Associating Multi-Modal Brain Imaging Phenotypes and Genetic Risk Factors via a Dirty Multi-Task Learning Method

Lei Du, *Member, IEEE*, Fang Liu, Kefei Liu, Xiaohui Yao, Shannon L. Risacher,
Junwei Han, *Senior Member, IEEE*, Andrew J. Saykin,
and Li Shen, *Senior Member, IEEE*

Abstract—Brain imaging genetics becomes more and more important in brain science, which integrates genetic variations and brain structures or functions to study the genetic basis of brain disorders. The multi-modal imaging data collected by different technologies, measuring the same brain distinctly, might carry complementary information. Unfortunately, we do not know the extent to which the phenotypic variance is shared among multiple imaging modalities, which further might trace back to the complex genetic mechanism. In this paper, we propose a novel dirty multi-task sparse canonical correlation analysis (SCCA) to study imaging genetic problems with multi-modal brain imaging quantitative traits (QTs) involved. The proposed method takes advantages of the multi-task learning and parameter decomposition. It can not only identify the shared imaging QTs and genetic loci across multiple modalities, but also identify the modality-specific imaging QTs and genetic loci, exhibiting a flexible capability of identifying complex multi-SNP-multi-QT associations. Using the state-of-the-art multi-view SCCA and multi-task SCCA, the proposed method shows better or comparable canonical correlation coefficients and canonical weights on both synthetic and real neuroimaging genetic data. In addition, the identified modality-consistent biomarkers, as well as the modality-specific biomarkers, provide meaningful and interesting information, demonstrating the dirty multi-task SCCA could be a powerful alternative method in multi-modal brain imaging genetics.

Index Terms—Brain imaging genetics, sparse canonical correlation analysis, multi-task learning, the dirty multi-task SCCA.

I. Introduction

RECENTLY, brain imaging genetics gains more and more attention in brain science. The primal aim of the imaging genetics is to uncover the genetic basis of brain structures, brain functions, and brain disorders such as Alzheimer's disease (AD) [1]–[3]. Therefore, the genetic variations such as single nucleotide polymorphisms (SNPs) and neuroimaging quantitative traits (QTs) are usually analyzed together [3]. Benefiting from the advances of imaging technology, different types of brain imaging data have been collected [4]. For example, the structural magnetic resonance imaging (sMRI) scans provide the morphometry of the brain such as the gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF), and the positron-emission tomography (PET) scans measure the metabolic processes of the brain. These imaging QTs obtained by different image technologies, measuring the same brain from different perspectives, might carry complementary information. As a result, combining multi-modal imaging QTs could help better identify those relevant imaging QTs and SNPs that are correlated to the brain disorder. Moreover, an imaging QT could be only relevant when it is measured by a specific imaging technology, while another QT could be relevant no matter which imaging technology is used. This might trace back to the complex genetic mechanism, and further complicate the identification of meaningful SNPs. Therefore, incorporating multi-modal imaging QTs and genetic variations into the imaging genetic framework, and studying the modality-consistent biomarkers, as well as the modality-specific biomarkers, could be beneficial to exploit meaningful genetic mechanism for brain disorders [5].

Lei Du, Fang Liu, and Junwei Han are with the School of Automation, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: dulei@nwpu.edu.cn; fangliu@mail.nwpu.edu.cn; jhan@nwpu.edu.cn).

Kefei Liu, Xiaohui Yao, and Li Shen are with the Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: kefei.liu@pennmedicine.upenn.edu; xiaohui.yao@pennmedicine.upenn.edu; li.shen@pennmedicine.upenn.edu).

Shannon L. Risacher and Andrew J. Saykin are with the Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN 46202 USA (e-mail: srisache@iupui.edu; asaykin@iupui.edu).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2020.2995510

The regression-oriented multi-task learning (MTL) is widely used in imaging genetics for its power in identifying complex multi-SNP-multi-QT associations [3], [6], [7]. These MTL methods usually preselect a few imaging QTs of interest as dependent variables and multiple SNPs as independent variables, and then reveal the joint effect of multi-locus genotype on a few phenotypes via multivariate multiple regression [7]. Obviously, they can select SNPs which are relevant to the candidate imaging QTs simultaneously. On the contrary, the joint effect of multiple imaging QTs on a few SNPs can also be studied by MTL [8]. We have known that the brain is comprised of multiple regions, thereby multiple imaging QTs [9]. Therefore, utilizing only a few of them might be inadequate because it may lose critical information conveyed by those excluded cerebral components [5].

In addition, the bi-multivariate learning methods such as the sparse canonical correlation analysis (SCCA) are also very popular in imaging genetics [10]–[18]. These SCCA methods could also identify complex multi-SNP-multi-QT associations [3]. And, they can conduct feature selection for both SNPs and imaging QTs, while those MTL methods cannot. Generally, they are two-view SCCA, indicating that these SCCA methods can only analyze the relationship between SNPs and QTs of an unimodal imaging data. To the best of our knowledge, most SCCA methods fall into this category. They cannot include multi-modal imaging QTs and SNPs in a unified model, making them suboptimal since using only one modality of imaging QTs is inadequate. In order to incorporate more than two data modalities, it is straightforward to extend the two-view SCCA to multi-view/multi-set SCCA (mSCCA), and a few efforts has been made in this direction. For example, Hao et al. [19] proposed the three-way SCCA to study the relationships among SNPs, imaging QTs and diagnosis status, and Fang et al. [20] proposed the joint SCCA to learn diverse associations among subtype populations. Both methods are the naive extension of the conventional two-view SCCA. As a result, they might not identify reasonable genetic loci since, unless the multiple modalities of imaging QTs are highly correlated, demanding SNPs to be associated with imaging QTs of multiple heterogeneous modalities simultaneously could be a too stringent requirement.

The multi-task SCCA (MTSCCA) is recently proposed in [5], [21], [22], which studies the multi-modal imaging genetic problem by constructing multiple SCCA tasks jointly, with each associating SNPs with imaing QTs of one modality. This joint bi-multivariate learning shows great success in multi-modal imaging genetics. The aforementioned imaging technologies could be quite different, and thus the multiple modalities of imaging QTs can be weakly correlated [23]. In other words, an imaging QT could be informative under one imaging modality, while another imaging QT could be informative under another imaging modality. At the same time, there are still imaging QTs which might be informative no matter which imaging technology is used. Therefore, identifying modality-consistent and modality-specific imaging QTs as well as revealing associated SNPs are quite essential and meaningful.

With these observations above, in this paper, we propose a novel learning method which is designed for multi-modal imaging data oriented imaging genetics. The proposed method absorbs the merits of both MTL and parameters decomposition. The MTL framework makes it easier and practical to integrate multiple modalities of imaging QTs, and the parameters decomposition makes a diverse regularization which is quite meaningful. We name it the *dirty* MTSCCA in accordance with the terminology in [24], [25]. The dirty MTSCCA decomposes the conventional canonical weights into two parts, i.e. the task-consistent component shared among all tasks, and the task-specific component that is closely related to a specific task. It then penalizes the task-consistent and task-specific components differently to encourage different sparse structures. Thus dirty MTSCCA can identify both SNPs and imaging QTs that are measured by all imaging technologies, and SNPs and imaging QTs that could be only revealed by a specific imaging technology. In order to solve this dirty model, we propose an efficient iteration algorithm which guarantees to converge to a local optimum. Compared with two state-of-the-art methods including the multi-view SCCA [16] and multi-task SCCA [5], the results on both synthetic data and real neuroimaging genetics data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [26] show that, our method obtains improved or comparable bi-multivariate associations. Moreover, our method could identify the modality-consistent imaging QTs and SNPs, as well as the modality-specific imaging QTs and SNPs, showing a flexible and meaningful identification ability. Therefore, the dirty multi-task SCCA model is very suitable for multi-modal imaging genetic association analysis, and can be a significant addition to the imaging genetic method library.

II. THE DIRTY MULTI-TASK SCCA

In this paper, we denote scalars as italic letters, column vectors as boldface lowercase letters, and matrices as boldface capitals. The i-th row and j-th column of $\mathbf{X} = (x_{ij})$ is denoted as $\mathbf{x}^i$ and $\mathbf{x}_j$ respectively. $\mathbf{Y}_c$ denotes the c-th matrix of $\{\mathbf{Y}_1, \cdots, \mathbf{Y}_C\}$. $\|\mathbf{x}\|_2$ denotes the Euclidean norm of the vector $\mathbf{x}$, $\|\mathbf{X}\|_{1,1}$ denotes the element-wise ℓ_1-norm of $\mathbf{X}$, i.e. $\|\mathbf{X}\|_{1,1} = \sum_i \sum_j |x_{ij}|$, and $\|\mathbf{X}\|_F = \sqrt{\sum_i \sum_j x_{ij}^2}$ denotes its Frobenius norm. $\mathbf{X} \in \mathbb{R}^{n \times p}$ represents the genetic data with n subjects and p SNPs, and $\mathbf{Y}_c \in \mathbb{R}^{n \times q}$ ($c = 1, \cdots, C$) represents the phenotype data with q imaging QTs of the c-th modality, where C is the number of imaging modalities (tasks).

1) The Multi-Task SCCA: According to [5], [27], we use $\mathbf{U} \in \mathcal{R}^{p \times C}$ to denote the canonical weight matrix associated with $\mathbf{X}$ and $\mathbf{V} \in \mathcal{R}^{q \times C}$ to denote that associated with imaging QTs, where each $\mathbf{v}_c$ corresponds to $\mathbf{Y}_c$. Then the multi-task SCCA (MTSCCA) model is defined as follows

$$\min_{\mathbf{u}_c, \mathbf{v}_c} \sum_{c=1}^{C} \|\mathbf{X}\mathbf{u}_c - \mathbf{Y}_c \mathbf{v}_c\|_2^2$$

$$s.t. \|\mathbf{X}\mathbf{u}_c\|_2^2 = 1, \|\mathbf{Y}_c \mathbf{v}_c\|_2^2 = 1, \Omega(\mathbf{U}) \leq b_1, \Omega(\mathbf{V}) \leq b_2, \forall c.$$
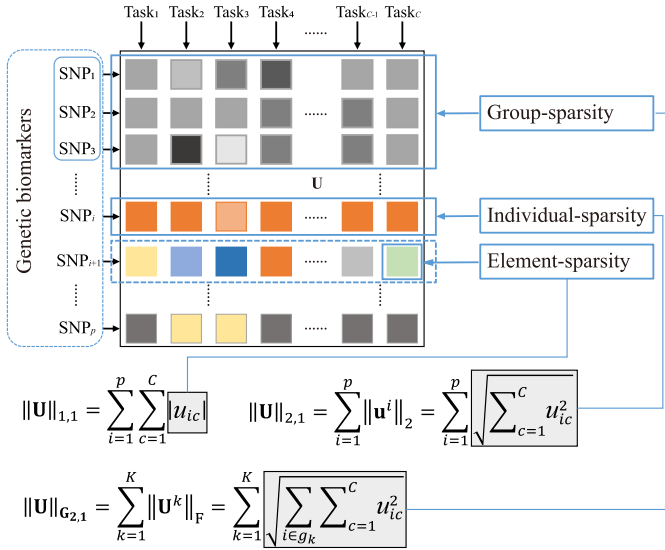
(1)

Fig. 1. Illustration of the group-sparsity, individual-sparsity and element-sparsity for canonical weight **U**. The group-sparsity indicates that SNPs in the same group are informative for all SCCA tasks simultaneously. The individual-sparsity across all tasks indicates that a SNP (imaging QT) is informative for all SCCA tasks. The element-sparsity indicates that a SNP (imaging QT) is only informative for a specific SCCA task.

This MTSCCA, overcoming the limitation of the conventional SCCA and multi-view/multi-set SCCA, can model the bi-association between SNPs and imaging QTs from multiple modalities [5]. However, in the multi-modal scenario, we usually desire both group-sparsity and individual-sparsity across multiple modalities, and element-sparsity that is only effective to a specific modality. Fig. 1 presents the group-sparsity, individual-sparsity and element-sparsity for canonical weight **U**. Generally, in multi-task learning, group-sparsity, individual-sparsity and element-sparsity penalties are imposed on **U** and **V** to achieve this aim. Obviously, group-sparsity, individual-sparsity and element-sparsity for the same weight matrix such as **U** are conflicting, and thus could harm the performance, and further the feature selection ability.

2) The Proposed Dirty MTSCCA Model: In order to construct a flexible and robust modeling method, and overcome the shortcomings of MTSCCA, we propose a novel dirty multi-task SCCA based on parameters decomposition [24], [25]. The dirty MTSCCA is formally defined as follows

$$\min_{\mathbf{S,W,B,Z}} \sum_{c=1}^{C} \|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c) - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 + \lambda_s \|\mathbf{S}\|_{G_{2,1}}$$
$$+ \beta_s \|\mathbf{S}\|_{2,1} + \lambda_w \|\mathbf{W}\|_{1,1} + \beta_b \|\mathbf{B}\|_{2,1} + \lambda_z \|\mathbf{Z}\|_{1,1}$$
$$s.t. \, \|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c)\|_2^2 = 1, \|\mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 = 1, \forall c. \quad (2)$$

In our model, the canonical weight **U** associated with the SNP data is decomposed into two components, i.e. **U** = **S**+**W**, where **S** is the task-consistent component being shared by all tasks, and **W** is the task-specific component being associated with a single task. Similarly, the canonical weight **V** associated with the imaging data is also decomposed into the task-consistent component **B** and the task-specific component

Z, i.e. **V** = **B** + **Z**. The λ_s, β_s, λ_w, β_b, λ_z are nonnegative tuning parameters.

Benefiting from the parameter decomposition, we can impose distinct penalties on different components of the canonical weight. Specifically, we use the $G_{2,1}$-norm [7], [27], i.e.

$$\|\mathbf{U}\|_{G_{2,1}} = \sum_{k=1}^{K} \left\|\mathbf{U}^k\right\|_F = \sum_{k=1}^{K} \sqrt{\sum_{i \in g_k} \sum_{c=1}^{C} (u_{ic})^2}, \quad (3)$$

to pursue a similar weight value for a group of SNPs, e.g. SNPs in the same linkage disequilibrium (LD), across multiple tasks. This group-sparsity, illustrated in Fig. 1, selects those relevant groups of SNPs shared among all tasks. However, a SNP in a relevant LD might be irrelevant to AD, while another SNP in an irrelevant LD could be informative. This prompts us to impose the popular $\ell_{2,1}$-norm, defined as

$$\|\mathbf{U}\|_{2,1} = \sum_{i=1}^{p} \left\|\mathbf{u}^i\right\|_2 = \sum_{i=1}^{p} \sqrt{\sum_{c=1}^{C} (u_{ic})^2}, \quad (4)$$

in multi-task learning, which helps accommodate the individual-sparsity shared by multiple tasks. It is worth mentioning that, to pursue the task-consistent feature selection, both $G_{2,1}$-norm and $\ell_{2,1}$-norm are imposed onto task-consistent component **S**, and the $\ell_{2,1}$-norm is used for **B**.

In addition to the task-consistent features, there are some features (SNPs or QTs) which could be relevant to only one specific task. This is a common situation in imaging genetics in which the imaging QTs are collected by different imaging technologies, thereby forming the heterogeneous multi-task learning. Then the element-sparsity is also of great importance in multi-modal brain imaging genetics. Finally, for the task-specific component **W** and **Z**, we use the $\ell_{1,1}$-norm which is defined previously to select the relevant feature for a specific SCCA task.

To sum up, the distinct regularization terms for different components encourage both task-consistent and task-specific feature selection instead of balancing between these two conflicting objectives as traditional SCCAs do. Therefore, this could assure an improved performance in terms of both the correlation and canonical weight profiles.

3) Extension to the Weighted Model: The model above equally treats each SCCA task regarding the SNPs and QTs of a specific imaging modality. In order to further make it practical and flexible, we introduce a weight vector $\kappa \in \mathbb{R}^{1 \times C} (0 \leq \kappa_c \leq 1, \sum_c \kappa_c = 1, c = 1, \cdots, C)$ to the loss function of the dirty MTSCCA, i.e.

$$\min_{\mathbf{S,W,B,Z}} \sum_{c=1}^{C} \kappa_c \|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c) - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2$$
$$+ \lambda_s \|\mathbf{S}\|_{G_{2,1}}$$
$$+ \beta_s \|\mathbf{S}\|_{2,1} + \lambda_w \|\mathbf{W}\|_{1,1} + \beta_b \|\mathbf{B}\|_{2,1} + \lambda_z \|\mathbf{Z}\|_{1,1}$$
$$s.t. \, \|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c)\|_2^2 = 1, \|\mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 = 1,$$
$$\sum_c \kappa_c = 1, \quad \forall c. \quad (5)$$

It is easy to verify that when all κ_c's are equal, Eq (5) will reduce to Eq (2). This model will also reduce to the conventional SCCA if only one of κ_c's is nonzero since the task-consistent components disappear.

Besides, the merits of the dirty MTSCCA are fourfold. First, it submerges both MTSCCA and mSCCA. For example, the dirty MTSCCA reduces to MTSCCA when $\mathbf{W} = \mathbf{0}$ and $\mathbf{Z} = \mathbf{0}$. Second, based on the parameter decomposition, it encourages the task-consistent (modality-consistent) sparsity [24] and task-specific (modality-specific) sparsity simultaneously in a unified model. On the contrary, the MTSCCA and mSCCA can only promote task-consistent sparsity. Third, the task-consistent component is jointly penalized by the group level regularization, such as the $G_{2,1}$-norm for SNPs to induce the task-consistent group-sparsity, and the $\ell_{2,1}$-norm for both SNPs and imaging QTs for the task-consistent individual-sparsity. This could help identify the SNPs and imaging QTs shared by multiple tasks, thereby by different imaging technologies. Fourth, our model penalizes the task-specific component differently via the $\ell_{1,1}$-norm to encourage element-wise sparsity for both SNPs and imaging QTs. This helps find out SNPs and imaging QTs that could only be identified by a specific imaging modality, i.e. imaging technology. In summary, thanks to the parameter decomposition, our method facilitates joint feature selection while allowing disparities as well [24], [25]. This makes our model flexible and practical since simultaneously demanding features to be task-consistent and task-specific is conflicting. In a word, this weighted model is very practical and powerful in multi-modal imaging genetics. Thereafter, we will use the dirty MTSCCA to refer to the weighted model.

4) The Optimization Algorithm: In this subsection, we will present how to solve the dirty MTSCCA efficiently. According to Eq. (5), the objective is not convex with joint consideration of $\mathbf{S}$, $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{Z}$. Thus it cannot be directly solved by gradient descent methods. The MTSCCA is a bi-convex problem indicating that $\mathbf{U}$ and $\mathbf{V}$ can be solved alternatively [5]. As a modified MTSCCA, our model is also bi-convex and thus could be handled by the alternative convex search (ACS) strategy. Specifically, the Eq. (5) is convex in $\mathbf{S}$ when fixing $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{Z}$ as constants. Similarly, Eq. (5) is also convex in $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{Z}$ alternately by fixing those remaining weight matrices. For this reason, the dirty MTSCCA can be solved via the alternative iteration method. Next, we first show how to solve $\mathbf{S}$ and $\mathbf{W}$ since they both originate from canonical weights associating with SNP data. Then we present the solution to $\mathbf{B}$ and $\mathbf{Z}$ which are associated with multiple modalities of imaging data.

a) Updating $\mathbf{S}$ and $\mathbf{W}$: If $\mathbf{B}$ and $\mathbf{Z}$ are fixed as constants, the objective with respect to $\mathbf{S}$ and $\mathbf{W}$ can be simplified as

$$\min_{\mathbf{S},\mathbf{W}} \sum_{c=1}^{C} \kappa_c \|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c) - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2$$
$$+ \lambda_s \|\mathbf{S}\|_{G_{2,1}} + \beta_s \|\mathbf{S}\|_{2,1} + \lambda_w \|\mathbf{W}\|_{1,1}$$
$$s.t. \|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c)\|_2^2 = 1, \quad \forall c. \quad (6)$$

In order to solve $\mathbf{S}$ and $\mathbf{W}$, we have the following theorem.

Theorem 1: The solution to Eq. (6) is attained by

$$\mathbf{s}_c^* = \frac{\hat{\mathbf{s}}_c}{\|\mathbf{X}(\hat{\mathbf{s}}_c + \hat{\mathbf{w}}_c)\|_2}, \text{ and } \mathbf{w}_c^* = \frac{\hat{\mathbf{w}}_c}{\|\mathbf{X}(\hat{\mathbf{s}}_c + \hat{\mathbf{w}}_c)\|_2}, \quad (7)$$

where $\hat{\mathbf{s}}_c$ is the solution of

$$\min_{\mathbf{S}} \sum_{c=1}^{C} \kappa_c \|\mathbf{X}\mathbf{s}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 + \lambda_s \|\mathbf{S}\|_{G_{2,1}} + \beta_s \|\mathbf{S}\|_{2,1}, \quad (8)$$

and $\hat{\mathbf{w}}_c$ is the solution of

$$\min_{\mathbf{W}} \sum_{c=1}^{C} \kappa_c \|\mathbf{X}\mathbf{w}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 + \lambda_w \|\mathbf{W}\|_{1,1}. \quad (9)$$

Proof: Following the same procedure in [28] (Appendix A.2), Eq. (7) in Theorem 1 can be proved straightforwardly. Thus we concentrate on the derivations of Eqs. (8-9).

We first expand the quadratic term in Eq. (6)

$$\min_{\mathbf{S},\mathbf{W}} \sum_{c=1}^{C} \kappa_c [\|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c)\|_2^2 - 2\mathbf{s}_c^\top \mathbf{X}^\top \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)$$
$$- 2\mathbf{w}_c^\top \mathbf{X}^\top \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c) + \|\mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2] + \lambda_s \|\mathbf{S}\|_{G_{2,1}}$$
$$+ \beta_s \|\mathbf{S}\|_{2,1} + \lambda_w \|\mathbf{W}\|_{1,1}$$
$$s.t. \|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c)\|_2^2 = 1, \forall c. \quad (10)$$

Given $\|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c)\|_2^2 = \|\mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 = 1$, we then minus $\frac{\kappa_c}{2} \|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c)\|_2^2$ and plus $\kappa_c \|\mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2$ into Eq. (10). At the same time, it is easier to derive

$$\frac{1}{2} \|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c)\|_2^2 = \frac{1}{2} \|\mathbf{X}\mathbf{s}_c\|_2^2 + \mathbf{s}_c^\top \mathbf{X}^\top \mathbf{X}\mathbf{w}_c$$
$$+ \frac{1}{2} \|\mathbf{X}\mathbf{w}_c\|_2^2 \leq \|\mathbf{X}\mathbf{s}_c\|_2^2 + \|\mathbf{X}\mathbf{w}_c\|_2^2.$$

Therefore, we easily have the upper bound of Eq. (10), thereby Eq. (6) as follows

$$\min_{\mathbf{S},\mathbf{W}} \sum_{c=1}^{C} \kappa_c \|\mathbf{X}\mathbf{s}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 + \kappa_c \|\mathbf{X}\mathbf{w}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2$$
$$+ \lambda_s \|\mathbf{S}\|_{G_{2,1}} + \beta_s \|\mathbf{S}\|_{2,1} + \lambda_w \|\mathbf{W}\|_{1,1}$$
$$s.t. \|\mathbf{X}(\mathbf{s}_c + \mathbf{w}_c)\|_2^2 = 1, \forall c. \quad (11)$$

Now by dropping the constraints, we have the objective function with respect to $\mathbf{S}$ as

$$\min_{\mathbf{S}} \sum_{c=1}^{C} \kappa_c \|\mathbf{X}\mathbf{s}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 + \lambda_s \|\mathbf{S}\|_{G_{2,1}} + \beta_s \|\mathbf{S}\|_{2,1}, \quad (12)$$

and that with respect to $\mathbf{W}$ as

$$\min_{\mathbf{W}} \sum_{c=1}^{C} \kappa_c \|\mathbf{X}\mathbf{w}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 + \lambda_w \|\mathbf{W}\|_{1,1}, \quad (13)$$

which completes the proof. ■

Since Eq. (8) is a multi-task regression problem, we can solve it using the off-the-shelf methods. We observe that the penalization of each $\mathbf{s}_c$ is different due to different κ_c's, and thus we separately solve each $\mathbf{s}_c$. Specifically, we first take the

derivative of Eq. (8) with respect to $\mathbf{s}_c$, and then let it be zero, viz,

$$(\mathbf{X}^\top \mathbf{X} + \frac{\lambda_s}{\kappa_c}\tilde{\mathbf{D}} + \frac{\beta_s}{\kappa_c}\mathbf{D})\mathbf{s}_c = \mathbf{X}^\top \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c), \qquad (14)$$

where $\tilde{\mathbf{D}}$ is a block diagonal matrix with the k-th block being $\frac{1}{2\|\mathbf{S}^k\|_{\mathrm{F}}}\mathbf{I}_k$, and $\mathbf{I}_k$ is an identity matrix which has the same size as the k-th group. The grouping information can be previously given based on the LD structure of SNPs. $\mathbf{D}$ is a diagonal matrix whose i-th diagonal element is $\frac{1}{2\|\mathbf{s}^i\|_2}(i = 1, \cdots, p)$. Then we can iteratively obtain $\hat{\mathbf{s}}_c$ as follows

$$\hat{\mathbf{s}}_c = (\mathbf{X}^\top \mathbf{X} + \frac{\lambda_s}{\kappa_c}\tilde{\mathbf{D}} + \frac{\beta_s}{\kappa_c}\mathbf{D})^{-1}\mathbf{X}^\top \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c). \qquad (15)$$

Attributing to the $\ell_{1,1}$-norm penalty, $\mathbf{w}_c$'s in Eq. (9) are not coupled closely, indicating that each $\mathbf{w}_c$ can be obtained separately. We take the derivative of Eq. (9) with respect to each $\mathbf{w}_c$ respectively, and let it be zero, i.e.

$$(\mathbf{X}^\top \mathbf{X} + \frac{\lambda_w}{\kappa_c}\check{\mathbf{D}}_c)\mathbf{w}_c = \mathbf{X}^\top \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c), \qquad (16)$$

where $\check{\mathbf{D}}_c$ is a diagonal matrix with its i-th element being $\frac{1}{2|w_{ic}|}(i = 1, \cdots, p)$. Further, the $\mathbf{w}_c$ can be attained by

$$\hat{\mathbf{w}}_c = (\mathbf{X}^\top \mathbf{X} + \frac{\lambda_w}{\kappa_c}\check{\mathbf{D}}_c)^{-1}\mathbf{X}^\top \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c). \qquad (17)$$

Now both $\mathbf{S}$ and $\mathbf{W}$ are attained based on Theorem 1, we proceed to solve $\mathbf{B}$ and $\mathbf{Z}$ by fixing $\mathbf{S}$ and $\mathbf{W}$.

b) Updating $\mathbf{B}$ and $\mathbf{Z}$: Firstly, $\mathbf{B}$ and $\mathbf{Z}$ can be solved using the same Theorem 1 as shown above. We further find that each $\mathbf{b}_c$ and $\mathbf{z}_c$ are associated with each modality of imaging QTs, i.e. $\mathbf{Y}_c$. Therefore, $\mathbf{b}_c$ and $\mathbf{z}_c$ should be solved separately. In particular, $\mathbf{b}_c$ and $\mathbf{z}_c$ can be solved iteratively by fixing the remaining $\mathbf{b}_{c'}$ and $\mathbf{z}_{c'}(c' \neq c)$, as well as $\mathbf{S}$ and $\mathbf{W}$.

Then following the same procedure of solving $\mathbf{w}_c$, we easily have

$$\hat{\mathbf{b}}_c = (\mathbf{Y}_c^\top \mathbf{Y}_c + \frac{\beta_b}{\kappa_c}\mathbf{Q})^{-1}\mathbf{Y}_c^\top \mathbf{X}(\mathbf{s}_c + \mathbf{w}_c), \qquad (18)$$

by taking the derivative with respect to every $\mathbf{b}_c$ separately, and letting them be zero. $\mathbf{Q}$ here is a diagonal matrix and its j-th element is $\frac{1}{2\|\mathbf{b}^j\|_2}(j = 1, \cdots, q)$.

Finally, the same procedure leads to

$$\hat{\mathbf{z}}_c = (\mathbf{Y}_c^\top \mathbf{Y}_c + \frac{\lambda_z}{\kappa_c}\check{\mathbf{Q}}_c)^{-1}\mathbf{Y}_c^\top \mathbf{X}(\mathbf{s}_c + \mathbf{w}_c), \qquad (19)$$

where $\check{\mathbf{Q}}_c$ is a diagonal matrix whose i-th diagonal element is $\frac{1}{2|z_{jc}|}(j = 1, \cdots, q)$.

Combining Eqs. (18)-(19) together, we finally have the solution to $\mathbf{B}$ and $\mathbf{Z}$ as follows

$$\mathbf{b}_c^* = \frac{\hat{\mathbf{b}}_c}{\left\|\mathbf{Y}_c(\hat{\mathbf{b}}_c + \hat{\mathbf{z}}_c)\right\|_2}, \text{ and } \mathbf{z}_c^* = \frac{\hat{\mathbf{z}}_c}{\left\|\mathbf{Y}_c(\hat{\mathbf{b}}_c + \hat{\mathbf{z}}_c)\right\|_2}. \qquad (20)$$

Eqs. (14)-(20) pave the way to solve the dirty MTSCCA problem, we then show the pseudo-code in Algorithm 1. To ensure efficiency, this algorithm iteratively updates $\mathbf{S}$, $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{Z}$ when the pre-defined stopping condition, such as

the maximum iterations or the tolerated error, is satisfied. Moreover, this algorithm is guaranteed to converge to a local optimum which is supported by the Theorem 2 in the next subsection.

Algorithm 1 The Dirty Multi-Task SCCA Algorithm

Require: $\mathbf{X} \in \mathcal{R}^{n \times p}$, $\mathbf{Y}_c \in \mathcal{R}^{n \times q}$, $c = 1, \cdots, C$;
$\quad \lambda_s$, β_s, λ_w, β_b, λ_z
Ensure: Output $\mathbf{S}$, $\mathbf{W}$, $\mathbf{B}$, $\mathbf{Z}$.
1: Initialize $\mathbf{S} \in \mathcal{R}^{p \times C}$, $\mathbf{W} \in \mathcal{R}^{p \times C}$, $\mathbf{B} \in \mathcal{R}^{q \times C}$ and $\mathbf{Z} \in \mathcal{R}^{q \times C}$;
2: **while** not convergence **do**
3: $\quad$ Update $\hat{\mathbf{s}}_c$ according to Eq. (15), and update $\hat{\mathbf{w}}_c$ according to Eq. (17);
4: $\quad$ Solve $\mathbf{S}^*$ and $\mathbf{W}^*$ according to Eq. (7);
5: $\quad$ Update $\hat{\mathbf{b}}_c$ according to Eq. (18), and update $\hat{\mathbf{z}}_c$ according to Eq. (19);
6: $\quad$ Solve $\mathbf{B}^*$ and $\mathbf{Z}^*$ according to Eq. (20);
7: **end while**

---

5) Convergence Analysis: We have the following theorem regarding the dirty MTSCCA algorithm.

Theorem 2: The Algorithm 1 decreases the objective value of Eq. (5) in each iteration.

Proof: (1) We first prove that the objective decreases after updating $\mathbf{S}$ and $\mathbf{W}$. We denote the updated $\mathbf{S}$, $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{Z}$ as $\bar{\mathbf{S}}$, $\bar{\mathbf{W}}$, $\bar{\mathbf{B}}$ and $\bar{\mathbf{Z}}$, respectively. From Eq. (15) we have

$$\begin{aligned}
&\sum_{c=1}^C \kappa_c \|\mathbf{X}\bar{\mathbf{s}}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 \\
&\quad + \lambda_s Tr(\bar{\mathbf{S}}^\top \tilde{\mathbf{D}}\bar{\mathbf{S}}) + \beta_s Tr(\bar{\mathbf{S}}^\top \mathbf{D}\bar{\mathbf{S}}) \\
&\leq \sum_{c=1}^C \kappa_c \|\mathbf{X}\mathbf{s}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 \\
&\quad + \lambda_s Tr(\mathbf{S}^\top \tilde{\mathbf{D}}\mathbf{S}) + \beta_s Tr(\mathbf{S}^\top \mathbf{D}\mathbf{S}). \qquad (21)
\end{aligned}$$

Based on the definitions of $\tilde{\mathbf{D}}$ and $\mathbf{D}$, we have

$$\begin{aligned}
&\sum_{c=1}^C \kappa_c \|\mathbf{X}\bar{\mathbf{s}}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 \\
&\quad + \lambda_s \sum_{k=1}^K \frac{\|\bar{\mathbf{S}}^k\|_{\mathrm{F}}^2}{2\|\mathbf{S}^k\|_{\mathrm{F}}} + \beta_s \sum_{i=1}^p \frac{\|\bar{\mathbf{s}}^i\|_2^2}{2\|\mathbf{s}^i\|_2} \\
&\leq \sum_{c=1}^C \kappa_c \|\mathbf{X}\mathbf{s}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c)\|_2^2 \\
&\quad + \lambda_s \sum_{k=1}^K \frac{\|\mathbf{S}^k\|_{\mathrm{F}}^2}{2\|\mathbf{S}^k\|_{\mathrm{F}}} + \beta_s \sum_{i=1}^p \frac{\|\mathbf{s}^i\|_2^2}{2\|\mathbf{s}^i\|_2}. \qquad (22)
\end{aligned}$$

Since $\|\bar{\mathbf{S}}^k\|_{\mathrm{F}} - \frac{\|\bar{\mathbf{S}}^k\|_{\mathrm{F}}^2}{2\|\mathbf{S}^k\|_{\mathrm{F}}} \leq \|\mathbf{S}^k\|_{\mathrm{F}} - \frac{\|\mathbf{S}^k\|_{\mathrm{F}}^2}{2\|\mathbf{S}^k\|_{\mathrm{F}}}$, and $\|\bar{\mathbf{s}}^i\|_2 - \frac{\|\bar{\mathbf{s}}^i\|_2^2}{2\|\mathbf{s}^i\|_2} \leq \|\mathbf{s}^i\|_2 - \frac{\|\mathbf{s}^i\|_2^2}{2\|\mathbf{s}^i\|_2}$ (Lemma 1 in [7]), we apply both inequations to Eq. (22) with respect to each group features

and individual one. This yields

$$
\begin{aligned}
&\sum_{c=1}^{C} \kappa_c \left\| \mathbf{X}\bar{\mathbf{s}}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c) \right\|_2^2 + \lambda_s \sum_{k=1}^{K} \left\| \bar{\mathbf{S}}^k \right\|_{\mathrm{F}} + \beta_s \sum_{i=1}^{p} \left\| \bar{\mathbf{s}}^i \right\|_2 \\
&\leq \sum_{c=1}^{C} \kappa_c \left\| \mathbf{X}\mathbf{s}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c) \right\|_2^2 + \lambda_s \sum_{k=1}^{K} \left\| \mathbf{S}^k \right\|_{\mathrm{F}} + \beta_s \sum_{i=1}^{p} \left\| \mathbf{s}^i \right\|_2 \\
&\Leftrightarrow \sum_{c=1}^{C} \kappa_c \left\| \mathbf{X}\bar{\mathbf{s}}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c) \right\|_2^2 + \lambda_s \left\| \bar{\mathbf{S}} \right\|_{\mathrm{G}_{2,1}} + \beta_s \left\| \bar{\mathbf{S}} \right\|_{2,1} \\
&\leq \sum_{c=1}^{C} \kappa_c \left\| \mathbf{X}\mathbf{s}_c - \mathbf{Y}_c(\mathbf{b}_c + \mathbf{z}_c) \right\|_2^2 + \lambda_s \left\| \mathbf{S} \right\|_{\mathrm{G}_{2,1}} + \beta_s \left\| \mathbf{S} \right\|_{2,1}.
\end{aligned}
\tag{23}
$$

Therefore, the objective value decreases when updating $\mathbf{S}$. After that, we can also prove that the objective value decreases in each iteration when updating $\mathbf{W}$. According to Theorem 1, the objective still decreases after scaling. This yields that the objective decreases after updating $\mathbf{S}$ and $\mathbf{W}$.

(2) Similarly, we can prove that the objective also decreases with each update of $\mathbf{B}$ and $\mathbf{Z}$.

The proof completes by combining conclusions (1) and (2). ∎

According to Eq. (5), we know that the objective is lower bounded by 0. Therefore, given the Theorem 2, the Algorithm 1 is guaranteed to converge to a local optimum.

III. EXPERIMENTS AND RESULTS

A. Experimental Setup

To evaluate the effectiveness of the proposed dirty MTSCCA, we choose two closely related methods as benchmarks. They are the multi-task SCCA (MTSCCA) [5] and the conventional multi-view/multi-set SCCA (mSCCA) [28]. Both methods can identify the complex bi-associations among three or more data sets, and thus could integrate multiple modalities of imaging QTs in one model, while those conventional two-view SCCA cannot [28].

The regularization parameters for each method should be fine tuned before experiments. In this paper, we employ the nested 5-fold cross-validation method. In particular, in the inner loop, parameters that generate the highest mean correlation coefficients will be selected as the optimal parameters, i.e. $\mathrm{CV}(\lambda, \beta) = \frac{1}{5} \sum_{j=1}^{5} \sum_{c=1}^{C} Corr((\mathbf{X}_j(\mathbf{u}_c)_j, (\mathbf{Y}_c)_j(\mathbf{v}_c)_j)$, where $\mathbf{X}_j$ and $(\mathbf{Y}_c)_j$ are the j-th testing sets in the inner loop, and $(\mathbf{u}_c)_j$ and $(\mathbf{v}_c)_j$ are the canonical weights learned from the inner training sets. Then the external loop calculates the final results based on the optimal parameters obtained from the inner loop. It is easy to know that too small parameters generate under-penalized results while too large ones generate over-penalized results. Therefore, we tune λ_s, β_s, λ_w, β_b and λ_z from a moderate interval 10^i ($i = -5, -4, \cdots, 0, \cdots, 4, 5$) via the grid search strategy.

Apart from the regularization parameters, task weight parameters κ_c's could affect the performance as well. Fortunately, they merely imply the priority of different tasks, and thus have less impact compared with those regularization parameters. For example, if the sMRI data is of good quality with high

resolution, we then prefer a high weight for sMRI-derived SCCA task (the SCCA task between sMRI data and SNPs) and small weights for those remaining tasks such as AV45-SNP and FDG-SNP. We, as is often the case, do not have the priori knowledge regarding the tasks' priorities, thus we use equal weights for different tasks. As a result, we use $\kappa_c = \frac{1}{C}(c = 1, \cdots, C)$ in this study. The average results from 100 repeated experiments are shown to assure a stable result. In this study, our method is terminated when both $\max_c |(\mathbf{s}_c + \mathbf{w}_c)^{t+1} - (\mathbf{s}_c + \mathbf{w}_c)^t| \leq \epsilon$ and $\max_c |(\mathbf{b}_c + \mathbf{z}_c)^{t+1} - (\mathbf{b}_c + \mathbf{z}_c)^t| \leq \epsilon$ are met, where ϵ is the pre-defined tolerable error and is set to $\epsilon = 10^{-5}$ according to experiments.

B. Results on Synthetic Data

1) Data Source: We simulate four synthetic data sets using different numbers of samples, features, and noise intensities. The first three data sets are generated from the same ground truth, however, they have different levels of noise. Specifically, the signal-to-noise ratio (SNR) in the first data set is the smallest, followed by the second and third one. We expect that this could show a method's performance under different noise levels. The fourth data set simulates a high-dimensional situation. The ground truths of these data sets are listed below, which are shown in Fig. 2 (top row).

- Data 1: $n = 100$, $\mathbf{u} = (\underbrace{0, \cdots, 0}_{50}, \underbrace{1, \cdots, 1}_{40}, \underbrace{0, \cdots, 0}_{60})^\top$,
  $\mathbf{v}_1 = (\underbrace{0, \cdots, 0}_{45}, \underbrace{1, \cdots, 1}_{30}, \underbrace{0, \cdots, 0}_{45})^\top$,
  $\mathbf{v}_2 = (\underbrace{0, \cdots, 0}_{20}, \underbrace{2, \cdots, 2}_{20}, \underbrace{0, \cdots, 0}_{20}, \underbrace{1, \cdots, 1}_{20}, \underbrace{0, \cdots, 0}_{40})^\top$,
  $\mathbf{v}_3 = (\underbrace{1, \cdots, 1}_{20}, \underbrace{0, \cdots, 0}_{30}, \underbrace{2, \cdots, 2}_{40}, \underbrace{0, \cdots, 0}_{30})^\top$,
  and $\mathbf{v}_4 = (\underbrace{0, \cdots, 0}_{45}, \underbrace{1.5, \cdots, 1.5}_{30}, \underbrace{0, \cdots, 0}_{45})^\top$.
  We generate a random latent vector μ of length n with unit norm. The data matrix $\mathbf{X}$ is generated from $x_{\ell,i} \sim N(\mu_\ell u_i, \sigma_x)$, where $\sigma_x = 5$ denotes the noise variance. $\mathbf{Y}_j$ is generated from $(y_{\ell,j})_c \sim N(\mu_\ell v_{j,c}, \sigma_{y_c})$ with $\sigma_{y_1} = \sigma_{y_2} = \sigma_{y_3} = \sigma_{y_4} = 5$.

- Data 2 ~ Data 3: These two data sets are generated using the same ground truth as Data 1 but with different noise levels, i.e. $\sigma_x = \sigma_{y_1} = \sigma_{y_2} = \sigma_{y_3} = \sigma_{y_4} = 0.5$ for Data 2, and $\sigma_x = \sigma_{y_1} = \sigma_{y_2} = \sigma_{y_3} = \sigma_{y_4} = 0.1$ for Data 3. Hence the true correlation coefficients of these three data sets are different. In particular, Data 1 has the lowest correlation coefficient, followed by Data 2, and Data 3 has the highest correlation coefficient.

- Data 4: $n = 500$, $\sigma_x = \sigma_{y_1} = \sigma_{y_2} = \sigma_{y_3} = \sigma_{y_4} = 0.1$,
  $\mathbf{u} = (\underbrace{0, \cdots, 0}_{400}, \underbrace{1, \cdots, 1}_{200}, \underbrace{0, \cdots, 0}_{300}, \underbrace{2, \cdots, 2}_{100}, \underbrace{0, \cdots, 0}_{1000})^\top$,
  $\mathbf{v}_1 = (\underbrace{0, \cdots, 0}_{300}, \underbrace{1.5, \cdots, 1.5}_{100}, \underbrace{0, \cdots, 0}_{200})^\top$,
  $\mathbf{v}_2 = (\underbrace{0, \cdots, 0}_{250}, \underbrace{1.5, \cdots, 1.5}_{150}, \underbrace{0, \cdots, 0}_{200})^\top$,
  $\mathbf{v}_3 = (\underbrace{0, \cdots, 0}_{250}, \underbrace{1.5, \cdots, 1.5}_{150}, \underbrace{0, \cdots, 0}_{200})^\top$,

TABLE I
TRAINING AND TESTING CCCS (MEAN ± STD) ESTIMATED FROM SYNTHETIC DATA SETS

		Training CCCs				Testing CCCs			
		Task1	Task2	Task3	Task4	Task1	Task2	Task3	Task4
Data1	mSCCA	0.87±0.02	0.87±0.02	0.86±0.02	0.87±0.02	0.19±0.13	0.22±0.16	0.19±0.14	0.18±0.13
	MTSCCA	0.88±0.02	0.89±0.02	0.88±0.02	0.88±0.02	0.20±0.14	0.19±0.13	0.18±0.14	0.18±0.13
	Our Method	0.95±0.01	0.95±0.01	0.95±0.01	0.95±0.01	0.18±0.13	0.18±0.13	0.29±0.18	0.18±0.13
Data2	mSCCA	0.89±0.01	0.92±0.01	0.92±0.01	0.90±0.01	0.42±0.15	0.59±0.12	0.61±0.12	0.52±0.14
	MTSCCA	0.90±0.01	0.92±0.01	0.92±0.01	0.89±0.01	0.50±0.15	0.53±0.13	0.67±0.10	0.46±0.14
	Our Method	0.96±0.01	0.96±0.00	0.96±0.00	0.96±0.00	0.48±0.16	0.55±0.13	0.66±0.10	0.42±0.15
Data3	mSCCA	0.97±0.00	0.97±0.00	0.97±0.00	0.97±0.00	0.90±0.04	0.91±0.03	0.90±0.04	0.91±0.04
	MTSCCA	0.97±0.00	0.98±0.00	0.97±0.00	0.98±0.00	0.91±0.04	0.92±0.03	0.94±0.03	0.93±0.03
	Our Method	0.98±0.00	0.98±0.00	0.98±0.00	0.98±0.00	0.92±0.04	0.92±0.03	0.93±0.03	0.93±0.03
Data4	mSCCA	0.99±0.00	0.99±0.00	0.81±0.01	0.99±0.00	0.98±0.00	0.98±0.00	0.63±0.07	0.98±0.00
	MTSCCA	0.99±0.00	0.99±0.00	0.84±0.01	0.99±0.00	0.99±0.00	0.99±0.00	0.69±0.05	0.99±0.00
	Our Method	0.99±0.00	0.99±0.00	0.87±0.01	0.99±0.00	0.99±0.00	0.99±0.00	0.72±0.05	0.99±0.00

and $\mathbf{v}_4 = (\underbrace{0, \cdots, 0}_{250}, \underbrace{1.5, \cdots, 1.5}_{150}, \underbrace{0, \cdots, 0}_{200})^\top$. The data matrices $\mathbf{X}$ is created by $x_{\ell,i} \sim N(\mu_\ell u_i, \sigma_x)$, and $\mathbf{Y}_c$ is generated by $(y_{\ell,j})_c \sim N(\mu_\ell v_{j,c}, \sigma_{y_c})$, with the random latent vector μ of length n with unit norm.

2) Bi-Multivariate Association Identification: We run all methods on four synthetic data sets, and show the training and testing canonical correlation coefficients (CCCs) in Table I. The CCCs of the first three data sets clearly show the effectiveness of each method under different noise levels. All three methods perform poorly on the first data set, since they are all overfitted. The results on Data 2 and Data 3 become better and better as the noise level decreases. We can also observe that both MTSCCA and our method outperform mSCCA owing to the multi-task modeling paradigm, and moreover, our method perform slightly better than MTSCCA, which is supported by the parameter decomposition. The CCCs of the fourth data set confirm this too. This reveals that, owing to the multi-task learning framework and parameter decomposition, the ability of identifying bi-multivariate associations could be improved.

3) Task-Consistent and Task-Specific Feature Selection: In addition to the CCC, selecting relevant features is also very important and meaningful. The heat maps in Fig. 2 present the decomposed feature selection results of our method, as well as those of the benchmarks. For our method, the identified features with non-zero weight of the task-consistent and task-specific components are quite interesting. The proposed method can not only show the features shared across multiple tasks, but also identify features that are only associated with a specific task. In contrast, both mSCCA and MTSCCA only return a single feature selection results with task-consistent or task-specific component fused. This could be insufficient when we care about which features are task-consistent or which ones are task-specific. It is interesting that a part of task-specific features are missed on Data 2 and Data 3. The reason is that, in this study, we only focus on the leading pair of canonical weights, which could omit those features with relative weak signals. We then keep on identifying the second pair of canonical weights, and our method successfully identifies these missed task-specific features. In summary, both CCCs and feature selection results demonstrate that our method is a powerful learning approach in this simulated multi-modal bi-association identification problem.

TABLE II
PARTICIPANT CHARACTERISTICS

	HC	SMC	EMCI	LMCI	AD
Num	182	75	217	184	97
Gender (M/F)	89/93	29/46	113/104	96/88	54/43
Handedness (R/L)	163/19	65/10	194/23	165/19	89/8
Age (mean±std)	73.93±5.51	71.77±5.76	70.59±7.16	71.89±7.92	73.99±8.44
Education (mean±std)	16.43±2.68	16.87±2.71	15.94±2.64	16.14±2.92	15.60±2.61

C. Results on Real Neuroimaging Genetics Data

1) Data Source: The genotyping and brain imaging data used in this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). One primary goal of ADNI has been 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 mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

The neuroimaging data of 755 non-Hispanic Caucasian participants were downloaded from the ADNI website (adni.loni.usc.edu), and the details of the participant characteristics are shown in Table II. There are five kinds of diagnostic groups, i.e. healthy control (HC), significant memory concern (SMC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI) and AD, and each of them have three modalities of imaging data, including 18F florbetapir (AV45) PET scans, fluorodeoxyglucose (FDG) PET scans, and sMRI scans. These multi-modal imaging data were aligned to each subject's same visit. The sMRI scans were processed with voxel-based morphometry (VBM) by SPM [29]. And, every scan had been aligned to a T1-weighted template image, segmented to the gray matter (GM), the white matter (WM) and the cerebrospinal fluid (CSF) maps, normalized to the standard Montreal Neurological Institute (MNI) space as $2 \times 2 \times 2$ mm^3 voxels, and smoothed with an 8mm FWHM kernel. Besides, the AV45-PET and FDG-PET scans were registered into the same MNI space. We further extracted region-of-interest (ROI) level measurements based on the MarsBaR automated anatomical labeling (AAL) atlas [30]. They were mean gray matter densities for VBM-sMRI scans, beta-amyloid depositions for AV45-PET scans and glucose utilizations for FDG-PET scans. In the experiments, the imaging measures were pre-adjusted
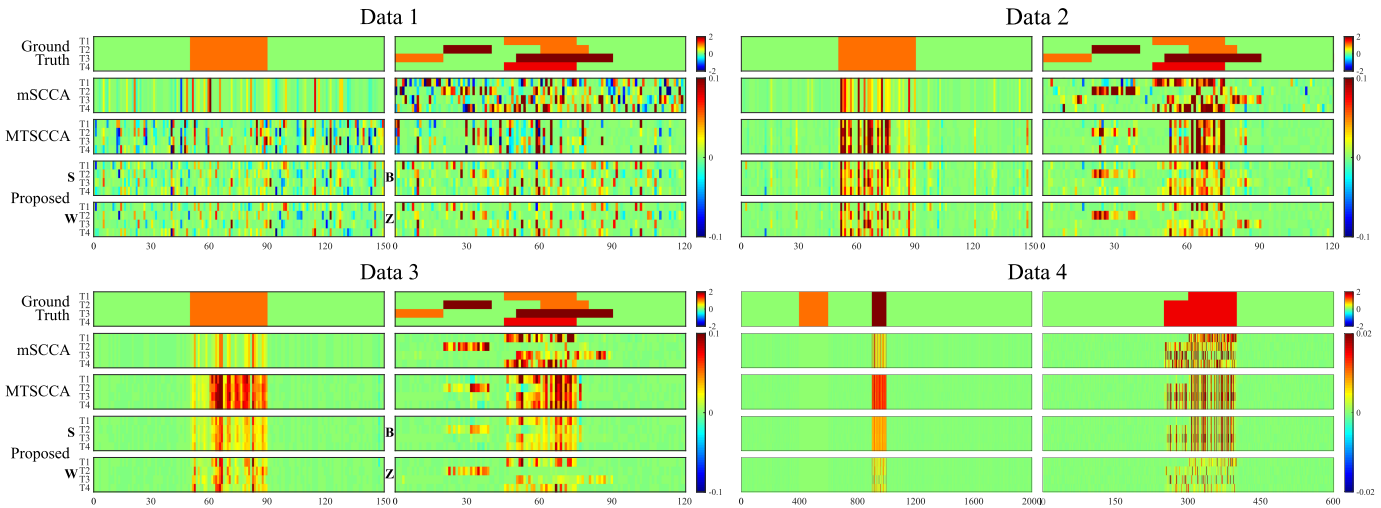
Fig. 2. Comparison of canonical weights in terms of each task for synthetic data sets. For each data set, the canonical weights **U** is shown on the left, and **V** is shown on the right. The top row shows the ground truth of **U** and **V**, and the remaining rows correspond to the SCCA methods: (1) mSCCA; (2) MTSCCA; (3) the proposed method. Our method has two weights for **X** and each $\mathbf{Y}_c$ owing to the parameter decomposition. Within each panel, there are four rows corresponding to four SCCA tasks (denoted as T1~T4) between **X** and each $\mathbf{Y}_c$.

to remove the effects of the baseline age, gender, education, and handedness by the regression weights derived from the HC subjects.

The genotyping data were also downloaded from the ADNI website. They were genotyped using the Human 610-Quad or OmniExpress Array (Illumina, Inc., San Diego, CA, USA), and preprocessed using the standard quality control (QC) and imputation steps. According to the quality-controlled SNPs, the missing genotypes were imputed by the MaCH software tool [31]. Among all human chromosomes, the chromosome 19 has the highest gene density of all human chromosomes, more than double the genome-wide average [32], [33]. In addition, this chromosome also includes the well-known AD risk genes such as *APOE*, *TOMM40* and *ABCA7*. Therefore, a bi-multivariate association study between this chromosome and whole brain imaging markers could be of great interest, and has potential to yield interesting AD risk factors. We investigated 1,011 SNPs from chromosome 19 with the well-known AD risk genes such as *APOE* included. The linkage disequilibrium (LD) block information, indicating the structure of highly correlated SNPs, was used as the prior knowledge. Our aim is to study the associations between multi-modal imaging QTs (GM densities for VBM-sMRI scans, amyloid values for AV45-PET scans and glucose utilizations for FDG-PET scans) and this segment of SNPs, and select those relevant imaging markers and genetic loci.

2) Bi-Multivariate Association Identification: We first show the training and testing CCCs in Table III, which indicates the strength of the identified bi-multivariate associations between SNPs and imaging QTs of three modalities. There are three CCCs for each method since we have three imaging modalities, thereby three SCCA tasks. In this table, we clearly observe that the proposed method obtains better CCCs than or comparable CCCs to both mSCCA and MTSCCA in terms of each task. The multi-task SCCA also performs better than mSCCA for most cases, confirming the superior modeling capability of multi-task learning in multi-modal imaging

TABLE III
CCCs (MEAN±STD) ESTIMATED BETWEEN SNPS AND
IMAGING QTS OF THREE MODALITIES

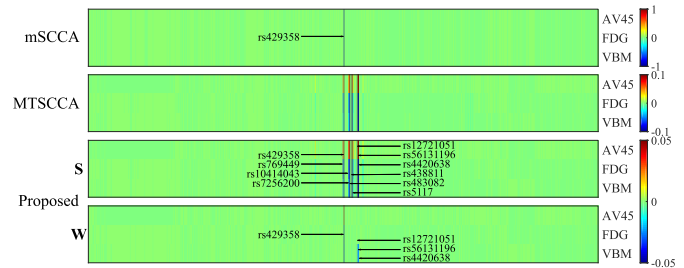|	Training CCCs		
	SNP-AV45	SNP-FDG	SNP-VBM
mSCCA	0.44±0.01	0.33±0.01	0.25±0.02
MTSCCA	0.47±0.01	0.35±0.01	0.29±0.01
Our Method	0.48±0.01	0.36±0.01	0.29±0.01
	Testing CCCs		
mSCCA	0.41±0.07	0.29±0.07	0.21±0.07
MTSCCA	0.43±0.07	0.30±0.06	0.19±0.08
Our Method	0.44±0.07	0.30±0.06	0.21±0.07

Fig. 3. Comparison of canonical weights of SNPs in terms of each task. Each row corresponds to an SCCA method: (1) mSCCA; (2) MTSCCA and (3) the proposed method. Our method has two weights for SNPs and imaging QTs owing to the parameter decomposition. Within each panel, there are three rows corresponding to three SCCA tasks.

genetic scenes. This demonstrates that, decomposing canonical weights into task-consistent and task-specific components, and penalizing them distinctly to pursue a diverse feature selection, the dirty MTSCCA exhibits improved bi-multivariate associations.

3) Modality-Consistent and Modality-Specific Feature Selection: Now we investigate the identified SNPs and imaging QTs based on the absolute values of canonical weights. The heat maps in Fig. 3 show the feature selection for SNPs. Since our model has two separate components for SNPs, i.e. the task-consistent component **S** and the task-specific component

W, we show both of them here. mSCCA yields one canonical weight vector for SNPs, and thus we repeatedly stack its weight vector for three times. We observe that all SNPs with non-zero values of our method have been shown to be relevant to the progression of AD. For example, rs429358 (*APOE*) is identified by both **S** and **W**, demonstrating its strong association with AD. In addition, the dirty MTSCCA shows a clear task-consistent pattern, indicating that these SNPs, e.g. rs12721051 (*APOC1*) [34], rs56131196 (*APOC1*) [34], rs438811 (*APOC1*) [35], rs483082 (*APOC1*), rs5117 (*APOC1*) etc., could be identified no matter which imaging technology is employed. Our method and MTSCCA identify more AD-related loci than mSCCA, demonstrating the multi-task modeling possesses comprehensive feature selection capacity. The heat maps of imaging QTs, shown in Fig. 4, exhibit interesting task-consistent and task-specific profiles. Our method shows that the left hippocampus, the left olfactory sulcus [36], the right inferior parietal lobule [37] and the left amygdala [38] exhibit clearly task-consistent patterns, indicating that these brain areas can be identified by all imaging technologies. Besides, task-specific **Z** shows that the beta-amyloid deposition in the left medial orbitofrontal cortex [39] and the left medial frontal gyrus could be identified using the AV45-PET scans. The left and right angular gyri [40], and the cingulum [41] are identified by using the FDG-PET scans. Both left and right of the eighth cerebellum [42] are highlighted when using the VBM-sMRI scans. MTSCCA and mSCCA can also identify several meaningful brain areas, however, they could not uncover the different types of complex associations between SNPs and imaging QTs of multiple modalities. This real data study demonstrates that the dirty MTSCCA could be very promising and meaningful in multi-modal brain imaging genetics.

IV. DISCUSSION

In this section, we investigate the selected features regarding the SNPs and imaging QTs, and their relationships to the diagnosis status. This could further demonstrate the stratified feature selection ability of the proposed method.

A. Top Selected Loci

We average the canonical weights across five folds to select the top ten SNPs and show them in Table IV where both the modality-consistent and modality-specific SNPs are contained. The top ten modality-consistent SNPs and those of the modality-specific are similar and the major difference is their SNPs's priorities. This is interesting since it reveals that imaging QTs from different scanning machines focus on different aspects of Alzheimer's disease, thereby leading to the identification of SNPs with distinct priorities. At the same time, as long as the relevant ROIs are correctly identified, the same sets of SNPs could be identified no matter which imaging technology is used. It is worth noting that rs429358, the well-known AD-risk locus, ranks the first within all three modality-specific results, while it is not the first one in the modality-consistent result. This seems unusual at first glance but the truth is not. In the task-consistent results, the first six

TABLE IV
TOP TEN MODALITY-CONSISTENT AND MODALITY-SPECIFIC SNPs BY AVERAGED CANONICAL WEIGHTS

Modality-consistent	AV45-specific	FDG-specific	VBM-specific
rs12721051	rs429358	rs429358	rs429358
rs56131196	rs12721051	rs12721051	rs12721051
rs4420638	rs56131196	rs56131196	rs56131196
rs438811	rs4420638	rs4420638	rs4420638
rs483082	rs769449	rs769449	rs438811
rs5117	rs10414043	rs10414043	rs483082
rs429358	rs7256200	rs7256200	rs5117
rs769449	rs438811	rs10119	rs10119
rs10414043	rs483082	rs438811	rs769449
rs7256200	rs73052335	rs483082	rs12721046

loci are from the same LD group and thus their combined effect might dominate rs429358 owing to the $G_{2,1}$-norm for consistent feature selection across multiple tasks.

To understand the modality-specific SNPs, we choose rs10119, rs73052335 and rs12721046 for further investigation because they are identified by only one or two SCCA tasks. The one-way analysis of variance (ANOVA) is applied to verify a SNP's effect on the diagnosis with age, gender, years of education and handedness being included as covariates. The *p*-values show that all three SNPs pass through the significance level (rs10119, $p = 1.42 \times 10^{-14}$; rs73052335, $p = 6.13 \times 10^{-12}$; rs12721046, $p = 8.41 \times 10^{-11}$), indicating their strong relationship to the AD. This demonstrates that the dirty MTSCCA could successfully find out meaningful modality-specific SNPs.

B. Top Selected Brain Imaging ROIs

The top ten brain imaging ROIs based on the averaged canonical weights are shown in Table V. We observe that distinct sets of modality-consistent and modality-specific ROIs are identified in our analyses. Within the modality-consistent ROIs, three types of imaging measurements show high consistency which is guaranteed by the $\ell_{2,1}$-norm. Meanwhile, there are still modality-specific ROIs such as the left inferior occipital lobe of FDG-PET and left insula gyrus of VBM-sMRI scans. The brain glucose hypometabolism in the occipital lobe revealed by the FDG-PET, and the atrophy in the left insula gyrus revealed by the VBM-sMRI have been shown to be related to AD [43], [44]. This complex and diverse neurodegenerative patterns of AD is successfully identified by our method, endorsing the necessity of the modality-specific feature selection, which further underpins the motivation and significance of this study.

We further investigate the selected modality-specific ROIs. The first ROI of each SCCA task is the left medial orbitofrontal gyrus (AV45-PET), the left posterior cingulate gyrus (FDG-PET), and the left hippocampus lobe (VBM-sMRI), respectively. The one-way ANOVA analysis shows that their main effects reach the significance level ($p < 2.2 \times 10^{-16}$) when including age, gender, education and handedness as covariates. This is very interesting since our method not only identifies significant imaging ROIs, but also assigns different priorities to different ROIs based on different imaging technologies. Therefore, our method provides a diverse and meaningful clue for AD diagnosis and monitoring.
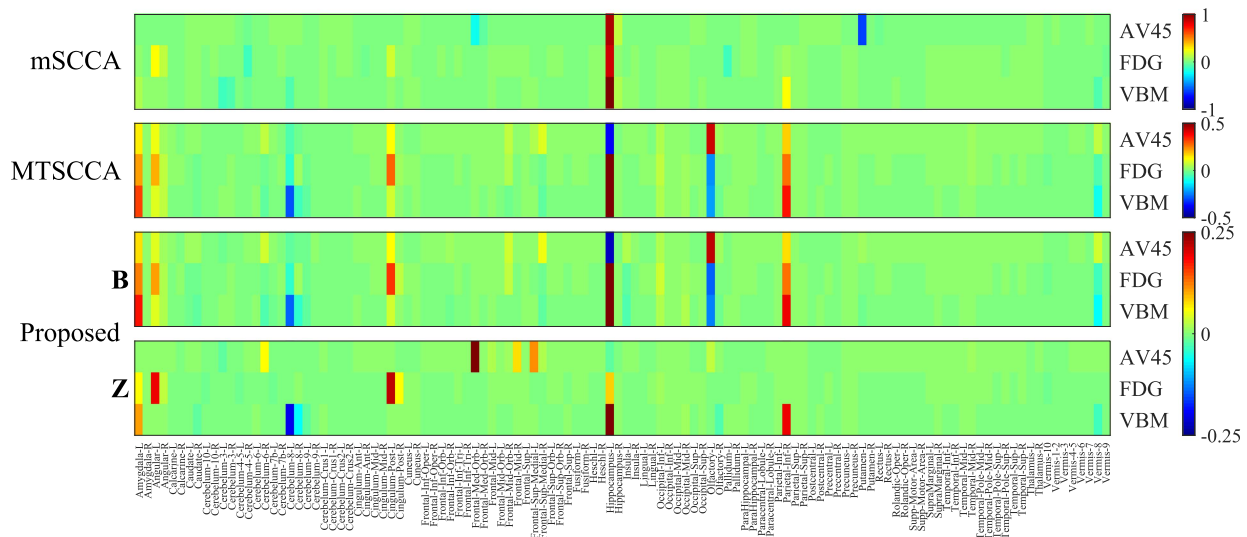
Fig. 4. Comparison of canonical weights of imaging QTs in terms of each task. Each row corresponds to an SCCA method: (1) mSCCA; (2) MTSCCA and (3) the proposed method. Our method has two weights for SNPs and imaging QTs owing to the parameter decomposition. Within each panel, there are three rows corresponding to three SCCA tasks.

TABLE V
TOP TEN MODALITY-CONSISTENT AND MODALITY-SPECIFIC ROIS BY AVERAGED CANONICAL WEIGHTS

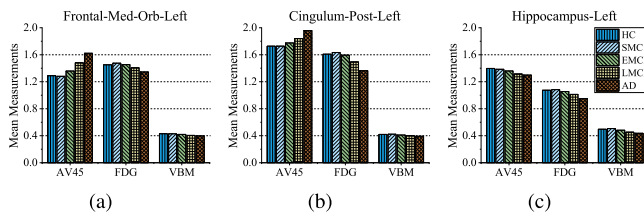| Modality-consistent			Modality-specific		
AV45	FDG	VBM	AV45	FDG	VBM
Hippocampus-Left	Hippocampus-Left	Hippocampus-Left	Frontal-Med-Orb-Left	Cingulum-Post-Left	Hippocampus-Left
Olfactory-Left	Cingulum-Post-Left	Parietal-Inf-Right	Frontal-Sup-Medial-Left	Angular-Left	Cerebelum-8-Left
Parietal-Inf-Right	Olfactory-Left	Amygdala-Left	Frontal-Mid-Right	Hippocampus-Left	Parietal-Inf-Right
Amygdala-Left	Parietal-Inf-Right	Cerebelum-8-Left	Cerebelum-6-Right	Cingulum-Post-Right	Amygdala-Left
Cingulum-Post-Left	Amygdala-Left	Olfactory-Left	Olfactory-Left	Amygdala-Left	Cerebelum-8-Right
Frontal-Sup-Medial-Right	Angular-Left	Vermis-8	Frontal-Sup-Medial-Right	Angular-Right	Vermis-8
Angular-Left	Cerebelum-8-Left	Cerebelum-8-Right	Hippocampus-Left	Occipital-Inf-Left	Olfactory-Right
Cerebelum-6-Right	Occipital-Inf-Left	Angular-Left	Cerebelum-3-Left	Vermis-10	Cerebelum-9-Left
Vermis-8	Vermis-8	Cingulum-Post-Left	Frontal-Mid-Left	Cerebelum-10-Left	Caudate-Right
Frontal-Mid-Orb-Right	Cingulum-Post-Right	Insula-Left	Temporal-Mid-Right	Cerebelum-4-5-Left	Occipital-Mid-Right

Fig. 5. The measurement distributions of imaging QTs (mean value the first ROI of each SCCA task) among different diagnostic groups and different imaging modalities. (a) The left medial orbitofrontal gyrus. (b) The left posterior cingulate gyrus. (c) The left hippocampus lobe.

C. Population Stratification Analysis

We here conduct the population stratification analysis to further evaluate the effectiveness of the modality-specific feature selection. For the sake of simplicity, we investigate the first ROI of each SCCA task associating with each imaging modality, since those remaining ROIs can be analyzed in the same way.

Fig. 5(a) presents distributions of the beta-amyloid deposition in the left medial orbitofrontal gyrus among different diagnostic groups and different imaging modalities. We clearly observe that the beta-amyloid deposition patterns exhibits

differently. In particular, the beta-amyloid deposition shows a significant increase (HCs vs. ADs: $p = 6.93 \times 10^{-19}$, SMCs vs. ADs: $p = 5.34 \times 10^{-16}$, EMCIs vs. ADs: $p = 8.73 \times 10^{-14}$, LMCIs vs. ADs: $p = 2.12 \times 10^{-4}$) in the AD group compared with other groups. This significance also exists in EMCIs and LMCIs compared with those other diagnostic groups. The AD group also shows a clear brain glucose hypometabolism (FDG-PET) in this ROI compared with those other groups. It is interesting that, still in the left medial orbitofrontal gyrus, the significant atrophy happens in dementia groups such as LMCIs and ADs, while no pronounced difference among preclinical and prodromal diagnostic groups such as EMCIs, SMCs and HCs. The proposed dirty MTSCCA, as expected, identifies this AD risk ROI in consistent with its diverse distributions among different groups. The similar patterns can also be observed in the left posterior cingulate gyrus in Fig. 5(b), where the hypometabolism, revealed by the FDG-PET, shows a significantly lower level in AD patients than non-AD subjects. In contrast, the beta-amyloid deposition in this ROI shows no significant difference among preclinical and prodromal groups, and also shows no significant difference between dementia groups. Interestingly, between preclinical or prodromal groups and dementia groups, the beta-amyloid deposition reaches the significance level. Fig. 5(c) shows that
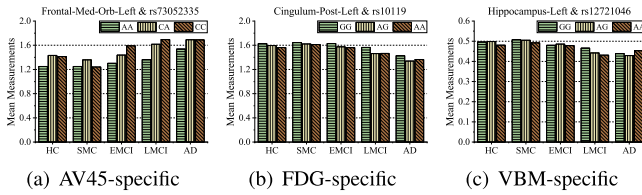
Fig. 6. Pairwise comparisons for modality-specific QT-SNP-diagnosis interactions within HCs, SMCs, EMCIs, LMCIs and ADs, respectively. Two-way ANOVA was applied to access the effects of genotype and baseline diagnosis on imaging QTs. Age, gender, years of education, handedness were included as covariates. (a) The beta-amyloid deposition in the left medial orbitofrontal gyrus, rs73052335 and diagnostic groups. (b) The glucose metabolism in the left posterior cingulate gyrus, rs10119 and diagnostic groups. (c) The atrophy in the left hippocampus lobe, rs12721046 and diagnostic groups.

more severe atrophy happens to the left hippocampus lobe in AD patients compared with those non-AD subjects. The severer the atrophy, the severer the dementia is. Moreover, both reduced beta-amyloid deposition and glucose hypometabolism happen to the left hippocampus lobe, which is different from those observations in Fig. 5(a) and Fig. 5(b). Combined three subfigures together, we obtain the similar results to previous works which show that regional beta-amyloid deposition and regional glucose metabolism have little to no association [45]. On the contrary, we cannot draw the same conclusion without identifying modality-specific imaging QTs.

Despite the pairwise comparisons among different diagnostic groups, it is also necessary to interpret the identified phenotype-genotype associations within each group in this imaging genetic study. On this account, we use the first modality-specific QT-SNP pair in this refined analysis. Certainly, those modality-consistent and modality-specific QT-SNP pairs can be analyzed in the same way. Two-way ANOVA results show that the main effects of rs73052335 genotype ($p = 2.55 \times 10^{-30}$) and diagnosis ($p = 2.64 \times 10^{-16}$) on beta-amyloid deposition in the left medial orbitofrontal gyrus reach the significant level, while their SNP-by-diagnosis interaction effect ($p = 0.46$) is not. In addition, Fig. 6(a) contains pairwise comparisons among the heterozygote CA, homozygous AA and CC within each groups respectively. We observe that, excluding SMCs, subjects with heterozygote CA and homozygous CC have higher deposition than those with homozygous AA. Furthermore, subjects with CC tends to hold higher deposition than heterozygote CA in EMCIs and LMCIs but not ADs. This reveals that subjects with heterozygote CA and homozygous CC in rs73052335 locus are vulnerable to have higher beta-amyloid deposition.

As for brain glucose metabolism, using measurements in the left posterior cingulate gyrus, two-way ANOVA results reveal that main effects of rs10119 genotype ($p = 1.85 \times 10^{-10}$), diagnosis ($p = 3.59 \times 10^{-24}$), as well as their SNP-by-diagnosis interaction ($p = 0.02$) are significantly different among distinct groups. The histogram, in Fig. 6(b), within each group indicates that subjects with the minor allele A, compared with ones without it, suffer from severer glucose hypometabolism in left posterior cingulate gyrus. In LMCIs and ADs, somewhat severer glucose hypometabolism happens

to patients with heterozygote AG than those with homozygous AA, while this is not for HCs, SMCs and EMCIs. These results indicate that subjects with heterozygote AG and minor homozygous are vulnerable to severer glucose hypometabolism.

The atrophy in the left hippocampus is a well-known AD hallmark. The main effects of rs12721046 genotype ($p = 6.38 \times 10^{-3}$) and diagnosis ($p = 6.91 \times 10^{-29}$) are pronounced among distinct groups. Fig. 6(c) shows that atrophy patterns for subjects with heterozygote AG, homozygous GG and AA exhibit distinctly across groups. In particular, subjects with homozygous AA suffer from heavy atrophy in left hippocampus lobe in HCs, SMCs, EMCIs and LMCIs, but not in ADs. This is interesting and further investigation should be warranted.

In summary, results above might be caused by the complicated pathogenesis that hallmarks of AD of different imaging technologies exhibit regional heterogeneity. On one hand, this diversity and complexity captured by different imaging technologies such as PET and sMRI, offers the opportunity to understand the pathogenesis of AD comprehensively. On the other hand, it makes an urgent request for the modality-consistent and the modality-specific feature selection, since modality-consistent (MTSCCA [5] and mSCCA [16]) or modality-specific (conventional independent SCCA [16]) methods alone are insufficient. Finally, these results demonstrate that the dirty MTSCCA can identify both modality-consistent and modality-specific SNPs, imaging QTs and their associations in an integrated model. Therefore, our method is of great importance and meaning for multi-modal brain imaging genetics benefitting from its novel multi-modal bi-multivariate learning and clever parameter decomposition strategy.

V. CONCLUSIONS

Imaging data collected by different technologies, measuring the same brain distinctly, might carry complementary information. In this paper, we propose a dirty multi-task SCCA method which incorporates multiple modalities of imaging data into a unified model. By decomposing the SCCA's canonical weights into the task-consistent component and the task-specific component, and penalizing them distinctly, our method has the ability of identifying diverse and meaningful bi-multivariate associations between SNPs and imaging QTs. We derive an efficient optimization algorithm to solve the dirty model, and it is guaranteed to converge.

We compared the dirty MTSCCA with the conventional multi-view SCCA (mSCCA) and mutli-task SCCA (MTSCCA) on both synthetic data sets and real neuroimaging genetic data. The four synthetic data sets have different numbers of samples, features, and noises. The results on synthetic data sets demonstrated that our method improved both correlation coefficients and feature selection results. The real neuroimaging genetic data were downloaded from the ADNI database. Our method also obtained better performance than the benchmarks with higher correlation coefficients and clearer canonical weight patterns. Besides, our method identified task-consistent and task-specific features

with respect to SNPs and imaging QTs. The post analysis showed that most of the top ten SNPs and ROIs, including both task-consistent and task-specific markers, are correlated with AD. The task-specific ROIs identified by our method showed promising consistency with previous studies that different ROIs could be the hallmark of AD if different imaging technologies were used. This demonstrated the effectiveness of the proposed dirty multi-task SCCA, and further demonstrated it could be a powerful tool in big brain imaging genetics. Since the diagnosis status could be helpful for identifying interesting SNPs and imaging QTs, we intend to incorporate the diagnosis status into the model to make it supervised.

REFERENCES

[1] S. G. Potkin *et al.*, "Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: Methodological considerations," *Cognit. Neuropsychiatry*, vol. 14, nos. 4–5, pp. 391–418, Jul. 2009.

[2] A. J. Saykin *et al.*, "Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans," *Alzheimer's Dementia*, vol. 11, no. 7, pp. 792–814, Jul. 2015.

[3] L. Shen *et al.*, "Genetic analysis of quantitative phenotypes in AD and MCI: Imaging, cognition and biomarkers," *Brain Imag. Behav.*, vol. 8, no. 2, pp. 183–207, Jun. 2014.

[4] V. D. Calhoun and J. Sui, "Multimodal fusion of brain imaging data: A key to finding the missing Link(s) in complex mental illness," *Biol. Psychiatry, Cognit. Neurosci. Neuroimag.*, vol. 1, no. 3, pp. 230–244, May 2016.

[5] L. Du *et al.*, "Fast multi-task SCCA learning with feature selection for multi-modal brain imaging genetics," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 356–361.

[6] S. Lee, J. Zhu, and E. P. Xing, "Adaptive multi-task lasso: With application to eQTL detection," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2010, pp. 1306–1314.

[7] H. Wang *et al.*, "Identifying quantitative trait loci via group-sparse multitask regression and feature selection: An imaging genetics study of the ADNI cohort," *Bioinformatics*, vol. 28, no. 2, pp. 229–237, Jan. 2012.

[8] H. Wang *et al.*, "From phenotype to genotype: An association study of longitudinal phenotypic markers to Alzheimer's disease relevant SNPs," *Bioinformatics*, vol. 28, no. 18, pp. i619–i625, Sep. 2012.

[9] J. K. Mai, M. Majtanik, and G. Paxinos, *Atlas of the Human Brain*. New York, NY, USA: Academic, 2015.

[10] J. Chen, F. D. Bushman, J. D. Lewis, G. D. Wu, and H. Li, "Structure-constrained sparse canonical correlation analysis with an application to microbiome data analysis," *Biostatistics*, vol. 14, no. 2, pp. 244–258, Apr. 2013.

[11] X. Chen and H. Liu, "An efficient optimization algorithm for structured sparse CCA, with applications to eQTL mapping," *Statist. Biosci.*, vol. 4, no. 1, pp. 3–26, May 2012.

[12] L. Du *et al.*, "Structured sparse canonical correlation analysis for brain imaging genetics: An improved GraphNet method," *Bioinformatics*, vol. 32, no. 10, pp. 1544–1551, May 2016.

[13] L. Du *et al.*, "A novel SCCA approach via truncated ℓ_1-norm and truncated group lasso for brain imaging genetics," *Bioinformatics*, vol. 34, no. 2, pp. 278–285, Jan. 2018.

[14] L. Du *et al.*, "A novel structure-aware sparse learning algorithm for brain imaging genetics," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Sep. 2014, pp. 329–336.

[15] L. Du *et al.*, "Identifying associations between brain imaging phenotypes and genetic factors via a novel structured SCCA approach," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Cham, Switzerland: Springer, Jun. 2017, pp. 543–555.

[16] D. M. Witten and R. J. Tibshirani, "Extensions of sparse canonical correlation analysis with applications to genomic data," *Stat. Appl. Genet. Mol. Biol.*, vol. 8, no. 1, pp. 1–27, Jan. 2009.

[17] I. Wilms and C. Croux, "Sparse canonical correlation analysis from a predictive point of view," *Biometrical J.*, vol. 57, no. 5, pp. 834–851, Sep. 2015.

[18] L. Du *et al.*, "Detecting genetic associations with brain imaging phenotypes in Alzheimer's disease via a novel structured SCCA approach," *Med. Image Anal.*, vol. 61, Apr. 2020, Art. no. 101656.

[19] X. Hao *et al.*, "Mining outcome-relevant brain imaging genetic associations via three-way sparse canonical correlation analysis in Alzheimer's disease," *Sci. Rep.*, vol. 7, no. 1, p. 44272, May 2017.

[20] J. Fang, D. Lin, S. C. Schulz, Z. Xu, V. D. Calhoun, and Y.-P. Wang, "Joint sparse canonical correlation analysis for detecting differential imaging genetics modules," *Bioinformatics*, vol. 32, no. 22, pp. 3480–3488, 2016.

[21] L. Du *et al.*, "Identifying progressive imaging genetic patterns via multi-task sparse canonical correlation analysis: A longitudinal study of the ADNI cohort," *Bioinformatics*, vol. 35, no. 14, pp. i474–i483, Jul. 2019.

[22] L. Du *et al.*, "Identifying diagnosis-specific genotype-phenotype associations via joint multi-task sparse canonical correlation analysis and classification," *Bioinformatics*, to be published, doi: 10.1093/bioinformatics/btaa434.

[23] J. Sui, T. Adali, Q. Yu, J. Chen, and V. D. Calhoun, "A review of multivariate methods for multimodal fusion of brain imaging data," *J. Neurosci. Methods*, vol. 204, no. 1, pp. 68–81, Feb. 2012.

[24] A. Jalali, P. Ravikumar, and S. Sanghavi, "A dirty model for multiple sparse regression," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7947–7968, Dec. 2013.

[25] L. Du *et al.*, "A dirty multi-task learning method for multi-modal brain imaging genetics," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, Oct. 2019, pp. 447–455.

[26] M. W. Weiner *et al.*, "The Alzheimer's disease neuroimaging initiative: Progress report and future plans," *Alzheimers. Dement.*, vol. 6, no. 3, pp. 202–211, 2010.

[27] L. Du *et al.*, "Multi-task sparse canonical correlation analysis with application to multi-modal brain imaging genetics," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Oct. 15, 2019, doi: 10.1109/TCBB.2019.2947428.

[28] D. M. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, vol. 10, no. 3, pp. 515–534, Jul. 2009.

[29] J. Ashburner and K. J. Friston, "Voxel-based morphometry—The methods," *NeuroImage*, vol. 11, no. 6, pp. 805–821, 2000.

[30] N. Tzourio-Mazoyer *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, Jan. 2002.

[31] Y. Li, C. J. Willer, J. Ding, P. Scheet, and G. R. Abecasis, "MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes," *Genetic Epidemiol.*, vol. 34, no. 8, pp. 816–834, Dec. 2010.

[32] J. Grimwood *et al.*, "The DNA sequence and biology of human chromosome 19," *Nature*, vol. 428, no. 6982, pp. 529–535, 2004.

[33] J. C. Venter *et al.*, "The sequence of the human genome," *Science*, vol. 291, no. 5507, pp. 1304–1351, 2001.

[34] X. Zhou *et al.*, "Non-coding variability at the APOE locus contributes to the Alzheimer's risk," *Nature Commun.*, vol. 10, no. 1, pp. 1–16, Dec. 2019.

[35] Q. Yan *et al.*, "Genome-wide association study of brain amyloid deposition as measured by Pittsburgh Compound-B (PiB)-PET imaging," *Mol. Psychiatry*, Oct. 2018, doi: 10.1038/s41380-018-0246-7.

[36] M. M. Vasavada *et al.*, "Olfactory cortex degeneration in Alzheimer's disease and mild cognitive impairment," *J. Alzheimer's Disease*, vol. 45, no. 3, pp. 947–958, Mar. 2015.

[37] S. J. Greene and R. J. Killiany, "Subregions of the inferior parietal lobule are affected in the progression to Alzheimer's disease," *Neurobiol. Aging*, vol. 31, no. 8, pp. 1304–1311, Aug. 2010.

[38] S. P. Poulin, R. Dautoff, J. C. Morris, L. F. Barrett, and B. C. Dickerson, "Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity," *Psychiatry Res., Neuroimag.*, vol. 194, no. 1, pp. 7–13, Oct. 2011.

[39] J. Sepulcre *et al.*, "Neurogenetic contributions to amyloid beta and tau spreading in the human cortex," *Nature Med.*, vol. 24, no. 12, pp. 1910–1918, Dec. 2018.

[40] K. B. Walhovd *et al.*, "Combining MR imaging, positron-emission tomography, and CSF biomarkers in the diagnosis and prognosis of Alzheimer disease," *Amer. J. Neuroradiol.*, vol. 31, no. 2, pp. 347–354, Feb. 2010.

[41] L. Mosconi, "Brain glucose metabolism in the early and specific diagnosis of Alzheimer'S disease," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 32, no. 4, pp. 486–510, Apr. 2005.

[42] J. Wegiel *et al.*, "Cerebellar atrophy in Alzheimer's disease—Clinicopathological correlations," *Brain Res.*, vol. 818, no. 1, pp. 41–50, 1999.

[43] V. K. N. Shivamurthy, A. K. Tahari, C. Marcus, and R. M. Subramaniam, "Brain FDG PET and the diagnosis of dementia," *Amer. J. Roentgenol.*, vol. 204, no. 1, pp. W76–W85, Jan. 2015.

[44] Y. Y. Fathy *et al.*, "Differential insular cortex sub-regional atrophy in neurodegenerative diseases: A systematic review and meta-analysis," *Brain Imag. Behav.*, pp. 1–18, Apr. 2019.

[45] A. Altmann, B. Ng, S. M. Landau, W. J. Jagust, and M. D. Greicius, "Regional brain hypometabolism is unrelated to regional amyloid plaque burden," *Brain*, vol. 138, no. 12, pp. 3734–3746, Dec. 2015.