SCIENTIFIC REPORTS

OPEN

Pattern Discovery in Brain Imaging Genetics via SCCA Modeling with a Generic Non-convex Penalty

Lei Du[1], Kefei Liu[2], Xiaohui Yao [2], Jingwen Yan[2], Shannon L. Risacher[2], Junwei Han[1], Lei Guo[1], Andrew J. Saykin [2], Li Shen [2] & the Alzheimer's Disease Neuroimaging Initiative*

Brain imaging genetics intends to uncover associations between genetic markers and neuroimaging quantitative traits. Sparse canonical correlation analysis (SCCA) can discover bi-multivariate associations and select relevant features, and is becoming popular in imaging genetic studies. The L1-norm function is not only convex, but also singular at the origin, which is a necessary condition for sparsity. Thus most SCCA methods impose ℓ_1-norm onto the individual feature or the structure level of features to pursuit corresponding sparsity. However, the ℓ_1-norm penalty over-penalizes large coefficients and may incurs estimation bias. A number of non-convex penalties are proposed to reduce the estimation bias in regression tasks. But using them in SCCA remains largely unexplored. In this paper, we design a unified non-convex SCCA model, based on seven non-convex functions, for unbiased estimation and stable feature selection simultaneously. We also propose an efficient optimization algorithm. The proposed method obtains both higher correlation coefficients and better canonical loading patterns. Specifically, these SCCA methods with non-convex penalties discover a strong association between the *APOE* e4 rs429358 SNP and the hippocampus region of the brain. They both are Alzheimer's disease related biomarkers, indicating the potential and power of the non-convex methods in brain imaging genetics.

By identifying the associations between genetic factors and brain imaging measurements, brain imaging genetics intends to model and understand how genetic factors influence the structure or function of human brain[1–14]. Both genetic biomarkers such as single nucleotide polymorphisms (SNPs), and brain imaging measurements such as imaging quantitative traits (QTs) are multivariate. To address this problem, bi-multivariate association models, such as multiple linear regression[15], reduced rank regression[16–18], parallel independent component analysis[19], partial least squares regression[20,21], canonical correlation analysis (CCA)[22] and their sparsity-inducing variants[23], have been widely used to uncover the joint effect of multiple SNPs on one or multiple QTs. Among them, SCCA (Sparse CCA), which can discover bi-multivariate relationships and extract relevant features, is becoming popular in brain imaging genetics.

The CCA technique has been introduced for several decades[24]. CCA can only perform well when the number of observations is larger than the combined feature number of the two views. Unfortunately, the problem usually is a *large-p-small-n* problem in the biomedical and biology studies. And it gets even worse because in CCA we are facing a *large-(p + q)-small-n* problem. In order to overcome this limitation, sparse CCA (SCCA)[25–36] employs a sparsity inducing regularization term to select a small set of relevant features and has received increasing attention. The ℓ_1-norm based SCCA method[25] has gained great success for its sparsity pursuing capability. After that, there are many SCCA variants based on the ℓ_1-norm. For examples, the fused lasso penalty imposes the ℓ_1-norm onto the ordered pairwise features[25], and the group lasso penalty imposes the ℓ_1-norm onto the group of features[29,32]. Further, the graph lasso or the graph guided lasso can be viewed as imposing the ℓ_1-norm onto the pairwise features defined by an undirected graph[29].

However, the ℓ_1-norm penalty shows the conflict of optimal prediction and consistent feature selection[37]. In penalized least squares modeling, Fan and Li[38] showed that a good penalty function should meet three properties.

[1]School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China. [2]Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, 46202, USA. *A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to L.D. (email: dulei@nwpu.edu.cn) or L.S. (email: shenli@iu.edu)

First, the penalty function should be singular at the origin to produce sparse results. Second, it should produce continuous models for stable model selection, and third, the penalty function should not penalize large coefficients to avoid estimation bias. The ℓ_1-norm penalty is successful in feature selection because it is singular at the origin. On the contrary, the ℓ_1-norm penalty over-penalizes large coefficients, and thus it may be suboptimal with respect to the estimation risk[39,40]. The ℓ_0-norm function which only involves the number of nonzero features is an ideal sparsity-inducing penalty. However, it is neither convex nor continuous, and thus solving ℓ_0-norm constrained problem is NP-hard[41].

A number of non-convex penalties are proposed as the surrogate of the ℓ_0-norm to handle this issue. These penalties includes the ℓ_γ-norm ($0 < \gamma < 1$) penalty[42], the Geman penalty[43], the Smoothly Clipped Absolute Deviation (SCAD) penalty[38], the Laplace penalty[44], the Minimax Concave Penalty (MCP)[45], the Exponential-Type Penalty (ETP)[46] and the Logarithm penalty[47]. These non-convex functions have attractive theoretical properties for they all are singular at the origin and leave those larger coefficients unpenalized. Though they have gained great success in generalized linear models (GLMs), it is an unexplored topic to apply them to the SCCA models for achieving sparsity and unbiased prediction simultaneously.

Therefore, it is essential and of great interest to investigate performances of various SCCA models based on these non-convex penalties. A major challenge of non-convex function is the computational complexity. The local quadratic approximation (LQA) technique is introduced to solve the SCAD penalizing problem[38]. LQA approximates the objective by a locally quadratic expression which can be solved like a ridge constrained problem. Inspired by this, in this paper, we propose a generic non-convex SCCA models with these non-convex penalties, and propose a unified optimization algorithm based on the LQA technique and the Alternate Convex Search (ACS) method[48]. Using both synthetic data and real imaging genetic data, the experimental results show that with appropriate parameters, the non-convex SCCA methods have better performance on both canonical loading patterns and correlation coefficients estimation than the ℓ_1-norm based SCCA methods.

Methods

Throughout this paper, scalars are denoted as italic letters, column vectors as boldface lowercase letters, and matrices as boldface capitals. The $\|\mathbf{u}\|$ denotes the Euclidean norm of a vector $\mathbf{u}$.

Preliminaries.   *Sparse Canonical Correlation Analysis (SCCA).*   Let $\mathbf{X} \in \mathcal{R}^{n \times p}$ be a matrix representing the SNP biomarkers data, where n is the number of participants and p is the number of SNPs. Let $\mathbf{Y} \in \mathcal{R}^{n \times q}$ be the QT data with q being the number of imaging measurements. A typical SCCA model is defined as

$$\min_{\mathbf{u},\mathbf{v}} - \mathbf{u}^\mathrm{T}\mathbf{X}^\mathrm{T}\mathbf{Y}\mathbf{v} \tag{1}$$

$$s.t. \quad \|\mathbf{X}\mathbf{u}\|^2 \leq 1, \|\mathbf{Y}\mathbf{v}\|^2 \leq 1, \Omega(\mathbf{u}) \leq c_1, \Omega(\mathbf{v}) \leq c_2,$$

where $\mathbf{X}\mathbf{u}$ and $\mathbf{Y}\mathbf{v}$ are the canonical variables, $\mathbf{u}$ and $\mathbf{v}$ are the corresponding canonical vectors we desire to estimate, and c_1, c_2 are the tuning parameters that control the sparsity level of the solution. The penalty function could be the ℓ_1-norm penalty, or its variants such as the fused lasso, group lasso and graph lasso[25,27,29,32,34].

Non-convex Penalty Functions for SCCA.   In this paper, we investigate seven non-convex surrogate penalties of ℓ_0-norm in the SCCA model. They are singular at the origin, which is essential to achieve sparsity in the solution. And they do not overly penalize large coefficients. In order to facilitate a unified description, we denote the non-convex penalty as

$$\Omega_\mathrm{nc}(\mathbf{u}) = \sum_{i=1}^{p} P_{\lambda,\gamma}(|u_i|), \tag{2}$$

where λ and γ are nonnegative parameters, and $P_{\lambda,\gamma}(|u_i|)$ is a non-convex function. We absorb λ into the penalty because it cannot be decoupled from several penalties, such as the SCAD function[38]. We here have seven penalties and they are described in Table 1 and visualized in Fig. 1, where for clarity we have dropped the subscript i in u_i. There is a sharp point at the origin for each of them, indicating that they are singular at the origin. This is essential to achieve sparseness in the solution. Besides, these curves are concave in $|u_i|$ and monotonically decreasing on $(-\infty, 0]$, and monotonically increasing on $[0, \infty)$. Therefore, though these penalties are not convex, they are piecewise continuously differentiable and their supergradients exist on both $(-\infty, 0]$ and $[0, \infty)$[49]. Table 1 also shows their supergradients $P'_{\lambda,\gamma}(|u_i|)$ with respect to $|u_i|$.

The Proposed Non-convex SCCA Model and Optimization Algorithm.   Replacing the ℓ_1-norm constraints in the SCCA model, we define the unified non-convex SCCA model as follows

$$\min_{\mathbf{u},\mathbf{v}} - \mathbf{u}^\mathrm{T}\mathbf{X}^\mathrm{T}\mathbf{Y}\mathbf{v} + \Omega_\mathrm{nc}(\mathbf{u}) + \Omega_\mathrm{nc}(\mathbf{v}) \tag{3}$$

$$s.t. \quad \|\mathbf{X}\mathbf{u}\|^2 \leq 1, \|\mathbf{Y}\mathbf{v}\|^2 \leq 1,$$

where $\Omega_\mathrm{nc}(\mathbf{u})$ and $\Omega_\mathrm{nc}(\mathbf{v})$ can be any of the non-convex functions listed in Table 1.

To solve the non-convex SCCA problem, we use the Lagrangian method,

Penalty Name	Function ($P_{\lambda,\gamma}(|u|)$)	Supergradient ($P'_{\lambda,\gamma}(|u|)$)
ℓ_γ-norm[42]	$\lambda|u|^\gamma$	$\begin{cases} \infty, & u = 0, \\ \lambda\gamma|u|^{\gamma-1}, & |u| > 0. \end{cases}$
Geman[43]	$\dfrac{\lambda|u|}{|u| + \gamma}$	$\dfrac{\lambda\gamma}{(|u| + \gamma)^2}$
SCAD[38]	$\begin{cases} \lambda|u|, & |u| \leq \lambda \\ \dfrac{-|u|^2 + 2\gamma\lambda|u| - \lambda^2}{2(\gamma - 1)}, & \lambda \leq |u| \leq \gamma\lambda \\ \dfrac{\lambda^2(\gamma + 1)}{2}, & |u| \geq \gamma\lambda. \end{cases}$	$\begin{cases} \lambda, & |u| \leq \lambda \\ \dfrac{\gamma\lambda - |u|}{\gamma - 1}, & \lambda \leq |u| \leq \gamma\lambda \\ 0, & |u| \geq \gamma\lambda. \end{cases}$
Laplace[44]	$\lambda\left(1 - \exp\left(-\dfrac{|u|}{\gamma}\right)\right)$	$\dfrac{\lambda}{\gamma}\exp\left(-\dfrac{|u|}{\gamma}\right)$
MCP[45]	$\begin{cases} \lambda|u| - \dfrac{|u|^2}{2\gamma}, & |u| \leq \gamma\lambda \\ \dfrac{1}{2}\gamma\lambda^2, & |u| \geq \gamma\lambda. \end{cases}$	$\begin{cases} \lambda - \dfrac{|u|}{\gamma}, & |u| \leq \gamma\lambda \\ 0, & |u| \geq \gamma\lambda. \end{cases}$
ETP[46]	$\dfrac{\lambda}{1 - \exp(-\gamma)}(1 - \exp(-\gamma|u|))$	$\dfrac{\lambda\gamma}{1 - \exp(-\gamma)}\exp(-\gamma|u|)$
Logarithm[47]	$\dfrac{\lambda}{\log(\gamma + 1)}\log(\gamma|u| + 1)$	$\dfrac{\lambda\gamma}{(\gamma|u| + 1)\log(\gamma + 1)}$

Table 1. The seven non-convex penalty functions and their supergradients.

$$\mathcal{L}(\mathbf{u}, \mathbf{v}) = -\mathbf{u}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{v} + \Omega_{\mathrm{nc}}(\mathbf{u}) + \Omega_{\mathrm{nc}}(\mathbf{v}) + \frac{\alpha_1}{2}(\|\mathbf{X}\mathbf{u}\|^2 - 1) + \frac{\alpha_2}{2}(\|\mathbf{Y}\mathbf{v}\|^2 - 1),$$

(4)

which is equivalent to

$$\mathcal{L}(\mathbf{u}, \mathbf{v}) = -\mathbf{u}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{v} + \Omega_{\mathrm{nc}}(\mathbf{u}) + \Omega_{\mathrm{nc}}(\mathbf{v}) + \frac{\alpha_1}{2}\|\mathbf{X}\mathbf{u}\|^2 + \frac{\alpha_2}{2}\|\mathbf{Y}\mathbf{v}\|^2$$

(5)

from the point of view of optimization. α_1, α_2, λ_1, λ_2 and γ are nonnegative tuning parameters. Next we will show how to solve this non-convex problem.

The first term $-\mathbf{u}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{v}$ on the right of equation (5) is biconvex in $\mathbf{u}$ and $\mathbf{v}$. $\|\mathbf{X}\mathbf{u}\|^2$ is convex in $\mathbf{u}$, and $\|\mathbf{Y}\mathbf{v}\|^2$ is convex in $\mathbf{v}$. It remains to approximate both $\Omega_{\mathrm{nc}}(\mathbf{u})$ and $\Omega_{\mathrm{nc}}(\mathbf{v})$ and transform them into convex ones.

The local quadratic approximation (LQA) technique was introduced to quadratically expresses the SCAD penalty[38]. Based on LQA, we here show how to represent these non-convex penalties in a unified way. First, we have the first-order Taylor expansion of $P_{\lambda_1,\gamma}(\sqrt{\mu})$ at μ_0 $P_{\lambda,\gamma}((\mu)^{1/2})$ at μ_0

$$P_{\lambda,\gamma}(\sqrt{\mu}) \approx P_{\lambda,\gamma}(\sqrt{\mu_0}) + P'_{\lambda,\gamma}(\sqrt{\mu_0})\frac{1}{2\sqrt{\mu_0}}(\mu - \mu_0),$$
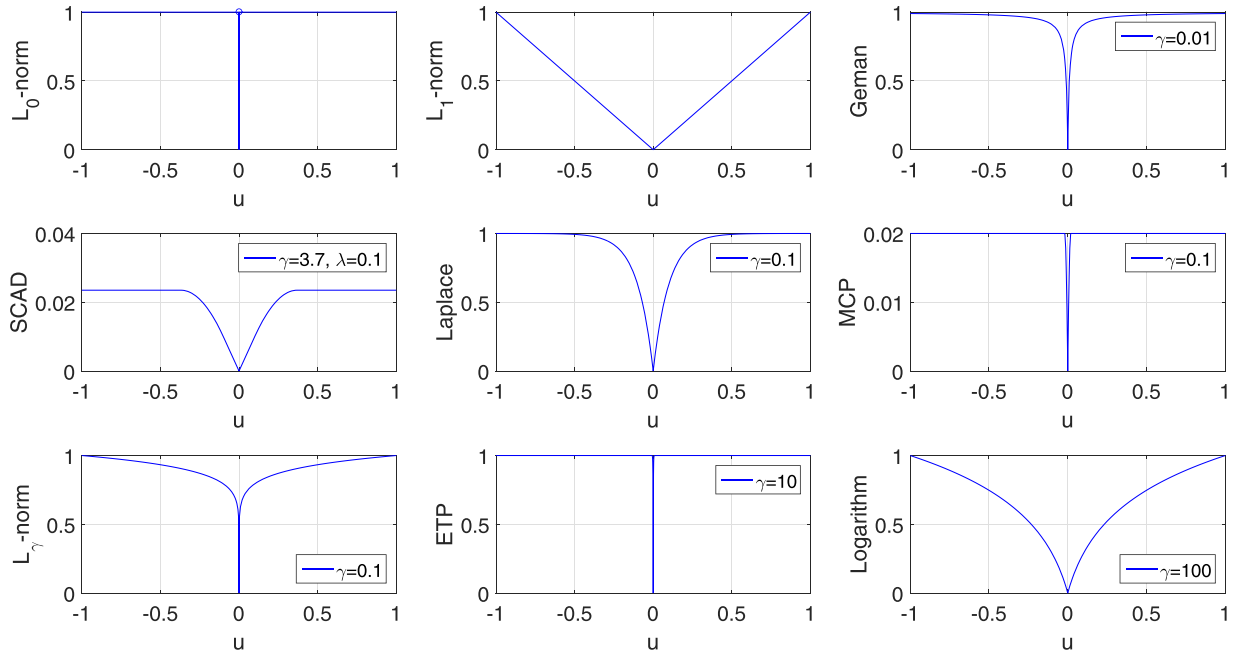
(6)

where μ_0 and μ are neighbors, e.g., the estimates at two successive iterations during optimization. Substituting $\mu = u_i^2$ and $\mu_0 = (u_i^t)^2$ into (6), we have

$$P_{\lambda,\gamma}(|u_i|) \approx P_{\lambda,\gamma}(|u_i^t|) + P'_{\lambda,\gamma}(|u_i^t|)\frac{1}{2|u_i^t|}(u_i^2 - (u_i^t)^2)$$

(7)

with $P'_{\lambda,\gamma}(|u_i^t|)$ being the supergradient of $P_{\lambda,\gamma}(|u_i^t|)$ (as shown in Table 1) at $|u_i^t|$.

Then we obtain a quadratic approximation to $\Omega_{\mathrm{nc}}(\mathbf{u})$:

$$\Omega_{\mathrm{nc}}(\mathbf{u}) = \sum_{i=1}^{p}P_{\lambda,\gamma}(|u_i|) \approx \sum_{i=1}^{p}\frac{P'_{\lambda,\gamma}(|u_i^t|)}{2|u_i^t|}u_i^2 + C_{\mathbf{u}},$$

(8)

Figure 1. Illustration of the ℓ_0, ℓ_1 and seven non-convex functions. All the non-convex penalty functions share two common properties: They are singular at origin, concave and monotonically decreasing on $(-\infty, 0]$, and concave and monotonically increasing on $[0, \infty)$.

where

$$C_{\mathbf{u}} = \sum_{i=1}^{p}\left[P_{\lambda,\gamma}(|u_i^t|) - \frac{1}{2}P'_{\lambda,\gamma}(|u_i^t|)|u_i^t|\right]$$

is not a function of $\mathbf{u}$ and thus will not contribute to the optimization.

In a similar way, we can construct a quadratic approximation to $\Omega_{nc}(\mathbf{v})$

$$\Omega_{nc}(\mathbf{v}) = \sum_{j=1}^{q}P_{\lambda,\gamma}(|v_j|) \approx \sum_{j=1}^{q}\frac{P'_{\lambda,\gamma}(|v_j^t|)}{2|v_j^t|}v_j^2 + C_{\mathbf{v}}, \tag{9}$$

where

$$C_{\mathbf{v}} = \sum_{j=1}^{q}\left[P_{\lambda,\gamma}(|v_j^t|) - \frac{1}{2}P'_{\lambda,\gamma}(|v_j^t|)|v_j^t|\right]$$

is not a function of $\mathbf{v}$ and makes no contribute towards the optimization.

Denote the estimates of $\mathbf{u}$ and $\mathbf{v}$ in the t-th iteration as $\mathbf{u}^t$ and $\mathbf{v}^t$, respectively. To update the estimates of $\mathbf{u}$ and $\mathbf{v}$ in the $(t+1)$-th iteration, we substitute the approximate functions of $\Omega_{nc}(\mathbf{u})$ and $\Omega_{nc}(\mathbf{v})$ in equations (8) and (9) into $\mathcal{L}(\mathbf{u}, \mathbf{v})$ in 5, and solve the resultant approximate version of the original problem:

$$\begin{aligned}
\arg\min \mathcal{L}(\mathbf{u}, \mathbf{v}) &= \arg\min \; -\mathbf{u}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{v} + \sum_{i=1}^{p}\frac{P'_{\lambda_1,\gamma}(|u_i^t|)}{2|u_i^t|}u_i^2 \\
&\quad + \sum_{j=1}^{q}\frac{P'_{\lambda_1,\gamma}(|v_j^t|)}{2|v_j^t|}v_j^2 + \frac{\alpha_1}{2}||\mathbf{X}\mathbf{u}||^2 + \frac{\alpha_2}{2}||\mathbf{Y}\mathbf{v}||^2
\end{aligned} \tag{10}$$

Obviously, the equation (10) is a quadratical expression, and is biconvex in $\mathbf{u}$ and $\mathbf{v}$. This means it is convex in terms of $\mathbf{u}$ given $\mathbf{v}$, and vice versa. Then according to the alternate convex search (ACS) method which is designed to solve biconvex problems[48], the $(t+1)$-th estimation of $\mathbf{u}$ and $\mathbf{v}$ can be calculated via

$$\begin{aligned}
\mathbf{u}^{t+1} &= \arg\min_{\mathbf{u}} -\mathbf{u}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{v}^t + \sum_{i=1}^{p}\frac{P'_{\lambda_1,\gamma}(|u_i^t|)}{2|u_i^t|}u_i^2 + \frac{\alpha_1}{2}||\mathbf{X}\mathbf{u}||^2, \\
\mathbf{v}^{t+1} &= \arg\min_{\mathbf{v}} -(\mathbf{u}^{t+1})^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{v} + \sum_{j=1}^{q}\frac{P'_{\lambda_2,\gamma}(|v_j^t|)}{2|v_j^t|}v_j^2 + \frac{\alpha_2}{2}||\mathbf{Y}\mathbf{v}||^2.
\end{aligned} \tag{11}$$
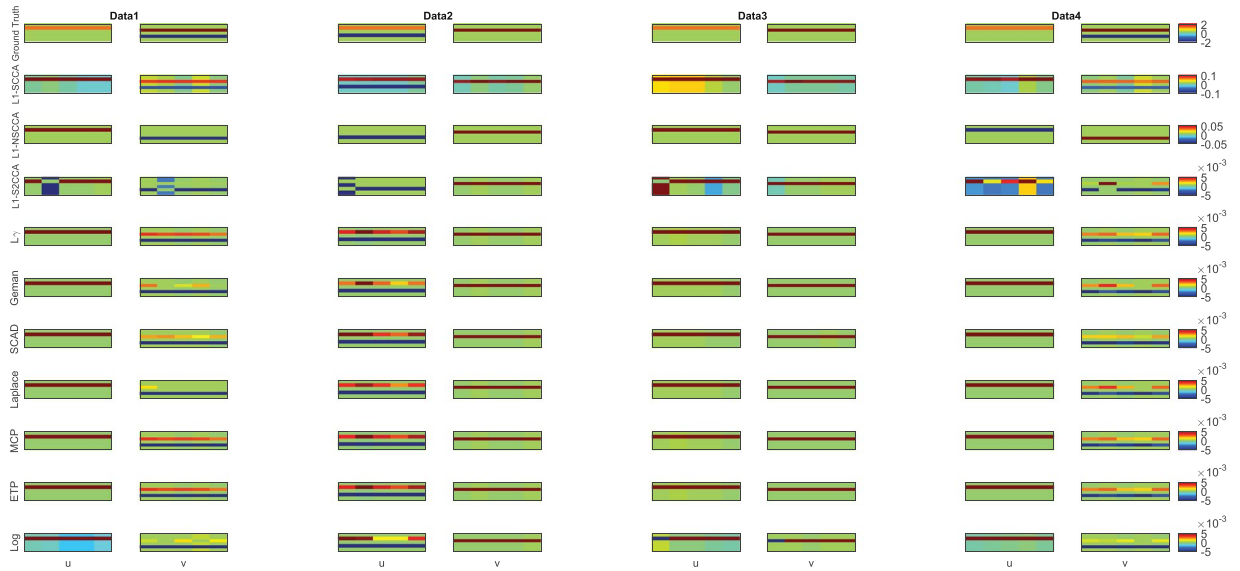
Figure 2. Canonical loadings estimated on four synthetic data sets. The first column shows results for Data1, and the second column is for Data2, and so forth. The first row is the ground truth, and each remaining one corresponds to an SCCA method: (1) Ground Truth. (2) L1-SCCA. (3) L1-NSCCA. (4) L1-S2CCA. (5) ℓ_γ-norm and so forth. For each data set and each method, the estimated weights of **u** is shown on the left panel, and **v** is on the right. In each individual heat map, the x-axis indicates the indices of elements in u or v; the y-axis indicates the indices of the cross-validation folds.

	HC	MCI	AD
Num	204	363	176
Gender(M/F)	111/93	235/128	95/81
Handedness(R/L)	190/14	329/34	166/10
Age(mean ± std)	76.07 ± 4.99	74.88 ± 7.37	75.60 ± 7.50
Education(mean ± std)	16.15 ± 2.73	15.72 ± 2.30	14.84 ± 3.12

Table 2. Participant characteristics.

Both equations above are quadratic, and thus their closed-form solutions exist. Taking the partial derivative of $\mathcal{L}(\mathbf{u}, \mathbf{v})$ in (5) with respect to **u** and **v** and setting the results to zero, we have

$$0 \in -\mathbf{X}^T\mathbf{Y}\mathbf{v} + (\mathbf{D}_1^t + \alpha_1\mathbf{X}^T\mathbf{X})\mathbf{u}, \tag{12}$$

$$0 \in -\mathbf{Y}^T\mathbf{X}\mathbf{u} + (\mathbf{D}_2^t + \alpha_2\mathbf{Y}^T\mathbf{Y})\mathbf{v}, \tag{13}$$

where $\mathbf{D}_1^t$ is a diagonal matrix with the i-th diagonal entry as $\frac{P'_{\lambda_1,\gamma}(|u_i^t|)}{|u_i^t|}$ ($i\in[1, p]$). It can be calculated by taking the partial derivative of equation (7) with respect to u_i. $\mathbf{D}_2^t$ is also a diagonal matrix with the j-th diagonal entry as $\frac{P'_{\lambda_1,\gamma}(|v_j^t|)}{|v_j^t|}$ ($j\in[1, q]$), and can be computed similarly. However, the i-th element of $\mathbf{D}_1^t$ does not exist if $u_i^t = 0$. According to perturbed version of LQA[50], we address this by adding a slightly perturbed term. Then the i-th element of $\mathbf{D}_1^t$ is

$$\mathbf{D}_1^t(i, i) = \frac{P'_{\lambda_1,\gamma}(|u_i|)}{|u_i| + \zeta} \tag{14}$$

where ζ is a tiny positive number. Hunter and Li[50] showed that this modification guarantees optimizing the equation (11). Then we have the updating expressions at the $(t+1)$-th iteration

$$\mathbf{u}^{t+1} = (\mathbf{D}_1^t + \alpha_1\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}\mathbf{v}^t, \tag{15}$$

$$\mathbf{v}^{t+1} = (\mathbf{D}_2^t + \alpha_2\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{X}\mathbf{u}^{t+1}. \tag{16}$$

	ℓ_γ-norm	SCAD	Geman, Laplace, MCP	ETP, Log
Range of γ	0.1, 0.2, 0.3	3.7	0.1, 0.01, 0.001	10, 100, 1000

Table 3. The searching range of optimal γ for each non-convex penalty.

We alternate between the above two equations to graduate refine the estimates for **u** and **v** until convergence. The pseudo code of the non-convex SCCA algorithm is described in Algorithm 1.

Algorithm 1 The Unified Non-convex SCCA Algorithm

Require:
  $\mathbf{X} \in \mathcal{R}^{n \times p}$, $\mathbf{Y} \in \mathcal{R}^{n \times p}$, pre-tuned $\lambda_1, \alpha_1, \lambda_2, \alpha_2$, and γ.
Ensure:
  Canonical loadings **u** and **v**.
 1: Initialize $t = 0$, $\mathbf{u}^t \in \mathbb{R}^{p \times 1}$, and $\mathbf{v}^t \in \mathbb{R}^{q \times 1}$;
 2: **while** not converged **do**
 3:    Calculate $\mathbf{D}_1$ with the i-th diagonal entry as $\frac{P'_{\lambda_1, \gamma}(|u_i^t|)}{|u_i^t|}$;
 4:    Calculate $\mathbf{u}^{t+1}$ according to equation (15);
 5:    Scale **u** so that $\|\mathbf{Xu}\|^2 = 1$;
 6:    Calculate $\mathbf{D}_2$ with the j-th diagonal entry as $\frac{P'_{\lambda_1, \gamma}(|v_j^t|)}{|v_j^t|}$;
 7:    Calculate $\mathbf{v}^{t+1}$ according to equation (16);
 8:    Scale **v** so that $\|\mathbf{Yv}\|^2 = 1$;
 9:    $t = t + 1$;
10: **end while**
11: **return** **u** and **v**.

Computational Analysis.   In Algorithm 1, Step 3 and Step 6 are linear in the dimension of **u** and **v**, and are easy to compute. Step 4 and Step 7 are the critical steps of proposed algorithm. Since we have closed-form updating expressions, they can be calculated via solving a system of linear equations with quadratic complexity which avoids computing the matrix inverse with cubic complexity. Step 5 and 8 are the re-scale step and very easy to calculate. Therefore, the whole algorithm is efficient.

Data Availability.   The synthetic data sets generated in this work are available from the corresponding authors' web sites, http://www.escience.cn/people/dulei/code.html and http://www.iu.edu/ shenlab/tools/ncscca/. The real data set is publicly available in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database repository, http://adni.loni.usc.edu.

Experiments and Datasets
Data Description.   *Synthetic Dataset.*   There are four data sets with sparse true signals for both **u** and **v**, i.e., only a small subset of features are nonzero. The number of features of both *u* and *v* are larger than the observations to simulate a *large-(p + q)-small-n* task. The generating process is as follows. We first generate **u** and **v** with most feature being zero. After that, the latent variable z is constructed from Gaussian distribution $N(\mathbf{0}, \mathbf{I}_{n \times n})$. Then we create the data **X** from $\mathbf{x}_i \sim N(z_i\mathbf{u}, \sum_x)$ and data $\mathbf{y}_i \sim N(z_i\mathbf{v}, \sum_y)$, where $(\sum_x)_{jk} = \exp(-|u_j - u_k|)$ and $(\sum_y)_{jk} = \exp(-|v_j - v_k|)$. The first three sets have 250 features for **u** and 600 ones for **v**, but they have different correlation coefficients. There are 500 features and 900 features in **u** and **v** respectively for the last data set. We show the true signal of every data set in Fig. 2 (top row).

Real Neuroimaging Genetics Dataset.   Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA) etc, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org. The study protocols were approved by the Institutional Review Boards of all participating centers (Northwestern Polytechnical University, Indiana University and ADNI (A complete list of ADNI sites is available at http://www.adni-info.org/)) and written informed consent was obtained from all participants or authorized representatives. All the analyses were performed on the de-identified ADNI data, and were determined by Indiana University Human Subjects Office as IU IRB Review Not Required.

The real neuroimaging genetics dataset were collected from 743 participants, and the details was presented in Table 2. There were 163 candidate SNP biomarkers from the AD-risk genes, e.g., *APOE*, in the genotyping data. The structural MRI scans were processed with voxel-based morphometry (VBM) in SPM8[51,52]. Briefly, scans were aligned to a T1-weighted template image, segmented into gray matter (GM), white matter (WM) and

	u				v			
	Data1	Data2	Data3	Data4	Data1	Data2	Data3	Data4
L1-SCCA	1.00±0.00	0.75±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.74±0.10	1.00±0.00	1.00±0.00
L1-S2CCA	1.00±0.00	0.38±0.00	1.00±0.00	1.00±0.00	0.75±0.00	0.75±0.00	1.00±0.00	0.75±0.00
L1-NSCCA	0.80±0.45	0.30±0.41	0.80±0.45	0.40±0.55	1.00±0.00	0.65±0.15	1.00±0.00	0.80±0.27
ℓ_γ-norm	1.00±0.00	0.75±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.76±0.04	1.00±0.00	1.00±0.00
Geman	1.00±0.00	0.75±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.74±0.01	1.00±0.00	1.00±0.00
SCAD	1.00±0.00	0.75±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.74±0.02	1.00±0.00	1.00±0.00
Laplace	1.00±0.00	0.75±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.75±0.02	1.00±0.00	1.00±0.00
MCP	1.00±0.00	0.75±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.76±0.04	1.00±0.00	1.00±0.00
ETP	1.00±0.00	0.75±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.76±0.04	1.00±0.00	1.00±0.00
Log	1.00±0.00	0.75±0.00	1.00±0.00	1.00±0.00	1.00±0.00	0.75±0.02	1.00±0.00	1.00±0.00

Table 4. Performance comparison on synthetic data sets. The AUC (area under the curve) values (mean ± std) of estimated canonical loadings **u** and **v**.

|	Training				Testing			
	data1	data2	data3	data4	data1	data2	data3	data4
L1-SCCA	0.65±0.03	0.83±0.03	0.65±0.05	0.66±0.04	0.59±0.14	0.82±0.05	0.59±0.25	0.62±0.08
L1-S2CCA	0.51±0.25	0.67±0.30	0.63±0.28	0.32±0.15	0.55±0.23	0.68±0.28	0.53±0.29	0.24±0.20
L1-NSCCA	0.62±0.04	0.80±0.01	0.75±0.01	0.65±0.02	0.61±0.17	0.80±0.04	0.73±0.13	0.65±0.10
ℓ_γ-norm	0.62±0.04	**0.83±0.01**	0.75±0.01	0.65±0.02	0.61±0.17	**0.84±0.02**	0.73±0.13	0.66±0.10
Geman	0.62±0.04	**0.83±0.01**	0.75±0.01	0.65±0.02	0.62±0.17	0.83±0.02	0.72±0.13	0.66±0.10
SCAD	0.62±0.04	**0.83±0.01**	0.75±0.01	0.65±0.03	0.61±0.17	**0.84±0.02**	0.73±0.13	0.66±0.10
Laplace	0.62±0.04	**0.83±0.01**	0.75±0.01	0.65±0.02	0.61±0.17	0.83±0.02	0.73±0.13	0.66±0.10
MCP	0.62±0.04	**0.83±0.01**	0.75±0.01	0.65±0.02	0.61±0.17	**0.84±0.02**	0.73±0.13	0.66±0.10
ETP	0.62±0.04	**0.83±0.01**	0.75±0.01	0.65±0.02	0.61±0.17	**0.84±0.02**	0.73±0.13	0.66±0.10
Log	**0.66±0.03**	**0.83±0.01**	**0.76±0.01**	**0.68±0.03**	**0.62±0.14**	0.83±0.03	**0.73±0.12**	**0.67±0.08**

Table 5. Training and testing correlation coefficients (mean ± std) of 5-fold cross-validation synthetic data sets. The best values are shown in boldface.

cerebrospinal fluid (CSF) maps, normalized to MNI space, and smoothed with an 8mm FWHM kernel. We sub-sampled the whole brain and generated 465 voxels spanning the whole brain ROIs. The regression technique was employed to remove the effects of the baseline age, gender, education, and handedness for these VBM measures. The aim of this study is to evaluate the correlation between the SNPs and the VBM measures, and further identify which SNPs and ROIs are associated.

Experimental Setup. *Benchmarks.* In this paper, we are mainly interested in whether these non-convex SCCA methods could enhance the performance of ℓ_1-SCCA method based on our motivation. It is reasonable to employ the ℓ_1-norm based methods in comparison. Therefore, the structure-aware SCCA methods such as[28,29,32,34] are not contained here as benchmark. Based on different mathematical techniques, there are three different ℓ_1-SCCA algorithms. They are the singular value decomposition based method[25], the primal-dual based method[29] and the LQA based method[32]. Though the latter two are proposed for capturing group or network structure, they can be easily reformulated to the ℓ_1-norm constrained methods, such as setting the parameters associated with the structure penalty to zero[29]. Therefore, to make the comparison fair and convincing, we choose all of them as benchmarks. With a slight abuse of notation, we use the penalty name to refer a non-convex SCCA method, e.g. ETP for ETP based SCCA method. For the ℓ_1-norm based methods, we call them L1-SCCA[25], L1-S2CCA[32], and L1-NSCCA[29].

Parameter Tuning. There are four parameters $\lambda_i (i = 1, 2)$ and $\alpha_i (i = 1, 2)$ associated with the non-convex SCCA methods, and one pivotal parameter γ. According to their equations, these non-convex penalties can approximate the ℓ_0-norm by providing an appropriate γ. In this situation, the λ_i and α_i play a very weak role because theoretically the ℓ_0-norm penalized problem does not rely on the parameters. Based on this consideration, we here only tune the γ other than tuning λ_i and α_i by a grid search strategy. This reduces the time consumption dramatically but does not affect the performance significantly. Further, we observe that two γ's perform similarly if they are not significantly different. Thus the tuning range of γ is not continuous. Besides, we set $\gamma = 3.7$ for SCAD penalty since[38] suggested that this is a very reasonable choice. The details of tuning range for each penalty are contained in Table 3. For λ_i and α_i, we simply set them to 1 in this study.

Termination Criterion. We use $\max_i |u_i^{t+1} - u_i^t| \leq \varepsilon$ and $\max_j |v_j^{t+1} - v_j^t| \leq \varepsilon$ as the termination condition for Algorithm 1, where ε is the user defined error bound. In this study, we set $\varepsilon = 10^{-5}$ according to experiments. All
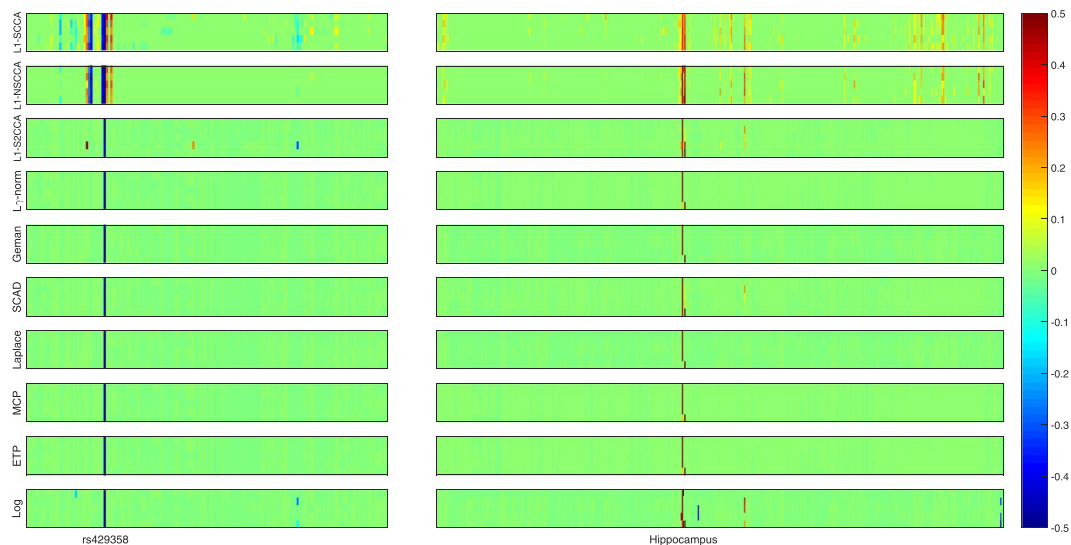
Figure 3. Canonical loadings estimated on real imaging genetics data. Each row corresponds to a SCCA method: (1) L1-SCCA, (2) L1-NSCCA, (3) L1-S2CCA, (4) ℓ_γ-norm and so forth. For each method, the estimated $\boldsymbol{u}$ is shown on the left panel, and $\boldsymbol{v}$ is on the right one. In each individual heat map, the x-axis indicates the indices of elements in u or v (i.e., SNPs or ROIs); the y-axis indicates the indices of the cross-validation folds.
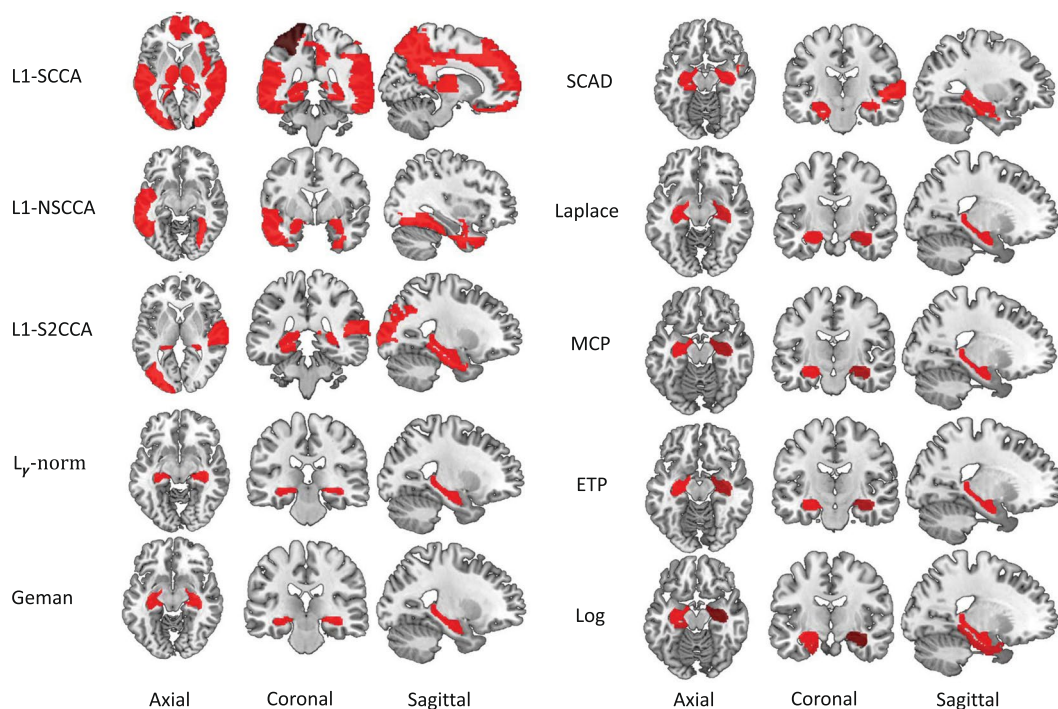
Figure 4. Mapping averaged canonical weight $\boldsymbol{v}$'s estimated by every SCCA method onto the brain. The left panel and right panel show five methods respectively, where each row corresponds to a SCCA method. The L1-SCCA identifies the most signals, followed by the L1-NSCCA and L1-S2CCA. All the proposed methods identify a clean signal that helps further investigation.

methods use the same setup, i.e., the same partition of the five-fold cross-validation, running on the same platform.

Results on Synthetic Data. Figure 2 shows the heat maps of canonical loadings estimated from all SCCA methods, where each row corresponds to an experimental method. We clearly observe that the non-convex SCCA methods and L1-SCCA correctly identify the identical signal positions to the ground truth across four data sets. Besides true signals, L1-SCCA introduces several undesired signals which makes it be inferior to our methods. As

	L1-SCCA	L1-S2CCA	L1-NSCCA	ℓ_γ-norm	Geman	SCAD	Laplace	MCP	ETP	Log
Training	0.27 ± 0.01	0.29 ± 0.02	0.27 ± 0.01	0.28 ± 0.02	0.27 ± 0.02	0.29 ± 0.02	0.27 ± 0.02	0.28 ± 0.02	0.28 ± 0.02	$\mathbf{0.33 \pm 0.03}$
Testing	0.18 ± 0.04	0.25 ± 0.10	0.22 ± 0.07	0.26 ± 0.09	0.26 ± 0.10	0.27 ± 0.09	0.26 ± 0.10	0.26 ± 0.09	0.26 ± 0.09	$\mathbf{0.27 \pm 0.11}$
Training-Testing Gap	0.09	0.04	0.05	0.02	0.01	0.02	0.01	0.02	0.02	0.06

Table 6. Performance comparison on real data set. Training and testing correlation coefficients (mean ± std) of 5-fold cross-validation are shown. The best value is shown in boldface.

a contrast, L1-NSCCA finds out an incomplete proportion of the ground truth, and L1-S2CCA performs unstably as it fails on some folds. Moreover, we also prioritize these methods using the AUC (area under ROC) criterion in Table 4, where a higher value indicates a better performance. The results exhibit that the non-convex SCCA methods have the highest score at almost every case. L1-SCCA scores similarly to the proposed methods, but later we can see it pays the price at a reduced prediction ability. Table 5 presents the estimated correlation coefficients on both training and testing data, where the best values are shown in boldface. The proposed SCCA methods alternatively gain the best value, and the Log method wins out for the most times. This demonstrates that the non-convex methods outperform ℓ_1-norm based SCCA methods in terms of the prediction power. In summary, the proposed methods identify accurate and sparse canonical loading patterns and obtain high correlation coefficients simultaneously, while those ℓ_1-norm based SCCA methods cannot.

Results on Real Neuroimaging Genetics Data. In this real data study, the genotyping data is denoted by **X**, and the imaging data is denoted by **Y**. The **u** is a vector of weights of all SNPs, and **v** is a vector of weights of all imaging markers. The canonical correlation coefficients are defined as Pearson correlation coefficient between **Xu** and **Yv**, i.e., $(\mathbf{Xu})^T \mathbf{Yv}/(\|\mathbf{Xu}\| \|\mathbf{Yv}\|)$.

Figure 3 presents the heat maps regrading the canonical loadings generated from the training set. In this figure, each row shows two weights of a SCCA method, where a larger weight stands for a more importance. The weight associated with the SNPs is on the left panel, and that associated with the voxels is on the right. The proposed non-convex SCCA methods obtain very clean and sparse weights for both **u** and **v**. The largest signal on the genetic side is the *APOE* e4 SNP rs429358, which has been previously reported to be related to AD[53]. On the right panel, the largest signal for all SCCA methods comes from the hippocampus region. This is one of the most notable biomarkers as an indicator of AD, since atrophy of hippocampus has been shown to be related to brain atrophy and neuron loss measured with MRI in AD cohort[53]. In addition, the L1-S2CCA and SCAD methods identify a weak signal from the parahippocampal gyrus, which is previously reported as an early biomarker of AD[54]. On some folds, the Log method also finds out the lingual region, parahippocampal gyrus, vermis region. Interestingly, all the three regions have shown to be correlated to AD, and could be further considered as an indicating biomarker that can be observed prior to a dementia diagnosis. For example, Sjöbeck and Englund reported that molecular layer gliosis and atrophy in the vermis are clearly severer in AD patients than in the health controls[55]. This is meaningful since the non-convex SCCA methods identify the correct clue for further investigation. On this account, both L1-SCCA and L1-NSCCA are not good choices since they identify too many signals, which may misguide subsequent investigation. The figure shows that L1-S2CCA could be an alternative choice for sparse imaging genetics analysis, but it performs unstably across the five folds. And, the non-convex methods is more consistent and stable than those ℓ_1-SCCA methods. To show the results more clearly, we map the canonical weights (averaged across 5 folds) regarding the imaging measurements from each SCCA method onto the brain in Fig. 4. The figure confirms that the L1-SCCA and L1-NSCCA find out many signals that are not sparse. The L1-S2CCA identifies fewer signals than both L1-SCCA and L1-NSCCA, but more than all these non-convex SCCA methods. All the non-convex SCCA only highlights a small region of the whole brain. This again reveals that the proposed methods have better canonical weights which reduces the effort of further investigation.

Besides, we include both training and testing correlation coefficients in Table 6, where their *mean* and *standard deviation* are shown. The training results of all methods are similar, with the Log method gains the highest value of 0.33 ± 0.03. As for the testing results, which is our primary interest, all the non-convex SCCA methods obtain better values than these ℓ_1-SCCA methods. Besides, the difference between the training and testing performance of the proposed methods is much smaller than that of three ℓ_1-SCCA methods. This means that the non-convex methods have better generalization performance as they are less likely to fall into overfitting issue. The result of this real imaging genetics data reveals that the proposed SCCA methods can extract more accurate and sparser canonical weights for both genetic and imaging biomarkers, and obtain higher correlation coefficients than those ℓ_1-SCCA methods.

Conclusion

We have proposed a unified non-convex SCCA model and an efficient optimization algorithm using a family of non-convex penalty functions. These penalties are concave and piecewise continuous, and thus piecewise differentiable. We approximate these non-convex penalties by an ℓ_2 function via the local quadratic approximation (LQA)[38]. Therefore, the proposed algorithm is effective and runs fast.

We compare the non-convex methods with three state-of-the-art ℓ_1-SCCA methods using both simulation data and real imaging genetics data. The simulation data have different ground truth structures. The results on the simulation data show that the non-convex SCCA methods identify cleaner and better canonical loadings than the three ℓ_1-SCCA methods, i.e. L1-SCCA[25], L1-S2CCA[32], and L1-NSCCA[29]. These non-convex methods also recover

higher correlation coefficients than ℓ_1-SCCA methods, demonstrating that ℓ_1-SCCA methods have suboptimal prediction capability as they may over penalize large coefficients. The results on the real data show that the proposed methods discover a pair of meaningful genetic and brain imaging biomarkers, while the ℓ_1-SCCA methods return too many irrelevant signals. The correlation coefficients show that the non-convex SCCA methods hold better testing values. This verifies our motivation that the non-convex penalty can improve the prediction ability, and thus has better generalization capability. Obviously, the parameter γ plays a key role in these non-convex penalties. In the future work, we will investigate how to choose a reasonable γ; and explore how to incorporate structure information into the model as structure information extraction is an important task for brain imaging genetics as well as biology studies.

References

1. Hibar, D. P., Kohannim, O., Stein, J. L., Chiang, M.-C. & Thompson, P. M. Multilocus genetic analysis of brain images. *Frontiers in Genetics* **2**, 73 (2011).
2. Hariri, A. R., Drabant, E. M. & Weinberger, D. R. Imaging genetics: perspectives from studies of genetically driven variation in serotonin function and corticolimbic affective processing. *Biological psychiatry* **59**, 888–897 (2006).
3. Viding, E., Williamson, D. E. & Hariri, A. R. Developmental imaging genetics: challenges and promises for translational research. *Development and Psychopathology* **18**, 877–892 (2006).
4. Mattay, V. S., Goldberg, T. E., Sambataro, F. & Weinberger, D. R. Neurobiology of cognitive aging: insights from imaging genetics. *Biological psychology* **79**, 9–22 (2008).
5. Bigos, K. L. & Weinberger, D. R. Imaging genetics - days of future past. *Neuroimage* **53**, 804–809 (2010).
6. Scharinger, C., Rabl, U., Sitte, H. H. & Pezawas, L. Imaging genetics of mood disorders. *Neuroimage* **53**, 810–821 (2010).
7. Potkin, S. G. *et al.* Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: methodological considerations. *Cognitive neuropsychiatry* **14**, 391–418 (2009).
8. Kim, S. *et al.* Influence of genetic variation on plasma protein levels in older adults using a multi-analyte panel. *PLoS One* **8**, e70269 (2013).
9. Shen, L. *et al.* 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**, 1051–63 (2010).
10. Winkler, A. M. *et al.* Cortical thickness or grey matter volume? the importance of selecting the phenotype for imaging genetics studies. *Neuroimage* **53**, 1135–1146 (2010).
11. Meda, S. A. *et al.* A large scale multivariate parallel ica method reveals novel imaging–genetic relationships for alzheimer's disease in the adni cohort. *Neuroimage* **60**, 1608–1621 (2012).
12. Nho, K. *et al.* Whole-exome sequencing and imaging genetics identify functional variants for rate of change in hippocampal volume in mild cognitive impairment. *Molecular psychiatry* **18**, 781 (2013).
13. Shen, L. *et al.* Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain imaging and behavior* **8**, 183–207 (2014).
14. Saykin, A. J. *et al.* Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. *Alzheimer's & Dementia* **11**, 792–814 (2015).
15. Wang, H. *et al.* Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. *Bioinformatics* **28**, 229–237 (2012).
16. Vounou, M., Nichols, T. E. & Montana, G. Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *NeuroImage* **53**, 1147–59 (2010).
17. Vounou, M. *et al.* Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in alzheimer's disease. *Neuroimage* **60**, 700–716 (2012).
18. Zhu, X., Suk, H.-I., Huang, H. & Shen, D. Structured sparse low-rank regression model for brain-wide and genome-wide associations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 344–352 (Springer, 2016).
19. Liu, J. *et al.* Combining fmri and snp data to investigate connections between brain function and genetics using parallel ica. *Human brain mapping* **30**, 241–255 (2009).
20. Geladi, P. & Kowalski, B. R. Partial least-squares regression: a tutorial. *Analytica chimica acta* **185**, 1–17 (1986).
21. Grellmann, C. *et al.* Comparison of variants of canonical correlation analysis and partial least squares for combined analysis of mri and genetic data. *NeuroImage* **107**, 289–310 (2015).
22. Hardoon, D., Szedmak, S. & Shawe-Taylor, J. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation* **16**, 2639–2664 (2004).
23. Hardoon, D. R. & Shawe-Taylor, J. Sparse canonical correlation analysis. *Machine Learning* **83**, 331–353 (2011).
24. Hotelling, H. Relations between two sets of variates. *Biometrika* **28**, 321–377 (1936).
25. Witten, D. M., Tibshirani, R. & Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10**, 515–34 (2009).
26. Witten, D. M. & Tibshirani, R. J. Extensions of sparse canonical correlation analysis with applications to genomic data. *Statistical applications in genetics and molecular biology* **8**, 1–27 (2009).
27. Parkhomenko, E., Tritchler, D. & Beyene, J. Sparse canonical correlation analysis with application to genomic data integration. *Statistical Applications in Genetics and Molecular Biology* **8**, 1–34 (2009).
28. Chen, X., Liu, H. & Carbonell, J. G. Structured sparse canonical correlation analysis. In *International Conference on Artificial Intelligence and Statistics*, 199–207 (2012).
29. Chen, X. & Liu, H. An efficient optimization algorithm for structured sparse cca, with applications to EQTL mapping. *Statistics in Biosciences* **4**, 3–26 (2012).
30. Chen, J. & Bushman, F. D. *et al.* Structure-constrained sparse canonical correlation analysis with an application to microbiome data analysis. *Biostatistics* **14**, 244–258 (2013).
31. Lin, D., Calhoun, V. D. & Wang, Y.-P. Correspondence between fMRI and SNP data by group sparse canonical correlation analysis. *Medical image analysis* **18**, 891–902 (2014).
32. Du, L. *et al.* A novel structure-aware sparse learning algorithm for brain imaging genetics. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, 329–336 (2014).
33. Yan, J. *et al.* Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm. *Bioinformatics* **30**, i564–i571 (2014).
34. Du, L. *et al.* Structured sparse canonical correlation analysis for brain imaging genetics: An improved graphnet method. *Bioinformatics* **32**, 1544–1551 (2016).
35. Du, L. *et al.* Sparse canonical correlation analysis via truncated l_1-norm-norm with application to brain imaging genetics. In *IEEE International Conference on Bioinformatics and Biomedicine*, 707–711 (IEEE, 2016).

36. Du, L. *et al*. Identifying associations between brain imaging phenotypes and genetic factors via a novel structured scca approach. In *International Conference on Information Processing in Medical Imaging*, 543–555 (Springer, 2017).

37. Meinshausen, N. & Bühlmann, P. High-dimensional graphs and variable selection with the lasso. *The annals of statistics* 1436–1462 (2006).

38. Fan, J. & Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360 (2001).

39. Zou, H. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418–1429 (2006).

40. Shen, X., Pan, W. & Zhu, Y. Likelihood-based selection and sharp parameter estimation. *Journal of the American Statistical Association* **107**, 223–232 (2012).

41. Fung, G. & Mangasarian, O. Equivalence of minimal l_0-and l_p-norm solutions of linear equalities, inequalities and linear programs for sufficiently small p. *Journal of optimization theory and applications* **151**, 1–10 (2011).

42. Frank, L. E. & Friedman, J. H. A statistical view of some chemometrics regression tools. *Technometrics* **35**, 109–135 (1993).

43. Geman, D. & Yang, C. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing* **4**, 932–946 (1995).

44. Trzasko, J. & Manduca, A. Highly undersampled magnetic resonance image reconstruction via homotopic l_1-minimization. *IEEE Transactions on Medical imaging* **28**, 106–121 (2009).

45. Zhang, C. Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics* **38**, 894–942 (2010).

46. Gao, C., Wang, N., Yu, Q. & Zhang, Z. A feasible nonconvex relaxation approach to feature selection. In *AAAI*, 356–361 (2011).

47. Friedman, J. H. Fast sparse regression and classification. *International Journal of Forecasting* **28**, 722–738 (2012).

48. Gorski, J., Pfeuffer, F. & Klamroth, K. Biconvex sets and optimization with biconvex functions: a survey and extensions. *Mathematical Methods of Operations Research* **66**, 373–407 (2007).

49. Lu, C., Tang, J., Yan, S. & Lin, Z. Generalized nonconvex nonsmooth low-rank minimization. In *IEEE Conference on Computer Vision and Pattern Recognition*, 4130–4137 (2014).

50. Hunter, D. R. & Li, R. Variable selection using mm algorithms. *Annals of statistics* **33**, 1617 (2005).

51. Ashburner, J. & Friston, K. J. Voxel-based morphometry–the methods. *Neuroimage* **11**, 805–21 (2000).

52. Risacher, S. L. & Saykin, A. J. *et al*. Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Current Alzheimer Research* **6**, 347–61 (2009).

53. Hampel, H. *et al*. Core candidate neurochemical and imaging biomarkers of alzheimer's disease. *Alzheimer's & Dementia* **4**, 38–48 (2008).

54. Echavarri, C. *et al*. Atrophy in the parahippocampal gyrus as an early biomarker of alzheimer's disease. *Brain Structure and Function* **215**, 265–271 (2011).

55. Sjöbeck, M. & Englund, E. Alzheimer's disease and the cerebellum: a morphologic study on neuronal and glial changes. *Dementia and geriatric cognitive disorders* **12**, 211–218 (2001).

Acknowledgements

Author Contributions

L.D., L.G. and L.S. conceived and designed the research. L.D., K.L. and J.H. carried out the study analysis. X.Y., J.Y, S.L.R. and A.J.S. collected the data from ADNI database. L.D., K.L., L.S. and A.J.S. analyzed the results and wrote the paper. 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.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Consortia
Alzheimer's Disease Neuroimaging Initiative

Michael W. Weiner[3], Paul Aisen[4], Ronald Petersen[5], Clifford R. Jack[5], William Jagust[6], John Q. Trojanowki[7], Arthur W. Toga[4], Laurel Beckett[8], Robert C. Green[9], John Morris[10], Leslie M. Shaw[7], Zaven Khachaturian[11], Greg Sorensen[12], Maria Carrillo[13], Lew Kuller[14], Marc Raichle[10], Steven Paul[15], Peter Davies[16], Howard Fillit[17], Franz Hefti[18], David Holtzman[10], M. Marcel Mesulam[19], William Potter[20], Peter Snyder[21], Adam Schwartz[22], Tom Montine[23], Ronald G. Thomas[24], Michael Donohue[24], Sarah Walter[24], Devon Gessert[24], Tamie Sather[24], Gus Jiminez[24], Archana B. Balasubramanian[24], Jennifer Mason[24], Iris Sim[24], Danielle Harvey[8], Matthew Bernstein[5], Nick Fox[25], Paul Thompson[26], Norbert Schuff[3], Charles DeCArli[8], Bret Borowski[5], Jeff Gunter[5], Matt Senjem[5], Prashanthi Vemuri[5], David Jones[5], Kejal Kantarci[5], Chad Ward[5], Robert A. Koeppe[27], Norm Foster[28], Eric M. Reiman[29], Kewei Chen[29], Chet Mathis[14], Susan Landau[6], Nigel J. Cairns[10], Erin Franklin[10], Lisa Taylor-Reinwald[10], Virginia Lee[7], Magdalena Korecka[7], Michal Figurski[7], Karen Crawford[4], Scott Neu[4], Tatiana M. Foroud[2], Steven Potkin[30], Kelley Faber[2], Sungeun Kim[2], Kwangsik Nho[2], Leon Thal[24], Neil Buckholtz[31], Marilyn Albert[32], Richard Frank[33], John Hsiao[31], Jeffrey Kaye[34], Joseph Quinn[34], Lisa Silbert[34], Betty Lind[34], Raina Carter[34], Sara Dolen[34], Lon S. Schneider[4], Sonia Pawluczyk[4], Mauricio Beccera[4], Liberty Teodoro[4], Bryan M. Spann[4], James Brewer[24], Helen Vanderswag[24], Adam Fleisher[24], Judith L. Heidebrink[27], Joanne L. Lord[27], Sara S. Mason[5], Colleen S. Albers[5], David Knopman[5], Kris Johnson[5], Rachelle S. Doody[35], Javier Villanueva-Meyer[35], Valory Pavlik[35], Victoria Shibley[35], Munir Chowdhury[35], Susan Rountree[35], Mimi Dang[35], Yaakov Stern[36], Lawrence S. Honig[36], Karen L. Bell[36], Beau Ances[10], Maria Carroll[10], Mary L. Creech[10], Erin Franklin[10], Mark A. Mintun[10], Stacy Schneider[10], Angela Oliver[10], Daniel Marson[37], David Geldmacher[37], Marissa Natelson Love[37], Randall Griffith[37], David Clark[37], John Brockington[37], Erik Roberson[37], Hillel Grossman[38], Effie Mitsis[38], Raj C. Shah[39], Leyla deToledo-Morrell[39], Ranjan Duara[40], Maria T. Greig-Custo[40], Warren Barker[40], Chiadi Onyike[32], Daniel D'Agostino[32], Stephanie Kielb[32], Martin Sadowski[41], Mohammed O. Sheikh[41], Anaztasia Ulysse[41], Mrunalini Gaikwad[41], P. Murali Doraiswamy[42], Jeffrey R. Petrella[42], Salvador Borges-Neto[42], Terence Z. Wong[42], Edward Coleman[42], Steven E. Arnold[7], Jason H. Karlawish[7], David A. Wolk[7], Christopher M. Clark[7], Charles D. Smith[43], Greg Jicha[43], Peter Hardy[43], Partha Sinha[43], Elizabeth Oates[43], Gary Conrad[43], Oscar L. Lopez[14], Mary Ann Oakley[14], Donna M. Simpson[14], Anton P. Porsteinsson[44], Bonnie S. Goldstein[44], Kim Martin[44], Kelly M. Makino[44], M. Saleem Ismail[44], Connie Brand[44], Adrian Preda[30], Dana Nguyen[30], Kyle Womack[45], Dana Mathews[45], Mary Quiceno[45], Allan I. Levey[46], James J. Lah[46], Janet S. Cellar[46], Jeffrey M. Burns[47], Russell H. Swerdlow[47], William M. Brooks[47], Liana Apostolova[26], Kathleen Tingus[26], Ellen Woo[26], Daniel H. S. Silverman[26], Po H. Lu[26], George Bartzokis[26], Neill R Graff-Radford[48], Francine Parfitt[48], Kim Poki-Walker[48], Martin R. Farlow[2], Ann Marie Hake[2], Brandy R. Matthews[2], Jared R. Brosch[2], Scott Herring[2], Christopher H. van Dyck[49], Richard E. Carson[49], Martha G. MacAvoy[49], Pradeep Varma[49], Howard Chertkow[50], Howard Bergman[50], Chris Hosein[50], Sandra Black[51], Bojana Stefanovic[51], Curtis Caldwell[51], Ging-Yuek Robin Hsiung[52], Benita Mudge[52], Vesna Sossi[52], Howard Feldman[52], Michele Assaly[52], Elizabeth Finger[53], Stephen Pasternack[53], Irina Rachisky[53], John Rogers[53], Dick Trost[53], Andrew Kertesz[53], Charles Bernick[54], Donna Munic[54], Emily Rogalski[19], Kristine Lipowski[19], Sandra Weintraub[19], Borna Bonakdarpour[19], Diana Kerwin[19], Chuang-Kuo Wu[19], Nancy Johnson[19], Carl Sadowsky[55], Teresa Villena[55], Raymond Scott Turner[56], Kathleen Johnson[56], Brigid Reynolds[56], Reisa A. Sperling[9], Keith A. Johnson[9], Gad Marshall[9], Jerome Yesavage[57], Joy L. Taylor[57], Barton Lane[57], Allyson Rosen[57], Jared Tinklenberg[57], Marwan N. Sabbagh[58], Christine M. Belden[58], Sandra A. Jacobson[58], Sherye A. Sirrel[58], Neil Kowall[59], Ronald Killiany[59], Andrew E. Budson[59], Alexander Norbash[59], Patricia Lynn Johnson[59], Thomas O. Obisesan[60], Saba Wolday[60], Joanne Allard[60], Alan Lerner[61], Paula Ogrocki[61], Curtis Tatsuoka[61], Parianne Fatica[61], Evan Fletcher[8], Pauline Maillard[8], John Olichney[8], Charles DeCarli[8], Owen Carmichael[8], Smita Kittur[62], Michael Borrie[63], T.-Y. Lee[63], Rob Bartha[63], Sterling Johnson[64], Sanjay Asthana[64], Cynthia M. Carlsson[64], Pierre Tariot[29], Anna Burke[29], Ann Marie Milliken[29], Nadira Trncic[29], Adam Fleisher[29], Stephanie Reeder[29], Vernice Bates[65], Horacio Capote[65], Michelle Rainka[65], Douglas W. Scharre[66], Maria Kataki[66], Brendan Kelly[66], Earl A. Zimmerman[67], Dzintra Celmins[67], Alice D. Brown[67], Godfrey D. Pearlson[68], Karen Blank[68], Karen Anderson[68], Laura A. Flashman[69], Marc Seltzer[69], Mary L. Hynes[69], Robert B. Santulli[69],

Kaycee M. Sink[70], Leslie Gordineer[70], Jeff D. Williamson[70], Pradeep Garg[70], Franklin Watkins[70], Brian R. Ott[71], Geoffrey Tremont[71], Lori A. Daiello[71], Stephen Salloway[72], Paul Malloy[72], Stephen Correia[72], Howard J. Rosen[3], Bruce L. Miller[3], David Perry[3], Jacobo Mintzer[73], Kenneth Spicer[73], David Bachman[73], Nunzio Pomara[74], Raymundo Hernando[74], Antero Sarrael[74], Susan K. Schultz[75], Karen Ekstam Smith[75], Hristina Koleva[75], Ki Won Nam[75], Hyungsub Shim[75], Norman Relkin[15], Gloria Chaing[15], Michael Lin[15], Lisa Ravdin[15], Amanda Smith[76], Balebail Ashok Raj[76] & Kristin Fargher[76]

[3]University of California, San Francisco, USA. [4]University of Southern California, Los Angeles, USA. [5]Mayo Clinic, Rochester, Minnesota, USA. [6]University of California, Berkeley, Berkeley, USA. [7]University of Pennsylvania, Philadelphia, USA. [8]University of California, Davis, Davis, USA. [9]Brigham and Women's Hospital/Harvard Medical School, Boston, USA. [10]Washington University St. Louis, St. Louis, USA. [11]Prevent Alzheimer's Disease, 2020, Rockville, USA. [12]Siemens, Munich, Germany. [13]Alzheimer's Association, Illinois, USA. [14]University of Pittsburgh, Pennsylvania, USA. [15]Cornell University, New York, USA. [16]Albert Einstein College of Medicine of Yeshiva University, New York, USA. [17]AD Drug Discovery Foundation, New York, USA. [18]Acumen Pharmaceuticals, California, USA. [19]Northwestern University, Illinois, USA. [20]National Institute of Mental Health, Maryland, USA. [21]Brown University, Rhode Island, USA. [22]Eli Lilly, Indiana, USA. [23]University of Washington, Washington, USA. [24]University of California, San Diego, California, USA. [25]University of London, London, UK. [26]University of California, Los Angeles, California, USA. [27]University of Michigan, Michigan, USA. [28]University of Utah, Utah, USA. [29]Banner Alzheimer's Institute, Arizona, USA. [30]University of California, Irvine, California, USA. [31]National Institute on Aging, Maryland, USA. [32]Johns Hopkins University, Maryland, USA. [33]Richard Frank Consulting, New Hampshire, USA. [34]Oregon Health and Science University, Oregon, USA. [35]Baylor College of Medicine, Texas, USA. [36]Columbia University Medical Center, New York, USA. [37]University of Alabama-Birmingham, Alabama, USA. [38]Mount Sinai School of Medicine, New York, USA. [39]Rush University Medical Center, Rush University, Illinois, USA. [40]Wien Center, Florida, USA. [41]New York University, New York, USA. [42]Duke University Medical Center, North Carolina, USA. [43]University of Kentucky, Kentucky, USA. [44]University of Rochester Medical Center, New York, USA. [45]University of Texas Southwestern Medical School, Texas, USA. [46]Emory University, Georgia, USA. [47]University of Kansas, Medical Center, Kansas, USA. [48]Mayo Clinic, Jacksonville, Florida, USA. [49]Yale University School of Medicine, Connecticut, USA. [50]McGill University, Montreal-Jewish General Hospital, Quebec, Canada. [51]Sunnybrook Health Sciences, Ontario, Canada. [52]U.B.C. Clinic for AD & Related Disorders, British Columbia, Canada. [53]Cognitive Neurology-St. Joseph's, Ontario, Canada. [54]Cleveland Clinic Lou Ruvo Center for Brain Health, Ohio, USA. [55]Premiere Research Inst (Palm Beach Neurology), Florida, USA. [56]Georgetown University Medical Center, Washington D.C, USA. [57]Stanford University, California, USA. [58]Banner Sun Health Research Institute, Arizona, USA. [59]Boston University, Massachusetts, USA. [60]Howard University, Washington D.C, USA. [61]Case Western Reserve University, Ohio, USA. [62]Neurological Care of CNY, New York, USA. [63]Parkwood Hospital, Pennsylvania, USA. [64]University of Wisconsin, Wisconsin, USA. [65]Dent Neurologic Institute, New York, USA. [66]Ohio State University, Ohio, USA. [67]Albany Medical College, New York, USA. [68]Hartford Hospital, Olin Neuropsychiatry Research Center, Connecticut, USA. [69]Dartmouth-Hitchcock Medical Center, New Hampshire, USA. [70]Wake Forest University Health Sciences, North Carolina, USA. [71]Rhode Island Hospital, Rhode Island, USA. [72]Butler Hospital, Rhode Island, USA. [73]Medical University South Carolina, Carolina, USA. [74]Nathan Kline Institute, New York, USA. [75]University of Iowa College of Medicine, Iowa, USA. [76]USF Health Byrd Alzheimer's Institute, University of South Florida, Florida, USA.