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

Lei Du1, Kefei Liu2, Xiaohui Yao3, Jingwen Yan2, Shannon L. Risacher2, Junwei Han1, Lei Guo1, Andrew J. Saykin2, Li Shen2 & 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 $\ell_1$-norm onto the individual feature or the structure level of features to pursue corresponding sparsity. However, the $\ell_1$-norm penalty over-penalizes large coefficients and may incur 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 brain1–14. 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 regression15, reduced rank regression16–18, parallel independent component analysis19, partial least squares regression20,21, canonical correlation analysis (CCA)22 and their sparsity-inducing variants23, 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 decades24. 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 $\ell_1$-norm based SCCA method25 has gained great success for its sparsity pursuing capability. After that, there are many SCCA variants based on the $\ell_1$-norm. For examples, the fused lasso penalty imposes the $\ell_1$-norm onto the ordered pairwise features29, and the group lasso penalty imposes the $\ell_1$-norm onto the group of features29,30. Further, the graph lasso or the graph guided lasso can be viewed as imposing the $\ell_1$-norm onto the pairwise features defined by an undirected graph26.

However, the $\ell_1$-norm penalty shows the conflict of optimal prediction and consistent feature selection37. In penalized least squares modeling, Fan and Li38 showed that a good penalty function should meet three properties.

1School of Automation, Northwestern Polytechnical University, Xi’an, 710072, China. 2Radiology 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 $\ell_0$-norm penalty is successful in feature selection because it is singular at the origin. On the contrary, the $\ell_1$-norm penalty over-penalizes large coefficients, and thus it may be suboptimal with respect to the estimation risk\(^{39,40}\). The $\ell_0$-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 $\ell_0$-norm constrained problem is NP-hard\(^{41}\).

A number of non-convex penalties are proposed as the surrogate of the $\ell_0$-norm to handle this issue. These penalties includes the $\ell_1$-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 unpunished. 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 $\ell_1$-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 $\|u\|$ denotes the Euclidean norm of a vector $u$.

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

$$\min_{u,v} - u^T X^T Y v$$

subject to

$$\|X u\|_2^2 \leq 1, \quad \|Y v\|_2^2 \leq 1, \quad \Omega(u) \leq c_1, \quad \Omega(v) \leq c_2,$$

where $X u$ and $Y v$ are the canonical variables, $u$ and $v$ 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 $\ell_1$-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 the $\ell_1$-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

$$\Omega_{\lambda, \gamma}(u) = \sum_{i=1}^{k} P_{\lambda, \gamma}(|u_i|),$$

where $\lambda$ and $\gamma$ are nonnegative parameters, and $P_{\lambda, \gamma}(|u_i|)$ is a non-convex function. We absorb $\lambda$ 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 $\ell_1$-norm constraints in the SCCA model, we define the unified non-convex SCCA model as follows

$$\min_{u,v} - u^T X^T Y v + \Omega_{\lambda}(u) + \Omega_{\gamma}(v)$$

subject to

$$\|X u\|_2^2 \leq 1, \quad \|Y v\|_2^2 \leq 1,$$

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

To solve the non-convex SCCA problem, we use the Lagrangian method,
Table 1. The seven non-convex penalty functions and their supergradients.

| Penalty Name | Function (P_\lambda,\gamma(u)) | Supergradient (P^*\lambda,\gamma(u)) |
|--------------|-------------------------------|-----------------------------------|
| \ell_1 norm | \lambda | \infty, u = 0, \lambda\gamma|u||^{-1}, |u| > 0. |
| German | \frac{\lambda |u|}{|u| + \gamma} | \lambda^2 \frac{|u|}{|u| + \gamma^2} |
| SCAD | \lambda | \frac{1}{2} | \sqrt{1 + \frac{2|u\gamma - |u||}{\lambda \gamma}} |
| Laplace | \lambda | \frac{1}{2} \exp(\frac{-|u|}{\gamma}) |
| MCP | \lambda | \frac{1}{2} \exp(\frac{-|u|}{\gamma}) |
| ETP | \lambda | \frac{1}{2} \exp(\frac{-|u|}{\gamma}) |
| Logarithm | \lambda | \frac{1}{2} \exp(\frac{-|u|}{\gamma}) |

The local quadratic approximation (LQA) technique was introduced to quadratically expresses the SCAD penalty. 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 \( P_{\lambda,\gamma}(\frac{\gamma}{\gamma'}) \) at \( \mu_0 P_{\lambda,\gamma}(\mu) \) at \( \mu_0 

\[ P_{\lambda,\gamma}(\frac{\gamma}{\gamma'}) \approx P_{\lambda,\gamma}(\mu_0) + P'_{\lambda,\gamma}(\mu_0) \left( \frac{\gamma}{\gamma'} - 1 \right) \]

which is equivalent to

\[ P_{\lambda,\gamma}(\frac{\gamma}{\gamma'}) \approx P_{\lambda,\gamma}(\mu_0) + \frac{\gamma}{\gamma'} \left( \frac{\gamma}{\gamma'} - 1 \right) \]

from the point of view of optimization. \( \alpha_1, \alpha_2, \lambda_1, \lambda_2 \) and \( \gamma \) are nonnegative tuning parameters. Next we will show how to solve this non-convex problem.

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

The local quadratic approximation (LQA) technique was introduced to quadratically expresses the SCAD penalty. 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 \( P_{\lambda,\gamma}(\frac{\gamma}{\gamma'}) \) at \( \mu_0 P_{\lambda,\gamma}(\mu) \) at \( \mu_0 

\[ P_{\lambda,\gamma}(\frac{\gamma}{\gamma'}) \approx P_{\lambda,\gamma}(\mu_0) + P'_{\lambda,\gamma}(\mu_0) \left( \frac{\gamma}{\gamma'} - 1 \right) \]

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

\[ P_{\lambda,\gamma}(|u_i|) \approx P_{\lambda,\gamma}(|u_i|) + P'_{\lambda,\gamma}(|u_i|) \left( \frac{1}{2 |u_i|} u_i^2 - u_i^2 \right) \]

with \( P'_{\lambda,\gamma}(|u_i|) \) being the supergradient of \( P_{\lambda,\gamma}(|u_i|) \) (as shown in Table 1) at \( |u_i| \).

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

\[ \Omega_{nc}(u) = \sum_{i=1}^p P_{\lambda,\gamma}(|u_i|) \approx \sum_{i=1}^p \frac{P'_{\lambda,\gamma}(|u_i|)}{2 |u_i|} u_i^2 + C_w, \]
where

$$C_u = \sum_{j=1}^{q} \lambda_j \gamma_j (|u_j|) - \frac{1}{2} \sum_{i=1}^{p} \lambda_i \gamma_i (|u_i|) |u_i|$$

is not a function of $u$ and thus will not contribute to the optimization.

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

$$\Omega_{nc}(v) = \sum_{j=1}^{q} \lambda_j \gamma_j (|v_j|) \approx \sum_{j=1}^{q} \lambda_j \gamma_j (|v_j|) \frac{1}{2} v_j^2 + C_v,$n

(9)

where

$$C_v = \sum_{j=1}^{q} \lambda_j \gamma_j (|v_j|) - \frac{1}{2} \sum_{i=1}^{p} \lambda_i \gamma_i (|v_i|) |v_i|$$

is not a function of $v$ and makes no contribution to the optimization.

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

$$\arg \min \mathcal{L}(u, v) = \arg \min \left[ -u^T X^T Y v + \frac{p}{2} \sum_{i=1}^{p} \lambda_i (|u_i|) u_i^2 \right. $$

$$+ \left. \sum_{j=1}^{q} \frac{\lambda_j \gamma_j (|v_j|)}{2} v_j^2 + \frac{\alpha_1}{2} |Xu|^2 + \frac{\alpha_2}{2} |Yv|^2 \right]$$

(10)

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

$$u^{t+1} = \arg \min_u \left[ -u^T X^T Y v^t + \frac{p}{2} \sum_{i=1}^{p} \lambda_i (|u_i|) u_i^2 \right. $$

$$+ \left. \sum_{j=1}^{q} \frac{\lambda_j \gamma_j (|v_j|)}{2} v_j^2 + \frac{\alpha_1}{2} |Xu|^2 \right]$$

(11)
Both equations above are quadratic, and thus their closed-form solutions exist. Taking the partial derivative \( \mathcal{L}(\mathbf{u}, \mathbf{v}) \) in (5) with respect to \( \mathbf{u} \) and \( \mathbf{v} \) and setting the results to zero, we have
\[
0 \in -X^T Y \mathbf{v} + (D_1' + \alpha_1 X^TX) \mathbf{u},
\]
\[
0 \in -Y^T X \mathbf{u} + (D_2' + \alpha_2 Y^TY) \mathbf{v},
\]
where \( D_1' \) is a diagonal matrix with the \( i \)-th diagonal entry as \( \frac{p'_{\lambda g_c}(\|u_i\|)}{|u_i|} \), and \( D_2' \) is also a diagonal matrix with the \( j \)-th diagonal entry as \( \frac{p'_{\lambda g_c}(\|v_j\|)}{|v_j|} \). It can be calculated by taking the partial derivative of equation (7) with respect to \( u_i \). However, the \( i \)-th element of \( D_1' \) does not exist if \( u_{i0} = 0 \).

According to perturbed version of LQA 50, we address this by adding a slightly perturbed term. Then the \( i \)-th element of \( D_1' \) is
\[
D_1'(i, i) = \frac{p'_{\lambda g_c}(\|u_i\|)}{|u_i|} + \zeta,
\]
where \( \zeta \) 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
\[
u^{t+1} = (D_2' + \alpha_2 Y^TY)^{-1} Y^TX \mathbf{u}^{t},
\]
\[
u^{t+1} = (D_2' + \alpha_2 Y^TY)^{-1} Y^TX \mathbf{u}^{t+1}.
\]
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:**
- \( X \in \mathbb{R}^{n \times p} \), \( Y \in \mathbb{R}^{n \times q} \), pre-tuned \( \lambda_1, \alpha_1, \lambda_2, \alpha_2 \), and \( \gamma \).

**Ensure:**
- Canonical loadings \( u \) and \( v \).

1. Initialize \( t = 0 \), \( u^t \in \mathbb{R}^{p \times 1} \), and \( v^t \in \mathbb{R}^{q \times 1} \);
2. while not converged do
   3. Calculate \( \mathbf{D}_1 \) with the \( i \)-th diagonal entry as \( \frac{P_{\text{SCCA}}[\|u_i\|]}{\|u_i\|} \);
   4. Calculate \( u^{t+1} \) according to equation (15);
   5. Scale \( u \) so that \( \|Xu\|^2 = 1 \);
   6. Calculate \( \mathbf{D}_2 \) with the \( j \)-th diagonal entry as \( \frac{P_{\text{SCCA}}[\|v_j\|]}{\|v_j\|} \);
   7. Calculate \( v^{t+1} \) according to equation (16);
   8. Scale \( v \) so that \( \|Yv\|^2 = 1 \);
   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 \( \mathcal{N}(0, I_{n \times n}) \). Then we create the data \( X \) from \( x_i \sim N(z_i, \sum_{z_i}) \) and data \( y_j \sim N(z_j, \sum_{z_j}) \), where \( \sum_{z_i} = \exp(-|u_i - u_i|) \) and \( \sum_{z_j} = \exp(-|v_j - v_j|) \). 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 [1,2]. Briefly, scans were aligned to a T1-weighted template image, segmented into gray matter (GM), white matter (WM) and

| \( \beta \)-norm | SCAD | Geman, Laplace, MCP | ETP, Log |
|-----------------|------|---------------------|---------|
| Range of \( \gamma \) | 0.1, 0.2, 0.3 | 3.7 | 0.1, 0.01, 0.001 | 10, 100, 1000 |

**Table 3.** The searching range of optimal \( \gamma \) for each non-convex penalty.
Table 4. Performance comparison on synthetic data sets. The AUC (area under the curve) values (mean ± std) of estimated canonical loadings \( \mathbf{u} \) and \( \mathbf{v} \).

|        | Training |        |        |
|--------|----------|--------|--------|
|        | data1    | data2  | data3  |
| u      |          |        |        |
| L1-SCCA| 0.65 ± 0.03 | 0.83 ± 0.03 | 0.65 ± 0.05 |
| L1-S2CCA| 0.51 ± 0.25 | 0.67 ± 0.30 | 0.63 ± 0.28 |
| L1-NSCCA| 0.62 ± 0.04 | 0.80 ± 0.01 | 0.75 ± 0.01 |
| \( \ell_1 \)-norm | 0.62 ± 0.04 | 0.83 ± 0.01 | 0.75 ± 0.01 |

|        | data4    |        |        |
|        |          |        |        |
| v      |          |        |        |
| L1-SCCA| 0.66 ± 0.14 | 0.82 ± 0.05 | 0.59 ± 0.14 |
| L1-S2CCA| 0.53 ± 0.29 | 0.68 ± 0.28 | 0.55 ± 0.23 |
| L1-NSCCA| 0.73 ± 0.13 | 0.84 ± 0.17 | 0.84 ± 0.17 |
| \( \ell_1 \)-norm | 0.73 ± 0.13 | 0.84 ± 0.17 | 0.84 ± 0.17 |

|        | Testing |        |        |
|--------|---------|--------|--------|
|        | data1   | data2  | data3  |
| u      |          |        |        |
| L1-SCCA| 0.62 ± 0.05 | 0.59 ± 0.14 | 0.82 ± 0.05 |
| L1-S2CCA| 0.53 ± 0.29 | 0.68 ± 0.28 | 0.55 ± 0.23 |
| L1-NSCCA| 0.73 ± 0.13 | 0.84 ± 0.17 | 0.84 ± 0.17 |
| \( \ell_1 \)-norm | 0.73 ± 0.13 | 0.84 ± 0.17 | 0.84 ± 0.17 |

|        | data4    |        |        |
|        |          |        |        |
| v      |          |        |        |
| L1-SCCA| 0.62 ± 0.05 | 0.59 ± 0.14 | 0.82 ± 0.05 |
| L1-S2CCA| 0.53 ± 0.29 | 0.68 ± 0.28 | 0.55 ± 0.23 |
| L1-NSCCA| 0.73 ± 0.13 | 0.84 ± 0.17 | 0.84 ± 0.17 |
| \( \ell_1 \)-norm | 0.73 ± 0.13 | 0.84 ± 0.17 | 0.84 ± 0.17 |

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

|        | Training |        |        |
|--------|----------|--------|--------|
|        | data1    | data2  | data3  |
| \( \ell_1 \)-norm | 0.66 ± 0.03 | 0.83 ± 0.01 | 0.76 ± 0.01 |

|        |          |        |        |
|        | data4    |        |        |
| \( \ell_1 \)-norm | 0.67 ± 0.08 | 0.75 ± 0.00 | 0.75 ± 0.00 |

Experimental Setup. Benchmarks. In this paper, we are mainly interested in whether these non-convex SCCA methods could enhance the performance of \( \ell_2 \)-SCCA method based on our motivation. It is reasonable to employ the \( \ell_1 \)-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 \( \ell_1 \)-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 \( \ell_1 \)-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 \( \ell_1 \)-norm based methods, we call them L1-SCCA\(^{25}\), L1-S2CCA\(^{32}\), and L1-NSCCA\(^{32}\).

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 \( \gamma \). According to their equations, these non-convex penalties can approximate the \( \ell_0 \)-norm by providing an appropriate \( \gamma \). In this situation, the \( \lambda_i \) and \( \alpha_i \) play a very weak role because theoretically the \( \ell_1 \)-norm penalized problem does not rely on the parameters. Based on this consideration, we here only tune the \( \gamma \) other than tuning \( \lambda_i \) and \( \alpha_i \) by a grid search strategy. This reduces the time consumption dramatically but does not affect the performance significantly. Further, we observe that two \( \gamma \)'s perform similarly if they are not significantly different. Thus the tuning range of \( \gamma \) 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 \( \lambda_i \) and \( \alpha_i \), 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 \( \varepsilon \) is the user defined error bound. In this study, we set \( \varepsilon = 10^{-5} \) according to experiments. All
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
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.

| Method          | Training | Testing | Training-Testing Gap |
|-----------------|----------|---------|----------------------|
| L1-SCCA         | 0.27 ± 0.01 | 0.25 ± 0.10 | 0.09 ± 0.02 |
| L1-S2CCA        | 0.29 ± 0.02 | 0.26 ± 0.09 | 0.03 ± 0.02 |
| L1-NSCCA        | 0.27 ± 0.01 | 0.26 ± 0.10 | 0.01 ± 0.02 |
| L1-norm         | 0.27 ± 0.02 | 0.27 ± 0.09 | 0.02 ± 0.02 |
| Geman           | 0.29 ± 0.02 | 0.26 ± 0.10 | 0.01 ± 0.02 |
| SCAD            | 0.27 ± 0.02 | 0.26 ± 0.10 | 0.02 ± 0.02 |
| Laplace         | 0.28 ± 0.02 | 0.26 ± 0.09 | 0.02 ± 0.02 |
| MCP             | 0.28 ± 0.02 | 0.26 ± 0.09 | 0.02 ± 0.02 |
| ETP             | 0.33 ± 0.03 | 0.27 ± 0.11 | 0.06 ± 0.06 |
| Log             | 0.33 ± 0.03 | 0.27 ± 0.11 | 0.06 ± 0.06 |

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., \( \frac{\langle Xu \rangle \langle Yv \rangle}{\|Xu\| \|Yv\|} \).

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\(^{33}\). 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\(^{36}\). 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\(^{34}\).

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\(^{35}\). 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 \( \ell_1 \)-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 methods 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 \textit{mean} and \textit{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 \( \ell_1 \)-SCCA methods. Besides, the difference between the training and testing performance of the proposed methods is much smaller than that of these \( \ell_1 \)-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 \( \ell_1 \)-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 \( \ell_q \)-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 \( \ell_1 \)-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., \( \frac{\langle Xu \rangle \langle Yv \rangle}{\|Xu\| \|Yv\|} \).

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\(^{33}\). 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\(^{36}\). 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\(^{34}\).

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\(^{35}\). 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 \( \ell_1 \)-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 methods 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 \textit{mean} and \textit{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 \( \ell_1 \)-SCCA methods. Besides, the difference between the training and testing performance of the proposed methods is much smaller than that of these \( \ell_1 \)-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 \( \ell_1 \)-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 \( \ell_q \)-function via the local quadratic approximation (LQA)\(^{30}\). Therefore, the proposed algorithm is effective and runs fast.

We compare the non-convex methods with three state-of-the-art \( \ell_1 \)-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 \( \ell_1 \)-SCCA methods, i.e. L1-SCCA\(^{26}\), L1-S2CCA\(^{27}\), and L1-NSCCA\(^{29}\). These non-convex methods also recover
higher correlation coefficients than $\ell_1$-SCCA methods, demonstrating that $\ell_1$-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 $\ell_1$-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 $\gamma$ plays a key role in these non-convex penalties. In the future work, we will investigate how to choose a reasonable $\gamma$ 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. Scharring, C., Rahl, U., Sitte, H. H. & Pertzaw, 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. Sapkin, 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 multistack 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. I. 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. Transciprome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm. *Bioinformatics* 30, 1564–1571 (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 $\ell_1$-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 $l_2$ 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, I. 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 minmax 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
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; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. L. Du was supported by the National Natural Science Foundation of China (61602384); the Natural Science Basic Research Plan in Shaanxi Province of China (2017JQ6001); the China Postdoctoral Science Foundation (2017M613202); and the Fundamental Research Funds for the Central Universities (3102016OQD0065) at Northwestern Polytechnical University. This work was also supported by the National Institutes of Health R01 EB022574, R01 LM011360, U01 AG024904, P30 AG10133, R01 AG19771, U11 TR001108, R01 AG 042437, R01 AG046171, R01 AG040770; the Department of Defense W81XWH-14-2-0151, W81XWH-13-1-0259, W81XWH-12-2-0012; the National Collegiate Athletic Association 14132004 at Indiana University.

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.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Kaycee M. Sink70, Leslie Gordineer70, Jeff D. Williamson70, Pradeep Garg70, Franklin Watkins70, Brian R. Ott71, Geoffrey Tremont72, Lori A. Daiello73, Stephen Salloway72, Paul Malloy72, Stephen Correia72, Howard J. Rosen9, Bruce L. Miller3, David Perry3, Jacobo Mintzer73, Kenneth Spicer73, David Bachman73, Nunzio Pomara75, Raymundo Hernando74, Antero Sarrael9, Susan K. Schultz75, Karen Ekstam Smith75, Hristina Koleva75, Ki Won Nam75, Hyungsub Shim75, Norman Relkin15, Gloria Chaing15, Michael Lin15, Lisa Ravdin15, Amanda Smith15, Balebail Ashok Raj76 & Kristin Fargher76

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