

Identifying diagnosis-specific genotype–phenotype associations via joint multitask sparse canonical correlation analysis and classification

Lei Du^{1,*}, Fang Liu¹, Kefei Liu², Xiaohui Yao², Shannon L. Risacher³, Junwei Han¹, Lei Guo¹, Andrew J. Saykin³ and Li Shen^{2,*}; for the Alzheimer’s Disease Neuroimaging Initiative[†]

¹Department of intelligent science and technology, School of Automation, Northwestern Polytechnical University, Xi’an 710072, China, ²Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA and ³Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN 46202, USA

*To whom correspondence should be addressed.†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/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Abstract

Motivation: Brain imaging genetics studies the complex associations between genotypic data such as single nucleotide polymorphisms (SNPs) and imaging quantitative traits (QTs). The neurodegenerative disorders usually exhibit the diversity and heterogeneity, originating from which different diagnostic groups might carry distinct imaging QTs, SNPs and their interactions. Sparse canonical correlation analysis (SCCA) is widely used to identify bi-multivariate genotype–phenotype associations. However, most existing SCCA methods are unsupervised, leading to an inability to identify diagnosis-specific genotype–phenotype associations.

Results: In this article, we propose a new joint multitask learning method, named MT–SCCALR, which absorbs the merits of both SCCA and logistic regression. MT–SCCALR learns genotype–phenotype associations of multiple tasks jointly, with each task focusing on identifying one diagnosis-specific genotype–phenotype pattern. Meanwhile, MT–SCCALR cannot only select relevant SNPs and imaging QTs for each diagnostic group alone, but also allows the selection of those shared by multiple diagnostic groups. We derive an efficient optimization algorithm whose convergence to a local optimum is guaranteed. Compared with two state-of-the-art methods, MT–SCCALR yields better or similar canonical correlation coefficients and classification performances. In addition, it owns much better discriminative canonical weight patterns of great interest than competitors. This demonstrates the power and capability of MTSCCAR in identifying diagnostically heterogeneous genotype–phenotype patterns, which would be helpful to understand the pathophysiology of brain disorders.

Availability and implementation: The software is publicly available at <https://github.com/dulei323/MTSCCALR>.

Contact: dulei@nwpu.edu.cn or Li.Shen@pennmedicine.upenn.edu

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Alzheimer’s disease (AD) is a severe neurodegenerative disorder incurring heavy economic and emotional costs to patients and their families (Alzheimer’s Association, 2013). Generally, AD, as well as other neurodegenerative and neuropsychiatric disorders, exhibits the heterogeneity and diversity (Lam *et al.*, 2013; Wang *et al.*, 2015), supported by mounting evidence that the sporadic AD could be multiple diseases instead of a single disease (Au *et al.*, 2015). Therefore, uncovering the diagnosis-specific, including subgroup-specific and normal ageing-specific, genetic factors, imaging

phenotypes and their interactions is an important and meaningful research topic. This could underpin AD subgroup identification, thereby help with mechanistic understanding of this neurodegenerative disorder.

Brain imaging genetics provides us a powerful opportunity to gain new in-depth insights into the genetic basis of the phenotypic characteristics of the brain (Shen and Thompson, 2020). Within this area, the genetic variations such as single nucleotide polymorphisms (SNPs) and brain imaging quantitative traits (QTs) are jointly analyzed with expectation to understand the normal and disordered brain function and behavior (Saykin *et al.*, 2015; Shen *et al.*, 2014).

Given a large number of SNPs and imaging QTs, the univariate methods and the regression based methods have limited capability. The univariate methods treat each marker (SNP or QT) independently and thus they inevitably overlook the relationship within SNPs and imaging QTs (Shen et al., 2010). The regression-based methods alleviate this issue by looking into multiple SNPs' (QTs') impact on a single imaging QT (SNP) or a few candidate imaging QTs (SNPs). Obviously, they cannot select features of interest for SNPs and imaging QTs simultaneously, which, however, is a critical issue in imaging genetics.

The above drawback can be addressed by bi-multivariate learning methods, such as sparse canonical correlation analysis (SCCA), partial least squares regression (Beaton et al., 2014) and reduced rank regression (Vounou et al., 2010). Among which SCCA and its variants become more and more popular due to its simplicity but powerful detection capability (Chen et al., 2013; Du et al., 2016, 2018, 2019a; Witten and Tibshirani, 2009). However, SCCA is unsupervised and thus the diagnosis information is usually overlooked. This might lead to discovering disease irrelevant SNP–QT associations even though the discriminative information among distinct diagnostic groups has been implied in imaging QTs. Many efforts have been made in this direction to enable supervised SCCA. For example, Yan et al. (2017) proposed discriminative SCCA (DSCCA) which considers the relationship within the same diagnostic group and that between groups. They later designed another outcome-relevant SCCA (Yan et al., 2018) to make use of the diagnosis information through regularization with the similarity matrix among subjects. Zille et al. (2018) proposed a fused method in which the diagnosis information is considered by the regression objective while the SNP–QT association is captured by the SCCA objective simultaneously. A common critical issue holding by methods above is that they find out only one feature subset in terms of SNP–QT associations for all diagnostic groups. According to the feature selection taxonomy in the machine learning community (Baggenstoss, 1999; Pineda-Bautista et al., 2011), these SCCAs fall into the traditional feature selection category, in contrast to the class-specific [or diagnosis-specific (The *diagnosis-specific* and *class-specific* are alternatively used throughout this article without distinction.)] feature selection (Pineda-Bautista et al., 2011), as they only select a single feature subset to discriminate all classes. Obviously, the identified SNP–QT associations are not diagnosis-specific. Nevertheless, identifying diagnosis-specific imaging genetic patterns could be of great interest and meaning, to which the targeted in-depth investigation, subgroup identification and personalized medication could be applied (Mukherjee et al., 2018).

On this account, the diagnosis-specific feature selection methods which select a feature subset (possibly different) for each diagnostic group (Pineda-Bautista et al., 2011) is more desirable and essential. This topic is in agreement with class-specific feature selection. For example, Wang et al. (2015) used the multitask support vector machine (SVM) to learn multiple heterogeneous classification tasks together. This method is somewhat difficult to interpret due to its nonobvious modeling strategy, and it only identifies imaging markers which is insufficient to subtype identification. The joint SCCA (JSCCA) (Fang et al., 2016) studied the imaging genetic associations within each diagnostic group via a modified multiview SCCA (mSCCA). However, its shortcoming is conspicuous as many undesirable associations could dominate the association of interest when conducting SCCA within a single diagnostic group (Lorena et al., 2008). Strictly speaking, both methods are beyond the diagnosis-specific feature selection due to their nonstandard modeling paradigm (Pineda-Bautista et al., 2011; Wang et al., 2016; Zhang and Wu, 2015). Thus, both of them are inadequate for diagnosis-specific identification. It is straightforward to employ well-studied diagnosis-specific algorithms for imaging genetics (Pineda-Bautista et al., 2011). Unfortunately, they are classification-based methods indicating that they can only identify label-relevant features. As a result, the primary mission of brain imaging genetics, i.e. identifying meaningful SNP–QT associations, is overlooked.

To address the issues above, we propose a novel multitask bi-multivariate learning method with feature selection to identify diagnosis-specific genotype–phenotype patterns for each patient group as well as normal controls. The proposed method, named MT–SCCALR, integrates multitask SCCA and multitask logistic

regression (LR) in a unified model. The advantages of MT–SCCALR are fourfold. First, different to existing unsupervised and supervised SCCAs, MT–SCCALR can identify diagnosis-specific SNP–QT associations by jointly learning multiple SCCA tasks and LR tasks. The identified diagnosis-specific feature set, including SNPs and imaging QTs, is exclusively held by a specific diagnostic group. Second, better than JSCCA which incurs undesirable diagnosis-irrelevant features, MT–SCCALR follows sophisticated class-specific modeling strategy, and thus could avoid the diagnosis-irrelevant QTs, SNPs and their associations. Third, using regularization techniques, MT–SCCALR not only selects features such as SNPs and QTs for each diagnostic group, but also those that are commonly carried by all subjects, enabling a hierarchical strategy for feature selection. Fourth, an efficient iteration optimization algorithm is derived, which is demonstrated to converge to a local optimum.

To evaluate the performance of MT–SCCALR, we use four synthetic datasets with distinct characters and a real neuroimaging genetic dataset downloaded from the Alzheimer's disease neuroimaging initiative (ADNI) database (Mueller et al., 2005). There are 755 non-Hispanic Caucasian participants with their 18-Fr florbetapir PET scans and genotyping data contained. We aim to detect the diagnosis-specific associations between these imaging QTs and SNPs. Compared with two state-of-the-art methods (Fang et al., 2016; Yan et al., 2017), the experimental results show that MT–SCCALR performs better than or similarly to benchmarks in terms of correlation coefficients and classification accuracies. Interestingly, the canonical weights reveal that our method successfully identifies QTs, SNPs and their associations being specific for each diagnostic group while those competitors cannot. In a word, the proposed integrated multitask SCCA and multitask LR offers a very promising new strategy for brain imaging genetics.

2 Materials and methods

In this article, lowercase letters denote vectors, and uppercase ones denote matrices. \mathbf{x}^i and \mathbf{x}_j denote the i th row and j th column of matrix $\mathbf{X} = (x_{ij})$. $\|\mathbf{x}\|_2$ denotes the Euclidean norm, $\|\mathbf{X}\|_{1,1} = \sum_i \sum_j |x_{ij}|$ denotes the element-wise ℓ_1 -norm, $\|\mathbf{X}\|_{2,1} = \sum_i \|\mathbf{x}^i\|_2$ denotes the $\ell_{2,1}$ -norm, and $\|\mathbf{X}\|_F = \sqrt{\sum_i \sum_j x_{ij}^2}$ denotes the Frobenius norm.

2.1 Overview

To identify diagnosis-specific features in imaging genetics, it is straightforward to train multiple classifiers independently with respect to different tasks. For example, we can build a classifier to discriminate HC from AD, and build another classifier to discriminate between MCI and AD. However, this strategy treats these tasks as independent and isolated models, hence the underlying interacting relationships among tasks might be overlooked. Another critical issue is that it is somewhat difficult to interpret the identified features as it is not a rigorous model for class-specific feature selection (Pineda-Bautista et al., 2011).

In this article, we formally define the diagnosis-specific feature selection model for imaging genetics following the strategy in traditional feature selection community (Pineda-Bautista et al., 2011; Zhang and Wu, 2015). Figure 1 presents the framework of the diagnosis-specific feature selection workflow. First, the whole target population is divided into different groups such as healthy control (HC), mild cognitive impairment (MCI) and AD. If possible, the AD patients can be further split into AD subgroups, e.g. the typical, limbic-predominant, or hippocampal-sparing groups (Ferreira et al., 2017). Second, the class binarization is applied to construct multiple classification tasks via the one-versus-all [OVA, or one-against-all (OAA)] decomposition strategy. The OVA has shown, in general, good performance for all datasets (Lorena et al., 2008). In particular, each task in our method is to classify a specific diagnostic group out of those subjects not in this group, e.g. HC versus non-HC or MCI versus non-MCI. This is of great importance and meaning as doing this is helpful to understand the in-depth and unique

characters for the diagnostic group of interest, making personalized clinical diagnosis and treatment possible. Besides, if necessary, the class balancing techniques such as oversampling will be used to overcome the class imbalance issue (Pineda-Bautista *et al.*, 2011). Third, the novel heterogeneous multitask method, i.e. the joint multitask SCCA and multitask LR, is proposed to simultaneously and systematically considering the different diagnostic groups' relatedness. The SCCA objective is used to learn the association between QTs and SNPs, and the LR objective is used to learn discriminating features. In addition, the class-specific as well as class-consistent characteristics of both imaging QTs and SNPs are also considered by newly designed penalties. Finally, we obtain the dementia-related, including both MCI and AD, and the normal ageing-related imaging QTs and SNPs. These diagnosis-specific feature sets would be of more interest than a single set for all diagnosis groups.

The class binarization and class balancing are easy to implement (Pineda-Bautista *et al.*, 2011). Thus we focus on discussing the MT–SCCALR algorithm which is also the major contribution of this study.

2.2 The diagnosis-specific feature selection model for imaging genetics

MT–SCCALR is a heterogeneous multitask method with class-specific feature selection which fuses the bi-multivariate (SCCA) associations identification and classification (multiclass LR). Suppose we are given n participants with p SNPs and q imaging QTs from C diagnostic groups, we then use $\mathbf{X}_c \in \mathbb{R}^{n \times p}$ to load the genetic data, $\mathbf{Y}_c \in \mathbb{R}^{n \times q}$ to collect the imaging QT data and $\mathbf{z}^l \in \mathbb{R}^{1 \times C}$ to collect the diagnostic label vector of subject l . Generally, in \mathbf{z}^l , only one element is one and others are zeros. Then we directly learn two canonical weight matrices as $\mathbf{U} = \begin{bmatrix} u_{11} & \cdots & u_{1C} \\ \cdots & \cdots & \cdots \\ u_{p1} & \cdots & u_{pC} \end{bmatrix} \in \mathbb{R}^{p \times C}$, and

$\mathbf{V} = \begin{bmatrix} v_{11} & \cdots & v_{1C} \\ \cdots & \cdots & \cdots \\ v_{q1} & \cdots & v_{qC} \end{bmatrix} \in \mathbb{R}^{q \times C}$, where u_{ic} is the weight of the i th

SNP for the c th task, and v_{jc} is the weight of the j th imaging QT for the c th task. Obviously, these two canonical weight matrices can capture the local feature importance for both SNPs and imaging QTs with respect to each class, indicating their class-specific influence. In contrast, the traditional SCCA learns a canonical weight vector for SNPs and QTs, resulting in inability to select diagnosis-specific features. To identify the class-specific features and account for their interrelationships, we fuse the LR objective and the SCCA objective, i.e.

$$\min_{\mathbf{U}, \mathbf{V}} \mathcal{L}_{\text{LR}}(\mathbf{V}) + \mathcal{L}_{\text{SCCA}}(\mathbf{U}, \mathbf{V}). \quad (1)$$

In this model, \mathcal{L}_{LR} identifies the discriminating imaging QTs by conducting multitask classification for C tasks. After that, $\mathcal{L}_{\text{SCCA}}$ jointly learns the bi-multivariate associations between imaging QTs and SNPs for multiple tasks. It is worth mentioning that we do not include the logistic term between SNPs and the class label as associations between QTs and SNPs will finally encourage the identified SNPs being discriminating. This is reasonable and will make our model concise.

The model above encounters severe overfitting problem as generally the number of SNPs or imaging QTs is much larger than the

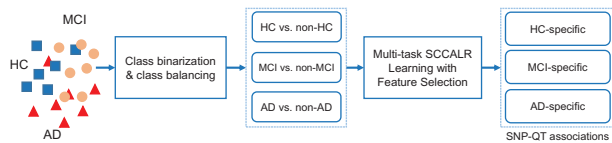


Fig. 1. Framework of diagnosis-specific imaging genetic pattern identification with three diagnostic groups: HC, MCI and AD. There certainly can be more than three diagnostic groups

sample size. Therefore, the sparsity regularization technique is utilized. On this account, our MT–SCCALR model becomes

$$\min_{\mathbf{U}, \mathbf{V}} \mathcal{L}_{\text{LR}}(\mathbf{V}) + \mathcal{L}_{\text{SCCA}}(\mathbf{U}, \mathbf{V}) + \Omega(\mathbf{U}) + \Omega(\mathbf{V}). \quad (2)$$

The $\Omega(\mathbf{U})$ is the sparsity-inducing penalty to identify those SNPs of interest, and $\Omega(\mathbf{V})$ is to identify relevant imaging QTs. In this class-specific multitask model, $\Omega(\mathbf{U})$ and $\Omega(\mathbf{V})$ are designed to incorporate three types of regularization methods for feature selection, i.e. the class-consistent sparsity, the class-specific sparsity in terms of selecting features jointly and individually for SNPs and imaging QTs.

To sum up, the multitask SCCA term captures the SNP–QT associations. The LR term captures the QT–diagnosis relationship. The sparsity-inducing terms help select relevant QTs and SNPs holding by a specific diagnosis group while accounting for those shared by multiple groups. Therefore, this novel fusion model is endowed with a diverse feature selection, especially the interesting class-specific feature selection. Next, we will present each term of MT–SCCALR in details.

2.3 The OVA multiclass classification via the LR

The LR is a popular classification method due to its simplicity but efficiency. In the proposed model, we regress the class label on imaging QTs by the LR objective to learn their associations. In the multiclass setting, we train multiple binary classifiers by the multitask modeling, i.e.

$$\mathcal{L}_{\text{LR}}(\mathbf{V}) = \sum_{c=1}^C \frac{1}{n_c} \sum_{l=1}^{n_c} [\log(1 + e^{\mathbf{y}_c^l \mathbf{v}_c}) - z_{lc} \mathbf{y}_c^l \mathbf{v}_c], \quad (3)$$

where n_c is the sample size for each classification task. n_c 's could be equal without class balancing or unequal after class balancing. z_{lc} is the corresponding class label of the l th subject for the c th task, and \mathbf{y}_c^l is the data vector of the l th subject for the c th task. This objective is usually called the negative log-likelihood and is convex (Zaidi and Webb, 2017).

2.4 The bi-multivariate association identification via the multitask SCCA

Conventional SCCA cannot identify class-specific SNP–QT associations as it only learns a single feature subset for all classes. The multitask SCCA (Du *et al.*, 2019b) systematically considers the relatedness among multiple SCCA tasks and thus can be applied to handle class-specific SNP–QT associations identification. Denoting data matrices with respect to the c th SCCA task as \mathbf{X}_c and \mathbf{Y}_c , the multitask SCCA is defined as

$$\min_{\mathbf{u}_c, \mathbf{v}_c} \sum_{c=1}^C -\mathbf{u}_c^T \mathbf{X}_c^T \mathbf{Y}_c \mathbf{v}_c \quad (4)$$

s.t. $\|\mathbf{X}_c \mathbf{u}_c\|_2^2 = 1, \|\mathbf{Y}_c \mathbf{v}_c\|_2^2 = 1, \forall c.$

According to (Du *et al.*, 2019b), this equation can be equivalently rewritten as

$$\min_{\mathbf{u}_c, \mathbf{v}_c} \sum_{c=1}^C \|\mathbf{X}_c \mathbf{u}_c - \mathbf{Y}_c \mathbf{v}_c\|_2^2 \quad (5)$$

s.t. $\|\mathbf{X}_c \mathbf{u}_c\|_2^2 = 1, \|\mathbf{Y}_c \mathbf{v}_c\|_2^2 = 1, \forall c.$

based on $\forall c, \|\mathbf{X}_c \mathbf{u}_c\|_2^2 = 1$ and $\|\mathbf{Y}_c \mathbf{v}_c\|_2^2 = 1$.

This objective jointly learns bi-multivariate associations and thus, in general, outperforms conventional SCCAs (Du *et al.*, 2019b). It is worth noting that in this model, \mathbf{X}_c and \mathbf{Y}_c correspond to the whole population but not only subjects in the c th class. For example, suppose the c th task is MCI versus non-MCI as shown in Figure 1, \mathbf{X}_c and \mathbf{Y}_c contain all subjects other than only MCI subjects. On the contrary, in JSCCA model, both \mathbf{X}_c and \mathbf{Y}_c come from only the MCI group (Fang *et al.*, 2016). As we analyzed earlier, features identified by our model possess stronger discriminating ability than that of JSCCA whose might be out of interest.

2.4.1 Regularization for imaging QTs via class-consistent and class-specific sparsity

The brain disorder such as AD usually exhibits the heterogeneity for multiple diagnostic groups (Lam et al., 2013; Wang et al., 2015). This diversity and complexity raise three critical questions. First, an imaging QT could present similar degenerative pattern among all diagnostic groups, because that both dementia brain and normal ageing brain suffer from functional and structural degeneration. Second, more commonly, an imaging QT probably exhibits diverse and different degenerative patterns across multiple diagnostic groups. The hippocampal-sparing AD patients show similar atrophy to normal aging in hippocampus (Murray and Dickson, 2008), while typical AD patients suffer from pronouncedly severer hippocampus atrophy compared to normal ageing subjects. The last but not the least, the network or graph structure has been clearly observed by autopsy or noninvasive imaging techniques (Bullmore and Sporns, 2009). Thus, a damage to the network structure might be a sign of dementia. On the contrary, this implies that an intact network might only exist in normal ageing subjects but not AD patients.

Therefore, to consider the complexity of brain disorders, we defined $\Omega(\mathbf{V})$ as follows

$$\Omega(\mathbf{V}) = \lambda_{v1} \|\mathbf{V}\|_{2,1} + \lambda_{v2} \|\mathbf{V}\|_{1,1} + \lambda_{v3} \sum_{c=1}^C \|\mathbf{v}_c\|_{GGL}, \quad (6)$$

where λ_{v1} , λ_{v2} and λ_{v3} are nonnegative parameters and can be obtained by cross-validation or holdout.

The first term is the $\ell_{2,1}$ -norm which is defined as follows

$$\|\mathbf{V}\|_{2,1} = \sum_{j=1}^q \|\mathbf{v}^j\|_2 = \sum_{j=1}^q \sqrt{\sum_{c=1}^C v_{jc}^2}. \quad (7)$$

This penalty encourages the task-consistent sparsity, indicating that elements of vector \mathbf{v}^j ($j = 1, \dots, q$) will be zeros or nonzeros simultaneously. As a result, an imaging QT presenting similar degenerative pattern for all groups will be selected or discarded jointly. Using this penalty is more practical since by finding out QTs shared among multiple diagnostic groups, the identified class-specific QTs could be more helpful.

The second regularizer is the $\ell_{1,1}$ -norm (ℓ_1 -norm for matrices) which is defined as

$$\|\mathbf{V}\|_{1,1} = \sum_{j=1}^q \|\mathbf{v}^j\|_1 = \sum_{j=1}^q \sum_{c=1}^C |v_{jc}| \quad (8)$$

This $\ell_{1,1}$ -norm penalty first prompts the individual sparsity for an imaging QT across all classes, and then prompts the sparsity for all imaging QTs. This is important and meaningful as it cannot only select relevant imaging QTs, but also determinate whether an imaging QT is relevant for a specific class. Similar to ℓ_1 -norm, although $\ell_{1,1}$ -norm is nonsmooth, it is convex and thus is easy to optimize.

The third term is the graph-guided pairwise group Lasso (GGL) (Du et al., 2017, 2020) whose definition is

$$\|\mathbf{v}_c\|_{GGL} = \sum_{(j,k) \in E} \sqrt{v_{jc}^2 + v_{kc}^2}, \quad (9)$$

where E is the edge set of the graph in which those highly correlated nodes are connected. Given the network structure of the brain (knowledge-guided), or considering the brain as a completed graph (data-driven), this penalty captures the high-level structure information among imaging QTs. As shown in Du et al. (2017, 2020), it holds the capability to assign similar weights for highly correlated imaging QTs, and thus can help identify the network structure. Moreover, using this penalty within each diagnostic group, we could capture the network holding by a specific group alone. Finally, the GGL penalty is convex, indicating that it can be easily solved.

In summary, combining these three regularizers together can well address three questions raised at the beginning of this

subsection. On this account, plugging $\Omega(\mathbf{V})$ into MT-SCCALR will make it a more reasonable model, and thus yield class-consistent and class-specific feature subsets of interest.

2.4.2 Regularization for SNPs via class-consistent and class-specific sparsity

Once those relevant imaging QTs, including both class-consistent and class-specific ones, are correctly identified, they could guide us to identify those class-consistent and class-specific SNPs as well. It is well known that SNPs usually affect the brain structure and function at both group level [linkage disequilibrium (LD) structure (Reich et al., 2001)] and individual level. In particular, at the individual level, a SNP could affect normal ageing brain and dementia brain at the same time, while another SNP might only influence the dementia one. This requires both class-consistent and class-specific feature selection for an individual SNP. In addition, at the group level, SNPs within the same LD or gene might jointly affect the brain structure and function (Wang et al., 2012b). An important thing is that, the genetic variation might happen to patients but not HCs, resulting in that an LD structure could only exist in healthy subjects. For this reason, we define the $\Omega(\mathbf{U})$ as

$$\Omega(\mathbf{U}) = \lambda_{u1} \|\mathbf{U}\|_{2,1} + \lambda_{u2} \|\mathbf{U}\|_{1,1} + \lambda_{u3} \sum_{c=1}^C \|\mathbf{u}_c\|_{FGL}, \quad (10)$$

with λ_{u1} , λ_{u2} and λ_{u3} being nonnegative tuning parameters. Both $\ell_{2,1}$ -norm and $\ell_{1,1}$ -norm are the same to that in Eqs. (7–8). They encourage class-consistent and class-specific feature selection for a single SNP.

In addition, the third term in Eq. (10) is the fused pairwise group Lasso (FGL) imposed on each \mathbf{u}_c (Du et al., 2020), i.e.

$$\|\mathbf{u}_c\|_{FGL} = \sum_{i=1}^{p-1} \sqrt{u_{ic}^2 + u_{(i+1)c}^2}, \quad (11)$$

This penalty endows the model with a chain of smoothness across all elements of \mathbf{u}_c , encouraging the selection of two adjacent and strongly correlated variables (Du et al., 2020). This penalty shows clearly grouping effects and thus could be used to automatically identify group structures. This plays a key role in our model as it considers the feature selection at a higher level. In this article, we use this penalty within each diagnostic group via the data-driven setup with expectation to identify the group structure that only exists in a specific group.

2.5 The optimization and convergence

Now we explicitly write both objectives and penalties with respect to imaging phenotypes and genotypes,

$$\begin{aligned} \min_{\mathbf{U}, \mathbf{V}} \sum_{c=1}^C \frac{1}{n_c} \sum_{l=1}^{n_c} [\log(1 + e^{z_l \mathbf{v}_c^l}) - z_l c \mathbf{y}_c^l \mathbf{v}_c] + \sum_{c=1}^C \|\mathbf{X}_c \mathbf{u}_c - \mathbf{Y}_c \mathbf{v}_c\|_2^2 \\ + \lambda_{v1} \|\mathbf{V}\|_{2,1} + \lambda_{v2} \|\mathbf{V}\|_{1,1} + \lambda_{v3} \sum_{c=1}^C \|\mathbf{v}_c\|_{GGL} \\ + \lambda_{u1} \|\mathbf{U}\|_{2,1} + \lambda_{u2} \|\mathbf{U}\|_{1,1} + \lambda_{u3} \sum_{c=1}^C \|\mathbf{u}_c\|_{FGL} \\ \text{s.t. } \|\mathbf{X}_c \mathbf{u}_c\|_2^2 = 1, \|\mathbf{Y}_c \mathbf{v}_c\|_2^2 = 1, \forall c. \end{aligned} \quad (12)$$

Mathematically, Eq. (12) is neither convex nor smooth. To handle this issue, by replacing $\|\mathbf{X}_c \mathbf{u}_c\|_2^2 = 1$ with $\|\mathbf{X}_c \mathbf{u}_c\|_2^2 \leq 1$ and $\|\mathbf{Y}_c \mathbf{v}_c\|_2^2 = 1$ with $\|\mathbf{Y}_c \mathbf{v}_c\|_2^2 \leq 1$, we analyze that Eq. (12) is convex if we consider \mathbf{V} as a constant, and vice versa. Fortunately, this biconvexity has been well studied previously (Gorski et al., 2007), based on which we can solve \mathbf{U} and \mathbf{V} alternatively after smoothing those nonsmooth penalties such as $\Omega(\mathbf{U})$ and $\Omega(\mathbf{V})$.

2.5.1 The solution to V

First, we fix \mathbf{U} to solve \mathbf{V} . The Lagrangian of Eq. (12) with respect to \mathbf{V} can be simplified as

$$\sum_{c=1}^C \frac{1}{n_c} \sum_{l=1}^{n_c} [\log(1 + e^{y_{lc}^l v_c}) - z_{lc} y_{lc}^l v_c] + \sum_{c=1}^C \|\mathbf{X}_c \mathbf{u}_c - \mathbf{Y}_c \mathbf{v}_c\|_2^2 + \lambda_{v1} \|\mathbf{V}\|_{2,1} + \lambda_{v2} \|\mathbf{V}\|_{1,1} + \lambda_{v3} \sum_{c=1}^C \|\mathbf{v}_c\|_{GGL} + \gamma_v \sum_{c=1}^C \|\mathbf{Y}_c \mathbf{v}_c\|_2^2 \quad (13)$$

by discarding those constants.

Using the subgradient of these penalties, we can obtain the derivative of Eq. (13) with respect to each \mathbf{v}_c . Letting it be zero yields [When $\|\mathbf{v}^j\|_2 = 0$, we regularize the j th diagonal element of $\tilde{\mathbf{D}}_1$ as $\frac{1}{2\|\mathbf{v}^j + \xi\|_2}$, where ξ is a very small positive value. $\tilde{\mathbf{D}}_1$ and \mathbf{D}_1 can be regularized similarly. It is easy to prove that when $\xi \rightarrow 0$, the regularized problem is equivalent to problem Eq. (13).]

$$\frac{\partial \mathcal{L}_{LR}(\mathbf{V})}{\partial \mathbf{v}_c} - 2\mathbf{Y}_c^T \mathbf{X}_c \mathbf{u}_c + 2\lambda_{v1} \tilde{\mathbf{D}}_1 \mathbf{v}_c + 2\lambda_{v2} \tilde{\mathbf{D}}_1 \mathbf{v}_c + 2\lambda_{v3} \mathbf{D}_1 \mathbf{v}_c + 2(\gamma_v + 1) \mathbf{Y}_c^T \mathbf{Y}_c \mathbf{v}_c = 0, \quad (14)$$

where $\tilde{\mathbf{D}}_1$ is a diagonal matrix with the j th element being $\frac{1}{2\|\mathbf{v}^j\|_2}$ ($j = 1, \dots, q$), and $\tilde{\mathbf{D}}_1$ is a diagonal matrix where the j th element is $\frac{1}{2\|\mathbf{v}_c^j\|_2}$ ($j = 1, \dots, q$; and $c = 1, \dots, C$). \mathbf{D}_1 is also a diagonal matrix with its j th element being $\sum_{i,(j,k) \in E} \frac{1}{2\sqrt{v_{jc}^2 + v_{kc}^2}}$ ($j = 1, \dots, q$; and $c = 1, \dots, C$) (Du et al., 2017). Obviously, $\tilde{\mathbf{D}}_1$, $\tilde{\mathbf{D}}_1$ and \mathbf{D}_1 depend on the independent variable \mathbf{V} , and thus they are unknown. On this account, the iterative algorithm can be a solver which first guesses an initial value of \mathbf{V} , and then calculates these diagonal matrices.

Once $\tilde{\mathbf{D}}_1$, $\tilde{\mathbf{D}}_1$ and \mathbf{D}_1 are available, we can solve Eq. (14) correspondingly. However, we cannot find a closed-form solution due to the nontrivial derivative of the logistic term. Thus we address this using the Newton's method which depends on the Hessian matrix. Based on this, solving Eq. (14) is equivalent to solve

$$\min_{\mathbf{V}} \mathcal{L}_{LR}(\mathbf{V}) + \sum_{c=1}^C \|\mathbf{X}_c \mathbf{u}_c - \mathbf{Y}_c \mathbf{v}_c\|_2^2 + \lambda_{v1} \text{Tr}(\mathbf{V}^T \tilde{\mathbf{D}}_1 \mathbf{V}) + \lambda_{v2} \text{Tr}(\mathbf{V}^T \tilde{\mathbf{D}}_1 \mathbf{V}) + \lambda_{v3} \text{Tr}(\mathbf{V}^T \mathbf{D}_1 \mathbf{V}) + \gamma_v \sum_{c=1}^C \mathbf{v}_c^T \mathbf{Y}_c^T \mathbf{Y}_c \mathbf{v}_c. \quad (15)$$

According to Krishnapuram et al. (2005) and Lee et al. (2006), we first obtain the first-order derivative of the logistic objective with respect to each v_{jc}

$$\frac{\partial \mathcal{L}_{LR}(\mathbf{V})}{\partial v_{jc}} = \frac{1}{n_c} \sum_{l=1}^{n_c} (P(z_{lc} = 1 | y_{lc}^l) - z_{lc}) y_{lj,c}, \quad (16)$$

where $P(z_{lc} = 1 | y_{lc}^l)$ is the class posterior probability (Zaidi and Webb, 2017). Here we use $y_{lj,c}$ denote the l th row and j th column element of matrix \mathbf{Y}_c .

Then we can calculate the second-order derivative based on Eq. (16),

$$\frac{\partial \mathcal{L}_{LR}(\mathbf{V})}{\partial v_{jc} \partial v_{kc}} = \frac{1}{n_c} \sum_{l=1}^{n_c} (1 - P(z_{lc} = 1 | y_{lc}^l)) P(z_{lc} = 1 | y_{lc}^l) y_{lj,c} y_{lk,c}. \quad (17)$$

After both the first-order and second-order derivatives have been obtained, we can easily calculate the gradient (or subgradient) vector $\mathbf{g}(\mathbf{v}_c)$ and the Hessian matrix $\mathbf{H}(\mathbf{v}_c)$ of Eq. (15) regarding \mathbf{v}_c . Therefore, the solution to each \mathbf{v}_c can be finally attained via

$$\mathbf{v}_c = \mathbf{v}_c - \mathbf{H}^{-1}(\mathbf{v}_c) \mathbf{g}(\mathbf{v}_c). \quad (18)$$

2.5.2 The solution to U

Given \mathbf{V} , the solution to \mathbf{U} can be attained as well. First, we write the Lagrangian of Eq. (12) with \mathbf{U} being unknown variable,

$$\sum_{c=1}^C \|\mathbf{X}_c \mathbf{u}_c - \mathbf{Y}_c \mathbf{v}_c\|_2^2 + \lambda_{u1} \|\mathbf{U}\|_{2,1} + \lambda_{u2} \|\mathbf{U}\|_{1,1} + \lambda_{u3} \sum_{c=1}^C \|\mathbf{u}_c\|_{FGL} + \gamma_u \sum_{c=1}^C \|\mathbf{X}_c \mathbf{u}_c\|_2^2 \quad (19)$$

where those constants are discarded. If we treat $\mathbb{Y} = [\mathbf{Y}_1 \mathbf{v}_1, \dots, \mathbf{Y}_C \mathbf{v}_C]$ as dependent variables and \mathbf{X}_c 's as independent variables, this equation becomes a multiregression task learning method with class-consistent and class-specific feature selection. To solve this multitask problem, we take its derivative with respect to \mathbf{u}_c and set it to zero, i.e.

$$-\mathbf{X}_c^T \mathbf{Y}_c \mathbf{v}_c + \lambda_{u1} \tilde{\mathbf{D}}_2 \mathbf{u}_c + \lambda_{u2} \tilde{\mathbf{D}}_2 \mathbf{u}_c + \lambda_{u3} \mathbf{D}_2 \mathbf{u}_c + (\gamma_u + 1) \mathbf{X}_c^T \mathbf{X}_c \mathbf{u}_c = 0. \quad (20)$$

Similarly to solving \mathbf{V} , \mathbf{D}_2 here is a diagonal matrix whose i th element is $\frac{1}{2\|\mathbf{u}^i\|_2}$ ($i = 1 \dots, p$), and $\tilde{\mathbf{D}}_2$ is a diagonal matrix with the i th element being $\frac{1}{2\|\mathbf{u}_c^i\|_2}$ ($i = 1, \dots, p$; and $c = 1, \dots, C$). Finally, \mathbf{D}_2 is a diagonal matrix whose i th diagonal entry is $\frac{1}{2\sqrt{u_{(i-1)c}^2 + u_{ic}^2}} + \frac{1}{2\sqrt{u_{ic}^2 + u_{(i+1)c}^2}}$ ($i = 1, \dots, p$; and $c = 1, \dots, C$) [The first element of \mathbf{D}_2 is $\frac{1}{2\sqrt{u_{1c}^2 + u_{2c}^2}}$, and the p th element is $\frac{1}{2\sqrt{u_{(p-1)c}^2 + u_{pc}^2}}$. See Du et al. (2020) for details.].

Using the iteration algorithm, we can obtain the closed-form equation regarding each \mathbf{u}_c , i.e.

$$\mathbf{u}_c = (\lambda_{u1} \tilde{\mathbf{D}}_2 + \lambda_{u2} \tilde{\mathbf{D}}_2 + \lambda_{u3} \mathbf{D}_2 + (\gamma_u + 1) \mathbf{X}_c^T \mathbf{X}_c)^{-1} \mathbf{X}_c^T \mathbf{Y}_c \mathbf{v}_c. \quad (21)$$

Now we have the building blocks for optimizing Eq. (12). The final solution can be attained via iteratively and alternatively optimizing \mathbf{V} and \mathbf{U} . The pseudocode is presented in Algorithm 1 which is guaranteed to converge to a local optimum. In this algorithm, Steps 1 and 5 are easily and very fast to calculate. Step 3 involves computing the Hessian matrix which can be obtained efficiently based on (Lin et al., 2007). At last, Steps 4 and 6 can be attained by efficiently solving a system of linear equations (Wang et al., 2012a,b).

2.5.3 Convergence analysis

We have the following theorem for the MT–SCCALR Algorithm.

THEOREM 1. The Algorithm 1 decreases the objective in each iteration.

Algorithm 1 The MT–SCCALR algorithm

Require:

The genotype data $\mathbf{X} \in \mathbb{R}^{n \times p}$ and imaging phenotype data $\mathbf{Y} \in \mathbb{R}^{n \times q}$ of C diagnostic groups. The pre-tuned λ_{v1} , λ_{v2} , λ_{v3} , γ_v , λ_{u1} , λ_{u2} , λ_{u3} , and γ_u .

Ensure:

Canonical weights \mathbf{V} and \mathbf{U} .

- 1: Class binarization and class balancing, initialize $\mathbf{U} \in \mathbb{R}^{p \times C}$ and $\mathbf{V} \in \mathbb{R}^{q \times C}$;
- 2: **while** not convergence **do**
- 3: Calculate the first-order and second-order derivatives of Eq. (15);
- 4: Solve \mathbf{v}_c according to Eq. (18), and scale \mathbf{v}_c so that $\|\mathbf{Y}_c \mathbf{v}_c\|_2^2 = 1$;
- 5: Update $\tilde{\mathbf{D}}_2$, \mathbf{D}_2 and \mathbf{D}_2 ;
- 6: Solve \mathbf{u}_c according to Eq. (21), and scale \mathbf{u}_c as that $\|\mathbf{X}_c \mathbf{u}_c\|_2^2 = 1$;
- 7: **end while**
- 8: Sorting each \mathbf{u}_c and \mathbf{v}_c in descending order based on their absolute value respectively.

The proof is contained in [Supplementary Material](#) due to space limitation. We have known that MT-SCCALR is biconvex, and obviously, [Eq. \(12\)](#) has the lower bound of zero, hence a local optimum can be attained finally by running Algorithm 1. To ensure efficiency, we stop our algorithm when both $\max_i \max_c |\hat{U} - U| \leq \epsilon$ and $\max_i \max_c |\hat{V} - V| \leq \epsilon$ hold. In addition, we empirically set the tolerance error $\epsilon = 10^{-5}$ in this article for all experiments, and certainly, ϵ can be obtained based on experiments.

3 Experimental results and discussions

3.1 Experimental design

We compared our method with two most related SCCA method, i.e. the discriminate SCCA (DSCCA) ([Yan et al., 2017](#)) and the JSCCA ([Fang et al., 2016](#)), which directly identify imaging QTs and SNPs with discriminating ability. DSCCA extended the traditional SCCA by the locality preserving projection penalty. Its identified imaging QTs and proteomic markers can successfully discriminate between every two diagnostic groups such as HC versus AD. JSCCA was a type of mSCCA ([Witten and Tibshirani, 2009](#)) which learns the association between QTs and SNPs within the same diagnostic group. Therefore, using DSCCA and JSCCA as benchmarks, our approach was compared with the state-of-the-art discriminating SCCA method, assuring a practical and meaningful performance evaluation. Of note, our primary aim in this study was to identify meaningful diagnosis-specific features other than just classification. Thus we did not compare to methods which use one versus one classification or a nonobvious modeling strategy ([Wang et al., 2015](#)).

To find out suitable parameters, we used the fivefold cross-validation strategy to fine-tune them. Generally, we set parameters to those which generate the highest testing canonical correlation coefficients (CCC) and classification accuracy. There were in total eight parameters in the original model which was time intensive. To alleviate this issue for both MT-SCCALR and benchmarks, we fixed $\gamma_v = 1$ and $\gamma_u = 1$ because that they mainly affect the amplitude of U and V ([Chen and Liu, 2012](#)). In addition, we also used several heuristic rules to further reduce the time effort. Specifically, if we prefer the class-specific feature selection as in this study, we could use large parameters for $\ell_{1,1}$ -norm, FGL-norm and GGL-norm. On the contrary, we might sometimes desire for the class-consistent feature selection, and then we could set large parameters for the $\ell_{2,1}$ -norm. This could yield reasonable results as only focusing on the CCC might lead to undesirable features. Besides, we used a two-stage tuning procedure which first tuned parameters from 10^i ($i = -2, -1, 0, 1, 2$) with a large interval, and then further tuned them from a relative smaller interval $\Gamma \pm [0.1, 0.2, \dots, 1]$, where Γ was the optimal parameters obtained from the first stage. Usually, this two-stage parameter tuning could yield better performance, compared to blindly grid search, for both correlation coefficients and feature selection. In experiments, we only tuned parameters in the first loop where the first fold was used for testing and the remaining folds were used for training, and these tuned parameters were used for all experiments to generate final results. According to experiments, this setup will not affect the performance significantly and could reduce the time consumption significantly. All methods used the same experimental setup to assure a fair comparison.

3.2 Simulation study

We generated four simulation datasets based on different ground truths to make a thorough comparison. We assumed three groups of imaging data Y and genotype data X. The first two datasets ($n = 100$, $p = 120$ and $q = 150$) contained the same true signal but different noise levels. The third dataset ($n = 100$, $p = 120$ and $q = 150$) mainly simulated a task-specific situation while the fourth one ($n = 200$, $p = 400$ and $q = 300$) primarily simulated a task-consistent situation. Specifically, in the fourth dataset, there was a successive relationship among classes to simulate the relationship between HC and MCI, and that between MCI and AD. All four datasets are generated as follows. First, we created two sparse

matrices $U \in \mathcal{R}^{p \times C}$ and $V \in \mathcal{R}^{q \times C}$. Within both U and V, there are features shared by all tasks and specifically hold by a single task. We then generated a latent vector $z \in \mathcal{R}^{m \times 1}$, based on which we generated three pairs of $X_c \sim N(z; u_c, I_{p \times p})$ and $Y_c \sim N(z; v_c, I_{q \times q})$ to form three diagnostic groups. We showed the ground truth in [Figure 2](#) (top row).

As we mainly focus on the identified class-specific features, in [Figure 2](#), we first showed the heatmaps of canonical weights which indicate the importance of features. DSCCA identified one canonical weight vector for all classes, and then we stacked it for C times to make its heatmap available. JSCCA generated one canonical weight vector for u and C canonical weight vectors for V, and thus we only stacked u for C times. In this figure, the features identified by MT-SCCALR were consistent to the ground truth, while DSCCA cannot. JSCCA performed slightly better than DSCCA, and they both had little capability in identifying class-specific features. From Data 4, we observed that if there were no class-specific features, all three methods could find out these class-consistent features. These results demonstrated that MT-SCCALR had more diverse feature selection ability than DSCCA and JSCCA. In [Figure 3](#), we showed the testing CCCs and testing classification accuracies which were obtained by SVM based on LIBSVM (<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>) software package. The CCCs and classification performance showed no significant difference between MT-SCCALR and DSCCA, and they both performed better than JSCCA, especially on the CCCs. This demonstrated that MT-SCCALR had similar CCCs and classification performance to benchmarks, but identified much better class-specific features while those benchmarks cannot.

3.3 Real neuroimaging genetic study

The real brain imaging and genotyping data were obtained from the ADNI (adni.loni.usc.edu) database. The primary goal of the initiative is to test whether serial magnetic resonance imaging (MRI), or other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org.

There are 755 non-Hispanic Caucasian participants, including 182 HC, 292 MCI and 281 AD, whose baseline 18-Fr florbetapir PET scans were collected. We used the pipeline to preprocess these PET scans such as average, alignment, resample, smoothness and normalization to obtain the standardized uptake value ratio (SUVR) images ([Jagust et al., 2010](#)). To reduce the time consumption and boost the statistical power, we extracted the region of interest level amyloid measurements instead of the voxel level measurements. We finally generated 116 mean amyloid measurements spanning the whole brain according to the MarsBaR AAL atlas ([Tzourio-Mazoyer et al., 2002](#)), and used them as imaging QTs. Moreover, these imaging QTs were preadjusted to remove the effects of the baseline age, gender, handedness and years of education ([Table 1](#)).

The genotyping data were genotyped by the Human 610-Quad or OmniExpress Array platform (Illumina, Inc., San Diego, CA, USA), and preprocessed following standard quality control and imputation procedures. There were 1692 SNPs included which were collected from the neighbor of AD risk gene APOE according to the

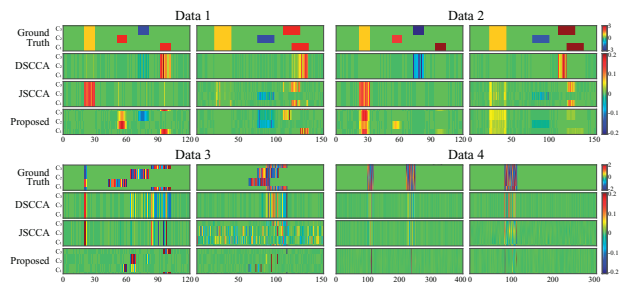


Fig. 2. Canonical weights on synthetic data. Row 1–4: Ground truth, DSCCA, JSCCA and MT-SCCALR respectively. For each data, canonical weights U is shown on the left, and V is shown on the right. In each panel, there are three rows (each row contains fivefold canonical weights) corresponding to three tasks

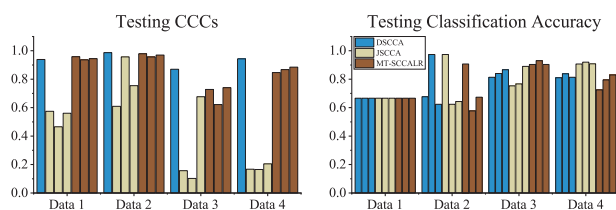


Fig. 3. Comparison of the mean CCCs and classification accuracy obtained from fivefold testing trials on synthetic data

Table 1. Participant characteristics

	HC	MCI	AD
Number	182	292	281
Gender (M/F, %)	48.90/51.10	48.63/51.37	53.38/46.62
Handedness (R/L, %)	89.56/10.44	88.70/11.30	90.39/9.61
Age (mean±SD)	73.93±5.51	70.90±6.84	72.61±8.15
Education (mean±SD)	16.43±2.68	16.18±2.68	15.95±2.82

ANNOVAR annotation. In this study, we intend to study bi-multivariate associations between regional imaging amyloid depositions and SNPs, and with focus on identifying diagnosis-specific amyloid depositions and SNPs, as well as their interactions.

3.3.1 Identification and interpretation of imaging QTs

We applied all three methods to this real neuroimaging genetic data. The selected imaging QTs are highlighted in Figure 4, in which there are three columns for both MT-SCCALR and JSCCA. We stacked the canonical weight vector of DSCCA for three times. We observed that DSCCA did not hold the diagnosis-specific feature selection capability due to the nature of its modeling method. JSCCA identified multiple weights corresponding to multiple diagnostic groups, showing a somewhat diverse feature selection. Our method performed the best among three methods as we could observe a much clear diagnosis-specific feature selection profiles. Besides, MT-SCCALR showed a clearly group feature selection patterns due to the structure identification penalties such as *FGL* and *GGL*.

We further investigated the meaning of the selected imaging QTs. MT-SCCALR identified three canonical weight vectors in accordance to three diagnostic groups, i.e. HC, MCI and AD. Most of the HC-specific signals were from the frontal areas, indicating that a subject's health condition could be determined based on the amyloid burden of these areas. Besides, the cingulum and the precuneus were also highlighted by MT-SCCALR. Interestingly, most of the AD-specific signals were from the frontal area too but with opposite signs, which matches our intuition. It is obvious that opposite weights of these areas for HCs and ADs tell us that there are an opposite amyloid deposition pattern for these two groups. In contrast, both DSCCA and JSCCA cannot draw this in-depth conclusion. For MCI-specific imaging QTs, our method reported signals from the bilateral calcarine, lingual, fusiform and temporal. It seems strange as the frontal areas were not identified, but the truth is not. In general, MCIs have intermediate amyloid burden and thus using the deposition measurement from the frontal might not discriminate MCI subjects from those non-MCI subjects. Moreover, these brain areas had been shown to be related to MCI (Pan *et al.*, 2017), confirming the detection power of our method. Of note, both DSCCA and JSCCA could not identify heterogeneous imaging QTs. The results together demonstrated that the diagnosis-specific feature selection is of great interest and meaning.

3.3.2 Identification and interpretation of SNPs

We presented the identified SNPs in Figure 5. In this figure, both DSCCA and JSCCA only identified one canonical weight vector for all tasks, and we stacked them for three times. We observed that all three methods identified rs429358 (*APOE*), i.e. the well-known

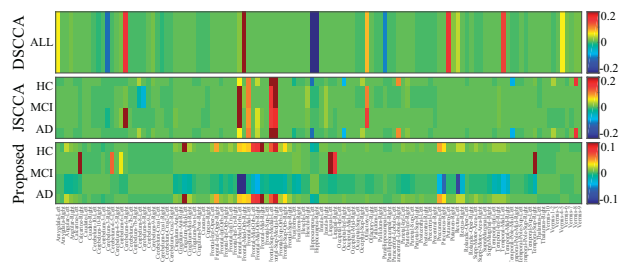


Fig. 4. Canonical weights (mean) of imaging QTs from fivefold cross-validation trials. Each row corresponds to an SCCA method: (1) DSCCA; (2) JSCCA and (3) MT-SCCALR

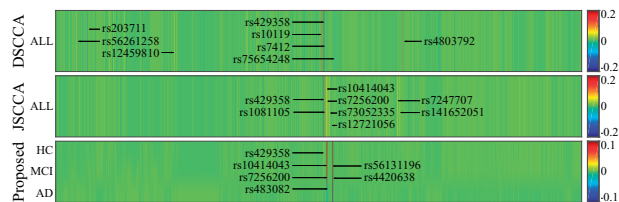


Fig. 5. Canonical weights (mean) of SNPs from fivefold cross-validation trials. Each row corresponds to an SCCA method: (1) DSCCA; (2) JSCCA and (3) MT-SCCALR

AD-risk loci. Taking the top five selected SNPs as examples, DSCCA identified two additional AD associated locus rs7412 (*APOE*) (Yi *et al.*, 2014) and rs10119 (*TOMM40*), with further investigation being warranted for rs4803792 and rs203711. JSCCA also identified SNPs from AD-associated genes such as *APOC1* (rs7247707, rs10414043 and rs7256200) and *APOE* (rs1081105), indicating its better identification than DSCCA. MT-SCCALR exhibited distinct patterns as it assigned different weight values (importance) for each SNP for different groups. At first glance, the SNPs for HC, MCI and AD were similar. However, similar to that for imaging QTs, our method yielded opposite feature signs for HC versus non-HC, and AD versus non-AD tasks. This is interesting as it reveals that HCs and ADs hold different genotypes which provide us a more in-depth clue. Compared to benchmarks which only tell us that a SNP is relevant or irrelevant, our method not only reveals whether a SNP is relevant, but also implies the directionality of the genetic effect. One may argue that we can swap both signs for imaging QTs and SNPs simultaneously, but the directionality of the imaging genetic correlation stays the same. Therefore, based on the heterogeneous multitasking, our method could successfully identify a diverse diagnosis-specific SNPs. These results suggest that MT-SCCALR is quite promising and might possess enhanced feature selection ability in imaging genetics.

3.3.3 Bi-multivariate association and classification

Finally, the testing CCC and classification accuracy are shown in Figure 6, in which a higher value indicated a stronger association or more accurate prediction. It is clear that DSCCA estimated the highest CCCs as it used all sample size to yield the CCC. Our method obtained better CCCs than JSCCA due to the OAA class binarization. Using the top ten selected features, including both imaging QTs and SNPs, of each method, we used LIBSVM (<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>) software package to implement SVM using the linear kernel with default setting. The classification results showed that all three methods performed similarly, implying that it is difficult to separate this real imaging genetic data (Wang *et al.*, 2012a). On the contrary, this confirms the necessity and meaning of the diagnosis-specific feature selection, which could be helpful to subgroups identification. As a result, following the multitask modeling, MT-SCCALR showed a promising performance in multiclass imaging genetics.

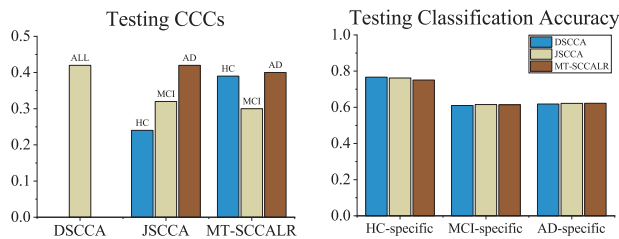


Fig. 6. Comparison of the mean CCCs and classification accuracy obtained from fivefold testing trials on ADNI

4 Conclusion

Identifying diagnosis-specific genetic markers and brain imaging measurements is an important task in precision medicine. AD is a severe neurodegenerative disorder presenting significant heterogeneity and diversity (Lam et al., 2013; Wang et al., 2015). Most existing bi-multivariate learning methods were unsupervised. Although a few studies made progress on supervised bi-multivariate learning, they could not identify diagnosis-specific biomarkers. We formally defined the framework of diagnosis-specific feature selection for imaging genetics, and proposed a computational method to identify diagnosis-specific, as well as diagnosis-consistent, feature subsets. Different to existing supervised SCCA, MT-SCCALR learned canonical weight matrices with each column corresponding to a diagnostic group. Using fused LR and SCCA, as well as the regularization, our model could identify meaningful and distinct imaging QTs and SNPs for each group. The algorithm was proved to converge to a local optimum.

Experiments on both synthetic data and real neuroimaging genetic data were conducted. Compared with two state-of-the-art methods [DSCCA (Yan et al., 2017) and JSCCA (Fang et al., 2016)], MT-SCCALR obtained similar correlation coefficients and classification accuracies to DSCCA and JSCCA. But it outperformed both benchmarks on revealing canonical weights on both synthetic and real data. In particular, our method successfully identified diagnosis-specific features including QTs and SNPs, while those benchmarks cannot. We also demonstrated that within each diagnostic group, the identified imaging QTs and SNPs were inconsistent, indicating that different diagnostic groups could carry different feature subsets. This is more reasonable and practical than existing methods as mounting evidences suggest that the sporadic AD might not be a single disease (Au et al., 2015). These results reveal that MT-SCCALR gains a promising success in diverse feature selection for multiple diagnostic groups in imaging genetics as well as multiple omics analysis. However, as the calculation of Hessian matrix and gradient of GGL penalty is time-intensive, we suggest using brain region-based imaging measurements other than voxel-based ones. An interesting future direction could be to make our method able to stratify different AD patients instead of only identifying diagnosis-specific features.

Acknowledgements

Data collection and sharing for this project was funded by the ADNI (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites

in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Funding

This work was supported by the National Natural Science Foundation of China [61973255, 61602384]; Natural Science Basic Research Program of Shaanxi [2020JM-142]; China Postdoctoral Science Foundation [2017M613202] and Postdoctoral Science Foundation of Shaanxi [2017BSHEDZZ81] at Northwestern Polytechnical University. This work was also supported by the National Institutes of Health [R01 EB022574, RF1 AG063481, U19 AG024904, P30 AG10133, R01 AG19771] at University of Pennsylvania and Indiana University.

Conflict of Interest: none declared.

References

- Alzheimer's Association (2013) 2013 Alzheimer's disease facts and figures. *Alzheimers Dement.*, 9, 208–245.
- Au, R. et al. (2015) Back to the future: Alzheimer's disease heterogeneity revisited. *Alzheimer's & Dementia: diagnosis, Assess. Dis. Monit.*, 1, 368–370.
- Baggenstoss, P.M. (1999) Class-specific feature sets in classification. *IEEE Trans. Signal Process.*, 47, 3428–3432.
- Beaton, D. et al. (2014) Imaging genetics with partial least squares for mixed-data types (MiMoPLS). In: *International Conference on Partial Least Squares and Related Methods*, pp. 73–91. Springer, Paris, France.
- Bullmore, E.T. and Sporns, O. (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.*, 10, 186–198.
- Chen, J. et al. (2013) Structure-constrained sparse canonical correlation analysis with an application to microbiome data analysis. *Biostatistics*, 14, 244–258.
- Chen, X. and Liu, H. (2012) An efficient optimization algorithm for structured sparse CCA, with applications to eQTL mapping. *Stat. Biosci.*, 4, 3–26.
- Du, L. et al.; for the Alzheimer's Disease Neuroimaging Initiative. (2016) Structured sparse canonical correlation analysis for brain imaging genetics: an improved GraphNet method. *Bioinformatics*, 32, 1544–1551.
- Du, L. et al. (2017) Identifying associations between brain imaging phenotypes and genetic factors via a novel structured SCCA approach. In: *International Conference on Information Processing in Medical Imaging*, pp. 543–555. Springer, Boone, North Carolina, USA.
- Du, L. et al.; for the Alzheimer's Disease Neuroimaging Initiative. (2018) A novel SCCA approach via truncated ℓ_1 -norm and truncated group lasso for brain imaging genetics. *Bioinformatics*, 34, 278–285.
- Du, L. et al.; Alzheimer's Disease Neuroimaging Initiative. (2019a) Identifying progressive imaging genetic patterns via multi-task sparse canonical correlation analysis: a longitudinal study of the ADNI cohort. *Bioinformatics*, 35, i474–i483.
- Du, L. et al. (2019b) Multi-task sparse canonical correlation analysis with application to multi-modal brain imaging genetics. *IEEE ACM Trans. Comput. Biol. Bioinf.*, doi: 10.1109/TCBB.2019.2947428.
- Du, L. et al. (2020) Detecting genetic associations with brain imaging phenotypes in Alzheimer's disease via a novel structured SCCA approach. *Med. Image Anal.*, 61, 101656.
- Fang, J. et al. (2016) Joint sparse canonical correlation analysis for detecting differential imaging genetics modules. *Bioinformatics*, 32, 3480–3488.
- Ferreira, D. et al. (2017) Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications. *Sci. Rep.*, 7, 46263.
- Gorski, J. et al. (2007) Biconvex sets and optimization with biconvex functions: a survey and extensions. *Math. Method Oper. Res.*, 66, 373–407.
- Jagust, W.J. et al. (2010) The Alzheimer's disease neuroimaging initiative positron emission tomography core. *Alzheimers Dement.*, 6, 221–229.
- Krishnapuram, B. et al. (2005) Sparse multinomial logistic regression: fast algorithms and generalization bounds. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27, 957–968.
- Lam, B. et al. (2013) Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. *Alzheimer's Res. Therapy*, 5, 1–14.
- Lee, S.-I. et al. (2006) Efficient L1 regularized logistic regression. *AAAI*, 6, 401–408.

- Lin, C.-J. *et al.* (2007) Trust region newton methods for large-scale logistic regression. In: *International Conference on Machine Learning*, pp. 561–568. ACM, Corvallis, Oregon, USA.
- Lorena, A.C. *et al.* (2008) A review on the combination of binary classifiers in multiclass problems. *Artif. Intell. Rev.*, **30**, 19–37.
- Mueller, S.G. *et al.* (2005) The Alzheimer’s disease neuroimaging initiative. *NeuroImage Clin. N. Am.*, **15**, 869–877.
- Mukherjee, S. *et al.* (2018) Genetic data and cognitively defined late-onset Alzheimer’s disease subgroups. *Mol. Psychiatr.* doi: 10.1038/s41380-018-0298-8.
- Murray, M.E. and Dickson, D. (2008) O1-01-02: Alzheimer’s disease with relative hippocampal sparing: a distinct clinicopathologic variant. *Alzheimers Dement.*, **4**, T106.
- Pan, P. *et al.* (2017) Aberrant spontaneous low-frequency brain activity in amnesic mild cognitive impairment: a meta-analysis of resting-state fMRI studies. *Ageing Res. Rev.*, **35**, 12–21.
- Pineda-Bautista, B.B. *et al.* (2011) General framework for class-specific feature selection. *Expert Syst. Appl.*, **38**, 10018–10024.
- Reich, D.E. *et al.* (2001) Linkage disequilibrium in the human genome. *Nature*, **411**, 199–204.
- Saykin, A.J. *et al.*; Alzheimer’s Disease Neuroimaging Initiative. (2015) Genetic studies of quantitative MCI and AD phenotypes in ADNI: progress, opportunities, and plans. *Alzheimers Dement.*, **11**, 792–814.
- Shen, L. and Thompson, P.M. (2020) Brain imaging genomics: integrated analysis and machine learning. *Proc. IEEE*, **108**, 125–162.
- Shen, 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.
- Shen, L. *et al.*; Alzheimer’s Disease Neuroimaging Initiative. (2014) Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain Imaging Behav.*, **8**, 183–207.
- Tzourio-Mazoyer, N. *et al.* (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, **15**, 273–289.
- Vounou, M. *et al.* (2010) Discovering genetic associations with high-dimensional neuroimaging phenotypes: a sparse reduced-rank regression approach. *NeuroImage*, **53**, 1147–1159.
- Wang, H. *et al.*; for the Alzheimer’s Disease Neuroimaging Initiative. (2012a) Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. *Bioinformatics*, **28**, i127–i136.
- Wang, H. *et al.* (2012b) Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. *Bioinformatics*, **28**, 229–237.
- Wang, X. *et al.* (2015) Classification of MRI under the presence of disease heterogeneity using multi-task learning: application to bipolar disorder. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 125–132. Springer, Munich, Germany.
- Wang, L. *et al.* (2016) Feature selection methods for big data bioinformatics: a survey from the search perspective. *Methods*, **111**, 21–31.
- Witten, D.M. and Tibshirani, R.J. (2009) Extensions of sparse canonical correlation analysis with applications to genomic data. *Stat. Appl. Genet. Mol.*, **8**, 1–27.
- Yan, J. *et al.* (2017) Identification of discriminative imaging proteomics associations in Alzheimer’s disease via a novel sparse correlation model. In: *Pacific Symposium on Biocomputing 2017*, pp. 94–104.
- Yan, J. *et al.* (2018) Joint exploration and mining of memory-relevant brain anatomic and connectomic patterns via a three-way association model. In: *International Symposium on Biomedical Imaging*, pp. 6–9.
- Yi, L. *et al.* (2014) A non-invasive, rapid method to genotype late-onset Alzheimer’s disease-related apolipoprotein E gene polymorphisms. *Neural Regen. Res.*, **9**, 69.
- Zaidi, N.A. and Webb, G.I. (2017) A fast trust-region newton method for softmax logistic regression. In: *SIAM International Conference on Data Mining*, pp. 705–713. SIAM, Houston, Texas, USA.
- Zhang, M.-L. and Wu, L. (2015) Lift: multi-label learning with label-specific features. *IEEE Trans. Pattern Anal. Mach. Intell.*, **37**, 107–120.
- Zille, P. *et al.* (2018) Enforcing co-expression within a brain-imaging genomics regression framework. *IEEE Trans. Med. Imaging*, **37**, 2561–2571.