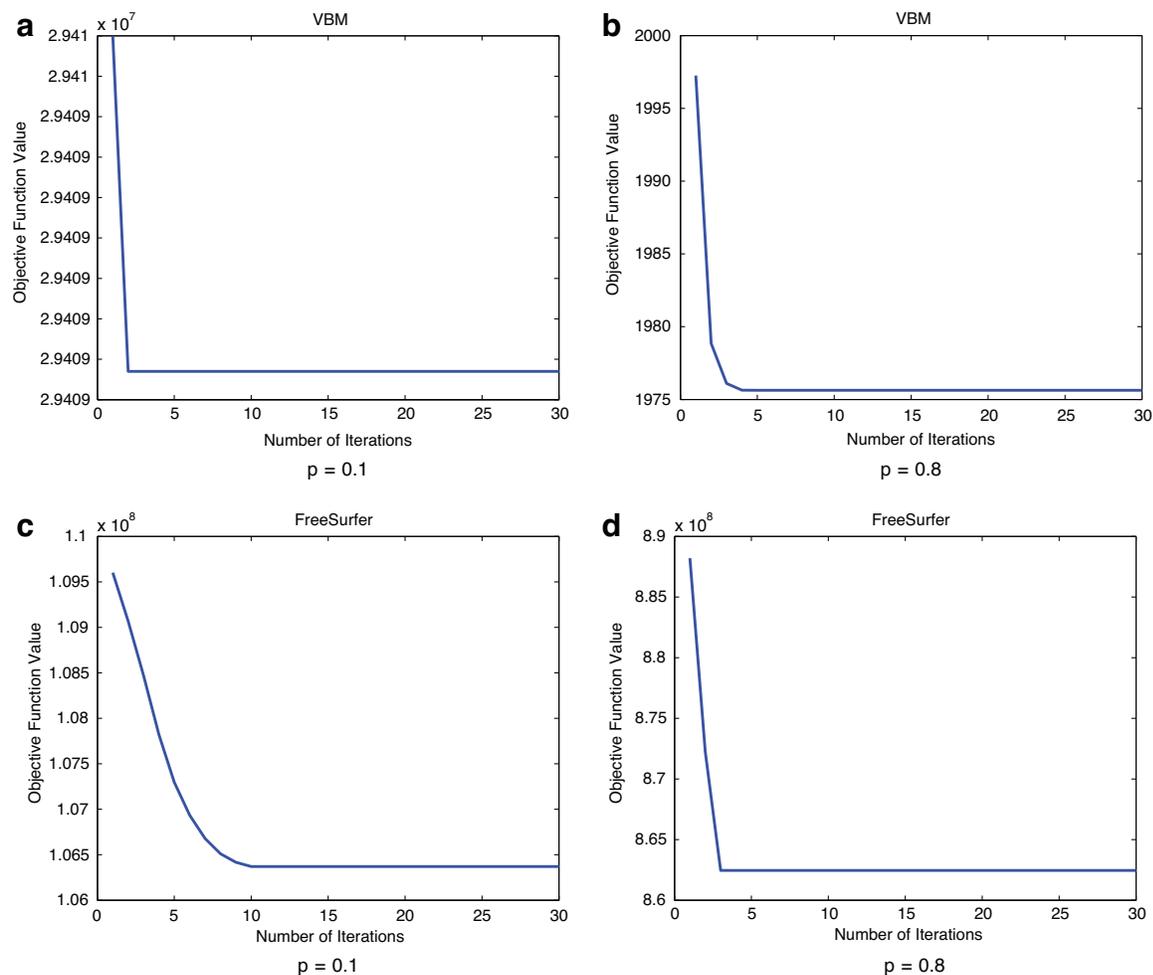


FIG. 3. Objective function value of our proposed model in Equation (12) in each iteration with different p -values in the experiments on “VBM” biomarker (a, b) and “FreeSurfer” biomarker (c, d). VBM, Voxel-Based Morphometry.

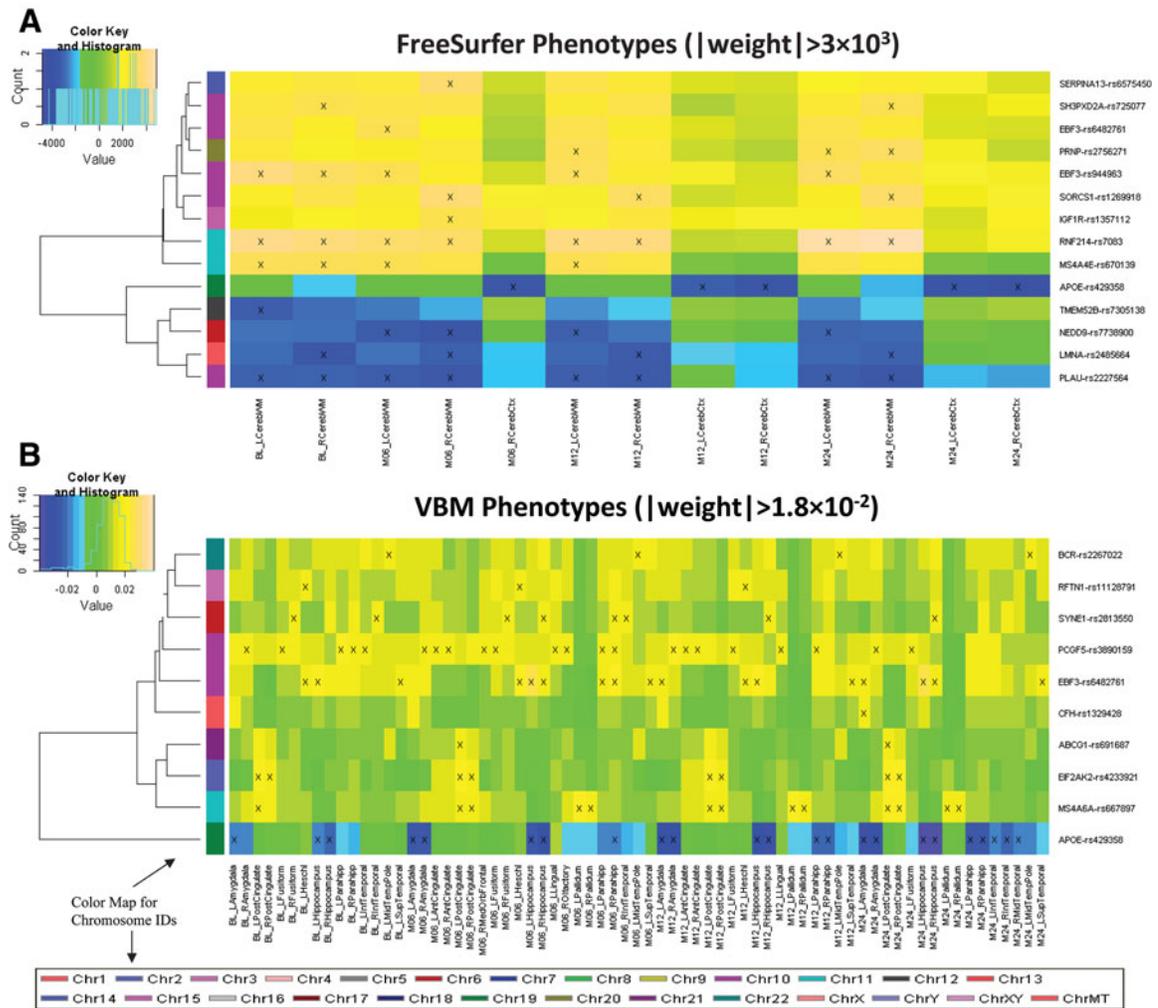


FIG. 4. Heat maps of top regression weights between QTs and SNPs filtered by a user-specified weight threshold: (A) FreeSurfer QTs and (B) VBM QTs. Weights from each regression analysis are color mapped and displayed in the heat maps. Heat map blocks labeled with “x” reach the weight threshold. Only top SNPs and QTs are included in the heat maps, and so each row (SNP) and column (QT) have at least one “x” block. Dendrograms derived from hierarchical clustering are plotted for SNPs. The color bar on the left side of the heat map codes the chromosome IDs for the corresponding SNPs. QTs, quantitative traits.

(2) MS4A6A-rs667897 predicts GM densities of posterior cingulate and pallidum at almost all time points. Another interesting finding is EBF3-rs482761, which is identified in both analyses: (1) In FreeSurfer analysis, it predicts left cerebral white matter volume at M06; and (2) in VBM analysis, it predicts GM densities of left Heschl’s gyri, left superior temporal gyri, hippocampus, and left amygdala at multiple time points. EBF3 (early B-cell factor 3) is a protein-coding gene and has been associated with neurogenesis and glioblastoma. In general, both FreeSurfer and VBM analyses picked up similar regions across different time points. These identified imaging genomic associations warrant further investigation in independent cohorts. If replicated, these findings can potentially contribute to biomarker discovery for diagnosis and drug design.

6. CONCLUSIONS

In this article, we proposed a novel temporal structure auto-learning model to study the associations between genetic variations and longitudinal imaging phenotypes. Our model can simultaneously uncover

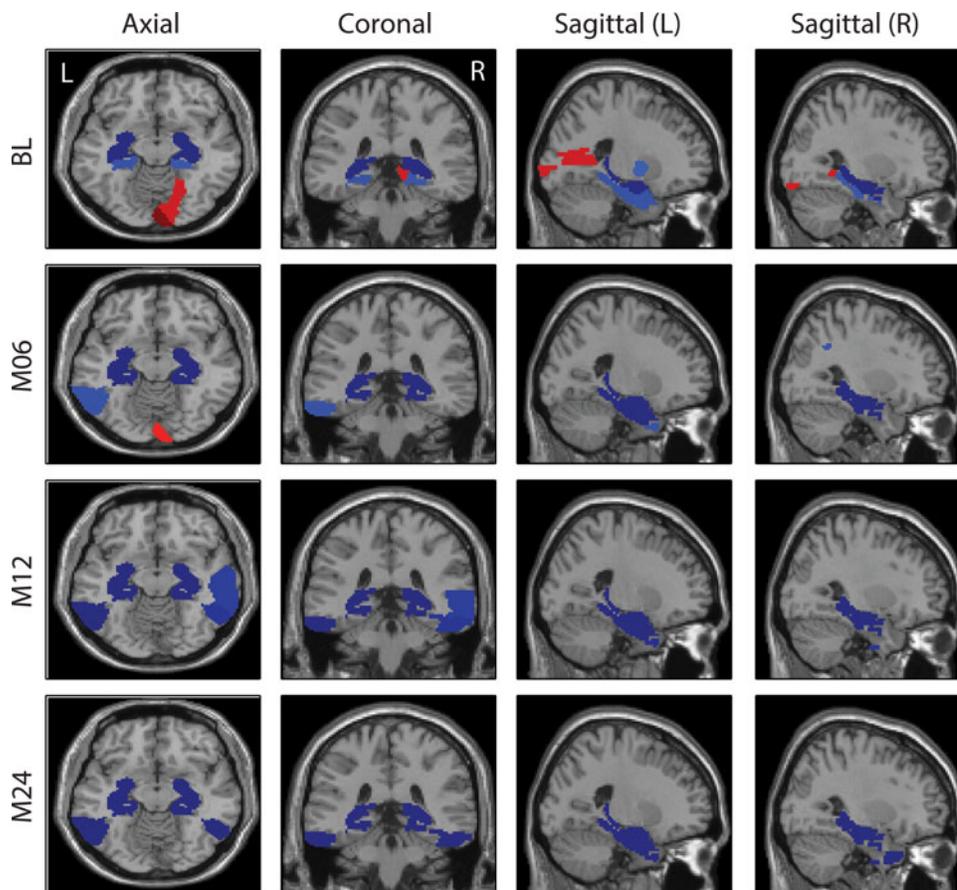


FIG. 5. Top 10 weights of APOE-rs429358 mapped on the brain for the VBM analysis.

the interrelation structures existing in different prediction tasks and utilize such learned interrelated structures to enhance the feature learning model. Moreover, we utilized the Schatten p -norm to extract the common subspace shared by the prediction tasks. Our new model is applied to ADNI cohort for neuroimaging phenotypes prediction through SNPs. We conducted experiments on both synthetic and real benchmark data. Empirical results validated the effectiveness of our model by demonstrating the improved prediction performance compared with related methods. In real data analysis, we also identified a set of interesting and biologically meaningful imaging genomic associations, showing the potential for biomarker discovery in disease diagnosis and drug design.

ACKNOWLEDGMENTS

This work was partially supported by the National Science Foundation [IIS 1302675 to H.H., IIS 1344152 to H.H., DBI 1356628 to H.H., IIS 1619308 to H.H., and IIS 1633753 to H.H.] at PITT and [IIS 1622526 to L.S.] at IU; and by the National Institutes of Health [R01 LM011360 to L.S. and A.S., U01 AG024904 to Michael Weiner and A.S., RC2 AG036535 to Michael Weiner and A.S., R01 AG19771 to A.S., P30 AG10133 to A.S., and UL1 TR001108 to Anantha Shekhar], and [R01 AG049371 to H.H.] at PITT.

AUTHOR DISCLOSURE STATEMENT

The authors declare that no competing financial interests exist.

REFERENCES

- Alzheimer's Association. 2016. 2016 Alzheimer's disease facts and figures. *Alzheimer's Dement.* 12, 459–509.
- Andrawis, J.P., Hwang, K.S., Green, A.E., et al. 2012. Effects of ApoE4 and maternal history of dementia on hippocampal atrophy. *Neurobiol. Aging* 33, 856–866.
- Ashburner, J., and Friston, K.J. 2000. Voxel-based morphometry—The methods. *Neuroimage* 11(6 Pt 1), 805–821.
- Ashford, J.W., and Schmitt, F.A. 2001. Modeling the time-course of Alzheimer dementia. *Curr. Psychiatry Rep.* 3, 20–28.
- Bentler, P.M., and Lee, S.-Y. 1978. Matrix derivatives with chain rule and rules for simple, Hadamard, and Kronecker products. *J. Math. Psychol.* 17, 255–262.
- Bertram, L., McQueen, M.B., Mullin, K., et al. 2007. Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database. *Nat. Genet.* 39, 17–23.
- Bezdek, J.C., and Hathaway, R.J. 2003. Convergence of alternating optimization. *Neural Parallel Sci. Comput.* 11, 351–368.
- Bis, J.C., DeCarli, C., Smith, A.V., et al. 2012. Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat. Genet.* 44, 545–551.
- Burns, A., and Iliffe, S. 2009. Alzheimer's disease. *Br. Med. J.* 338, b158.
- Cai, T.T., and Zhang, A. 2013. Sharp RIP bound for sparse signal and low-rank matrix recovery. *Appl. Comput. Harmon. Anal.* 35, 74–93.
- Candès, E.J. 2008. The restricted isometry property and its implications for compressed sensing. *C. R. Math.* 346, 589–592.
- Evgeniou, A., and Pontil, M. 2007. Multi-task feature learning. *Adv. Neural Inf. Process. Syst.* 19, 41.
- Fischl, B., Salat, D.H., Busa, E., et al. 2002. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.
- Hariri, A.R., Drabant, E.M., and Weinberger, D.R. 2006. Imaging genetics: Perspectives from studies of genetically driven variation in serotonin function and corticolimbic affective processing. *Biol. Psychiatry* 59, 888–897.
- Harold, D., Abraham, R., Hollingworth, P., et al. 2009. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat. Genet.* 41, 1088–1093.
- Ji, S., and Ye, J. 2009. An accelerated gradient method for trace norm minimization, 457–464. In *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM.
- Khachaturian, Z.S. 1985. Diagnosis of Alzheimer's disease. *Arch. Neurol.* 42, 1097–1105.
- Kim, S., Sohn, K.-A., and Xing, E.P. 2009. A multivariate regression approach to association analysis of a quantitative trait network. *Bioinformatics* 25, i204–i212.
- Li, Y., Willer, C.J., Ding, J., et al. 2010. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* 34, 816–834.
- Ma, J., Yu, J.T., and Tan, L. 2014. MS4A cluster in Alzheimer's disease. *Mol. Neurobiol.* 51, 1240–1248.
- Nie, F., Huang, H., and Ding, C.H.Q. 2012. Low-rank matrix recovery via efficient Schatten p-norm minimization. In *AAAI Conference on Artificial Intelligence*. AAAI Press.
- Obozinski, G., Taskar, B., and Jordan, M. 2006. Multi-task feature selection. Technical Report. Statistics Department, UC Berkeley.
- Pereira, J.B., Cavallin, L., Spulber, G., et al. 2014. Influence of age, disease onset and ApoE4 on visual medial temporal lobe atrophy cut-offs. *J. Intern. Med.* 275, 317–330.
- Petrella, J.R., Coleman, R.E., and Doraiswamy, P.M. 2003. Neuroimaging and early diagnosis of Alzheimer disease: A look to the future 1. *Radiology* 226, 315–336.
- Potkin, S.G., Guffanti, G., Lakatos, A., et al. 2009a. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. *PLoS One* 4, e6501.
- Potkin, S.G., Turner, J.A., Guffanti, G., et al. 2009b. Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: Methodological considerations. *Cogn. Neuropsychiatry* 14, 391–418.
- Risacher, S.L., Kim, S., Shen, L., et al. 2013. The role of apolipoprotein E (APOE) genotype in early mild cognitive impairment (E-MCI). *Front. Aging Neurosci.* 5, 11.
- Rockafellar, R.T. 1970. *Convex Analysis (Princeton Mathematical Series)*. Princeton University Press, 46:49.
- Ruhe, A. 1970. Perturbation bounds for means of eigenvalues and invariant subspaces. *BIT Numer. Math.* 10, 343–354.
- Sabatti, C., Service, S.K., Hartikainen, A.L., et al. 2008. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.* 41, 35–46.
- Saykin, A. J., Shen, L., Foroud, T.M., et al. 2010. Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimer's Dement.* 6, 265–273.
- Shen, L., Kim, S., Risacher, S.L., et al. 2010. 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, 1051–1063.

- Wang, X. 2017. https://github.com/littleq1991/sparse_lowRank_regression.
- Wang, H., Nie, F., Huang, H., et al. 2012a. From phenotype to genotype: An association study of longitudinal phenotypic markers to Alzheimer's disease relevant SNPs. *Bioinformatics* 28, i619–i625.
- Wang, H., Nie, F., Huang, H., et al. 2012b. High-order multi-task feature learning to identify longitudinal phenotypic markers for Alzheimer's disease progression prediction, 1277–1285. In *Advances in Neural Information Processing Systems*. NIPS.
- Wenk, G.L. 2003. Neuropathologic changes in Alzheimer's disease. *J. Clin. Psychiatry*. 64, 7–10.

Address correspondence to:

Prof. Heng Huang
Department of Electrical and Computer Engineering
University of Pittsburgh
Pittsburgh, PA 15213

E-mail: henghuanghh@gmail.com