

METHODOLOGY

Open Access



Identification of multimodal brain imaging association via a parameter decomposition based sparse multi-view canonical correlation analysis method

Jin Zhang, Huiai Wang, Ying Zhao, Lei Guo, Lei Du*  and Alzheimer's Disease Neuroimaging Initiative

From International Conference on Intelligent Biology and Medicine (ICIBM 2021)
Philadelphia, PA, USA. 8-10 August 2021

*Correspondence:
dulei@nwpu.edu.cn
School of Automation,
Northwestern Polytechnical
University, Xi'an, China

Abstract

Background: With the development of noninvasive imaging technology, collecting different imaging measurements of the same brain has become more and more easy. These multimodal imaging data carry complementary information of the same brain, with both specific and shared information being intertwined. Within these multimodal data, it is essential to discriminate the specific information from the shared information since it is of benefit to comprehensively characterize brain diseases. While most existing methods are unqualified, in this paper, we propose a parameter decomposition based sparse multi-view canonical correlation analysis (PDSMCCA) method. PDSMCCA could identify both modality-shared and -specific information of multimodal data, leading to an in-depth understanding of complex pathology of brain disease.

Results: Compared with the SMCCA method, our method obtains higher correlation coefficients and better canonical weights on both synthetic data and real neuroimaging data. This indicates that, coupled with modality-shared and -specific feature selection, PDSMCCA improves the multi-view association identification and shows meaningful feature selection capability with desirable interpretation.

Conclusions: The novel PDSMCCA confirms that the parameter decomposition is a suitable strategy to identify both modality-shared and -specific imaging features. The multimodal association and the diverse information of multimodal imaging data enable us to better understand the brain disease such as Alzheimer's disease.

Keywords: Sparse learning, Multi-view canonical correlation analysis, Parameter decomposition



Background

Alzheimer's Disease (AD) [1–5], the most common type of dementia, is a terrible neurodegenerative but its pathology is still unclear. And with the advance of imaging technologies, we can obtain multimodal imaging data of brain structure and function easily [6]. For example, the structural changes of the brain can be measured by structural magnetic resonance imaging (sMRI) scans, and the positron emission tomography (PET) scans can capture the brain activities such as the metabolic rate of glucose (FDG-PET) and amyloid depositions (AV45-PET) [7–10]. These different types of imaging data, including both modality-shared and -specific information, are collected simultaneously. As a result, it is essential to discriminate the modality-specific information from the modality-shared information, which could enable a better understanding of multimodal data and prompt reasonable multimodal brain imaging data integration [11–17].

The statistical pairwise correlation analysis has been widely used for medical image analysis. For example, researchers use both PET and functional magnetic resonance imaging (fMRI) data to study the relationship between brain and genes metabolism indicators [18]. With the deepening of research, researchers begin to use machine learning instead to focus on prediction tasks. However, they ignore the complex relationships in multimodal data. In contrast, exploring the correlation between multimodal brain imaging helps to reveal the pathogenesis of AD, thereby promotes the advancement of early diagnosis technology of the disease and the development of pharmaceutical research.

The existing correlation methods are mostly designed for two views [19, 20]. For instance, sparse canonical correlation analysis (SCCA) [21–27] has been widely used in brain imaging analysis. However, they cannot analyze multimodal imaging in a unified model. Although the multi-step strategy can be used to analyze the pairwise association between multiple modalities [2], it will inevitably cause the loss of potentially effective information. Thus these methods are sub-optimal. In order to analyze more than two modalities, SCCA can be directly and simply extended to multi-view paradigm [28] which has gained a lot of attentions. For example, based on sparse multiple/multi-view/multi-set canonical correlation analysis (SMCCA) [28, 29], researchers explored the association between multi-view data sets such as brain imaging data, genetic data and cognitive scores [30, 31]. However, SMCCA suffers from serious shortcomings which limit its application. First and foremost, SMCCA employs the ℓ_1 -norm, and thus cannot clearly report the modality-shared and -specific imaging features due to its overlook of the shared features of multiple modalities. In addition, the independent assumption of the in-set covariance of SMCCA makes the Pearson correlation coefficient break the range of $[-1, 1]$, and there is no measure to avoid the additional risk caused by this assumption. According to [27], this independent assumption may not guarantee the convergence and consistency. Therefore, SMCCA is insufficient and inadequate in multimodal brain imaging analysis problem.

With above observations, to better identify the complex multi-way correlations among multimodal imaging data, we propose a novel sparse multi-view canonical correlation analysis (PDSMCCA) method based on the parameter decomposition. On the one hand, to improve interpretability, PDSMCCA contains two types of regularization (ℓ_1 -norm and $\ell_{2,1}$ -norm). The ℓ_1 -norm penalizes each imaging feature of each modality separately [32], and $\ell_{2,1}$ -norm penalizes imaging features of multiple modalities jointly to obtain

the modality-shared features [33, 34]. Using ℓ_1 -norm and $\ell_{2,1}$ -norm together could offer a diverse feature selection. On the other hand, PDSMCCA decomposes the canonical weight into view-shared and -private components, which correspond to the modality-shared and -specific imaging features respectively. Owing to the decomposition strategy, PDSMCCA is able to obtain flexible imaging features. In addition, we relax the independent assumption of traditional SMCCA which treats the in-set covariance $\mathbf{X}^\top \mathbf{X}$ to be an identity [23]. Moreover, we introduce an efficient algorithm to solve the PDSMCCA model which converges to a local optimum. The results on synthetic data and real neuroimaging data show that, compared with the SMCCA method, our method obtains better or comparable canonical correlation coefficients (CCCs) and canonical weights. This indicates that our method is a powerful tool for multimodal brain imaging data association identification with diverse and desirable feature selection.

The contents of this article are arranged as follows. First, the SMCCA method is briefly introduced. Then, we describe the PDSMCCA in detail. Furthermore, we present the iterative optimization algorithm and prove its convergence, which is followed by experiments and results. Finally, the discussion and conclusion are provided.

Experimental results

We use synthetic data and real data to evaluate the performance of our method and employ the state-of-the-art method (SMCCA) as the benchmark method. The experiment adopts the nested fivefold cross-validation and the grid search strategy to tune suitable λ_B and λ_S , and the candidate parameter set is [0.01, 0.1, 1, 10, 100] which makes an appropriate feature selection since too large parameters and too small ones could incur undesirable features of interest. Besides, all methods are terminated when $\max |(\mathbf{b}_k + \mathbf{s}_k)^{\ell+1} - (\mathbf{b}_k + \mathbf{s}_k)^\ell| \leq 10^{-5}$ is met. The canonical correlation coefficient (CCC) and the feature selection (heatmap) are utilized as the evaluation criteria. The CCC is defined as

$$\text{CCC} = \frac{\mathbf{v}_i^\top \mathbf{X}_i^\top \mathbf{X}_j \mathbf{v}_j}{\sqrt{\mathbf{v}_i^\top \mathbf{X}_i^\top \mathbf{X}_i \mathbf{v}_i} \sqrt{\mathbf{v}_j^\top \mathbf{X}_j^\top \mathbf{X}_j \mathbf{v}_j}}, \quad (1)$$

where \mathbf{X} assumed to have been centered (zero mean), and $\mathbf{v} = \mathbf{b} + \mathbf{s}$. For CCC, a larger score indicates a better performance of identifying the bi-associations among multiple modalities.

Results on synthetic data

In this simulation study, we use two synthetic data sets which contain different ground truth and noise intensity. We first generate three canonical weight vectors $\mathbf{v}_j \in \mathbb{R}^{200 \times 1}$ and a latent vector μ with unit norm. The data matrix \mathbf{X}_k is generated by $(\mathbf{x}_{i,j})_k \sim N(\mu_i \mathbf{v}_{j,k}, e \cdot \mathbf{I}_{200 \times 200})$, where e denotes the noise level.

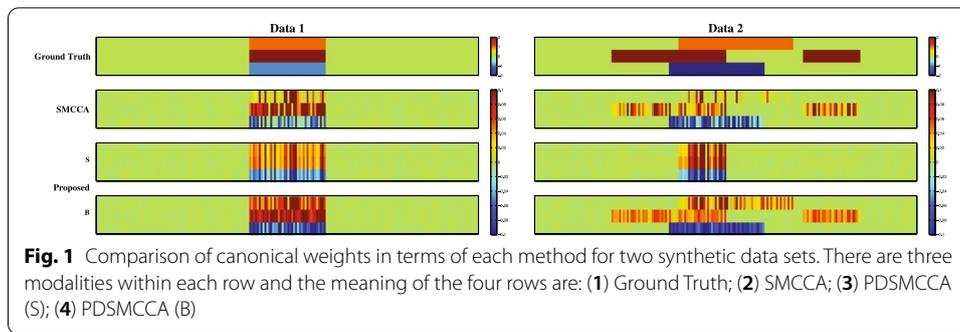


Table 1 CCCs (mean ± SD) comparison on synthetic data

	Training CCCs		Testing CCCs	
	SMCCA	PDSMCCA	SMCCA	PDSMCCA
Data1				
$X_1 - X_2$	0.97 ± 0.00	0.98 ± 0.00	0.96 ± 0.03	0.98 ± 0.01
$X_1 - X_3$	0.95 ± 0.01	0.98 ± 0.00	0.93 ± 0.01	0.97 ± 0.01
$X_2 - X_3$	0.97 ± 0.00	0.98 ± 0.00	0.96 ± 0.01	0.98 ± 0.01
Data2				
$X_1 - X_2$	0.92 ± 0.03	0.98 ± 0.00	0.81 ± 0.09	0.97 ± 0.01
$X_1 - X_3$	0.91 ± 0.03	0.98 ± 0.00	0.82 ± 0.09	0.98 ± 0.00
$X_2 - X_3$	0.99 ± 0.03	0.99 ± 0.00	0.98 ± 0.00	0.99 ± 0.00

The highest values are shown in bold

- Data 1: $n = 120$, $\mathbf{v}_1 = (0, \dots, 0, \underbrace{1, \dots, 1}_{80}, \underbrace{0, \dots, 0}_{40}, \underbrace{0, \dots, 0}_{80})^\top$, $\mathbf{v}_2 = (0, \dots, 0, \underbrace{2, \dots, 2}_{80}, \underbrace{2, \dots, 2}_{40}, \underbrace{0, \dots, 0}_{80})^\top$,
 $\mathbf{v}_3 = (0, \dots, 0, \underbrace{-1, \dots, -1}_{80}, \underbrace{0, \dots, 0}_{40}, \underbrace{0, \dots, 0}_{80})^\top$.
- Data 2: $n = 120$, $\mathbf{v}_1 = (0, \dots, 0, \underbrace{1, \dots, 1}_{75}, \underbrace{0, \dots, 0}_{60}, \underbrace{0, \dots, 0}_{65})^\top$, $\mathbf{v}_2 = (0, \dots, 0, \underbrace{2, \dots, 2}_{40}, \underbrace{2, \dots, 2}_{60}, \underbrace{2, \dots, 2}_{40}, \underbrace{2, \dots, 2}_{30}, \underbrace{0, \dots, 0}_{30})^\top$,
 $\mathbf{v}_3 = (0, \dots, 0, \underbrace{-2, \dots, -2}_{70}, \underbrace{-2, \dots, -2}_{50}, \underbrace{0, \dots, 0}_{80})^\top$.

In summary, we construct simulation data under different conditions to compare the proposed algorithm with the benchmark method.

Figure 1 shows the feature selection of the two methods on both synthetic data. It is worth noting that the intensity of the color reflects the relative importance of features. On the first data which only contains modality-shared features, both PDSMCCA and SMCCA can successfully identify these shared features. On the second data where both modality-shared and -specific features exist, SMCCA mixes these two types of features which is undesirable. On the contrary, PDSMCCA yields two types of features, including the modality-shared and -specific ones, which is more meaningful and practical. Table 1 presents the estimated canonical correlation coefficients between every two modalities. PDSMCCA obtains higher CCCs than SMCCA on both training and testing sets for two data sets. Therefore, PDSMCCA outperforms SMCCA in this simulation study.

Results on real data

The brain imaging data were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (<https://adni.loni.usc.edu>). and the primary goal of ADNI is 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 AD. For up-to-date information, see www.adni-info.org.

There were 755 samples including 281 ADs, 292 MCIs and 182 normal controls (NCs) non-Hispanic Caucasian participants. Three modalities of brain imaging data, including sMRI, FDG-PET and AV45-PET were used in this paper. FDG-PET and AV45-PET scans were co-registered to the standard MNI space. sMRI scans were processed with voxel-based morphometry (VBM) [35, 36] by the SPM software, and aligned to a T1-weighted template, then segmented to white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) maps, finally normalized to the same MNI space, and smoothed with an 8 mm³ FWHM kernel. According to the automated anatomical labeling (AAL) atlas, we obtained 116 regions of interest (ROI) measurements. In order to eliminate the influence of baseline age, gender, habit, and education level, we used regression weights obtained from NC subjects to pre-adjust these imaging QTs. We aim to improve the interpretability of multimodal data for complex pathogenesis mechanisms, as well as select imaging QTs of interest.

Figure 2 shows the feature selection results on real neuroimaging data. According to the intensity of the color, we can determine the relative importance of features. It is clear that PDSMCCA identifies more diverse imaging QTs than SMCCA. For the modality-shared features conveyed by **S**, PDSMCCA identifies the left and right hippocampus [4, 37], the left and right middle temporal [38], the left and right precuneus as the most relevant shared ROIs. Besides, PDSMCCA also identifies the modality-specific features which is shown in weight **B**. It is clear that the left and right medial orbitofrontal [9] are relevant only in AV45 scans [20, 39]. Meanwhile, the left post cingulum is relevant in FDG scans, and both the left and right hippocampus are relevant in sMRI scans. In contrast, SMCCA misses the brain regions shared by multiple modalities, since it cannot

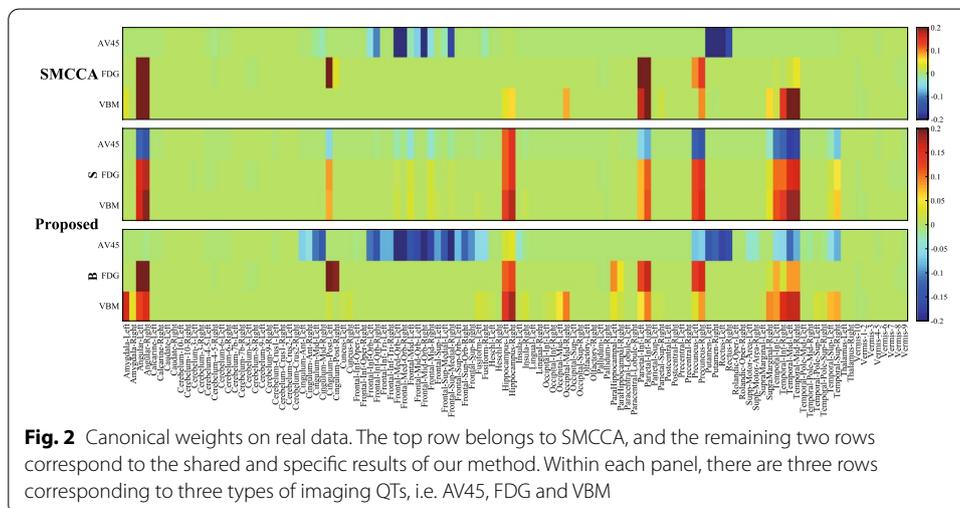


Fig. 2 Canonical weights on real data. The top row belongs to SMCCA, and the remaining two rows correspond to the shared and specific results of our method. Within each panel, there are three rows corresponding to three types of imaging QTs, i.e. AV45, FDG and VBM

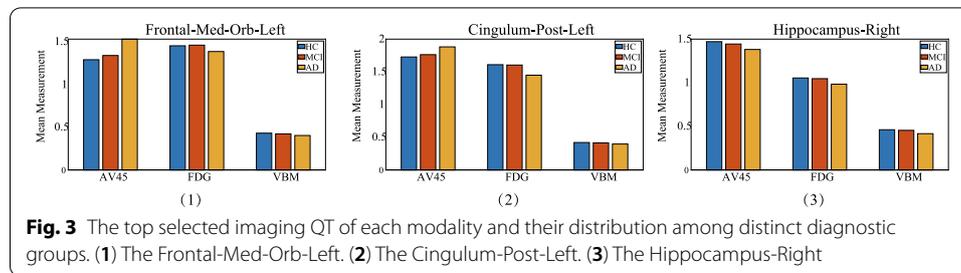


Table 2 CCCs (mean \pm SD) estimated between three types of imaging QTs

	AV45-FDG	AV45-VBM	FDG-VBM
Training			
SMCCA	0.33 \pm 0.01	0.28 \pm 0.02	0.50 \pm 0.02
PDSMCCA	0.35 \pm 0.01	0.31 \pm 0.01	0.49 \pm 0.01
Testing			
SMCCA	0.32 \pm 0.01	0.24 \pm 0.01	0.49 \pm 0.02
PDSMCCA	0.33 \pm 0.01	0.28 \pm 0.01	0.48 \pm 0.01

The highest values are shown in bold

obtain the diverse feature selection results. It mixes both modality-shared features and modality-specific ones which is insufficient in real applications. We also present the CCCs of both methods in Table 2. Our method obtains better CCCs than SMCCA, which indicates that our method can identify stronger bi-multivariate associations. In summary, PDSMCCA holds the capability to identify the multi-way correlations between multiple modalities of data, and can identify more meaningful features.

To further show the meaning of these selected imaging QTs, the ANOVA and population stratification analysis were conducted. The one-way ANOVA results showed that the top selected imaging QTs reached the level of significance ($p < 0.01$). This indicated imaging QTs were significantly related to the diagnosis. Moreover, in order to verify the biological effects of the selected imaging QTs. We further analyzed the prominent imaging QT of each modality, which were Frontal-Med-Orb-Left in AV45 [40], Cingulum-Post-Left [41] in FDG and Hippocampus-Right [37] in VBM. Since there were three diagnostic groups, we decided to investigate whether they were significantly different among different groups. Figure 3 showed that Frontal-Med-Orb-Left and Cingulum-Post-Left exhibited significant changes in FDG and AV45 which was consistent with the decline of metabolic rates of cerebral glucose and the variety of extracellular amyloid deposition. Besides, the Hippocampus-Right showed consistent patterns that decreased measurement were observed in all modalities. This may be attributed to its high correlation to AD. In summary, benefiting from the parameter decomposition strategy, our proposed method can obtain interesting and meaningful biomarkers in multimodal brain imaging analysis.

Discussion

Generally, different techniques yield different measurements of the same brain, and could carry shared and specific information simultaneously. In this paper, PDSMCCA is proposed to explore the multi-way relationship among multiple brain imaging modalities, and it can identify both modality-shared and -specific imaging features through the parameter decomposition technology. Importantly, this decomposition technology is flexible via balancing between two contradictory constraints (ℓ_1 -norm and $\ell_{2,1}$ -norm), and thus assures a better performance [42]. This improves the interpretability of traditional SMCCA method. Of note, similar to SMCCA, PDSMCCA is also unsupervised which could be a limitation. The future work is to incorporate the diagnostic labels into the PDSMCCA model, and build a supervised method to better mine the brain imaging association with selecting relevant imaging features.

Conclusion

To improve the interpretability of multimodal data for complex pathogenesis mechanisms, we proposed a novel sparse multi-view canonical correlation analysis method (PDSMCCA) based on parameter decomposition. In our model, the canonical weights were decomposed into modality-shared and modality-specific components, resulting in a flexible and meaningful interpretability. We also introduced an efficient optimization algorithm to solve PDSMCCA, and proved the convergence. The results on both synthetic and real neuroimaging data showed that compared with SMCCA, PDSMCCA accurately selected the modality-shared and -specific features, and obtained higher or comparable correlation coefficients. The diverse feature selection might provide a new insight for revealing AD pathology.

Method

In this paper, italic letters indicate scalars, boldface lowercase letters and boldface capitals represents column vectors and matrices respectively. Specifically, the i -th row and j -th column of \mathbf{V} is denoted as \mathbf{v}^i and \mathbf{v}_j . $\|\mathbf{V}\|_{2,1} = \sum_i \|\mathbf{v}^i\|_2$ is the $\ell_{2,1}$ -norm. In addition, $\|\mathbf{V}\|_{1,1}$ denotes the element-wise ℓ_1 -norm of \mathbf{V} , i.e., $\|\mathbf{V}\|_{1,1} = \sum_j \|\mathbf{v}_j\|_1 = \sum_i \|\mathbf{v}^i\|_1 = \sum_i \sum_j |v_{ij}|$.

SMCCA

SMCCA extends the conventional two-view SCCA model to multi-view oriented, which can handle the association identification among multiple data sets. Generally, the definition of SMCCA is as follows:

$$\begin{aligned} \min_{\mathbf{v}_k} \sum_{1 \leq k, j \leq K} & \left(-\mathbf{v}_k^T \mathbf{X}_k^T \mathbf{X}_j \mathbf{v}_j + \lambda \|\mathbf{v}_k\|_1 \right) \\ \text{s.t. } & \|\mathbf{v}_k\|_2^2 = 1 \quad (k = 1, \dots, K). \end{aligned} \quad (2)$$

According to [43, 44], (2) can be rewritten as a multivariate multiple regression model.

$$\begin{aligned} \min_{\mathbf{v}_k} \sum_{1 \leq k, j \leq K} & \left(\frac{1}{2} \|\mathbf{X}_k \mathbf{v}_k - \mathbf{X}_j \mathbf{v}_j\|_2^2 + \lambda \|\mathbf{v}_k\|_1 \right) \\ \text{s.t. } & \|\mathbf{v}_k\|_2^2 = 1 \quad (k = 1, \dots, K), \end{aligned} \tag{3}$$

where $\mathbf{X}_k \in \mathbb{R}^{n \times p}$ ($k = 1, \dots, K$) represents the k -th modality of imaging data with n samples and p imaging quantitative traits (QTs) and K is the number of imaging modalities. $\mathbf{v}_k \in \mathbb{R}^{p \times 1}$ represents the canonical weight corresponding to the k -th modality, and $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_K]$. These weights yielded by SMCCA show the importance of each imaging feature in associating multiple brain imaging modalities. However, SMCCA supposes $\mathbf{X}_k^\top \mathbf{X}_k = \mathbf{I}$ which weakens the performance of the model [23]. What's worse, the modality-shared imaging features mix up with those modality-specific ones, resulting in poor interpretability.

PDSMCCA

In order to better identify the relationship between multimodal brain imaging data and overcome the drawbacks of SMCCA, we propose a novel SMCCA (PDSMCCA) model. PDSMCCA is defined as follows:

$$\begin{aligned} \min_{\mathbf{b}_k, \mathbf{s}_k} \sum_{1 \leq k, j \leq K} & \frac{1}{2} \|\mathbf{X}_k(\mathbf{b}_k + \mathbf{s}_k) - \mathbf{X}_j(\mathbf{b}_j + \mathbf{s}_j)\|_2^2 + \lambda_B \|\mathbf{B}\|_{1,1} \\ & + \lambda_S \|\mathbf{S}\|_{2,1}, \quad \text{s.t. } \|\mathbf{X}_k(\mathbf{b}_k + \mathbf{s}_k)\|_2^2 = 1 \quad (k = 1, \dots, K), \end{aligned} \tag{4}$$

where λ_B and λ_S are two nonnegative tuning parameters, and $\mathbf{V} = \mathbf{B} + \mathbf{S}$. The decomposition of \mathbf{V} is interesting and meaningful.

Specifically, by using different regularization functions for \mathbf{B} and \mathbf{S} , we can enable them to select different types of features, e.g. the modality-shared and -specific features. In this paper, we impose the $\ell_{2,1}$ -norm [33] on \mathbf{S} to select the shared features across multiple modalities, and this penalty is defined as $\|\mathbf{S}\|_{2,1} = \sum_i \|\mathbf{s}^i\|_2$. In addition, we use the ℓ_1 -norm for an imaging QT across all imaging modalities. This might identify features that can only be recognized under certain technologies. And the penalty is defined as $\|\mathbf{B}\|_{1,1} = \sum_j \|\mathbf{b}_j\|_1 = \sum_i \|\mathbf{b}^i\|_1 = \sum_i \sum_j |b_{ij}|$.

The merits of PDSMCCA are as follows. First of all, our model directly calculates the multi-way association among multiple data modalities, which holds a powerful modeling capability. Besides, we use ℓ_1 -norm to identify related QTs that may only change in a single imaging modality, and use $\ell_{2,1}$ -norm to identify related imaging QTs that change together due to the co-varying effects of AD, which demonstrates a diverse and desirable feature selection capability. Most importantly, attributing to the parameter decomposition and diverse regularization, the modality-shared features and modality-specific features can be obtained in a unified model, which could provide a better interpretation for biomedical studies.

The optimization algorithm

According to Lemma 2.2 in [45], the optimum \mathbf{b}_k and \mathbf{s}_k can be obtained by $\mathbf{b}_k^* = \frac{\hat{\mathbf{b}}_k}{\|\mathbf{X}_k(\hat{\mathbf{b}}_k + \hat{\mathbf{s}}_k)\|_2}$ and $\mathbf{s}_k^* = \frac{\hat{\mathbf{s}}_k}{\|\mathbf{X}_k(\hat{\mathbf{b}}_k + \hat{\mathbf{s}}_k)\|_2}$ respectively. Further, $\hat{\mathbf{b}}_k$ and $\hat{\mathbf{s}}_k$ are solutions to the following objective,

$$\min_{\mathbf{b}_k, \mathbf{s}_k} \sum_{1 \leq k, j \leq K} \frac{1}{2} \|\mathbf{X}_k(\mathbf{b}_k + \mathbf{s}_k) - \mathbf{X}_j(\mathbf{b}_j + \mathbf{s}_j)\|_2^2 + \lambda_B \|\mathbf{B}\|_{1,1} + \lambda_S \|\mathbf{S}\|_{2,1}. \tag{5}$$

Equation (5) is a typical bi-convex function, and we can use the alternating convex search (ACS) method [46] to solve this objective. That is, we update one variable and fix all the remaining ones at each step. Since $\|\mathbf{X}_k(\mathbf{b}_k + \mathbf{s}_k)\|_2^2 = 1$, (5) is processed as follows:

$$\begin{aligned} \min_{\mathbf{b}_k, \mathbf{s}_k} \sum_{1 \leq k, j \leq K} \frac{1}{2} \|\mathbf{X}_k(\mathbf{b}_k + \mathbf{s}_k) - \mathbf{X}_j(\mathbf{b}_j + \mathbf{s}_j)\|_2^2 \\ - \frac{1}{4} \|\mathbf{X}_k(\mathbf{b}_k + \mathbf{s}_k)\|_2^2 + \frac{1}{2} \|\mathbf{X}_j(\mathbf{b}_j + \mathbf{s}_j)\|_2^2 \\ + \lambda_B \|\mathbf{B}\|_{1,1} + \lambda_S \|\mathbf{S}\|_{2,1}, \end{aligned} \tag{6}$$

according to inequality $\frac{1}{4} \|\mathbf{X}_k(\mathbf{b}_k + \mathbf{s}_k)\|_2^2 \leq \frac{1}{2} \|\mathbf{X}_k \mathbf{b}_k\|_2^2 + \frac{1}{2} \|\mathbf{X}_k \mathbf{s}_k\|_2^2$, we equivalently have the following objective with respect to \mathbf{b}_k and \mathbf{s}_k ,

$$\begin{aligned} \min_{\mathbf{b}_k, \mathbf{s}_k} \sum_{1 \leq k, j \leq K} \frac{1}{2} \|\mathbf{X}_k \mathbf{b}_k - \mathbf{X}_j(\mathbf{b}_j + \mathbf{s}_j)\|_2^2 \\ + \frac{1}{2} \|\mathbf{X}_k \mathbf{s}_k - \mathbf{X}_j(\mathbf{b}_j + \mathbf{s}_j)\|_2^2 + \lambda_B \|\mathbf{B}\|_{1,1} + \lambda_S \|\mathbf{S}\|_{2,1}. \end{aligned} \tag{7}$$

Equation (7) is convex in \mathbf{b}_k when fixing \mathbf{s}_k as constants.

$$\min_{\mathbf{b}_k} \sum_{1 \leq k, j \leq K} \frac{1}{2} \|\mathbf{X}_k \mathbf{b}_k - \mathbf{X}_j(\mathbf{b}_j + \mathbf{s}_j)\|_2^2 + \lambda_B \|\mathbf{b}_k\|_1, \tag{8}$$

Then based on the ACS strategy, we take the derivative with respect to each \mathbf{b}_k , and letting it be zero, we obtain

$$\mathbf{b}_k = (\lambda_B \mathbf{D}_b + (K - 1) \mathbf{X}_k^T \mathbf{X}_k)^{-1} \sum_{1 \leq k, j \leq K} \mathbf{X}_k^T \mathbf{X}_j(\mathbf{b}_j + \mathbf{s}_j), \tag{9}$$

where \mathbf{D}_b is a diagonal matrix with the i th diagonal element being $\frac{1}{|\mathbf{b}_{ik}|}$.

Similarly, the optimal \mathbf{s}_k can be obtained by solving (10)

$$\min_{\mathbf{s}_k} \sum_{1 \leq k, j \leq K} \frac{1}{2} \|\mathbf{X}_k \mathbf{s}_k - \mathbf{X}_j(\mathbf{b}_j + \mathbf{s}_j)\|_2^2 + \lambda_S \|\mathbf{S}\|_{2,1}. \tag{10}$$

then we have the closed-form updating rule for each \mathbf{s}_k ,

$$\mathbf{s}_k = (\lambda_S \mathbf{D}_s + (K - 1) \mathbf{X}_k^T \mathbf{X}_k)^{-1} \sum_{1 \leq k, j \leq K} \mathbf{X}_k^T \mathbf{X}_j(\mathbf{b}_j + \mathbf{s}_j), \tag{11}$$

where \mathbf{D}_s is a diagonal matrix, and its i th diagonal element is $\frac{1}{\|\mathbf{s}^i\|_2}$ ($i = 1, \dots, p$).

Once every \mathbf{b}_k and \mathbf{s}_k is attained, \mathbf{B} and \mathbf{S} can be attained as well. Finally, we present the pseudo-code in Algorithm 1. The input of PDSMCCA is the neuroimaging

quantitative trait data from multiple modalities, and the output is the canonical weight (absolute value) showing the relative importance of each imaging feature. Step 1 initializes \mathbf{B} and \mathbf{S} . Step 3 to 6 are iteration procedure to seek the final solutions.

Algorithm 1 Parameter Decomposition Based SMCCA (PDSMCCA)

Require:

$$\mathbf{X}_k \in \mathbb{R}^{n \times p}, \lambda_B, \lambda_S$$

Ensure:

Canonical weights \mathbf{B} and \mathbf{S} .

- 1: Initialize $\mathbf{B} \in \mathbb{R}^{p \times K}$, $\mathbf{S} \in \mathbb{R}^{p \times K}$, and normalize $\|\mathbf{X}_k(\mathbf{b}_k + \mathbf{s}_k)\|_2^2 = 1$;
 - 2: **while** not convergence **do**
 - 3: Solve (8) according to (9);
 - 4: $\mathbf{b}_k^* = \frac{\mathbf{b}_k}{\|\mathbf{X}_k(\mathbf{b}_k + \mathbf{s}_k)\|_2}$;
 - 5: Solve (10) according to (11);
 - 6: $\mathbf{s}_k^* = \frac{\mathbf{s}_k}{\|\mathbf{X}_k(\mathbf{b}_k + \mathbf{s}_k)\|_2}$;
 - 7: **end while**
-

Convergence analysis

Theorem 1 will prove that Algorithm 1 converge to a local optimum.

Theorem 1 *The value of (4) keeps decreasing throughtout the iteration of Algorithm 1.*

We use $\{\mathbf{b}_k^{(t)}, \mathbf{s}_k^{(t)}\}$ to represent the estimate of $\{\mathbf{b}_k, \mathbf{s}_k\}$ in the t th iteration. Next, we will prove that the value of (8) is continuously decreasing when solving \mathbf{b}_k . To facilitate understanding, we denote the objective of (8) as $F(\mathbf{b}_k)$:

$$F(\mathbf{b}_k) = \sum_{1 \leq k, j \leq K} \frac{1}{2} \|\mathbf{X}_k \mathbf{b}_k - \mathbf{X}_j(\mathbf{b}_j^{(t)} + \mathbf{s}_j^{(t)})\|_2^2 + \lambda_B \|\mathbf{b}_k\|_1. \tag{12}$$

Then we define

$$\begin{aligned} G(\mathbf{b}_k) &= \sum_{1 \leq k, j \leq K} \frac{1}{2} \|\mathbf{X}_k \mathbf{b}_k - \mathbf{X}_j(\mathbf{b}_j^{(t)} + \mathbf{s}_j^{(t)})\|_2^2 + \lambda_B \sum_{i=1}^p \left(\frac{b_{ki}^2}{2|b_{ki}^{(t)}|} + \frac{|b_{ki}^{(t)}|}{2} \right) \\ &= \sum_{1 \leq k, j \leq K} \frac{1}{2} \|\mathbf{X}_k \mathbf{b}_k - \mathbf{X}_j(\mathbf{b}_j^{(t)} + \mathbf{s}_j^{(t)})\|_2^2 + \lambda_B \left(\frac{1}{2} \mathbf{b}_k^\top \mathbf{D}_b \mathbf{b}_k + \frac{1}{2} \mathbf{b}_k^{(t)\top} \mathbf{D}_b \mathbf{b}_k^{(t)} \right), \end{aligned} \tag{13}$$

where \mathbf{D}_b is defined in (9), and (14) can be easily proved. It is obvious that $G(\mathbf{b}_k)$ is a convex quadratic function that satisfies

$$G(\mathbf{b}_k^{(t)}) = F(\mathbf{b}_k^{(t)}), \quad G(\mathbf{b}_k) \geq F(\mathbf{b}_k), \quad \forall \mathbf{b}_k \in \mathbb{R}^p. \tag{14}$$

Since the estimate of \mathbf{b}_k at the next iteration $t+1$, expressed in (8) and denoted as $\mathbf{b}_k^{(t+1)}$, is the minimizer of $G(\mathbf{b}_k)$, we have

$$G(\mathbf{b}_k^{(t+1)}) \leq G(\mathbf{b}_k^{(t)}). \quad (15)$$

Putting (13)–(15) together, we have

$$F(\mathbf{b}_k^{(t+1)}) \leq G(\mathbf{b}_k^{(t+1)}) \leq G(\mathbf{b}_k^{(t)}) = F(\mathbf{b}_k^{(t)}). \quad (16)$$

This formula shows that the objective decreases by fixing \mathbf{s}_k , which guarantees the convergence. And after the rescaling, the conclusion is still valid. Thus, for \mathbf{s}_k , we can get the same conclusion in the same way. By denoting the objective as $\mathcal{L}(\mathbf{b}_k, \mathbf{s}_k)$, then according to the conclusions above, we have

$$\mathcal{L}(\mathbf{b}_k^{(t+1)}, \mathbf{s}_k^{(t+1)}) \leq \mathcal{L}(\mathbf{b}_k^{(t+1)}, \mathbf{s}_k^{(t)}) \leq \mathcal{L}(\mathbf{b}_k^{(t)}, \mathbf{s}_k^{(t)}), \quad (17)$$

We further know $\mathcal{L}(\mathbf{b}_k, \mathbf{s}_k)$ is lower bounded by zero. Therefore, we combine (16)–(17), Algorithm 1 will converge to the optimum.

Abbreviations

CCC: Canonical correlation coefficient; PDSMCCA: Parameter decomposition based sparse multi-view canonical correlation analysis; AD: Alzheimer's Disease; ADNI: Alzheimer's Disease Neuroimaging Initiative; ACS: Alternating convex search; MRI: Magnetic resonance imaging; MCI: Mild cognitive impairment; sMRI: Structural magnetic resonance imaging; FDG: Fluorodeoxyglucose; QTs: Quantitative traits; ROI: Regions of interest; AAL: Automated anatomical labeling; CSF: Cerebrospinal fluid; GM: Gray matter; WM: White matter; SCCA: Sparse canonical correlation analysis; SMCCA: Sparse multiple/multi-view/multi-set canonical correlation analysis; NC: Normal controls; AV45: 18-F florbetapir PET; VBM: Voxel-based morphometry; PET: Positron emission tomography; fMRI: Functional magnetic resonance imaging.

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. Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

About this supplement

This article has been published as part of BMC Bioinformatics Volume 23 Supplement 3, 2022: Selected articles from the International Conference on Intelligent Biology and Medicine (ICIBM 2021): bioinformatics. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-23-supplement-3>.

Author contributions

JZ: methodology, software, writing-original draft. HW: visualization, writing-review. YZ: software, investigation. LG: validation, writing-review and editing. LD: conceptualization, writing-review and editing. All authors read and approved the final manuscript.

Funding

The data collection and sharing of 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). Publication costs are funded in part by the Key-Area Research and Development Program of Guangdong Province [2019B010110001]; the National Natural Science Foundation of China [61973255, 62136004, 61936007, 62036011, U1801265, 62027813]; the Key R&D Program of Shaanxi Province [2021ZDLGY01-08]; the Natural Science Basic Research Program of Shaanxi [2020JM-142]; the China Postdoctoral Science Foundation [2020T130537]

at Northwestern Polytechnical University. This work was also supported in part by the Shanghai Municipal Science and Technology Major Project [2018SHZDZX01] at LCNBI and ZJLab.

Availability of data and materials

Data used in this study are publicly available at the ADNI database (www.adni-info.org). They are from the ADNI1, ADNI GO and ADNI2 standard data sets that are considered in experiments with the MRI and PET data sets.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 4 April 2022 Accepted: 4 April 2022

Published online: 12 April 2022

References

- Goedert M, Spillantini MG. A century of Alzheimer's disease. *Science*. 2006;314(5800):777–81.
- Grellmann C, Bitzer S, Neumann J, Westlye LT, Andreassen OA, Villringer A, Horstmann A. Comparison of variants of canonical correlation analysis and partial least squares for combined analysis of MRI and genetic data. *Neuroimage*. 2015;107:289–310.
- Association A. Alzheimer's disease facts and figures. *Alzheimer's Dement*. 2019;15(3):321–87.
- Xiao E, Chen Q, Goldman AL, Tan HY, Healy K, Zoltick B, et al. Late-onset Alzheimer's disease polygenic risk profile score predicts hippocampal function. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2017;2(8):673–9.
- Bakkour A, Morris JC, Wolk DA, Dickerson BC. The effects of aging and Alzheimer's disease on cerebral cortical anatomy: specificity and differential relationships with cognition. *Neuroimage*. 2013;76:332–44.
- Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *Neuroimage*. 2017;155:530–48.
- Shivamurthy VK, Tahari AK, Marcus C, Subramaniam RM. Brain FDG PET and the diagnosis of dementia. *Am J Roentgenol*. 2015;204(1):76–85.
- Nakao T, Radua J, Rubia K, Mataix-Cols D. Gray matter volume abnormalities in ADHD: voxel-based meta-analysis exploring the effects of age and stimulant medication. *Am J Psychiatry*. 2011;168(11):1154–63.
- Woodward M, Rowe CC, Jones G, Villemagne VL, Varos T. Differentiating the frontal presentation of Alzheimer's disease with FDG-PET. *J Alzheimers Dis*. 2015;44(1):233–42.
- Stuss DT, Gow CA, Hetherington CR. "no longer gage": frontal lobe dysfunction and emotional changes. *J Consult Clin Psychol*. 1992;60(3):349.
- Lorenzi M, Simpson IJ, Mendelson AF, Vos SB, Cardoso MJ, Modat M, Schott JM, Ourselin S. Multimodal image analysis in Alzheimer's disease via statistical modelling of non-local intensity correlations. *Sci Rep*. 2016;6(1):1–8.
- Sui J, Adali T, Yu Q, Chen J, Calhoun VD. A review of multivariate methods for multimodal fusion of brain imaging data. *J Neurosci Methods*. 2012;204(1):68–81.
- Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage*. 2011;55(3):856–67.
- Ball G, Aljabar P, Nongena P, Kennea N, Gonzalez-Cinca N, Falconer S, Chew AT, Harper N, Wurie J, Rutherford MA, et al. Multimodal image analysis of clinical influences on preterm brain development. *Ann Neurol*. 2017;82(2):233–46.
- Oxtoby NP, Alexander DC. Imaging plus x: multimodal models of neurodegenerative disease. *Curr Opin Neurol*. 2017;30(4):371.
- Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage*. 2012;59(2):895–907.
- Fan C, Cheng Y, Gou H, Liu C, Deng S, Liu C, Chen X, Bu J, Zhang X. Neuroimaging and intervening in memory reconsolidation of human drug addiction. *Sci China Inf Sci*. 2020;63(7):1–11.
- Xu C, Wang Z, Fan M, Liu B, Song M, Zhen X, Jiang T, Initiative ADN, et al. Effects of BDNF val66met polymorphism on brain metabolism in Alzheimer's disease. *NeuroReport*. 2010;21(12):802.
- Timmers T, Ossenkoppele R, Wolters EE, Verfaillie SC, Visser D, Golla SS, Barkhof F, Scheltens P, Boellaard R, Van Der Flier WM, et al. Associations between quantitative [18 F] flortaucipir tau pet and atrophy across the Alzheimer's disease spectrum. *Alzheimer's Res Ther*. 2019;11(1):60.
- Hedden T, Mormino EC, Amariglio RE, Younger AP, Schultz AP, Becker JA, Buckner RL, Johnson KA, Sperling RA, Rentz DM. Cognitive profile of amyloid burden and white matter hyperintensities in cognitively normal older adults. *J Neurosci*. 2012;32(46):16233–42.
- Hardoon DR, Shawe-Taylor J. Sparse canonical correlation analysis. *Mach Learn*. 2011;83(3):331–53.
- Du L, Huang H, Yan JE. Structured sparse canonical correlation analysis for brain imaging genetics: an improved graphnet method. *Bioinformatics*. 2016;32(10):1544–51.
- Shen L, Risacher SL, Du L, Moore JH, Huang H, Inlow M, Kim S, Saykin AJ, Yan J. A novel structure-aware sparse learning algorithm for brain imaging genetics. *Med Image Comput Comput Assist Interv*. 2014;17(3):329–36.
- 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. Springer. 2017. p. 543–55.
- Wilms I, Croux C. Sparse canonical correlation analysis from a predictive point of view. *Biom J*. 2015;57(5):834–51.

26. Du L, Liu K, Yao X, Risacher SL, Shen L. Detecting genetic associations with brain imaging phenotypes in Alzheimer's disease via a novel structured SCCA approach. *Med Image Anal.* 2020;61:101656.
27. Chen M, Gao C, Ren Z, Zhou HH. Sparse cca via precision adjusted iterative thresholding. 2013. arXiv preprint. [arXiv:1311.6186](https://arxiv.org/abs/1311.6186).
28. Witten DM, Tibshirani RJ. Extensions of sparse canonical correlation analysis with applications to genomic data. *Stat Appl Genet Mol Biol.* 2009;8(1):28.
29. Du L, Zhang J, Liu F, Wang H, Guo L, Han J, Initiative ADN, et al. Identifying associations among genomic, proteomic and imaging biomarkers via adaptive sparse multi-view canonical correlation analysis. *Med Image Anal.* 2021;70:102003.
30. Hao X, et al. Mining outcome-relevant brain imaging genetic associations via three-way sparse canonical correlation analysis in Alzheimer's disease. *Sci Rep.* 2017;7:44272.
31. Fang J, Lin D, Schulz C, Xu Z, Calhoun VD, Wang YP. Joint sparse canonical correlation analysis for detecting differential imaging genetics modules. *Bioinformatics.* 2011;32:3480–8.
32. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc.* 1996;58(1):267–88.
33. Liu J, Ji S, Ye J. Multi-task feature learning via efficient $l_{2,1}$ -norm minimization. 2012. arXiv preprint. [arXiv:1205.2631](https://arxiv.org/abs/1205.2631).
34. Wang H, Nie F, Huang H, Kim S, Nho K, Risacher SL, Saykin AJ, Shen L, Initiative ADN. Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. *Bioinformatics.* 2012;28(2):229–37.
35. Harper L, Bouwman F, Burton EJ, Barkhof F, Scheltens P, O'Brien J, et al. Patterns of atrophy in pathologically confirmed dementias: a voxelwise analysis. *Journal of Neurology, Neurosurgery, & Psychiatry*, 908–916 (2017).
36. Wang W, Yu J, Liu Y, Yin R, Wang H, Wang J, Tan L, Radua J, Tan L. Voxel-based meta-analysis of grey matter changes in Alzheimer's disease. *Transl Neurodegener.* 2015;4(1):6.
37. Potkin SG, Guffanti G, Lakatos A, Turner JA, Kruggel F, Fallon JH, et al. Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease. *PLoS ONE.* 2009;4(8):e6501.
38. Galton C, Gomez-Anson B, Antoun N, Scheltens P, Patterson K, Graves M, Sahakian B, Hodges J. Temporal lobe rating scale: application to Alzheimer's disease and frontotemporal dementia. *J Neurol Neurosurg Psychiatry.* 2001;70(2):165–73.
39. Altmann A, Ng B, Landau SM, Jagust WJ, Greicius MD. Regional brain hypometabolism is unrelated to regional amyloid plaque burden. *Brain.* 2015;138(12):3734–46.
40. Sepulcre J, Grothe MJ, Quillias FD, Ortiz-Terán L, Diez I, Yang H-S, Jacobs HI, Hanseeuw BJ, Li Q, El-Fakhri G, et al. Neurogenetic contributions to amyloid beta and tau spreading in the human cortex. *Nat Med.* 2018;24(12):1910–8.
41. Mosconi L. Brain glucose metabolism in the early and specific diagnosis of Alzheimer's disease. *Eur J Nucl Med Mol Imaging.* 2005;32(4):486–510.
42. Jalali A, Sanghavi S, Ruan C, Ravikumar PK. A dirty model for multi-task learning. In: *NIPS*, 2010. p. 964–72.
43. Du L, Liu K, Yao X, Risacher S, Han J, Saykin A, Guo L, Shen L. Multi-task sparse canonical correlation analysis with application to multi-modal brain imaging genetics. *IEEE/ACM Trans Comput Biol Bioinf.* 2021;18(1):227–39.
44. Du L, Liu K, Zhu L, Yao X, Risacher SL, Guo L, Saykin AJ, Shen L, Initiative ADN. Identifying progressive imaging genetic patterns via multi-task sparse canonical correlation analysis: a longitudinal study of the ADNI cohort. *Bioinformatics.* 2019;35(14):474–83.
45. Witten DM, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics.* 2009;10(3):515–34.
46. Gorski J, Pfeuffer F, Klamroth K. Biconvex sets and optimization with biconvex functions: a survey and extensions. *Math Methods Oper Res.* 2007;66(3):373–407.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

