Multi-Task Sparse Canonical Correlation Analysis with Application to Multi-Modal Brain Imaging Genetics

Lei Du*, Member, IEEE, Kefei Liu, Xiaohui Yao, Shannon L. Risacher, Junwei Han, Senior Member, IEEE, Andrew J. Saykin, Lei Guo, Li Shen*, Senior Member, IEEE, for the Alzheimer’s Disease Neuroimaging Initiative †

Abstract—Brain imaging genetics studies the genetic basis of brain structures and functionalities via integrating genotypic data such as single nucleotide polymorphisms (SNPs) and imaging quantitative traits (QTs). In this area, both multi-task learning (MTL) and sparse canonical correlation analysis (SCCA) methods are widely used since they are superior to those independent and pairwise univariate analysis. MTL methods generally incorporate a few of QTs and could not select features from multiple QTs; while SCCA methods typically employ one modality of QTs to study its association with SNPs. Both MTL and SCCA are computational expensive as the number of SNPs increases. In this paper, we propose a novel multi-task SCCA (MTSCCA) method to identify bi-multivariate associations between SNPs and multi-modal imaging QTs. MTSCCA could make use of the complementary information carried by different imaging modalities. MTSCCA enforces sparsity at the group level via the $C_{2,1}$-norm, and jointly selects features across multiple tasks for SNPs and QTs via the $\ell_{2,1}$-norm. A fast optimization algorithm is proposed using the grouping information of SNPs. Compared with conventional SCCA methods, MTSCCA obtains better correlation coefficients and canonical weights patterns. In addition, MTSCCA runs very fast and easy-to-implement, indicating its potential power in genome-wide brain-wide imaging genetics.

Index Terms—Brain Imaging Genetics, Sparse Canonical Correlation Analysis, Multi-Task Sparse Canonical Correlation Analysis

1 INTRODUCTION

Brain Imaging Genetics

Brain imaging genetics is an emerging and important topic which integrates both the genetic factors and neuroimaging phenotypic measurements in brain science. This integration research of combining diverse genetic and genomic data is expected to uncover the genetic basis of brain structures and functionalities, and further offers new opportunities to interpret the causality of relationships between genetic variations and brain disorders such as the Alzheimer’s disease (AD) [1], [2]. Modern neuroimaging techniques, such as magnetic resonance imaging (MRI) and positron-emission tomography (PET), image the morphometry and metabolic processes of the brain based on different techniques, and generate different imaging data describing the brain from different perspectives. These multi-modal imaging data provide complementary information and have been demonstrated to offer comprehensive understandings of the brain structures, functionalities, and brain disorders [3]. Moreover, in biomedical studies, we usually face a huge size of genotyping biomarkers such as the single nucleotide polymorphisms (SNPs), which is a type of high-resolution markers in genome-wide association studies (GWAS). Therefore, developing the fast and efficient GWAS-oriented imaging genetics method which integrates multi-modal imaging data simultaneously is of great importance and meaning.

The multivariate learning methods are very popular in brain imaging genetics since both imaging data and genetic data are multivariate. The multi-task learning (MTL) techniques are of this kind and widely used in brain imaging genetics [4], [5]. Generally, these methods choose a few important imaging QTs relevant to their aim as dependent variables and SNPs as independent variables. Then joint effect of multi-locus genotype on few phenotypes is studied. This paradigm can select SNPs that are simultaneously relevant to the candidate brain phenotypes. However, the brain is demonstrated to be comprised of multiple regions. Then using only a small proportion of them could be lack of power since they may lose important information carried by cerebral components which are not included.

Although a brain-wide MTL model can be used, they are still insufficient since they cannot select relevant brain phenotypes from multiple brain cerebral components. Therefore, bi-multivariate methods become more and more popular in brain imaging genetics recently. Sparse canonical correlation analysis (SCCA) is such a technique which usually identifies the relationship between two views of data with sparse output induced by different regularization techniques [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. These two-
view SCCA methods have limited power since they only utilize one modal imaging QTs. Given multi-modal imaging data, incorporating them together could make use of the information carried by different modalities and would be beneficial to uncover interesting findings that using one modality cannot. Therefore, jointly analyzing the relationship between all the imaging phenotypes from different modalities and genetic factors via one single integrative SCCA model is desirable and of great interest. And this integrative model would be helpful to elucidate the shared mechanism of genetic factors on the brain.

One possible solution is the multi-view SCCA modelling which considers the pairwise relationship between all omics data involved. This multi-view SCCA is a naive extension to existing two-view SCCA models, and a three-view one has been introduced in [13]. It learns only one single canonical weight for genetic loci which is overstrict thus cannot make full use of the complementary information embedded in different modalities of imaging phenotypes.

Using brain-wide imaging QTs from multiple modalities, in this paper, we propose a Multi-Task learning based SCCA (MTSCCA) framework [16], [17] which can study bi-multivariate associations between these phenotypes and genotypes simultaneously. MTSCCA treats each SNP and QT as a feature, and then models the association between each imaging modality and SNPs as a learning task. Different from those conventional SCCA, including both two-view and three-view methods, MTSCCA learns one canonical weight matrix for SNPs, in which each column vector corresponds to one canonical weight of one SCCA task. In contrast, only one canonical weight vector is associated from those conventional SCCA, including both two-view and three-view methods, MTSCCA learns one canonical weight of SNPs to be associated with all modalities of imaging phenotypes.

To distinguish from the notation in mSCCA, in this section, we use $X \in \mathbb{R}^{n \times p}$ to represent the genetic data with $n$ participants and $p$ SNPs, and $Y_j \in \mathbb{R}^{n \times q} (j = 1, \ldots, c)$ to represent the phenotype data with $q$ imaging measurements, where $c$ is the number of imaging modalities (tasks). Let $U \in \mathbb{R}^{p \times c}$ be the canonical weight matrix associated with $X$ and $V \in \mathbb{R}^{q \times c}$ be that associated with imaging QTs with each $v_j$ corresponding to $Y_j$, we propose the novel multi-task based SCCA (MTSCCA) model as follows

$$\begin{align*}
\min_{u_j, v_j} & \sum_{i=1}^{n} -u_j^\top X_{ij} \Omega(Y_j v_j) \\
\text{s.t.} & \|X_{ij}\|_2 = 1, \|Y_j v_j\|_2 = 1, \Omega(U) \leq b_1, \Omega(V) \leq b_2, \forall j.
\end{align*}$$

2.2 MTSCCA

2.2.1 The MTSCCA Model

To distinguish from the notation in mSCCA, in this section, we use $X \in \mathbb{R}^{n \times p}$ to represent the genetic data with $n$ participants and $p$ SNPs, and $Y_j \in \mathbb{R}^{n \times q} (j = 1, \ldots, c)$ to represent the phenotype data with $q$ imaging measurements, where $c$ is the number of imaging modalities (tasks). Let $U \in \mathbb{R}^{p \times c}$ be the canonical weight matrix associated with $X$ and $V \in \mathbb{R}^{q \times c}$ be that associated with imaging QTs with each $v_j$ corresponding to $Y_j$, we propose the novel multi-task based SCCA (MTSCCA) model as follows

$$\begin{align*}
\min_{u_j, v_j} & \sum_{i=1}^{n} -u_j^\top X_i \Omega(Y_j v_j) \\
\text{s.t.} & \|X_{ij}\|_2 = 1, \|Y_j v_j\|_2 = 1, \Omega(U) \leq b_1, \Omega(V) \leq b_2, \forall j.
\end{align*}$$

Obviously, our model is distinct from those mCCA. First, MTSCCA employs the multi-task framework which learns a series of related SCCA tasks together. This simultaneous learning has been empirically [21], [22] and theoretically [21], [23] shown to improve performance dramatically compared with learning each task independently [24]. Second, our model learns a canonical weight matrix $U$ for SNPs, in which each column $u_j$ corresponds to an individual SCCA task. This is helpful since it does not require a unique canonical weight of SNPs to be associated with all modalities of imaging QTs at the same time. Third, MTSCCA learns one canonical weight corresponding to each imaging modality separately, indicating that we do not need to calculate multiple canonical weights for a specific imaging modality. This helps the model focus on the identification of markers from the genetic data, indicating it is quite suitable for imaging genetics analysis. Finally, our mode can be well scalable in terms of both modeling and computation. According to Eqs.
Since numerous SNPs inherently exhibit group structure in the genome, a realistic modeling method should take this information into consideration. In Eq. (2), a canonical weight matrix is associated with SNPs, and thus the conventional group Lasso which is used to penalize a vector cannot be employed directly. To tackle this issue, we use the $G_{2,1}$-norm function [5] which is formulated as

$$
\|U\|_{G_{2,1}} = \sum_{k=1}^K \|U^k\|_2 = \sum_{k=1}^K \left( \sum_{i \in g_k} \sum_{j=1}^p u_{ij}^k \right),
$$

where the SNPs are partitioned into $K$ groups $G = \{g_k\}_{k=1}^K$. This regularization penalizes the SNPs in the same group as a whole and expects to estimate equal or similar coefficients for them. According to [5], this penalty has two major merits. First, it incorporates the group structural knowledge into the model via packaging all SNPs in the same group together. This makes the model practical because it is in accordance with the genetic mechanism. Second, it penalizes the canonical weight coefficients of a group of variables across all SCCA tasks jointly. This setup can mutually promote each individual task.

Although using $G_{2,1}$-norm regularization is meaningful, there is a lack of feature selection at individual level. For those disease related SNPs, they could hardly be located in the same group. Generally, within a specific group, an individual variable could be relevant to the QTS and those remaining ones could be irrelevant. Therefore, we also model this via the $L_2,1$-norm regularization which is the Lasso regularization adjusted for multi-task feature selection,

$$
\|U\|_{2,1} = \sum_{i=1}^p \|u_i\|_2 = \sum_{i=1}^p \left( \sum_{j=1}^p u_{ij}^2 \right),
$$

Using both $G_{2,1}$-norm and $L_2,1$-norm regularization, MTSCCA can not only select features at the group level in accordance with the biological knowledge, but also jointly select feature at the individual level across all SCCA tasks.

### 2.2.2 Group-sparsity and Joint Individual Feature Selection for SNPs

At first glance this is similar to that used to jointly select individual features for SNPs, but it is employed here based on a different motivation. Although collected based on different imaging technologies, all modalities of imaging QTSs are measured from the same brain geography and have been mapped onto the same brain atlas via the segmentation and registration. Thus it is reasonable to assume equal or similar weights for those imaging QTSs associated with the same brain area but attributed to different modalities. Therefore, the $L_2$-norm imposed on $v$ penalizes those QTSs from the same brain area but different modalities together, and then the $L_1$-norm is utilized to select them jointly.
identity matrix of size equaling to the k-th group; $D_k$ is also a diagonal matrix with diagonal entries being $\frac{1}{2 ||x_i^T||_2} (i \in [1, p])$; and $\gamma_i = \gamma + 1$.

Then we can easily have

$$ (\beta \hat{D} + \lambda_1 D_k + \gamma_i x^T x) U = X^T Y, \tag{11} $$

and further

$$ U = (\beta \hat{D} + \lambda_1 D_k + \gamma_i x^T x)^{-1} X^T Y. \tag{12} $$

According to [5], this linear system in terms of $U$ can be efficiently solved via an iterative algorithm by alternatively first updating $\hat{D}$ and $D_k$, and then $U$. However, if the number of SNPs becomes larger and larger, this iterative algorithm is still computationally expensive.

**A fast implementation:** The primary difficulty of the $U$-update is the calculation of the covariance matrix $X^T X$ when there are a large number of features of $X$. In this paper, we use an approximation method to assure a fast implementation of $X^T X$ via making use of the prior information, i.e., the inherent structure of the SNPs within the genome. Fig. 1 is the illustration of the pairwise correlation coefficients and LD values in $r^2$ among a segment of SNPs at different loci from chromosome 19. The SNPs naturally form block structure along the diagonal, indicating a clear pattern of intra-block high correlation and inter-block low correlation. Since $X$ is centered and normalized, $X^T X$ is the same as the pairwise correlation coefficients as shown in Fig. 1. This indicates that $X^T X$ holds block diagonal structure too, and its off-block-diagonal elements are nearly zero, i.e., $x_i^T x_i \approx 0$ ($i \neq t$). In a word, the information of the covariance matrix are mainly carried by a series of block matrices along the diagonal. Most importantly, the size of these blocks are quite small compared with the original covariance matrix attributing to the fact that the LD block is usually much smaller than the number of SNPs ($p_k \ll p$) in human genome [25].

![Fig. 1. Illustration of the pairwise correlation coefficients and LD values ($r^2 \geq 0.2$) of SNPs from Chromosome 19 of an ADNI database.](image)

This structure has been widely used to guide the recovery of group relationship among SNPs via the group Lasso [26] or $G_{2, 1}$-norm [5]. However, they suffer from heavy computational issues caused by the enormous SNPs, and only when artificially assuming that $X^T X$ is an identity matrix could alleviate this issue [6], [7], [13]. From Fig. 1, we know that the identity assumption will inevitably lose information carried by those blocks along the diagonal [11]. In this study, we not only make use of this grouping information to identify relationships among SNPs, but also explore a fast and easy-to-implement method to handle the computational issues.

![Fig. 2. Illustration of the simplified covariance matrix $X^T X$, where $X_{g_k}$ and $X_{g_{k+1}}$ are two LD blocks, and $X_{g_k}^T X_{g_k}$ is abbreviated as $(X^T X)_{g_k}$. Since the correlation between the two blocks are very low ($X_{g_k}^T X_{g_{k+1}} \approx 0$ and $X_{g_k}^T X_{g_k} \approx 0$), their covariance can be ignored.](image)

Based on the analysis above, we propose that $X^T X$ can be computationally simplified by a series of $(X^T X)_{g_k}$ (abbreviated from $X_{g_k}^T X_{g_k}$) along the diagonal. We only omit those off-block-diagonal elements which has little influence on the performance. Fig. 2 is the illustration of the approximation where the off-block-diagonal elements are replaced by zero. It is clear that the primary information of $X^T X$ are well preserved since we take into consideration the LD structure. Therefore, compared with those methods using identity assumption, our method preserves more information of the data, and could be useful in identification of important genetic markers [11]. Most importantly, other than those methods calculating $X^T X$ via the brute force [11], we have a very fast implementation which is supported by the following theorem.

**Theorem 1.** If $X^T X$ is a block diagonal matrix, Eq.(11) can be solved by

$$ U = \oplus_{k=1}^K U_k^k, \quad U_k^k = \left[ \begin{array}{c} \vdots \\ U_k^1 \\ \vdots \end{array} \right], \tag{13} $$

$$ U_k^k = (\beta \hat{D}_{g_k} + \lambda_1 D_{g_k} + \gamma_i x^T x)^{-1} U_k^1 X_{g_k}^T \left[ Y_1 v_1, \ldots, Y_c v_c \right], $$

where $\hat{D}_{g_k}$ is the $k$-th block of the diagonal matrix $\hat{D}$; $D_{g_k}$ is the $k$-th block of $D_k$; and $\oplus$ denotes the operation that concatenates matrices vertically.

**Proof 1.** Since SNPs exhibit group structures, we denote $X = (\ldots X_{g_k} \ldots)$ with $k$ being the index of the $k$-th group. Then the covariance matrix $X^T X$ can be represented as

$$ X^T X = \begin{bmatrix} \ddots & \vdots \\ \vdots & (X^T X)_{g_k} \end{bmatrix}. $$

We have known that $D_1$ and $\hat{D}_1$ are diagonal matrices, indicating they both are diagonally separable. Then according to Eq. (12), we have

$$ U = \left( \beta \hat{D}_{g_k} + \lambda_1 D_{g_k} + \gamma_i x^T x \right)^{-1} U_k^1 X_{g_k}^T \left[ Y_1 v_1, \ldots, Y_c v_c \right]. $$

$$ U_k^k = \left( \begin{array}{c} \vdots \\ \beta \hat{D}_{g_k} + \lambda_1 D_{g_k} + \gamma_i x^T x \\ \vdots \end{array} \right)^{-1} U_k^1 X_{g_k}^T \left[ Y_1 v_1, \ldots, Y_c v_c \right], $$

$$ U_k^k = \left( \begin{array}{c} \vdots \\ \beta \hat{D}_{g_k} + \lambda_1 D_{g_k} + \gamma_i x^T x \\ \vdots \end{array} \right)^{-1} U_k^1 X_{g_k}^T \left[ Y_1 v_1, \ldots, Y_c v_c \right]. $$
The advantages of this theorem are threefold. (1) The time complexity of Eq. (13) is $O(np^2K)$ compared with that of Eq. (12) being $O(np^3)$, where $p_k$ is the size of the $k$-th group, and $p = \sum_{k=1}^{K} p_k$. This is a significant improvement because that the LD block size is usually quite small, i.e. $p_k \ll p$. (2) Benefiting from the computation effort reduction, the memory requirement is also saved a lot because storing $X^T X$ is very memory expensive than storing several $(x^T x)_{pk}$. (3) According to the proof, Eq. (13) is quite easy to implement, demonstrating it is very promising in big imaging genetic analysis. This is one of the contributions of this study and might provide a powerful tool for genome-wide and brain-wide bi-multivariate analysis.

### 2.3.2 Updating $v_j$

Note that each $v_j$ is associated with each $Y_j$ separately. This means that these $v_j$’s are not closely coupled such as $u_i$’s and should be tackled with separately. Next we will show how to solve $v_j$ with $v_k (k \neq j)$ and $U$ being fixed. Based on Eq. (9), we take the derivative with respect to $v_j$ and set it to zero

$$-Y_j^T X u_j + \lambda_2 D_2 v_j + \gamma_2 Y_j^T Y_j v_j = 0,$$

which can be rewritten as

$$(\lambda_2 D_2 + \gamma_2 Y_j^T Y_j) v_j = Y_j^T X u_j,$$

i.e.

$$v_j = (\lambda_2 D_2 + \gamma_2 Y_j^T Y_j)^{-1} Y_j^T X u_j,$$

where $D_2$ is a diagonal matrix with its $i$-th entry being $\frac{1}{2} v_{ij}^2 (i \in [1, q])$ on the diagonal; and $\gamma_2 = \gamma_2 + 1$. Therefore, each $v_j$ can also be solved alternatively through an iterative algorithm.

Now that the building blocks regarding updating $U$ and each individual $v_j$ are created, we present the pseudocode in Algorithm 1.

### 2.4 Convergence Analysis

We have the following theorem for Algorithm 1.

**Theorem 2.** Algorithm 1 decreases the objective value of Eq. (9) in each iteration.

**Proof 2.** In order to prove this theorem, we need two essential conclusions: (1) Eq. (12) decreases the objective Eq. (9) in each iteration; and (2) Eq. (16) decreases the objective Eq. (9) in each iteration.

**Algorithm 1** Algorithm to solve Eq. (9)

**Require:**

$X \in \mathbb{R}^{n \times p}$, $Y_j \in \mathbb{R}^{n \times q}$, $j = 1, \ldots, c$, $\beta, \lambda_1, \lambda_2, \gamma_1, \gamma_2$

**Ensure:**

Canonical weights $U$ and $V$.

1. Initialize $U \in \mathbb{R}^{P \times c}$, $V \in \mathbb{R}^{n \times c}$;
2. **while** not convergence **do**
   3. Update $\overline{D}_{y_k}$ and $D_1 v_k$;
   4. Solve $U$ according to Eq. (13), and normalize $u_j$ to $\|X u_j\|_2 = 1$;
   5. Update $D_2$;
   6. Solve $v_j (j = 1, \ldots, c)$ according to Eq. (16), and normalize $v_j$ to $\|Y_j v_j\|_2 = 1$;
7. **end while**

We first prove the conclusion (1). According to Eq. (12), we have

$$\sum_{j=1}^c \left( \|X u_j(t+1) - Y_j v_j(t)\|_2^2 + \gamma_1 \text{Tr}(U(t+1)^T X^T X U(t+1)) + \beta \text{Tr}(U(t+1)^T \overline{D} U(t+1)) + \lambda_1 \text{Tr}(U(t+1)^T D_1 U(t+1)) \right) \leq \sum_{j=1}^c \left( \|X u_j(t) - Y_j v_j(t)\|_2^2 + \gamma_1 \text{Tr}(U(t)^T X^T X U(t)) + \beta \text{Tr}(U(t)^T \overline{D} U(t)) + \lambda_1 \text{Tr}(U(t)^T D_1 U(t)) \right)$$

$$= \sum_{j=1}^c \left( \|X u_j(t+1) - Y_j v_j(t)\|_2^2 + \gamma_1 \|X u_j(t+1)\|_2^2 \right)$$

According to Lemma 1 in [5], $\left\|U^{(k+1)}\right\|_F - \left\|U^{(k+1)}\right\|_F - \frac{1}{2} \left\|U^{(k)}\right\|_F \leq \frac{1}{2} \left\|U^{(k)}\right\|_F$, and $\left\|u^{(t+1)}\right\|_2^2 - \frac{1}{2} u^{(t+1)}_i \leq \left\|u^{(t)}\right\|_2^2 - \frac{1}{2} u^{(t)}_i$. Then applying two inequations to Eq. (17) with respect to each group and individual feature, we have

$$\sum_{j=1}^c \left( \|X u_j(t+1) - Y_j v_j(t)\|_2^2 + \gamma_1 \|X u_j(t+1)\|_2^2 \right)$$

$$= \sum_{j=1}^c \left( \|X u_j(t) - Y_j v_j(t)\|_2^2 + \gamma_1 \|X u_j(t)\|_2^2 \right)$$

which can be rewritten in matrix form as

$$\sum_{j=1}^c \left( \|X u_j(t+1) - Y_j v_j(t)\|_2^2 + \gamma_1 \|X u_j(t+1)\|_2^2 \right)$$

+ $\beta \left\|U^{(t+1)}\right\|_{G_{2,1}} + \lambda_1 \left\|U^{(t+1)}\right\|_{2,1}$

$$\leq \sum_{j=1}^c \left( \|X u_j(t) - Y_j v_j(t)\|_2^2 + \gamma_1 \|X u_j(t)\|_2^2 \right)$$

+ $\beta \left\|U^{(t)}\right\|_{G_{2,1}} + \lambda_1 \left\|U^{(t)}\right\|_{2,1}$

$$\leq \sum_{j=1}^c \left( \|X u_j(t+1) - Y_j v_j(t)\|_2^2 + \gamma_1 \|X u_j(t+1)\|_2^2 \right)$$

+ $\beta \left\|U^{(t+1)}\right\|_{G_{2,1}} + \lambda_1 \left\|U^{(t+1)}\right\|_{2,1}$.

**1545-5983 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.**
\( \gamma_i \) is replaced by \( \gamma_1 \) since \( \| Xu_{(t+1)} \|_2^2 \) and \( \| Xu_t \|_2^2 \) have been normalized to 1. Thus the objective value is decreased in each iteration regarding updating \( u \).

Similarly, we have the following inequality.

\[
\sum_{j=1}^k \left( \| Xu_{(t+1)}^j - Y_j v_{(t+1)}^j \|_2^2 + \gamma_2 \| Y_j v_{(t+1)}^j \|_2^2 + \lambda_2 \| V^{(t+1)} \|_{2,1} \right) 
\leq \sum_{j=1}^k \left( \| Xu_{(t)}^j - Y_j v_j^j \|_2^2 + \gamma_2 \| Y_j v_j^j \|_2^2 + \lambda_2 \| V^{(t)} \|_{2,1} \right) .
\]

Now based on Eqs. (19-20), we have \( \mathcal{L}(U^{(t+1)}, V^{(t+1)}) \leq \mathcal{L}(U^{(t)}, V^{(t)}) \), which completes the proof.

According to Eq. (9), the objective is lower bounded by 0, and thus iteratively decreasing the objective value will converge to a local optimum. The proposed algorithm runs very fast owning to (1) its closed-form solution for each update; and (2) the divide-and-conquer strategy which is supported by Theorem 1.

### 3 RESULTS

#### 3.1 Benchmarks and Experimental Setup

In order to evaluate the performance of the proposed multi-task SCCA method, we choose the closely related mSCCA [13] and the conventional two-view SCCA as benchmark methods. A common problem of the two-view SCCA and mSCCA is that they suffer from heavy computational and memory requirement issues because they cannot handle the large covariance matrix calculation. To make the comparison available, based on Theorem 1, we implement the fast two-view SCCA and the fast mSCCA. This yields the three benchmark methods in this study, confirming another contribution of this study.

All the methods contain parameters that should be fine tuned before running experiments. We apply the nested 5-fold cross-validation in this work. Specifically, those tuning parameters were determined in the inner loop where a group of them generating the highest mean correlation coefficients, i.e. \( CV(\lambda, \beta, \gamma) = \frac{1}{3} \sum_{j=1}^3 Corr(X_{-j} u_j, Y_{-j} v_j) \), will be chosen as the optimal parameters, where \( X_{-j} \) and \( Y_{-j} \) are the \( j \)-th subset of the inner testing set, and \( u_j \) and \( v_j \) are the canonical weights estimated from the inner training set. Once determined, those parameters are used in the external loop to generate the final results. Before tuning parameters, we use some heuristic strategy to reduce the computation burden since blindly tuning them by grid search is computational intensive. For all methods, \( \gamma_1 \) and \( \gamma_2 \) are used to address the scaling issue when calculating the correlation coefficient. On this account, fixing the denominator to be 1 or other integers will just affect the magnitude of \( U \) and \( V \), and the relative relationship among each element remains the same. For example, suppose \( u_{1,1} = 5, u_{1,2} = 1 \) and \( \| Xu_1 \|_2^2 = 20 \) (\( \| u_1 \|_2^2 = 20 \) for two-view SCCA and mSCCA), tuning \( \| Xu_1 \|_2^2 = 1 \) will lead to \( u_{1,2} = 0.25 \) and \( u_{1,2} = 0.05 \); while tuning \( \| Xu_1 \|_2^2 = 10 \) will lead to \( u_{1,1} = 2.5 \) and \( u_{1,2} = 0.5 \). This will not affect the feature selection as \( u_{1,1} \) will always be selected with higher priority than \( u_{1,2} \). Therefore, we set \( \gamma_1 = \gamma_2 = 1 \) in this paper. Generally, too large parameters yield over-penalized results while too small ones yield less-penalized results. To avoid this issue, we tune the remaining parameters \( \lambda_1, \lambda_2, \beta \) from a moderate range \( 10^i (i = -5, -4, \ldots, 0, \ldots, 4, 5) \) via the grid search strategy. Finally, in order to make the results stable, we repeat each experiment 100 times and show the average results. In the experiments, all methods are stopped when both \( \max_i |u_{i(t+1)} - u_i(t)| \leq \epsilon (\forall u_i) \) and \( \max_j |v_{j(t+1)} - v_j(t)| \leq \epsilon (\forall v_j) \) are satisfied, where \( \epsilon \) is the tolerable error. We empirically set \( \epsilon = 10^{-5} \) from experiments in this paper.

#### 3.2 Simulation Study

This section presents the comparison results on the synthetic data. We generate four data sets with different number of samples and features, sparsity levels and noise levels to assure a thorough comparison. The first three data sets are generated using the same ground truth but with different noise strengths. The \( X, Y_j (j = 2) \) and \( z \) of them are all with \( n = 80, p = 120, q_1 = 100 \) and \( q_2 = 100 \). This could help show the performance when treating with different noises. The fourth data set is created to access the performance under high-dimensional situation, and \( n = 500, p = 2,000, q_1 = 1,000, q_2 = 1,000 \) respectively. The details of each data set are described as follows.

- **Data 1**: We first set \( u = (0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0)^T \), \( v_1 = (0, \ldots, 0, 0, \ldots, 0)^T \) and \( v_2 = (0, \ldots, 0, 0, \ldots, 0)^T \). Then we generate a random latent vector \( \mu \) of length \( n \) and normalize it to unit length. The data matrix \( X \) is created by \( x_{i,j} \sim N(\mu u_j, \sigma_x) \), where \( \sigma_x = 5 \) denotes the noise strength. Similarly, \( Y_j \) is created by \( y_{j,i} \sim N(\mu v_j, \sigma_y) \) with \( \sigma_y = 5 \) and \( \sigma_y = 5 \).

- **Data 2 - Data 3**: These two data sets are created with the same ground truth as the first one with different noises, i.e. \( \sigma_x = \sigma_{y_1} = \sigma_{y_2} = 1 \) for Data 2 and \( \sigma_x = \sigma_{y_1} = \sigma_{y_2} = 0.1 \) for Data 3. Therefore, the correlation coefficients of these three data sets are different, and that of the first data set is the smallest and that of the third one is the highest.

---

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2019.2947428, IEEE/ACM Transactions on Computational Biology and Bioinformatics.
TABLE 1

<table>
<thead>
<tr>
<th>Training Results</th>
<th>Testing Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>SCCA</td>
<td>mSCCA</td>
</tr>
<tr>
<td>Data 1</td>
<td>0.28±0.06</td>
</tr>
<tr>
<td>Data 2</td>
<td>0.59±0.06</td>
</tr>
<tr>
<td>Data 3</td>
<td>0.95±0.01</td>
</tr>
<tr>
<td>Data 4</td>
<td>0.89±0.02</td>
</tr>
</tbody>
</table>

Fig. 4. Canonical weights \( V \) (mean value) estimated on synthetic data. The first row is the ground truth, and each remaining row corresponds to an SCCA method: (1) Two-view SCCA, (2) mSCCA (Multi-view SCCA), (3) MTSCCA (Multi-task SCCA). In each subfigure, the horizontal axis represents the indices of \( v_j \) \((j = 1, 2)\), and the vertical axis represents the estimated weight value.

We first show the training and testing canonical correlation coefficients (CCCs) in Table 1. On the first three data sets, all methods obtain a good score when the true CC is high, while perform poorly (overfitting) when the true CC is excessively low due to the high percentage of noise. MTSCCA identifies the highest training CCCs among all three unsupervised methods, including both two-view SCCA and mSCCA. This demonstrates that MTSCCA performs better than the two single-task based SCCA methods. In Data 4, we observe that MTSCCA obtains higher training and testing CCCs than two-view SCCA and mSCCA in this high-dimensional data set. This indicates that, owing to the multi-task modeling strategy, the ability of identifying bi-multiplicative association can be improved.

In addition, the feature selection ability is also of great interest and importance. In Fig. 3, we show the scatter of the estimated \( u \) and \( v_j \)'s. For the two-view SCCA, each \( u_j \) is calculated independently from each single-task SCCA, and \( u \) is obtained by averaging \( u_j \)'s. The \( u \) of MTSCCA is also obtained by averaging \( u_j \)'s associating with three SCCA tasks. There are two estimated \( v_j \)'s for all methods and we show them separately in Fig. 4. In order to show the performance clearly, the ground truths are also presented in the figure (first row). Within each subfigure, the horizontal axis represents the indices and the vertical axis represents the weight values. A feature with a larger canonical weight (in absolute value) contributes more to the bi-multivariate correlation. We observe that all methods cannot find out the correct locations of signals in the first data owing to the low signal-to-noise ratio. Combining the results of the first three data sets together, all their performances improve from the first data set to the third one. MTSCCA holds the best canonical profiles being consistent with the ground truths, showing its better performance in feature selection than the two-view and multiple-view SCCA. In Data 4 where the feature dimensionality is high, MTSCCA always identifies correct signal locations. To make the comparison more formal, Table 2 and 3 show the sensitivity and specificity in terms of canonical weights \( u \) and \( v_j \)'s. Both metrics are calculated as follows. Features are selected based on their absolute weight values, and the larger the \( |u_i| \) (or \( |v_i| \)) is, the more relevant to the canonical correlation. Generally, given a predefined threshold, those features with larger-than-threshold values are selected. However, it is hard to predefine an appropriate threshold. To overcome this issue, in this paper, the sensitivity is calculated via

\[ \text{#true positive in the top } K \text{ selected features} \]

where \( K \) is the number of non-zero features of the ground truth. Similarly, the specificity is calculated by

\[ \text{#selected features required to cover the ground truth} \]

The results show that all methods obtain good sensitivity and specificity across these simulated data sets. MTSCCA performs slightly better than those single-task based SCCA methods owing to the multi-task modeling strategy. It is worth noting that in the original implementations, both two-view SCCA and multi-view SCCA fail since they cannot treat the large matrix calculation on the same platform as MTSCCA does. By incorporating Theorem 1, the two methods become fea-
sible to high-dimensional data sets. The runtime of each method is shown in Table 4, and there is no significant difference between these methods based on Theorem 1. This again demonstrates the effectiveness and practice of our fast implementation strategy.

### 3.3 Real Neuroimaging Genetics Study

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.

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Comparison of the sensitivity of canonical weights on synthetic data.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>u</td>
</tr>
<tr>
<td></td>
<td>Data 1 Data 2 Data 3 Data 4</td>
</tr>
<tr>
<td>SCCA</td>
<td>0.23 0.25 1.00 1.00</td>
</tr>
<tr>
<td>mSCCA</td>
<td>0.21 0.25 0.95 1.00</td>
</tr>
<tr>
<td>MTSCCA</td>
<td>0.83 0.90 1.00 1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Comparison of the specificity of canonical weights on synthetic data.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>v1</td>
</tr>
<tr>
<td></td>
<td>Data 1 Data 2 Data 3 Data 4</td>
</tr>
<tr>
<td>SCCA</td>
<td>0.85 0.89 1.00 1.00</td>
</tr>
<tr>
<td>mSCCA</td>
<td>0.85 0.89 1.00 1.00</td>
</tr>
<tr>
<td>MTSCCA</td>
<td>0.85 0.89 1.00 1.00</td>
</tr>
</tbody>
</table>

In summary, this simulation study using data sets with diverse characteristics demonstrates that MTSCCA is effective in bi-multivariate association identification with multiple data modalities. Moreover, MTSCCA identifies the best canonical loading profiles which is consistent with the ground truth compared to the single-task SCCA methods. In addition, it also reveals that the group structure can not only help prompt the identification performance, but also help reduce the time effort in high-dimensional scenario in multi-modal bi-multivariate association analysis.

The neuroimaging data were from 755 non-Hispanic Caucasian participants, including 281 AD, 292 MCI and 182 healthy control (HC). They were 18-F florbetapir PET (AV-45) scans, fluorodeoxyglucose positron emission tomography (FDG) scans, and structural MRI scans which were downloaded from the ADNI database (adni.loni.usc.edu). Details of this data set are exhibited in Table 5. The multi-modality imaging data were aligned to each participant’s same visit. The structural MRI scans were processed with voxel-based morphometry (VBM) via SPM [27]. Generally, all scans had been aligned to a T1-weighted template image, segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) maps, normalized to the standard Montreal Neurological Institute (MNI) space as $2 \times 2 \times 2 \text{mm}^3$ voxels, and had been smoothed with an 8mm FWHM kernel. The FDG-PET and AV45-PET scans were also registered into the same MNI space by SPM. We then subsampled the whole brain and generated 116 regions of interest (ROI) level measurements based on the MarsBaR automated anatomical labeling (AAL) atlas. They were the mean gray matter densities for structural MRI, amyloid values for AV45 scans and glucose utilization for FDG scans. Using the regression weights derived from the healthy control participants, these imaging measures were pre-adjusted to remove the effects of the baseline age, gender, education, and handedness.

The genotyping data of the same population were downloaded from the LONI 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. The QC criteria for the SNP data include (1) call rate check per subject and per SNP marker, (2) gender check, (3) sibling pair identification, (4) the Hardy-Weinberg equilibrium test, (5) marker removal by the minor allele frequency and (6) population stratification. In second pre-processing step, following the quality-controlled SNPs, those missing genotypes were imputed using the MaCH software [28]. Among all human chromosomes, the chromosome 19 sequence contains the most number of genes, in which the gene density is more than double the genome-wide average [29], [30]. 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. As
a result, all the SNPs from chromosome 19 were included, i.e., 152,787 SNPs were involved in this study. Among these enormous SNPs, most of them might be irrelevant to AD, while only a few of them could be relevant via influencing the intermediate brain imaging measurements. The aim is to identify this small subset of SNPs in chromosome 19 correlating to imaging markers and AD.

### 3.4 Improved Bi-multivariate Association

In this subsection we evaluate the proposed method in identifying the bi-multivariate associations between one genetic data and three sets of imaging phenotypes. Thus there will be three pairs of associations, and we denote them as SNPs-AV45, SNPs-FDG and SNPs-VBM for the sake of description. For the three SCCA tasks, the proposed MTSCCA learns them together and generate a canonical weights matrix $U$ for SNPs and one canonical weight vector $v_j$ for AV45, FDG and VBM. We then calculate three canonical correlation coefficients (CCCs) in terms of SNPs-AV45, SNPs-FDG and SNPs-VBM separately. The two-view SCCA naturally yields three CCCs for these three tasks. Though the mSCCA only learns one canonical weight vector for SNPs, we use it three times to generate three CCCs with respect to the three tasks.

![Performance Comparison](image)

**Fig. 5.** Performance comparison: The mean and standard deviation (SD) of the canonical correlation coefficients (CCCs) obtained from 5-fold cross-validation trials are plotted, where each error bar indicates $\pm0.5$SD. The subtitle SNPs-AV45 means the CCCs are calculated between the SNPs data and the AV45-PET data.

### 3.5 Genetic Marker Selection

Apart from the CCCs, the selected features in terms of SNPs are a major concern. This can help reveal those SNPs being highly related to imaging QTs and AD status at the

<table>
<thead>
<tr>
<th>Number of selected biomarkers</th>
<th>20</th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training correlation coefficients</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNP-AV45</td>
<td>1.12E-08</td>
<td>1.13E-07</td>
<td>1.14E-07</td>
<td>1.15E-07</td>
<td>1.16E-07</td>
</tr>
<tr>
<td>SNP-FDG</td>
<td>1.12E-08</td>
<td>1.13E-07</td>
<td>1.14E-07</td>
<td>1.15E-07</td>
<td>1.16E-07</td>
</tr>
<tr>
<td>SNP-VBM</td>
<td>1.12E-08</td>
<td>1.13E-07</td>
<td>1.14E-07</td>
<td>1.15E-07</td>
<td>1.16E-07</td>
</tr>
</tbody>
</table>

**Table 8**

Top ten SNPs selected by integrated canonical weights.

<table>
<thead>
<tr>
<th>Two-view SCCA</th>
<th>mSCCA</th>
<th>MTSCCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs429358</td>
<td>rs138339429</td>
<td>rs429358</td>
</tr>
<tr>
<td>rs10414043</td>
<td>rs141308467</td>
<td>rs56131196</td>
</tr>
<tr>
<td>rs147711004</td>
<td>rs85150143</td>
<td>rs12721051</td>
</tr>
<tr>
<td>rs146291812</td>
<td>rs17363184</td>
<td>rs4420638</td>
</tr>
<tr>
<td>rs625264</td>
<td>rs625264</td>
<td>rs117789331</td>
</tr>
<tr>
<td>rs73052335</td>
<td>rs1181833</td>
<td>rs66626994</td>
</tr>
<tr>
<td>rs18263561</td>
<td>rs253376</td>
<td>rs146275714</td>
</tr>
<tr>
<td>rs73052335</td>
<td>rs1749316</td>
<td>rs41289512</td>
</tr>
<tr>
<td>rs66626994</td>
<td>rs139402102</td>
<td>rs14771004</td>
</tr>
<tr>
<td>rs415966</td>
<td>rs4605289</td>
<td>rs10119</td>
</tr>
</tbody>
</table>

In addition, we calculate the $p$-values between our method and two competing methods and show them in Table 6, where the ‘-’ in parenthesis indicates that MTSCCA fails. The $p$-values are all reach the significance level which means that our method is significantly better than both competing methods. These results in terms of CCCs indicate that the proposed joint bi-multivariate learning method indeed has better association identification capability than those SCCA methods, including both two-view and multiple-view ones. Table 7 shows the runtime in seconds of each method, where that of the two-view SCCA are summation of three two-view tasks. The runtime results indicate that all three methods run fast on this large data set. This attributes to the grouping strategy used in the implementation according to Theorem 1. In contrast, both competing methods are incapacitated in their original implementations since they cannot manipulate a big matrix with hundreds of thousands of features included. This again assures our contribution to accelerate both our method and conventional methods via making use of the grouping structures.
same time. We show the top ten selected SNPs according to the canonical weight values of each individual method. In order to make the selection results stable, we average the canonical weight matrix into a vector and then choose the top ten SNPs based on their absolute values for MTSCCA. The top ten markers of two-view SCCA method are calculated via averaging the three separate canonical weights. Those of mSCCA are obtained by its canonical weight vector. The results of selected SNPs are shown in Table 8. Owing to the jointly learning paradigm, the proposed MTSCCA yields a surprisingly meaningful result with respect to selected features (SNPs). As expected, the notable AD risk markers rs429358 gains the highest weight value, and all of the remaining nine SNPs of MTSCCA, i.e. rs56131196 (APOC1), rs12721051 (APOC1), rs4420638 (APOC1), rs111789331 (4.5 kb of APOC1), rs66626994 (5.6 kb of APOC1), rs146275714 (PVRL2), rs41289512, rs147711004 (71 kb of APOE) and rs10119 (TOMM40), have been reported to show increasing risk of AD in previous studies [31], [32], [33]. This indicates the ability of MTSCCA in identifying meaningful SNPs from massive genetic markers. The two-view SCCA also identifies the rs429358 as its first important SNPs, and five other AD related SNPs (rs10414043, rs147711004, rs7256200, rs73052335 and rs66626994) have been reported previously. But it identifies four SNPs that are not reported by now and thus further investigation should be taken place. The mSCCA performs unacceptably in this comparison since it does not find out rs429358. Moreover, except for the marker rs623264, all identified SNPs of mSCCA have not been reported yet in the current stage. In summary, the results in terms of selected SNPs show that MTSCCA performs better than both competing methods. This reveals that MTSCCA could be a suitable tool and very helpful in discovering meaningful genetic markers in a very large scenario.

![Fig. 6. Comparison of canonical weights in terms of each imaging modality across five trials. Each row corresponds to a SCCA method: (1) Two-view SCCA; (2) mSCCA; (3) MTSCCA. Within each panel, there are three rows corresponding to three type of imaging QTs. i.e. AV45, FDG and VBM.](image)

### TABLE 9

<table>
<thead>
<tr>
<th>Top ten imaging QTs selected by canonical weights of each imaging modality of MTSCCA.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV45</td>
</tr>
<tr>
<td>FDG</td>
</tr>
<tr>
<td>VBM</td>
</tr>
<tr>
<td>Frontal_Med_Orb_Left</td>
</tr>
<tr>
<td>Cingulum_Post_Left</td>
</tr>
<tr>
<td>Postcentral_Left</td>
</tr>
<tr>
<td>Angular_Right</td>
</tr>
<tr>
<td>Angular_Right</td>
</tr>
<tr>
<td>Precentral_Left</td>
</tr>
<tr>
<td>Cingulum_Post_Left</td>
</tr>
<tr>
<td>Hippocampus_8</td>
</tr>
<tr>
<td>Cingulum_Post_Left</td>
</tr>
<tr>
<td>Cingulum_Ant_Left</td>
</tr>
<tr>
<td>Amygdala_Left</td>
</tr>
<tr>
<td>Thalamus_Right</td>
</tr>
<tr>
<td>Frontal_Mid_Right</td>
</tr>
<tr>
<td>Temporal_Mid_8</td>
</tr>
<tr>
<td>Temporal_Mid_8</td>
</tr>
<tr>
<td>Temporal_Mid_8</td>
</tr>
<tr>
<td>Vermis_8</td>
</tr>
<tr>
<td>Vermis_8</td>
</tr>
<tr>
<td>Vermis_8</td>
</tr>
</tbody>
</table>

### 4 Conclusion

High-throughput genotyping technique and neuroimaging techniques provide us a large amount of biomedical data, and finding out their bi-multivariate associations is important. In this paper, we have proposed a novel multi-task sparse canonical correlation analysis (MTSCCA) framework and apply it to imaging genetics with multi-modal brain imaging QTs. Different from existing SCCA, MTSCCA can incorporate multiple sets of imaging modalities data into a single integrative model. Furthermore, MTSCCA is a multiple bi-multivariate method and thus has better modeling capability than both SCCA and MTL regression. A fast optimization algorithm is proposed which avoids calculating the large covariance and its inverse. The algorithm is guaranteed to converge to a local optimum, and runs very fast with hundreds of thousands of features involved.

We compared MTSCCA with the conventional two-view and multi-view SCCA on an ADNI cohort. Our method obtained better performance than the benchmarks with higher...
correlation coefficients and clearer canonical weight patterns. MTSCCA succeeds in identifying a small set of SNPs from enormous genetic markers from the 19th chromosome. It is worth noting that all top ten selected SNPs of MTSCCA are AD risk factors. In addition, the canonical weight patterns of imaging QTs were also of great success. The identified imaging QTs were highly correlated to AD or MCI. These promising results demonstrated that the proposed multi-task SCA framework could be a powerful tool in big brain imaging genetics. Since a GWAS based bi-multivariate analysis is of much concern, in the future work, we will keep looking into the merit of MTSCCA and use it to genomewide brain-wide imaging analysis.

ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (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; CereSpire, 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.

This work was supported by the National Natural Science Foundation of China [61973255, 61602384]; NSF in Shaanxi Province of China [2017JQ6001]; CPSF [2017M613202]; PSF of Shaanxi [2017BSHEDZ281]; and FRFCU [3102018y029] at Northwestern Polytechnical University. This work was also supported by the National Institutes of Health [R01 EB022574, R01 LM011360, R01 AG063481, U01 AG024904, P30 AG10133, R01 AG19771]; and the National Science Foundation [IIS 1837964] at University of Pennsylvania and Indiana University.

References


Kefei Liu is a postdoctoral research associate of the Department of Biostatistics, Epidemiology and Informatics at the University of Pennsylvania. He received B.Sc. in mathematics from Wuhan University and Ph.D. in electronic engineering from the City University of Hong Kong. He is interested in developing machine learning and statistical methods for the analysis of large-scale heterogeneous biological data.

Xiaohui Yao received a B.S. degree in Computer Science and Technology from Qing Dao University, an M.S. degree in Computer Software and Theory from University of Science and Technology of China, and a Ph.D. degree in Bioinformatics from Indiana University. She is a Post-doctoral Fellow in the Department of Biostatistics, Epidemiology and Informatics at University of Pennsylvania. Her research interests include imaging genetics, multidimensional data mining, systems biology and information visualization.

Shannon L. Risacher received a B.S. degree in Psychology from Indiana University-Purdue University Indianapolis, and a Ph.D. degree in Medical Neuroscience from Indiana University School of Medicine. She is an Assistant Professor of Radiology and Imaging Sciences at Indiana University School of Medicine. Her main research interest is in identifying biomarkers for early detection of Alzheimer’s disease pathology before clinical symptoms.

Junwei Han received his Ph.D degrees in pattern recognition and intelligent systems from the School of Automation, Northwestern Polytechnical University, Xi’an, China, in 2000. He is currently a professor in School of Automation, Northwestern Polytechnical University. His research interests include computer vision and multi-media processing.

Andrew J. Saykin received a B.A. degree in Psychology from University of Massachusetts Amherst, and an M.S. degree in Clinical Psychology and a Psy.D. degree in Clinical Neuropsychology from Hahnemann Medical College. He is the Raymond C. Beeler Professor of Radiology and Professor of Medical and Molecular Genetics at Indiana University School of Medicine. His expertise is in the areas of multimodal neuroradiology research, human genetics, and neuropsychology/cognitive neuroscience. He has a longstanding interest in the structural, functional, and molecular substrates of cognitive deficits in Alzheimer’s disease, cancer, brain injury, schizophrenia, and other neurological and neuropsychiatric disorders.

Lei Guo received B.S., M.S., and Ph.D. degrees in 1982, 1986, and 1993, respectively. He is currently a professor of pattern recognition at Northwestern Polytechnical University, Xi’an, China. His research interests include computer vision, image processing, image segmentation, object detection and tracking.

Li Shen received a B.S. degree from Xi’an Jiaotong University, an M.S. degree from Shanghai Jiao Tong University, and a Ph.D. degree from Dartmouth College, all in Computer Science. He is a Professor of Informatics at University of Pennsylvania Perelman School of Medicine. He is an elected fellow of the American Institute for Medical and Biological Engineering (AIMBE). His research interests include medical image computing, bioinformatics, machine learning, brain imaging genomics, and big data science in biomedicine.