

REVIEW ARTICLE

WILEY

# A technical review of canonical correlation analysis for neuroscience applications

Xiaowei Zhuang<sup>1</sup>  | Zhengshi Yang<sup>1</sup>  | Dietmar Cordes<sup>1,2,3</sup> 

<sup>1</sup>Cleveland Clinic Lou Ruvo Center for Brain Health, Las Vegas, Nevada

<sup>2</sup>University of Colorado, Boulder, Colorado

<sup>3</sup>Department of Brain Health, University of Nevada, Las Vegas, Nevada

## Correspondence

Dietmar Cordes, Cleveland Clinic Lou Ruvo Center for Brain Health, 888 W. Bonneville Ave, Las Vegas, NV 89106.  
Email: cordesd@ccf.org

## Funding information

National Institute of Health, Grant/Award Number: 1R01EB014284; Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health, Grant/Award Number: 5P20GM109025; The Keep Memory Alive Foundation Young Scientist Award; A private grant from the Peter and Angela Dal Pezzo funds; A private grant from Lynn and William Weidner; A private grant from Stacie and Chuck Matthewson

## Abstract

Collecting comprehensive data sets of the same subject has become a standard in neuroscience research and uncovering multivariate relationships among collected data sets have gained significant attentions in recent years. Canonical correlation analysis (CCA) is one of the powerful multivariate tools to jointly investigate relationships among multiple data sets, which can uncover disease or environmental effects in various modalities simultaneously and characterize changes during development, aging, and disease progressions comprehensively. In the past 10 years, despite an increasing number of studies have utilized CCA in multivariate analysis, simple conventional CCA dominates these applications. Multiple CCA-variant techniques have been proposed to improve the model performance; however, the complicated multivariate formulations and not well-known capabilities have delayed their wide applications. Therefore, in this study, a comprehensive review of CCA and its variant techniques is provided. Detailed technical formulation with analytical and numerical solutions, current applications in neuroscience research, and advantages and limitations of each CCA-related technique are discussed. Finally, a general guideline in how to select the most appropriate CCA-related technique based on the properties of available data sets and particularly targeted neuroscience questions is provided.

## KEYWORDS

canonical correlation analysis, multivariate analysis, neuroscience

## 1 | INTRODUCTION

Recently in neuroscience research, multiple types of data are usually collected from the same individual, including demographics, clinical symptoms, behavioral and neuropsychological measures, genetic information, structural and functional magnetic resonance imaging (fMRI) data, position emission tomography (PET) data, functional near-infrared spectroscopy (fNIRS) data, and electrophysiological data.

Each of these data types, termed modality here, contains multiple measurements and provides a unique view of the subject. These measurements can be the raw data (e.g., neuropsychological tests) or derived information (e.g., brain regional volume and thickness measures derived from T1-weighted MRI).

Neuroscience research has been focused on uncovering associations between measurements from multiple modalities. Conventionally, a single measurement is selected from each modality, and their one-to-one univariate association is analyzed. Multiple correction is then performed to guarantee statistically meaningful results. These

Xiaowei Zhuang and Zhengshi Yang contributed equally to this manuscript.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

univariate associations have illuminated numerous findings in various neurological diseases, such as association between gray-matter density and Mini Mental State Examination score in Alzheimer's disease (Baxter et al., 2006), correlation between brain network temporal dynamics and Unified Parkinson Disease Rating Scale part III motor scores in Parkinson's disease subjects (Zhuang et al., 2018), and relationship between imaging biomarkers and cognitive performances in fighters with repetitive head trauma (Mishra et al., 2017).

However, the one-to-one univariate association overlooks the multivariate joint relationship among multiple measurements between modalities. Furthermore, when dealing with brain imaging data, highly correlated noise further decreases the effectiveness and sensitivity of mass-univariate voxel-wise analysis (Cremers, Wager, & Yarkoni, 2017; Zhuang et al., 2017), and different methods of multiple corrections might lead to various statistically meaningful results. Multivariate analysis, alternatively, uncovers the joint covariate patterns among different modalities and avoids multiple correction steps, which would be more appropriate to disentangle joint relationship between modalities and guarantees full utilization of all common information.

Canonical correlation analysis (CCA) is one candidate to uncover these joint multivariate relationships among different modalities. CCA is a statistical method that finds linear combinations of two random variables so that the correlation between the combined variables is maximized (Hotelling, 1936). CCA can identify the source of common statistical variations among multiple modalities, without assuming any particular form of directionality, which suits neuroscience applications. In practice, CCA has been mainly implemented as a substitute for univariate general linear model (GLM) to link different modalities, and therefore, is a major and powerful tool in multimodal data fusion. Multiple CCA variants, including kernel CCA, constrained CCA, deep CCA, and multiset CCA, also have been applied in neuroscience research. However, the complicated multivariate formulations and obscure capabilities remain obstacles for CCA and its variants to being widely applied.

In this study, we review CCA applications in neuroscience research from a technical perspective to improve the understanding of the CCA technique itself and to provide neuroscience researchers with guidelines of proper CCA applications. We briefly discuss studies through December 2019 that have utilized CCA and its variants to uncover the association between multiple modalities. We explain the existing CCA method and its variants for their formulations, properties, relationships to other multivariate techniques, and advantages and limitations in neuroscience applications. We finally provide a flowchart and an experimental example to assist researchers to select the most appropriate CCA technique based on their specific applications.

## 2 | INCLUSION/EXCLUSION OF STUDIES

Using the PubMed search engine in December 2019, we searched neuroimaging or neuroscience articles using CCA with the following string: ("canonical correlation" analysis) AND (neuroscience OR neuroimaging). This search yielded 192 articles; 11 additional articles were

included based on authors' preidentification. We excluded non-English articles, conference abstracts and duplicated studies, yielding 188 articles assessed for eligibility. We further identified 160 studies that met the following criteria: (a) primarily focused on a CCA or CCA-variant technique and (b) with an application to neuroimaging or neuroscience modalities. Reasons for exclusion and numbers of articles meeting exclusion criteria at each stage are shown in Figure 1.

The remaining articles were full-text reviewed and divided into five categories based on the applied CCA technique (Figure 2a): CCA ( $N = 67$ ); constrained CCA ( $N = 53$ ); nonlinear CCA ( $N = 7$ ); multiset CCA ( $N = 29$ ); and CCA-other ( $N = 7$ ). Three articles applied constrained multiset CCA, thus are categorized into both constrained CCA and multiset CCA. Numbers of articles of every year from 1990 to 2019 are plotted in Figure 2 (B).

In the following sections, we present technical details (Section 3) and neuroscience applications for each category (Section 4). In Section 5, we discuss technical differences and summarize advantages and limitations of each CCA-related technique. We finally provide an experimental example and guidance in Section 6 to researchers who are interested in applying multivariate CCA-related techniques in their work.

## 3 | TECHNICAL DETAILS

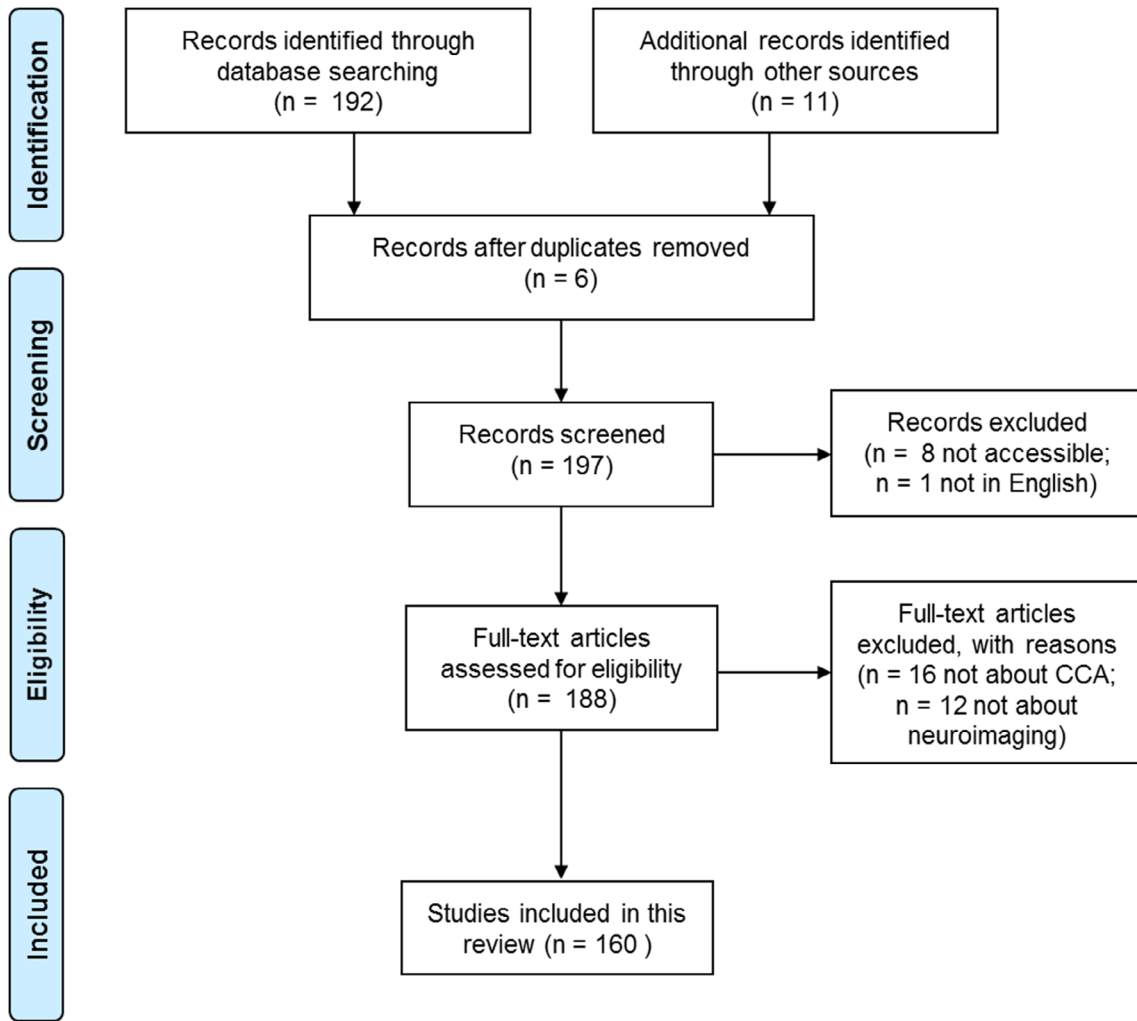
Figure 3 shows the detailed CCA equations (red box) and linkages between CCA and its variants. Constrained CCA (yellow boxes), nonlinear CCA (gray boxes), and multiset CCA (orange boxes) are focused, and linkages between CCA and other univariate (light green boxes) and multivariate (dark green boxes) techniques are also included. Here, we provide basic formulations and solutions of each CCA and its variants. We also discuss how CCA is mathematically linked to its variants and to other multivariate or univariate techniques. Researchers interested in further details can refer to the corresponding references.

### 3.1 | Conventional CCA

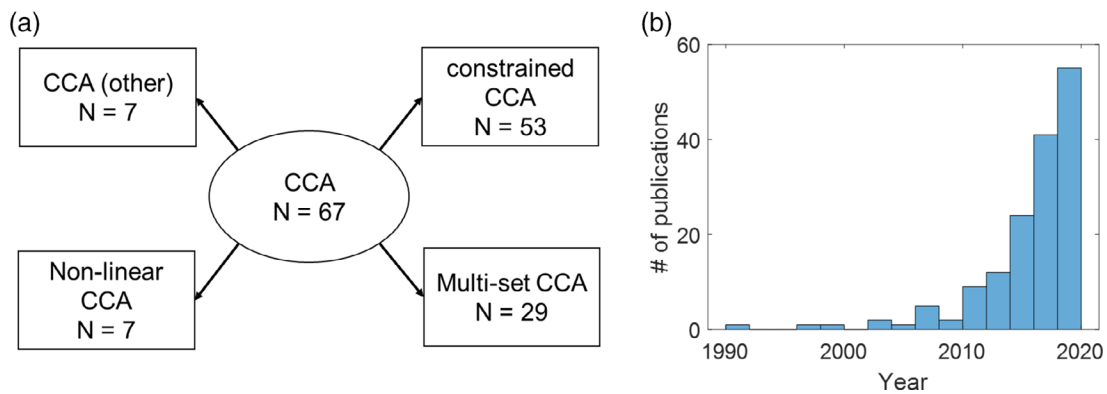
**Formulations.** CCA is designed to maximize the correlation between two latent variables  $\mathbf{y}_1 \in \mathbb{R}^{p_1 \times 1}$  and  $\mathbf{y}_2 \in \mathbb{R}^{p_2 \times 1}$ , which are also being referred to as modalities. Here, we denote  $\mathbf{Y}_k \in \mathbb{R}^{N \times p_k}$ ,  $k = 1, 2$  as collected samples of these two variables, where  $N$  represents the number of observations (samples) and  $p_k$ ,  $k = 1, 2$  represent the number of features in each variable. CCA determines the canonical coefficients  $\mathbf{u}_1 \in \mathbb{R}^{p_1 \times 1}$  and  $\mathbf{u}_2 \in \mathbb{R}^{p_2 \times 1}$  for  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$ , respectively, by maximizing the correlation between  $\mathbf{Y}_1 \mathbf{u}_1$  and  $\mathbf{Y}_2 \mathbf{u}_2$ :

$$\text{CCA} : \max_{\mathbf{u}_1, \mathbf{u}_2} \rho = \text{corr}(\mathbf{Y}_1 \mathbf{u}_1, \mathbf{Y}_2 \mathbf{u}_2) = \frac{\mathbf{u}_1^T \boldsymbol{\Sigma}_{12} \mathbf{u}_2}{\sqrt{\mathbf{u}_1^T \boldsymbol{\Sigma}_{11} \mathbf{u}_1} \sqrt{\mathbf{u}_2^T \boldsymbol{\Sigma}_{22} \mathbf{u}_2}}. \quad (1)$$

In Equation (1),  $\boldsymbol{\Sigma}_{11}$  and  $\boldsymbol{\Sigma}_{22}$  are the within-set covariance matrices and  $\boldsymbol{\Sigma}_{12}$  is the between-set covariance matrix. The denominator in Equation (1) is used to normalize within-set covariance, which guarantees that CCA is invariant to the scaling of coefficients.



**FIGURE 1** Inclusion and exclusion criteria for this review

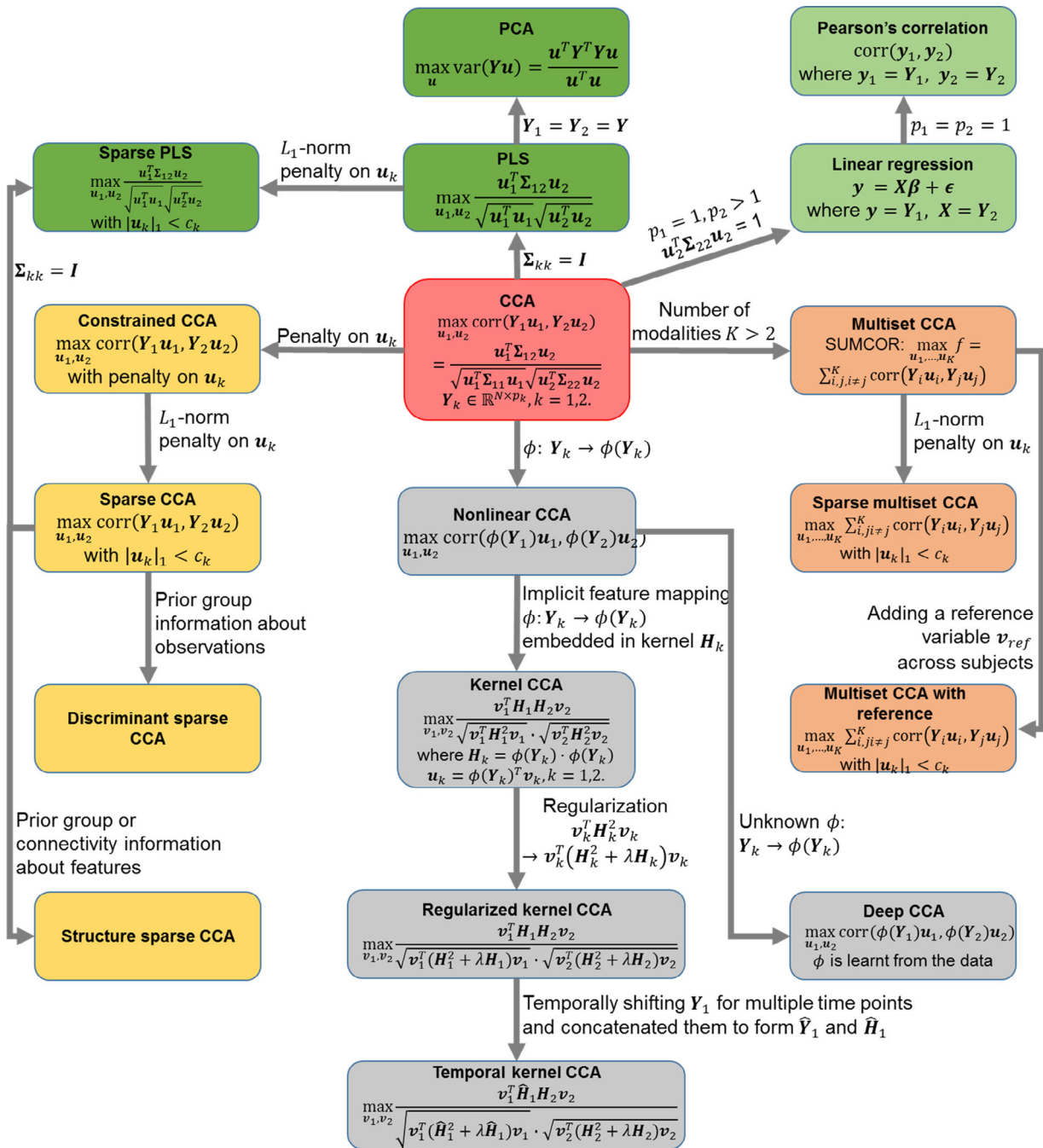


**FIGURE 2** Number of articles summarized by category (a) and year (b)

*Solutions.* Canonical coefficients  $u_1$  and  $u_2$  can be found by setting the partial derivative of the objective function (Equation (1)) with respect to  $u_1$  and  $u_2$  to zero, respectively, leading to:

$$\Sigma_{12}u_2 = \rho\Sigma_{11}u_1 \text{ and } \Sigma_{21}u_1 = \rho\Sigma_{22}u_2. \quad (2)$$

Equation (2) can be further reduced to a classical eigenvalue problem, if  $\Sigma_{kk}$  is invertible, as follows:



**FIGURE 3** Technical details of CCA and relationship between CCA and its variants. Background color indicates different techniques: red: conventional CCA; gray: nonlinear CCA; yellow: constrained CCA; orange: multiset CCA; green: other techniques related to CCA. CCA, canonical correlation analysis; PCA, principle component analysis; PLS, partial least square

$$\begin{aligned} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} u_1 &= \rho^2 u_1 \\ \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} u_2 &= \rho^2 u_2 \end{aligned} \quad (3)$$

Each pair of canonical coefficients  $\{u_1, u_2\}$  are the eigenvectors of  $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$  and  $\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$ , respectively with the same eigenvalue  $\rho^2$ . Following Equation (3), up to  $M = \min(p_1, p_2)$  pairs of canonical coefficients can be achieved through singular

value decomposition (SVD), and every pair of canonical variables  $\{Y_1 u_1^{(m)}, Y_2 u_2^{(m)}\}, m = 1, 2, \dots, M$ , are uncorrelated with another pair of canonical variables. Corresponding  $M$  canonical correlation values are in descending order as  $\rho^{(1)} > \rho^{(2)} > \dots > \rho^{(M)}$ .

As we stated above, one requirement for solving the CCA problem (Equation (1)) through this eigenvalue problem (Equation (3)) is that within-set covariance matrices  $\Sigma_{11}$  and  $\Sigma_{22}$  must be invertible.

To satisfy this requirement, the number of observations in  $Y_1$  and  $Y_2$  should be greater than the number of features, that is,  $N > p_k$ ,  $k = 1, 2$ . Furthermore, since the square of canonical correlation values ( $\rho^2$ ) are the eigenvalues of matrices  $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$  and  $\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ , both matrices are required to be positive definite.

*Statistical inferences.* Parametric inferences exist for CCA if both variables strictly follow the Gaussian distribution. The null hypothesis is that no (zero) canonical correlation exists between  $Y_1$  and  $Y_2$ , that is,  $\rho^{(1)} = \rho^{(2)} = \dots = \rho^{(M)} = 0$ . The alternative hypothesis is that at least one canonical correlation value is nonzero. A test statistic based on Wilk's  $\Lambda$  is (Bartlett, 1939):

$$\Lambda = -\left(N - \frac{p_1 + p_2 + 3}{2}\right) \log \prod_{i=1}^M (1 - \rho^{(i)}), \quad (4)$$

which follows a chi-square distribution  $\chi^2_{p_1 \times p_2}$  with degree of freedom of  $p_1 \times p_2$ . It is also of interest to test if a specific canonical correlation value ( $\rho^{(m)}$ ,  $1 \leq m \leq M$ ) is different from zero. In this case, the test statistic in Equation (4) becomes:

$$\Lambda^{(m)} = -\left(N - \frac{p_1 + p_2 + 3}{2}\right) \log \prod_{i=m+1}^M (1 - \rho^{(i)}), \quad (5)$$

which follows  $\chi^2_{(p_1-m)(p_2-m)}$ .

In practice, this parametric inference is not commonly used since it requires variables to strictly follow the Gaussian distribution and is sensitive to outliers (Bartlett, 1939). Instead, permutation-based nonparametric statistics have been widely used in CCA applications. In general, observations of one variable are randomly shuffled ( $Y_1$  becomes  $\widehat{Y}_1$ ) while observations of the other variable are kept intact ( $Y_2$  remains). A new set of canonical correlation values are then computed for  $\widehat{Y}_1$  and  $Y_2$  following Equation (3). This random shuffling is repeated multiple times, and the null distribution of canonical correlation values is generated. Statistical significance ( $p$ -values) for the true canonical correlation values are finally obtained from this null distribution.

### 3.2 | CCA variants

The conventional CCA (Equation (1)) can be modified for different purposes. Constrained CCA penalizes canonical coefficients  $u_1$  and  $u_2$  to satisfy certain requirements and more specifically, to avoid overfitting and unstable results caused by insufficient observations in  $Y_1$  or  $Y_2$ . Kernel and deep CCA are designed to uncover nonlinear correlations between modalities by projecting the original variables to new nonlinear feature spaces. Multiset CCA is proposed to find multivariate associations among more than two modalities. In this section, we systematically review constrained CCA, nonlinear CCA, multiset CCA, and other special CCA cases.

#### 3.2.1 | Constrained CCA

##### Generalized constrained CCA

*Formulation.* Constrained CCA is implemented by adding penalties to coefficients  $u_k$  in Equation (1). Penalties can be either equality constraints or inequality constraints, and based on researcher's own considerations, penalties can be added to either  $u_1$  or  $u_2$ , or to both  $u_1$  and  $u_2$ . Therefore, in general, the constrained CCA problem can be formulated in terms of the constrained optimization problem as:

$$\begin{aligned} \max_{u_1, u_2} \rho &= \text{corr}(Y_1 u_1, Y_2 u_2) \\ &= \frac{u_1^T \Sigma_{12} u_2}{\sqrt{u_1^T \Sigma_{11} u_1} \sqrt{u_2^T \Sigma_{22} u_2}}; \quad \text{s.t.} \begin{cases} \text{con}_i(u_1, u_2) = 0, \forall i \in E; \\ \text{con}_j(u_1, u_2) > 0, \forall j \in \text{In}E; \end{cases} \end{aligned} \quad (6)$$

where  $E$  represents the set of equality constraints and  $\text{In}E$  represents the set of inequality constraints.

*Solution.* Analytical solutions usually do not exist for constrained CCA problems, and solving Equation (6) requires numerical solutions through iterative optimization techniques. Multiple optimization techniques can be applied, such as the Broyden-Fletcher-Goldfarb-Shanno algorithm, augmented-Lagrangian algorithm, reduced gradient method and sequential quadratic programming. Examples and details of solving constrained CCA problems through above optimization techniques can be found in Yang, Zhuang, et al. (2018) and Zhuang et al. (2017).

##### Special case: $L_1$ -norm penalty and sparse CCA

*Formulation.* The most commonly implemented penalty in constrained CCA is the  $L_1$ -norm penalty added to either  $u_1$  or  $u_2$ , and is termed sparse CCA:

$$\begin{aligned} \text{sparse CCA: } \max_{u_1, u_2} \rho &= \text{corr}(Y_1 u_1, Y_2 u_2) \\ &= \frac{u_1^T \Sigma_{12} u_2}{\sqrt{u_1^T \Sigma_{11} u_1} \sqrt{u_2^T \Sigma_{22} u_2}}; \quad \text{s.t. } |u_1|_1 < c_1, |u_2|_1 < c_2, \end{aligned} \quad (7)$$

where  $|u_i|_1 < c_i$  are inequality constraints.

The  $L_1$ -norm penalty induces sparsity on canonical coefficients, and therefore sparse CCA can be implemented to high-dimensional variables. When dealing with high-dimensional variables, the within-set covariance matrices  $\Sigma_{11}$  and  $\Sigma_{22}$  in Equation (7) are also high-dimensional matrices, which are memory intensive. In addition, when the number of observations is less than the number of features, the covariance matrices cannot be estimated reliably from the sample. In these cases, within-set covariance matrices are usually replaced by identity matrices, and sparse CCA is then equivalent to sparse PLS. Please note that researchers may still name this technique as sparse CCA even after this replacement (Witten, Tibshirani, & Hastie, 2009).

With known prior information about features or observations, sparse CCA can be further modified to *structure sparse CCA* or *discriminant sparse CCA*, respectively. If the known prior information is

about features, such as categorizing features into different groups (Lin et al., 2014) or characterizing connections between features (Kim et al., 2019), the prior information will be implemented as an additional penalty on features, leading to *structure sparse CCA*. Alternatively, if the known prior information is about observations, such as diagnostic group of each subject, the prior information will be implemented as additional constraint on observations, leading to *discriminant sparse CCA* (Wang et al., 2019).

**Solutions.** Sparse CCA, structure sparse CCA, and discriminant sparse CCA can all be considered as special cases of a generalized constrained CCA (Equation (6)) problem with different equality and inequality constraint sets. Iterative optimization techniques used to solve the generalized constrained CCA problem are also applicable here to solve these special cases.

### 3.2.2 | Nonlinear CCA

Both CCA and constrained CCA assume linear intervariable relationships, however, this assumption does not hold in general for all variables in real data. Nonlinear CCA uncovers the joint nonlinear relationship between different variables, which is a complementary tool to conventional CCA methods. Kernel CCA, temporal kernel CCA, and deep CCA are the foremost techniques in this category.

#### Kernel CCA and temporal kernel CCA

**Formulation.** Kernel CCA uncovers the joint nonlinear relationship between two variables by mapping the original feature space in  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  on to a new feature space through a *predefined kernel function*. However, this new feature space is not explicitly defined. Instead, the original feature space for each observation in  $\mathbf{Y}_k$  is implicitly projected to a higher dimensional feature space  $\mathbf{Y}_k \rightarrow \phi(\mathbf{Y}_k)$  embedded in a prespecified kernel function  $\mathbf{H}_k \in \mathbb{R}^{N \times N}$ , which is independent of the number of features in the projected space. After transforming  $\mathbf{u}_k$  to  $\phi(\mathbf{Y}_k)^T \mathbf{v}_k$ , the CCA form in Equation (1) in the higher dimensional feature space, namely kernel CCA can be written as:

$$\text{Kernel CCA: } \max_{\mathbf{u}_1, \mathbf{u}_2} \rho = \text{corr}(\phi(\mathbf{Y}_1)\mathbf{u}_1, \phi(\mathbf{Y}_2)\mathbf{u}_2) = \max_{\mathbf{v}_1, \mathbf{v}_2} \frac{\mathbf{v}_1^T \mathbf{H}_1 \mathbf{H}_2 \mathbf{v}_2}{\sqrt{\mathbf{v}_1^T \mathbf{H}_1^2 \mathbf{v}_1} \sqrt{\mathbf{v}_2^T \mathbf{H}_2^2 \mathbf{v}_2}}; \\ \text{where } \mathbf{H}_k = \text{dot}(\phi(\mathbf{Y}_k), \phi(\mathbf{Y}_k)) \in \mathbb{R}^{N \times N} \text{ and } \mathbf{u}_k = \phi(\mathbf{Y}_k)^T \mathbf{v}_k, k = 1, 2, \quad (8)$$

where  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are unknowns to estimate, instead of  $\mathbf{u}_1$  and  $\mathbf{u}_2$ .

Temporal kernel CCA is a kernel CCA variant that is specifically designed for two time series with temporal delays. In temporal kernel CCA, one variable, for example,  $\mathbf{Y}_1$ , is shifted for multiple different time points and a new variable  $\tilde{\mathbf{Y}}_1$  is formed by concatenating the original  $\mathbf{Y}_1$  and the temporally shifted  $\mathbf{Y}_1$ . The new variable  $\tilde{\mathbf{Y}}_1$  and the original  $\mathbf{Y}_2$  are then input to kernel CCA as in Equation (8).

**Solution.** Closed-form analytical solution exists for kernel CCA (Equation (8)). By setting the partial derivatives of the objective

function in Equation (8) with respect to  $\mathbf{v}_1$  and  $\mathbf{v}_2$  to zero separately, kernel CCA can be converted to the following problem:

$$\mathbf{H}_1 \mathbf{H}_2 \mathbf{v}_2 = \rho \mathbf{H}_1^2 \mathbf{v}_1 \text{ and } \mathbf{H}_2 \mathbf{H}_1 \mathbf{v}_1 = \rho \mathbf{H}_2^2 \mathbf{v}_2. \quad (9)$$

Note that the kernel CCA problem defined in Equation (9) always holds true when  $\rho = 1$ . To avoid this trivial solution, a penalty term needs to be introduced to the norm of original canonical coefficients  $\mathbf{u}_k$ , such that  $\mathbf{v}_k^T \mathbf{H}_k^2 \mathbf{v}_k$  become  $\mathbf{v}_k^T \mathbf{H}_k^2 \mathbf{v}_k + \lambda \|\mathbf{u}_k\|^2 = \mathbf{v}_k^T (\mathbf{H}_k^2 + \lambda \mathbf{H}_k) \mathbf{v}_k$ , where  $\lambda$  is a regularization parameter. This regularized kernel CCA problem can be further represented as an eigenvalue problem (Hardoon, Szedmak, & Shawe-Taylor, 2004):

$$\begin{aligned} (\mathbf{H}_1 + \lambda \mathbf{I})^{-1} \mathbf{H}_2 (\mathbf{H}_2 + \lambda \mathbf{I})^{-1} \mathbf{H}_1 \mathbf{v}_1 &= \rho^2 \mathbf{v}_1 \\ (\mathbf{H}_2 + \lambda \mathbf{I})^{-1} \mathbf{H}_1 (\mathbf{H}_1 + \lambda \mathbf{I})^{-1} \mathbf{H}_2 \mathbf{v}_2 &= \rho^2 \mathbf{v}_2 \end{aligned} \quad (10)$$

where a closed-form solution exists in the new feature space.

#### Deep CCA

**Formulation.** Kernel CCA requires a predefined kernel function for the feature mapping to uncover the joint nonlinear relationship between two variables. Alternatively, recent development of deep learning makes it possible to learn the feature mapping from data itself. The deep learning variant of CCA, deep CCA (Andrew, Birmes, & Livescu, 2013), provides a more flexible and robust way to learn and search the nonlinear association between two variables. More specifically, deep CCA first passes the original  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  through multiple stacked layers of nonlinear transformations. Let  $\theta_1$  and  $\theta_2$  represent vectors of all parameters through all layers for  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$ , respectively, deep CCA can be represented as:

$$\text{Deep CCA: } \max_{\theta_1, \theta_2} \rho = \text{corr}(f(\mathbf{Y}_1; \theta_1), f(\mathbf{Y}_2; \theta_2)). \quad (11)$$

**Solution.** Deep CCA is solved through a deep learning schema by dividing the original data into training and testing sets.  $\theta_1$  and  $\theta_2$  are optimized by following the gradient of the correlation objective as estimated on the training data (Andrew et al., 2013). The number of unknown parameters in deep CCA is much higher than the number of unknowns in other CCA variants; therefore, a large number of training samples (in tens of thousands) are required for deep CCA to produce meaningful results. In most studies, it is unlikely to have enough observations (e.g. subjects) as training samples for deep CCA algorithms. Instead, in neuroscience applications, treating each brain voxel as a training sample, similar to Yang et al. (2020, 2019), would be more promising in deep CCA applications.

### 3.2.3 | Multiset CCA

Multiset CCA extends the conventional CCA from uncovering associations between two variables to finding common patterns among more than two variables. Constraints can also be incorporated in multiset CCA for various purposes.



### Multiset CCA

**Formulation.** The most intuitive formulation of multiset CCA is to optimize canonical coefficients of all variables by maximizing pairwise canonical correlations, named as SUMCOR multiset CCA:

$$\text{SUMCOR multiset CCA: } \max_{u_1, \dots, u_K} \sum_{i,j,i \neq j}^K \text{corr}(\mathbf{Y}_i \mathbf{u}_i, \mathbf{Y}_j \mathbf{u}_j), \quad (12)$$

where  $K > 2$  is the number of variables. A new matrix  $\hat{\Sigma} \in \mathbb{R}^{K \times K}$  is defined where each element  $\hat{\Sigma}_{ij}$  is a canonical correlation between two variables  $\mathbf{Y}_i$  and  $\mathbf{Y}_j$ :

$$\hat{\Sigma} = \begin{bmatrix} \mathbf{u}_1^T \Sigma_{11} \mathbf{u}_1 & \mathbf{u}_1^T \Sigma_{12} \mathbf{u}_2 & \dots & \mathbf{u}_1^T \Sigma_{1K} \mathbf{u}_K \\ \mathbf{u}_2^T \Sigma_{21} \mathbf{u}_1 & \mathbf{u}_2^T \Sigma_{22} \mathbf{u}_2 & & \mathbf{u}_2^T \Sigma_{2K} \mathbf{u}_K \\ \vdots & & \ddots & \vdots \\ \mathbf{u}_K^T \Sigma_{K1} \mathbf{u}_1 & \mathbf{u}_K^T \Sigma_{K2} \mathbf{u}_2 & \dots & \mathbf{u}_K^T \Sigma_{KK} \mathbf{u}_K \end{bmatrix}, \quad (13)$$

and  $\mathbf{u}_k^T \Sigma_{kk} \mathbf{u}_k, k = 1, \dots, K$  is set to 1 for normalization.

Besides maximizing SUMCOR, Kettenring (1971) summarizes four other possible objective functions in multiset CCA optimization: (a) SSQCOR, maximizing sum of squared pairwise correlations  $\sum_{i,j}^K \hat{\Sigma}_{ij}^2$ ; (b) MAXVAR, maximizing largest eigenvalue of correlation matrix  $\lambda_{\max}(\hat{\Sigma})$ ; (c) MINVAR, minimizing smallest eigenvalue of correlation matrix  $\lambda_{\min}(\hat{\Sigma})$ ; and (d) GENVAR, minimizing the determinant of correlation matrix  $\det(\hat{\Sigma})$ . In practice, SUMCOR multiset CCA is most commonly used followed by MAXVAR and SSQCOR multiset CCA.

**Solution.** Analytical solutions of multiset CCA are obtained by calculating the partial derivatives of the objective function with respect to each  $\mathbf{u}_i$ . Since SUMCOR and SSQCOR are linear and quadratic functions of each  $\mathbf{u}_i$ , respectively, closed-form analytical solutions can be obtained for these two cost functions by setting the partial derivatives equal to 0, which leads to generalized eigenvalue problems. Multiset CCA with all these five objective functions can also be solved by means of the general algebraic modeling system (Brooke, Kendrick, Meeraus, & Rama, 1998) and NLP solver CONOPT (Drud, 1985).

### Multiset CCA with constraints

In constrained multiset CCA, penalty terms can be added to each  $\mathbf{u}_i$  individually. Here we give examples of two commonly incorporated constraints in multiset CCA: sparse multiset CCA and multiset CCA with reference.

**Formulation: Sparse multiset CCA.** Similar to sparse CCA, sparse multiset CCA applies the  $L_1$ -norm penalty to one or more  $\mathbf{u}_i$  in Equation (12), and therefore induces sparsity on canonical coefficient(s) and can be applied to high-dimensional variables. Here, we give the equation of SUMCOR sparse multiset CCA as an example:

SUMCOR sparse multiset CCA:

$$\max_{u_1, \dots, u_K} \sum_{i,j,i \neq j}^K \text{corr}(\mathbf{Y}_i \mathbf{u}_i, \mathbf{Y}_j \mathbf{u}_j), \quad \text{s.t. } |\mathbf{u}_i|_1 < c_i. \quad (14)$$

**Formulation: Multiset CCA with reference.** Multiset CCA with reference enables the discovery of multimodal associations with a specific reference variable across subjects, such as a neuropsychological measurement (Qi, Calhoun, et al., 2018). In multiset CCA with reference, additional constraints of correlations between each canonical variable and the reference variable ( $\mathbf{v}_{\text{ref}}$ ) are added:

SUMCOR multiset CCA with ref:

$$\max_{u_1, \dots, u_K} \sum_{i,j,i \neq j}^K \left( \text{corr}(\mathbf{Y}_i \mathbf{u}_i, \mathbf{Y}_j \mathbf{u}_j) + \lambda \|\text{corr}(\mathbf{Y}_i \mathbf{u}_i, \mathbf{v}_{\text{ref}})\|_2^2 \right), \quad (15)$$

where  $\lambda > 0$  is the tuning parameter and  $\|\cdot\|_2^2$  is the  $L_2$ -norm. Therefore, multiset CCA with reference is a supervised multivariate technique that can extract common components across multiple variables that are associated with a specific prior reference.

**Solution.** Both Equations (14) and (15) can be viewed as constrained optimization problems with an objective function and multiple equality and inequality constraints. In this case, iterative optimization techniques are required to solve constrained multiset CCA problems.

## 3.2.4 | Other CCA-related techniques

There are many other CCA-related techniques developed, and here we only included three that have been applied in the neuroscience field: supervised local CCA, Bayesian CCA, and tensor CCA.

### Supervised local CCA

CCA by formulation is an unsupervised technique that uncovers joint relationships between two variables. Meanwhile, CCA can become a supervised technique by (a) adding additional constraints such as CCA (multiset CCA) with reference discussed in the section "Multiset CCA with constraints," or (b) directly incorporating group information into the objective function as in the supervised local CCA technique (Zhao et al., 2017).

Supervised local CCA is based on locally discriminant CCA (Peng, Zhang, & Zhang, 2010), which uses local group information to construct a between-set covariance matrix  $\tilde{\Sigma}_{12}$ , as a replacement of  $\Sigma_{12}$  in Equation (1). More specifically,  $\tilde{\Sigma}_{12}$  is defined as the covariance matrix from  $d$  nearest neighboring within-class samples ( $\Sigma_w$ ) penalized by the covariance from  $d$  nearest neighboring between-class samples ( $\Sigma_b$ ) with a tuning parameter  $\lambda$ ,

$$\tilde{\Sigma}_{12} = \Sigma_w - \lambda \Sigma_b. \quad (16)$$

However, this technique only considers the local group information with the global discriminating information ignored. To address this issue, Fisher discrimination information together with local group information is considered in supervised local CCA, which can be written as:

$$\text{Supervised local CCA: } \max_{u_1, u_2} \rho = \frac{u_1^T \tilde{\Sigma}_{12} u_2 + u_1^T S_1 u_1 + u_2^T S_2 u_2}{\sqrt{u_1^T \Sigma_{11} u_1} \sqrt{u_2^T \Sigma_{22} u_2}}, \quad (17)$$

$$S_k = Y_k^T U Y_k, k = 1, 2, U \in \mathbb{R}^{N \times N},$$

where  $S_k$  denote the between-group scatter matrices of the dataset  $k$ . If samples  $i$  and  $j$  belong to  $c$ th class,  $U_{ij}$  is set to  $\frac{1}{n_c}$ , where  $n_c$  denotes the number of samples in  $c$ th class; otherwise,  $U_{ij}$  is set to 0. Supervised local CCA is usually applied sequentially with gradually decreased  $d$  (named as hierarchical supervised local CCA) to reduce the influence of the neighborhood size and improve classification performance.

### Bayesian CCA

Bayesian CCA is another technique that overcomes the overfitting problem when applying CCA to variables with small sample sizes. Bayesian CCA is also proposed to complement CCA by providing a principal component analysis (PCA)-like description of variations that are not captured by the correlated components (Klami, Virtanen, & Kaski, 2013). Input to CCA in Equation (1),  $Y_1$  and  $Y_2$ , can be considered as  $N$  observations of one-dimensional random variables  $y_1 \in \mathbb{R}^{p_1 \times 1}$  and  $y_2 \in \mathbb{R}^{p_2 \times 1}$ . Using the same notations, Bayesian CCA can be formulated as a latent variable model (with latent variable  $z$ ) between  $y_1$  and  $y_2$  (Klami & Kaski, 2007; Wang, 2007):

$$\begin{aligned} z &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \\ y_k &\sim \mathcal{N}(\mathbf{A}_k z + \mathbf{B}_k z_k, \mathbf{D}_k), k = 1, 2, \end{aligned} \quad (18)$$

where  $\mathcal{N}(\mathbf{0}, \mathbf{I})$  denotes the multivariate Gaussian distribution with mean vector  $\mathbf{0}$  and identity covariance matrix  $\mathbf{I}$ .  $\mathbf{D}_k$  are diagonal covariance matrices and indicate features in  $y_k$  with independent noise. The latent variable  $z \in \mathbb{R}^{q \times 1}$ , where  $q$  represents the number of shared components, captures the shared variation between  $y_1$  and  $y_2$ , and can be linearly transformed back to the original space of  $y_k$  through  $\mathbf{A}_k z$ ,  $k = 1, 2$ . Similarly, the latent variable, where  $q_k$  represents the number of variable-specific components, captures the variable  $k$ -specific variation not shared between  $y_1$  and  $y_2$ , and can be linearly transformed back to the original space in  $y_k$  by  $\mathbf{B}_k z_k$ .

Browne (1979) demonstrated that Equation (18) was equivalent to CCA in Equation (1) by showing that maximum likelihood solutions to both Equations (1) and (18) share the same canonical coefficients with an unknown rotational transform, that is, Equation (18) is equivalent to conventional CCA (Equation (1)) in the aspect that their solutions share the same subspace. However, unlike conventional CCA (Equation (1)) that uses two variables  $u_1$  and  $u_2$  to project  $y_1$  and  $y_2$  to this subspace, Bayesian CCA maintains the shared variation between  $y_1$  and  $y_2$  in a single variable  $z$ .

The formulation of  $y_k$  in Equation (18) can be rewritten as  $y_k \sim \mathcal{N}(\mathbf{A}_k z, \mathbf{B}_k \mathbf{B}_k^T + \mathbf{D}_k)$ ,  $k = 1, 2$  after algebra operations. With  $\Psi_k = \mathbf{B}_k \mathbf{B}_k^T + \mathbf{D}_k$ , the model in Equation (18) can be transformed to

$$\begin{aligned} z &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \\ y_k &\sim \mathcal{N}(\mathbf{A}_k z, \Psi_k), k = 1, 2. \end{aligned} \quad (19)$$

In Equation (19), prior knowledge of the parameters (e.g.,  $\mathbf{A}_k$  and  $\Psi_k$ ) are required to construct the latent variable model for Bayesian CCA. For instance, the inverse Wishart distribution as a prior for the covariance  $\Psi_k$  and the automatic relevance determination (ARD; Neal, 2012) prior for the linear mappings  $\mathbf{A}_k$  are used when Bayesian CCA is proposed (Klami & Kaski, 2007; Wang, 2007). Since then, multiple Bayesian inference techniques have been developed, however, the early work of Bayesian CCA is limited to low-dimensional data (not more than eight dimensions in Klami & Kaski, 2007 and Wang, 2007) due to the computational complexity to estimate the posterior distribution over the  $p_k \times p_k$  covariance matrices  $\Psi_k$  (Klami et al., 2013). A group-wise ARD prior (Klami et al., 2013) was recently introduced for Bayesian CCA, which automatically identifies variable-specific and shared components. More importantly, this change made Bayesian CCA applicable for high-dimensional data. More technical details about Bayesian CCA can be found in Klami et al. (2013).

### Tensor CCA

*Two-dimensional CCA and tensor CCA for high-dimensional variables.* Variables input to CCA ( $Y_k \in \mathbb{R}^{N \times p_k}$ ,  $k = 1, 2, \dots$ ) are usually required to be 2D matrices with a dimension of number of observations ( $N$ ) times number of features ( $p_k$ ) in each variable.  $Y_k$  can be considered as  $N$  observations of the 1D variable  $y_k \in \mathbb{R}^{p_k \times 1}$ . In practice, tensor data, such as 3D images or 4D time series, are commonly involved in neuroscience applications, and these variables are required to be vectorized before inputting to CCA algorithms. This vectorization could potentially break the feature structures. In this case, to analyze 3D data, such as  $N$  samples of 2D variables ( $N \times p_1 \times p_2$ ), without breaking the 2D feature structure, two-dimensional CCA (2DCCA) has been proposed by Lee and Choi (2007).

Mathematically, 2DCCA maximizes the canonical correlation between two variables with  $N$  observations of 2D features:  $Y_1: \{Y_1^n \in \mathbb{R}^{p_{11} \times p_{12}}, n = 1, \dots, N\}$  and  $Y_2: \{Y_2^n \in \mathbb{R}^{p_{21} \times p_{22}}, n = 1, \dots, N\}$ . For each variable, 2DCCA searches left transforms  $l_1 \in \mathbb{R}^{p_{11} \times 1}$  and  $l_2 \in \mathbb{R}^{p_{21} \times 1}$  and right transforms  $r_1 \in \mathbb{R}^{p_{12} \times 1}$  and  $r_2 \in \mathbb{R}^{p_{22} \times 1}$  in order to maximize the correlation between  $l_1^T Y_1 r_1$  and  $l_2^T Y_2 r_2$ :

$$2D \text{ CCA: } \max_{l_1, l_2, r_1, r_2} \rho = \text{cov}(l_1^T Y_1 r_1, l_2^T Y_2 r_2), \text{ s.t. } \text{var}(l_1^T Y_1 r_1) = 1, \text{var}(l_2^T Y_2 r_2) = 1. \quad (20)$$

In Equation (20), for fixed  $l_1$  and  $l_2$ ,  $r_1$  and  $r_2$  can be obtained with the SVD algorithm similar to the one used in conventional CCA, and  $l_1$  and  $l_2$  can be obtained for fixed  $r_1$  and  $r_2$ , alternately. Therefore, an iterative alternating SVD algorithm (Lee & Choi, 2007) has been developed to solve Equation (20).

Above described 2DCCA can be treated as a constrained optimization problem with low-rank restrictions on canonical coefficients, similar



restrictions are used in (Chen, Kolar, & Tsay, 2019), where 2DCCA has been extended to higher dimensional tensor data, termed tensor CCA. The tensor CCA (Chen et al., 2019) searches two rank-one tensors  $\mathbf{u}_1 = \mathbf{u}_{11} \circ \dots \circ \mathbf{u}_{1m} \in \mathbb{R}^{p_{11} \times \dots \times p_{1m}}$  and  $\mathbf{u}_2 = \mathbf{u}_{21} \circ \dots \circ \mathbf{u}_{2m} \in \mathbb{R}^{p_{21} \times \dots \times p_{2m}}$  to maximize the correlation between  $\mathbf{Y}_1 : \{\mathbf{Y}_1^n \in \mathbb{R}^{p_{11} \times \dots \times p_{1m}}, n = 1, \dots, N\}$  and  $\mathbf{Y}_2 : \{\mathbf{Y}_2^n \in \mathbb{R}^{p_{21} \times \dots \times p_{2m}}, n = 1, \dots, N\}$ , where " $\circ$ " denotes outer product and  $\mathbf{u}_{k1}, \dots, \mathbf{u}_{km}$  are vectors. Chen et al. (2019) also introduced an efficient optimization algorithm to solve tensor CCA for high dimensional data sets.

*Tensor CCA for multiset data.* Another way to handle input variables with high-dimensional feature spaces is to generalize conventional CCA by analyzing constructed covariance tensors (Luo, Tao, Ramamohanarao, Xu, & Wen, 2015). This method requires random variables to be vectorized and is similar to multiset CCA since both of them deal with more than two input modalities. The differences between tensor CCA and multiset CCA in this case lie in that tensor CCA constructs a high-order covariance tensor for all input variables (Luo et al., 2015), whereas multiset CCA finds pair-wise covariance matrices. In addition, tensor CCA (Luo et al., 2015) does not maximize the pairwise correlation as in multiset CCA; instead, it directly maximizes the correlation over all canonical variables,

$$\begin{aligned} \max_{\mathbf{u}_1, \dots, \mathbf{u}_K} \rho = \text{Corr}(\mathbf{Y}_1 \mathbf{u}_1, \dots, \mathbf{Y}_K \mathbf{u}_K) &= (\mathbf{Y}_1 \mathbf{u}_1 \odot \dots \odot \mathbf{Y}_K \mathbf{u}_K)^T \\ &\times \mathbf{1}; \text{s.t. } (\mathbf{Y}_k \mathbf{u}_k)^T \mathbf{Y}_k \mathbf{u}_k = 1, k = 1, \dots, K, \end{aligned} \quad (21)$$

where  $\odot$  denotes element-wise product and  $\mathbf{1} \in \mathbb{R}^{N \times 1}$  is an all ones vector. The problem formulated in Equation (21) can be solved by using the alternating least square algorithm (Kroonenberg & de Leeuw, 1980).

### 3.2.5 | Statistical inferences of CCA variants

Nonparametric permutation tests have been widely performed in CCA variant techniques to determine the statistical significance of each canonical correlation value and the corresponding canonical coefficients. In these permutation tests, as we described in Section 3.1, observations of one variable are randomly shuffled ( $\mathbf{Y}_1$  becomes  $\widehat{\mathbf{Y}}_1$ ), while observations of the other variable are kept intact ( $\mathbf{Y}_2$  remains). This random shuffling is repeated multiple times ( $\sim 5,000$ ), and the exact same CCA variant technique is applied to each shuffled data. The obtained canonical correlation values from these randomly shuffled data form the null distribution. Statistical significances ( $p$ -values) of true canonical correlation values are determined by comparing true values to this null distribution.

Besides permutation tests, a null distribution can also be built by creating null data input to CCA variant techniques. The null data are usually generated based on the physical properties of input variables. For instance, when applying CCA-variant technique to link task fMRI data and the task stimuli, the null data of task fMRI can be obtained by applying wavelet-resampling to resting-state fMRI data (Breakspear,

Brammer, Bullmore, Das, & Williams, 2004; Zhuang et al., 2017). The null hypothesis here is that task fMRI data are not multivariately correlated with task stimuli, and the wavelet resampled resting-state fMRI data fits the requirements of the null data in this case.

## 3.3 | Technical differences

### 3.3.1 | Technical differences among CCA-related techniques

There are three prominent CCA techniques: conventional CCA shares the simplest formulation and can be easily applied to uncover multivariate linear relationships between two variables; nonlinear CCA by definition can extract multivariate nonlinear relationship between two variables through feature mapping with known predefined functions; and multiset CCA are able to find common covariates patterns among more than two variables. These three methods can be efficiently solved with closed-form analytical solutions, which are obtained by taking the partial derivatives of the objective function with respective to each unknown, separately.

Constrained (multiset) CCA incorporates prior information about input variables into each of the three CCA methods, in terms of equality and inequality constraints on the unknowns. Prior knowledge about the data or specific hypothesis are required for its applications. Closed-form solutions are no longer available for constrained (multiset) CCA and iterative optimization techniques are required to solve these problems.

Recently developed deep CCA is different from all other CCA-related techniques as it learns the optimum feature mapping from the data itself through deep learning with training and testing data being specified. Machine learning and deep learning expertise are required to solve this problem.

### 3.3.2 | Relationship between CCA and other multivariate and univariate techniques

#### Relationship with other multivariate techniques

In general, CCA can be directly rewritten in terms of the multivariate multiple regression (MVMR) model:

$$\mathbf{Y}_1 \mathbf{u}_1 = \mathbf{Y}_2 \mathbf{u}_2 + \boldsymbol{\varepsilon}, \quad (22)$$

where  $\mathbf{u}_1$  and  $\mathbf{u}_2$  are obtained by minimizing the residual term  $\boldsymbol{\varepsilon} \in \mathbb{R}^{N \times 1}$ . Since CCA is scale-invariant, a solution to Equation (22) is also a solution of Equation (1). Furthermore, with normalization terms of  $\mathbf{u}_1^T \boldsymbol{\Sigma}_{11} \mathbf{u}_1 = 1$  and  $\mathbf{u}_2^T \boldsymbol{\Sigma}_{22} \mathbf{u}_2 = 1$ , the MVMR model is exactly equivalent to CCA, that is, maximizing the canonical correlation between  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$  is equivalent to minimizing the residual term  $\boldsymbol{\varepsilon}$ :

$$\begin{aligned} \max_{\mathbf{u}_1, \mathbf{u}_2} (\text{corr}(\mathbf{Y}_1 \mathbf{u}_1, \mathbf{Y}_2 \mathbf{u}_2)) &\Leftrightarrow \max_{\mathbf{u}_1, \mathbf{u}_2} \mathbf{u}_1^T \boldsymbol{\Sigma}_{12} \mathbf{u}_2 \Leftrightarrow \min_{\mathbf{u}_1, \mathbf{u}_2} -\mathbf{u}_1^T \boldsymbol{\Sigma}_{12} \mathbf{u}_2 \\ &\Leftrightarrow \min_{\mathbf{u}_1, \mathbf{u}_2} \|\mathbf{Y}_1 \mathbf{u}_1 - \mathbf{Y}_2 \mathbf{u}_2\|_2^2. \end{aligned} \quad (23)$$

In addition, by replacing the covariance matrices  $\Sigma_{11}$  and  $\Sigma_{22}$  in the denominator in Equation (1) with the identity matrix  $I$ , conventional CCA is converted to partial least square (PLS), which maximizes the covariance between latent variables. If  $Y_1$  is the same as  $Y_2$ , the PLS will maximize the variance within a single variable, which is equivalent to PCA.

#### Relationship with univariate techniques

If one variable in CCA, for example,  $Y_1$ , only has a single feature, that is,  $y \in \mathbb{R}^{N \times 1}$ ,  $u_1$  can then be defined as 1 and CCA becomes a linear regression problem:

$$y = X\beta + \epsilon, \quad (24)$$

where  $Y_1$  is renamed as  $y$  and  $Y_2$  is renamed as  $X$  to follow conventional notations.  $\epsilon \in \mathbb{R}^{N \times 1}$  denotes the residual term. If both variables  $Y_1$  and  $Y_2$  contain only one feature, the canonical correlation between  $Y_1$  and  $Y_2$  becomes the Pearson's correlation between  $Y_1$  and  $Y_2$  as in the univariate analysis.

## 4 | NEUROSCIENCE APPLICATIONS

### 4.1 | CCA: Finding linear relationships

#### 4.1.1 | Direct application of CCA

##### Combine phenotypes and brain activities

To date, the most common CCA application in neuroscience is to find joint multivariate linear associations between phenotypic features and neurobiological activities. Phenotypic features usually include one or more measurements from demographics, genetic information, behavioral measurements, clinical symptoms, and performances of neuropsychological tests. Neurobiological activities are generally summarized with brain structural measurements, functional activations during specific tasks, both static and dynamic resting-state functional connectivity measurements, network topological measurements, and electrophysiological recordings (Table 1).

In normal healthy subjects, using CCA, multiple studies have delineated the joint multivariate relationships between the above imaging-derived features and nonimaging measurements, which have boosted our understandings of healthy development and healthy aging (Irimia & van Horn, 2013; Kuo et al., 2019; Shen et al., 2016; Tsvetanov et al., 2016). Furthermore, using multivariate CCA to combine imaging and nonimaging features have provided new insights to understand the joint relationship between brain activities and subjects' clinical symptoms, behavioral measurements, and performances of neuropsychological tests in various diseased populations, such as psychosis disease spectrum (Adhikari et al., 2019; Bai et al., 2019; Kottaram et al., 2019; Laskaris et al., 2019; Palaniyappan et al., 2019; Rodrigue et al., 2018; Tian et al., 2019; Viviano et al., 2018), Alzheimer's disease spectrum (Brier et al., 2016; Liao et al., 2010; McCrory & Ford, 1991; Zhu et al., 2016), neurodevelopmental diseases (Chenauksy et al., 2017; Lin,

Cocchi, et al., 2018; Zille et al., 2018), depression (Dinga et al., 2019), Parkinson's disease (Lin, Baumeister, Garg, and McKeown, 2018; Liu et al., 2018), multiple sclerosis (Leibach et al., 2016; Lin et al., 2017), epilepsy (Kucukboyaci et al., 2012) and drug addictions (Dell'Osso et al., 2014).

##### Brain activation in response to task stimuli

CCA has also been applied to detect brain activations in responses to stimuli during task-based fMRI experiments. Compared to the most commonly general linear regression model, local neighboring voxels are considered simultaneously in CCA to determine activation status of the central voxel (Friman, Cedefamn, Lundberg, Borga, & Knutsson, 2001; Nandy & Cordes, 2003; Nandy & Cordes, 2004; Rydell et al., 2006; Shams et al., 2006). In addition, in task-based electrophysiological experiments, Dmochowski et al. (2018) and de Cheveigne et al. (2018) have maximized the canonical correlation between an optimally transformed stimulus and properly filtered neural responses to delineate the stimulus-response relationship in electroencephalogram (EEG) data.

##### Denoising neuroscience data

Another application of CCA in neuroscience research is to remove noises from signals in the raw data. Through a blind source separation (BSS) framework, von Luhmann et al. (2019) extract comodulated canonical components between fNIRS signals and accelerometer signals, and consider those components above a canonical correlation threshold to be motion artifact. Through BSS-CCA algorithms, multiple studies demonstrate that muscle artifact can be efficiently removed from EEG signals (Hallez et al., 2009; Janani et al., 2020; Somers & Bertrand, 2016; Vergult et al., 2007). Furthermore, Churchill et al. (2012) remove physiological noise from fMRI signals through a CCA-based split-half resampling framework, and Li et al. (2017) remove gradient artifacts in concurrent EEG/fMRI recordings through maximizing the temporal autocorrelations of the time series.

##### Canonical granger causality

CCA has also been used to determine the causal relationship among regions of interest (ROIs) in fMRI functional connectivity analysis. Instead of using the mean ROI time series directly for analysis, multiple time series are specified for each ROI and CCA searches the optimally weighted mean time series during the analysis. Sato et al. (2010) compute multiple eigen-time series for each ROI and determine the granger causality between two ROIs by maximizing the canonical correlation between eigen-time series at time point  $t$  and  $t-1$  of the two ROIs. In a more recent work, instead of using eigen-time series of each ROI, Gulin et al. (2014) compute an optimized linear combination of signals from each ROI in CCA to enable a more accurate causality measurement.

### 4.1.2 | Practical considerations and data reduction steps

As we stated in Section 3.1, only if numbers of observations are more than numbers of features in both  $Y_1$  and  $Y_2$ , that is,  $N \gg p_k$ ,  $k = 1, 2$ ,

TABLE 1 CCA application

| CCA variant                | Modality 1   | Modality 2   | References  |
|----------------------------|--|--|---|
| CCA                        | Brain imaging data   | Clinical/behavioral/neuropsychological measurements              | Adhikari et al. (2019); Chenausky, Kernbach, Norton, and Schlaug (2017); Drysdale et al. (2017); Kottaram et al. (2019); Kucukboyaci et al. (2012); Kuo, Kutch, and Fisher (2019); Liao et al. (2010); Lin, Cocchi, et al. (2018); Lin, Vavasour, et al. (2018); Palaniyappan et al. (2019); Rodrigue et al. (2018); Shen et al. (2016); Tian, Zalesky, Bousman, Everall, and Pantelis (2019); Tsvetanov et al. (2016); Wee et al. (2017) |
|                            | Brain imaging data   | Brain imaging data   | Ashrafulla et al. (2013); Brier et al. (2016); Irimia and van Horn (2013); Li et al. (2017); Liu et al. (2018); Neumann et al. (2006); Palaniyappan et al. (2019); Viviano et al. (2018); Zhu, Suk, Lee, and Shen (2016)  |
|                            | Brain imaging data   | Task design  | El-Shabrawy et al. (2007); Nandy and Cordes (2003); Nandy and Cordes, (2004); Rydell, Knutsson, and Borga (2006); Shams, Hossein-Zadeh, and Soltanian-Zadeh (2006)  |
|                            | Electrophysiological data  | Clinical/behavioral measurements                                 | Abraham et al. (1996)   |
|                            | Electrophysiological data  | Electrophysiological data  | Brookes et al. (2014) (windowed-CCA), Ji (1999), McCrory and Ford (1991), Somers and Bertrand (2016), and Soto et al. (2016)  |
|                            | Electrophysiological data  | Stimulus   | de Cheveigne et al. (2018); Dmochowski, Ki, DeGuzman, Sajda, and Parra (2018)   |
|                            | Genetic information  | Clinical/behavioral measurements                                 | Laskaris et al. (2019); Kim, Won, Youn, and Park (2019);  |
|                            | Clinical/behavioral/demographics/neuropsychological measurements | Clinical/behavioral/demographics/neuropsychological measurements | Bedi et al. (2015); Dell'Oso et al. (2014); Gulin et al. (2014); Leibach, Stern, Arelis, Islas, and Barajas (2016); Lin et al. (2017); Lin, Cocchi, et al. (2018); Lin, Vavasour, et al. (2018); Lopez et al. (2017); Mirza et al. (2018); Valakos et al. (2018); Will et al. (2017)  |
|                            | Blind-source separation to denoise electrophysiological data     |  | Hallez et al. (2009); Janani et al. (2020); von Luhmann, Boukouvelas, Muller, and Adali (2019); Vergult et al. (2007)   |
|                            | Brain imaging data   | Clinical/behavioral/neuropsychological measurements              | Churchill et al. (2012); Hackmack et al. (2012); Li et al. (2019); Mihalik et al. (2019); Smith et al. (2015); Zarnani et al. (2019)  |
|                            | Brain imaging data   | Brain imaging data   | Abrol, Rashid, Rachakonda, Damaraju, and Calhoun (2017); Hirjak et al. (2019); Ouyang et al. (2015); Yang, Cao, et al. (2018); Yang, Zhuang, et al. (2018); Sato et al. (2010); Sui et al. (2010, 2011)   |
|                            | Brain imaging data   | Genetic data   | Bai, Zille, Hu, Calhoun, and Wang (2019); Zille, Calhoun, and Wang (2018)   |
|                            | Electrophysiological data  | Clinical/behavioral measurements                                 | Bologna et al. (2018)   |
| PCA/LASSO/regression + CCA |  |  |   |

Abbreviations: CAA, canonical correlation analysis; LASSO, least absolute shrinkage and selection operator; PCA, principal component analysis.

conventional CCA can produce statistically stable and meaningful results. However, in neuroscience applications, this requirement is not always fulfilled, especially when  $Y_1$  or  $Y_2$  represents brain activities where each brain voxel is considered a feature individually. In this case, any feature can be picked up and learned by the CCA process and directly applying Equation (1) to two sets will produce overfitted and unstable results. Therefore, additional data-reduction steps applied before CCA or constraints incorporated in the CCA algorithm are necessary to avoid overfitting in CCA applications. In this section, we focus on data reduction steps applied before conventional CCA.

The most commonly used data reduction technique is the PCA method applied to  $Y_1$  and  $Y_2$  separately. Through orthogonal transformation, PCA converts  $Y_1$  and  $Y_2$  into sets of linearly uncorrelated principal components. The principal components that do not pass certain criteria are discarded, leading to dimension-reduced variables:  $\tilde{Y}_1 \in \mathbb{R}^{N \times q_1}$  and  $\tilde{Y}_2 \in \mathbb{R}^{N \times q_2}$ , where  $N \gg q_k$ ,  $k = 1, 2$ . Equation (1) can then be applied to  $\tilde{Y}_1$  and  $\tilde{Y}_2$ . Multiple studies applied PCA to reduce data dimensions before applying CCA to find joint multivariate correlations between two high-dimensional variables (Abrol et al., 2017; Churchill et al., 2012; Hackmack et al., 2012; Li et al., 2019; Mihalik et al., 2019; Ouyang et al., 2015; Sato et al., 2010; Smith et al., 2015; Sui et al., 2010; Sui et al., 2011; Zarnani et al., 2019).

In addition, the least absolute shrinkage and selection operator (LASSO) algorithm (Tibshirani, 1996) has also been applied prior to CCA as a feature selection step to eliminate less informative features. For instance, in delineating the association between neurophysiological measures, which are derived from transcranial magnetic stimulation and electromyographic recordings, and kinematic-clinical-demographic measurements in Parkinson's disease subjects, Bologna et al. (2018) first perform logistic regression with LASSO penalty to determine the most predictive features for the disease in both variables. CCA is then applied to link the most predictive features from each variable. Similarly, sparse regression techniques have also been applied before CCA to genetic data in a neurodevelopmental cohort (Zille et al., 2018). Furthermore, feature selection can also be implemented in PCA as done in  $L_1$ -norm penalized sparse PCA (sPCA; Witten & Tibshirani, 2009; Yang, Zhuang, Bird, et al., 2019), which removes noninformative features during the dimension reduction step.

There is no single "correct" way or "gold standard" of the feature reduction step before applying CCA. Decisions should be made based on the data itself and the specific question that researchers are interested in.

## 4.2 | Constrained CCA: Removing noninformative features and stabilizing results

The other common solution in practice for  $N \ll p_k$ ,  $k = 1, 2$  is to incorporate constraints into the CCA algorithm directly, and consequently noninformative features can be removed and overfitting problems can be avoided (Table 2).

### 4.2.1 | Constraints in CCA algorithms: Sparse CCA to remove noninformative features

Most studies apply the sparse CCA method (detailed in the section "Special case:  $L_1$ -norm penalty and sparse CCA"), which maximizes canonical correlations between  $Y_1$  and  $Y_2$ , and suppresses noninformative features in  $Y_1$  and  $Y_2$  simultaneously (Badea et al., 2019; Lee et al., 2019; Moser et al., 2018; Pustina et al., 2018; Thyne & Mirman, 2018; Vatansever et al., 2017; Wang et al., 2018; Xia et al., 2018). The features determined to be noninformative are assigned with zero coefficients. Therefore, sparse CCA is particularly appropriate to combine modalities with large noise or substantial noninformative features, such as voxel-wise, regional-wise or connectivity-based brain features and genetic sequences (Avants et al., 2010; Deligianni et al., 2014; Du et al., 2017; Du, Liu, Yao, et al., 2019; Du, Zhang, et al., 2016; Duda et al., 2013; Gossmann et al., 2018; Grellmann et al., 2015; Jang et al., 2017; Kang et al., 2018; McMillan et al., 2014; Sheng et al., 2014; Sintini, Schwarz, Martin, et al., 2019; Sintini, Schwarz, Senjem, et al., 2019; Szefer et al., 2017; Wan et al., 2011). Rosa et al. (2015) further induce nonnegativity in the  $L_1$ -norm penalty in sparse CCA to investigate multivariate similarities between the effects of two antipsychotic drugs on cerebral blood flow using collected arterial spin labeling data.

Prior knowledge about  $Y_1$  and  $Y_2$  might also be available in neuroscience data. With known prior information of the feature dimension, structure-sparse CCA has been applied to associate brain activities with genetic information (Du et al., 2014; Du et al., 2015; Du, Huang, et al., 2016a; Du, Huang, et al., 2016b; Du, Liu, Zhang, et al., 2017; Kim et al., 2019; Lin et al., 2014; Liu et al., 2017; Yan et al., 2014), and to link structural and functional brain activities (Lisowska & Rekik, 2019; Mohammadi-Nejad et al., 2017). If prior knowledge is available of the observation dimension, such as memberships of diagnostic groups, discriminant sparse CCA is applied to investigate joint relationship between brain activities and genetic information for subjects with Schizophrenia disease spectrum (Fang et al., 2016) or Alzheimer's disease spectrum (Wang et al., 2019; Yan et al., 2017). Longitudinal data could also be collected in neuroscience research and are useful to monitor disease progression. Temporal constrained sparse CCA has been proposed to uncover how single nucleotide polymorphisms affect brain gray matter density across multiple time points in subjects with Alzheimer's disease spectrum (Du, Liu, Zhu, et al., 2019; Hao, Li, Yan, et al., 2017).

### 4.2.2 | Constraints in CCA algorithm: Constrained CCA to stabilize results

Multiple constraints have also been proposed in CCA applications to stabilize CCA coefficients between brain activities and clinical symptoms. For instance, to avoid overfitting between fNIRS signals during a moral judgment task and psychopathic personality inventory scores in healthy adults, Dashtestani et al. (2019) introduce a regularization parameter  $\lambda$  to keep the canonical coefficients small and to avoid high

**TABLE 2** Constrained CCA application

| CCA variant                  | Modality 1                      | Modality 2  | Reference   |
|------------------------------|---------------------------------|---|---|
| Sparse CCA (L1-norm penalty) | Brain imaging data              | Clinical/behavioral/neuropsychological measurements | Badea et al. (2019); Lee, Moser, Ing, Doucet, and Frangou (2019); Moser et al. (2018); Pustina, Avants, Faseyitan, Medaglia, and Coslett (2018); Thyne and Mirman (2018); Vatansever et al. (2017); Wang et al. (2018); Xia et al. (2018)   |
|                              | Brain imaging data              | Brain imaging data                                  | Avants, Cook, Ungar, Gee, and Grossman (2010); Deligianni, Carmichael, Zhang, Clark, and Clayden (2016); Deligianni, Centeno, Carmichael, and Clayden (2014); Duda, Detre, Kim, Gee, and Avants (2013); Jang et al. (2017); Kang, Kwak, Yoon, and Lee (2018); Rosa et al. (2015); Sintini, Schwarz, Martin et al. (2019); Sintini, Schwarz, Senjem, et al. (2019) |
|                              | Brain imaging data              | Genetic information                                 | Du et al. (2016); Du, Liu, Yao, et al. (2019); Du, Liu, Zhu, et al. (2019); Grellmann et al. (2015); Gossmann, Zille, Calhoun, and Wang (2018); McMillan et al. (2014); Sheng et al. (2014); Szefer, Lu, Nathoo, Beg, and Graham (2017); Wan et al. (2011)  |
|                              | Genetic information             | Clinical/behavioral/measurements                    | Leonenko et al. (2018)  |
| Structure-sparse CCA         | Brain imaging data              | Brain imaging data                                  | Lisowska and Rekik (2019); Mohammadi-Nejad, Hossein-Zadeh, and Soltanian-Zadeh (2017)   |
|                              | Brain imaging data              | Genetic information                                 | Du et al. (2014, 2015, 2016a, 2016b); Du et al. (2017); Kim et al. (2019); Liu et al. (2017); Lin, Calhoun, and Wang, 2014; Yan et al. (2014)   |
| Discriminant sparse CCA      | Brain imaging data              | Genetic information/blood data                      | Fang et al. (2016); Wang, Shao, Hao, Shen, and Zhang (2019); Yan, Risacher, Nho, Saykin, and Shen (2017)  |
| Constrained CCA              | Brain imaging data              | Clinical/behavioral/neuropsychological measurements | Grosenick et al. (2019); Dashtestani et al. (2019)  |
|                              | Brain imaging data              | Task design   | Cordes, Jin, Curran, and Nandy, (2012a, 2012b); Dong et al. (2015); Friman, Borge, Lundberg, and Knutsson (2003); Zhuang et al. (2017); Zhuang et al. (2019)  |
| Other constraints in CCA     | Longitudinal brain imaging data | Genetic information                                 | Du, Liu, Zhu, et al. (2019) (temporal multitask sparse CCA); Hao et al. (2017) (temporal group sparse CCA);   |

Abbreviation: CCA, canonical correlation analysis.

bias problem. Similarly, in preclinical research, Grosenick et al. (2019) uses two regularization parameters  $\lambda_1$  and  $\lambda_2$  to penalize the estimated covariance matrices for the resting-state functional connectivity features and Hamilton Rating Scale for Depression clinical symptoms, respectively.

Furthermore, as we stated in Section 4.1.1, CCA has been applied to detect brain activations in response to task stimuli during fMRI experiments. In these type of applications,  $Y_1$  represents time series from local neighborhood that is considered simultaneously in determining the activation status of the central voxels, and  $Y_2$  represents

the task design matrix. CCA is applied to find optimized coefficients  $u_1$  and  $u_2$ , such that the correlation between combined local voxels and task design is maximized. In this case, even though the central voxel may be inactivated in the task, activated neighboring voxels would lead to a high canonical correlation and thus produce falsely activated status of the central voxel, which is termed as smoothing artifact (Cordes et al., 2012a). To eliminate this artifact and to uncover real activation status, multiple constraints have been applied to  $u_1$  to guarantee the dominant effect of the central voxel in a local neighborhood (Cordes et al., 2012b; Dong et al., 2015; Friman et al., 2003;

Zhuang et al., 2017; Zhuang et al., 2019). Yang, Zhuang, et al. (2018) further extend the constraints from two-dimensional local neighborhood to three-dimensional neighboring voxels.

### 4.3 | Kernel CCA: Focusing on a nonlinear relationship between two modalities

Above CCA applications assume joint linear relationships between two modalities; however, this assumption might not always hold in neuroscience research. Kernel CCA has been proposed to uncover the nonlinear relationship between modalities without explicitly specifying the nonlinear feature space (Equation (8)). In human research, kernel CCA has been applied to investigate the joint nonlinear relationship between simultaneously collected fMRI and EEG data (Yang, Cao, et al., 2018), to uncover gene–gene co-association in Schizophrenia subjects (Ashad Alam et al., 2019), and to detect brain activations in response to fMRI tasks (Hardoon et al., 2007; Yang, Zhuang, et al., 2018). In preclinical research, temporal kernel CCA has been proposed to investigate the temporal-delayed nonlinear relationship between simultaneously recorded neural (electrophysiological recording in frequency-time space) and hemodynamic (fMRI in voxel space) signals in monkeys (Murayama et al., 2010), and to investigate a nonlinear predictive relationship between EEG signals from two different brain regions in macaques (Rodu et al., 2018) (Table 3).

### 4.4 | Multiset CCA: More than two modalities

Multiset CCA has been specifically proposed to find common multivariate patterns across  $K$  modalities, with  $K > 2$ . The widest application of multiset CCA in neuroscience research is to uncover *covariates* patterns among demographics, clinical characteristics, behavioral measurements and multiple brain activities, including structural MRI derived measurements (gray matter, white matter, and cerebrospinal fluid densities), diffusion weighted MRI derived measurements (myelin water fraction and white matter tracts), fMRI derived measurements (static and dynamic functional connectivity, task fMRI activations, amplitude of low frequency contributions) and PET derived measurements (standardized uptake values) (Baumeister et al., 2019; Langers et al., 2014; Lerman-Sinkoff et al., 2017; Lerman-Sinkoff et al., 2019;

Lin, Vavasour, et al., 2018; Lottman et al., 2018; Stout et al., 2018; Sui et al., 2013; Sui et al., 2015) (Table 4).

Multiset CCA has also been applied to group analysis, which combines data from multiple subjects within a single modality. In this type of applications, data from each subject are treated as one modality, and multiset CCA is used to uncover common patterns in fMRI data (Afshin-Pour et al., 2012; Afshin-Pour et al., 2014; Correa, Adali, et al., 2010; Varoquaux et al., 2010), consistent signals in electrophysiological recordings (Koskinen & Seppa, 2014; Lankinen et al., 2014; Lankinen et al., 2016; Lankinen et al., 2018; Zhang et al., 2017), covaried components in fNIRS data (Liu & Ayaz, 2018), and correlated fMRI and EEG signals (Correa, Eichele, et al., 2010) across multiple subjects.

Sparse multiset CCA has been applied to combine more than two variables and remove noninformative features simultaneously. Specifically, sparse multiset CCA has been applied to combine multiple brain imaging modalities with genetic information (Hao et al., 2017; Hu et al., 2016; Hu et al., 2018).

Multiset CCA with reference is specifically proposed as a supervised multimodal fusion technique in neuroscience research. Using neuropsychological measurements such as working memory or cognitive measurements as the reference, studies have uncovered stable covariates patterns among fractional amplitude of low frequency contribution maps derived from resting-state fMRI, gray matter volumes derived from structural MRI and fractional anisotropy maps derived from diffusion-weighted MRI that are linked with and can predict core cognitive deficits in schizophrenia (Qi, Calhoun, et al., 2018; Sui et al., 2018). Using genetic information as a prior reference, multiset CCA with reference has also uncovered multimodal covariates MRI biomarkers that are associated with microRNA132 in medication-naïve major depressive patients (Qi, Yang, et al., 2018). Furthermore, with clinical depression rating score as guidance, Qi et al. (2020) have demonstrated that the electroconvulsive therapy Hdepressive disorder patients produces a covariates remodeling in brain structural and functional images, which is unique to an antidepressant symptom response. As a supervised technique, multiset CCA can be applied to uncover covariates patterns across multiple variables of special interest.

### 4.5 | Other applications

CCA has also been applied in a supervised and hierarchical fashion. Zhao et al. (2017) have performed supervised local CCA with

**TABLE 3** Nonlinear Kernel CCA applications

| CCA variant         | Modality 1                                  | Modality 2          | Reference   |
|---------------------|---|---------------------|---|
| Kernel CCA          | Brain imaging data                          | Brain imaging data  | Yang, Cao, et al. (2018)  |
|                     | Brain imaging data                          | Task design         | Hardoon, Mourão-Miranda, Brammer, and Shawe-Taylor (2007); Yang, Zhuang, et al. (2018)    |
|                     | Genetic information                         | Genetic information | Ashad Alam, Komori, Deng, Calhoun, and Wang (2019)  |
| Temporal kernel CCA | Simultaneously recorded multiple modalities |                     | John et al. (2017); Murayama et al. (2010); Rodu, Klein, Brincat, Miller, and Kass (2018) |

Abbreviation: CCA, canonical correlation analysis.



**TABLE 4** Multiset CCA applications

| CCA variant                 |  | Detailed modalities   | Reference  |
|-----------------------------|--|---|--|
| Multiset CCA                | Combine multiple brain imaging data                | rsfMRI + task fMRI + sMRI   | Lerman-Sinkoff et al. (2017); Lerman-Sinkoff, Kandala, Calhoun, Barch, and Mamah (2019)  |
|                             |  | sMRI (WM + GM + CSF) + rsfMRI   | Lottman et al. (2018)  |
|                             |  | sMRI + fMRI + dMRI  | Sui et al. (2013, 2015)  |
|                             |  | Multiple task fMRI  | Langers, Krumbholz, Bowtell, and Hall (2014)   |
|                             |  | sMRI + fMRI + EEG   | Correa, Adali, Li, and Calhoun (2010)  |
|                             | Combine brain imaging data and other information   | Brain imaging data (sMRI/fMRI) + neuropsychological measurements + clinical/behavioral measurements | Baumeister et al. (2019); Lin, Cocchi, et al. (2018); Lin, Vavasour, et al. (2018)   |
|                             |  | Brain imaging data (PET + sMRI + fMRI) + neuropsychological measurements                            | Stout et al. (2018)  |
|                             | Combine multiple subjects within a single modality | Sub1 + Sub2 + ... + SubN within a single modality   | Afshin-Pour, Hossein-Zadeh, Strother, and Soltanian-Zadeh (2012); Afshin-Pour, Grady, and Strother (2014); Correa, Adali, et al. (2010); Gaebler et al. (2014); Koskinen and Seppa (2014); Lankinen, Saari, Hari, and Koskinen (2014); Lankinen et al. (2016, 2018); Liu and Ayaz (2018); Varoquaux et al. (2010); Zhang, Borst, Kass, and Anderson (2017) |
|                             | Combine multiple subjects from two modalities      | Sub1 + Sub2+ ... + SubN from fMRI and EEG   | Correa, Eichele, Adali, Li, and Calhoun (2010)   |
|                             | Combine multiple ROIs within a single modality     | ROI1 + ROI2 + ... + ROIN within a single modality   | Deleus et al. (2011)   |
| Constraints in multiset CCA | Sparse multiset CCA                                | Brain imaging data + genetic information + clinical measurements                                    | Hu, Lin, Calhoun, and Wang (2016); Hu et al. (2018); Yu et al. (2015)  |
|                             | Multiset CCA with reference                        | Brain imaging data (fMRI + sMRI + dMRI) with neuropsychological measurements as reference           | Qi et al. (2020), Qi, Calhoun, et al. (2018); Sui et al. (2018)  |
|                             |  | Brain imaging data (fMRI + sMRI + dMRI) with genetic information as reference                       | Qi, Yang, et al. (2018)  |

Abbreviations: CCA, canonical correlation analysis; CSF, cerebrospinal fluid; dMRI, diffusion-weighted MRI; EEG, electroencephalogram; GM, gray matter; MRI, magnetic resonance imaging; PET, position emission tomography; ROI, regions of interest; rsfMRI, resting-state functional MRI; sMRI, structural MRI; Sub, subject; WM, white matter.

gradually varying neighborhood sizes in early autism diagnosis, and in each iteration, CCA is used to combine canonical variates from the previous step (Table 5).

Bayesian CCA has been used to realign fMRI activation data between actors and observers during simple motor tasks to investigate whether seeing and performing an action activates similar brain areas (Smirnov et al., 2017). The Bayesian CCA assigns brain activations to one of three types (actor-specific, observer-specific and shared) via a group-wise sparse ARD prior. Furthermore, using Bayesian CCA, Fujiwara et al. (2013) establish mappings between the stimulus and the brain by automatically extracting modules from measured fMRI data, which can be used to generate effective prediction models for encoding and decoding.

More recently, in network neuroscience, Graa and Rekik (2019) propose a multiview learning-based data proliferator that enables the classification of imbalanced multiview representations. In their proposed approach, tensor-CCA is used to align all original and proliferated views into a shared subspace for the target classification.

**TABLE 5** Other CCA applications

| CCA variant          | CCA application                          | Reference   |
|----------------------|--|---|
| Supervised local CCA | Combine two modalities                   | Zhao, Qiao, Shi, Yap, and Shen (2017)                               |
| Tensor CCA           | Morphological networks                   | Graa and Rekik (2019)   |
| Bayesian CCA         | Realign fMRI data from multiple subjects | Smirnov et al. (2017)   |
|                      | Task fMRI activation detection           | Fujiwara, Miyawaki, and Kamitani (2013)                             |
| Others               | Toolbox                                  | Bilenko and Gallant (2016)  |
|                      | Reviews                                  | Liu and Calhoun (2014) and Sui, Adali, Yu, Chen, and Calhoun (2012) |

Abbreviations: CCA, canonical correlation analysis; fMRI, functional magnetic resonance imaging.

## 5 | ADVANTAGES AND LIMITATIONS OF EACH CCA TECHNIQUE IN NEUROSCIENCE APPLICATIONS

Table 6 explains the advantages and limitations of each CCA and its variant techniques.

### 5.1 | Canonical correlation analysis

#### 5.1.1 | Advantages

CCA can be applied easily to two variables and solved efficiently in closed-form using algebraic methods (Equation (3)). In CCA, the intermodality relationship is assumed to be linear and both modalities are exchangeable and treated equally. Canonical correlations are invariant to linear transforms of features in  $Y_1$  or  $Y_2$ . In neuroscience research, CCA uncovers the joint multivariate linear relationship between two modalities and has proven to be an effective multivariate and data-driven analysis method.

#### 5.1.2 | Limitations

CCA assumes and uncovers only a linear intermodality relationship, which might not hold for neuroscience data. Furthermore, directly applying CCA requires sufficient observation support of the variables (detailed in Section 3.1). For neuroscience data, especially voxel-wise brain imaging data, it is usually difficult to have more observations (e.g., subjects) than features (e.g., voxels). In this case, any feature in  $Y_1$  and  $Y_2$  can be picked up and learned by the CCA process, and directly applying CCA will produce overfitted and unstable results. ROI-based analysis, data reduction (e.g., PCA), and feature selection (e.g., LASSO) steps are commonly applied to reduce the number of features in neuroscience data prior to CCA.

Another limitation of CCA in general is that signs of the canonical correlations and canonical coefficients are indeterminate. Solving the eigenvalue problem in Equation (3) will always give a positive canonical correlation value, and reversing the signs of  $u_1$  and  $u_2$  simultaneously will lead to the same canonical correlation value. Therefore, with CCA, we can only conclude that two modalities are linearly and

**TABLE 6** Advantages and limitations of each CCA-related technique

| Category        | CCA variant                 | Advantages   | Limitations   |
|-----------------|-----------------------------|--|---|
| CCA             | CCA                         | 1) Has closed-form analytical solution<br>2) Easy to apply<br>3) Invariant to scaling                                    | 1) Requires $N \gg p_k, k = 1, 2$<br>2) Signs of canonical correlations are indeterminate   |
| Constrained CCA | Sparse CCA                  | 1) Removes noninformative features and solves $N \ll p_k$<br>2) Performs reasonably with high-dimensional-co-linear data | Requires optimization expertise   |
|                 | Structure sparse CCA        | Removes noninformative features, solving $N \ll p_k$ with prior information about the data                               | 1) Improves effectiveness of sparse CCA.<br>2) Produces biological meaningful results   |
|                 | Discriminant sparse CCA     |  | Discovers group discriminant features   |
|                 | Generalized constrained CCA | 1) Reduces false positives<br>2) Maintains most of the variance in a stable model  | 1) Requires optimization expertise<br>2) Requires predefined constraints  |
| Nonlinear CCA   | Kernel CCA                  | 1) Finds nonlinear relationship among modalities<br>2) Has analytical solution   | 1) Requires predefined kernel functions   |
|                 | Temporal kernel CCA         | Most appropriate to simultaneously collect data from two modalities with time delay                                      | 2) Difficult to project from kernel space back to original feature space, leading to difficulties in interpretation<br>3) Only linear kernel space can be projected back to the original feature space. |
|                 | Deep CCA                    | 1) Finds unknown nonlinear relationship<br>2) Purely data-driven   | 1) Requires deep learning expertise<br>2) Requires large number of training samples (in tens of thousands)  |
| Multiset CCA    | Multiset CCA                | 1) Good for more than two modalities<br>2) Good for group analysis   | 1) Requires predefined objective functions  |
|                 | Sparse multiset CCA         | 1) Good for more than two modalities<br>2) Removes noninformative features and solves $N \ll p_k$                        | 2) The number of final canonical components does not represent the intersected common patterns across all modalities  |
|                 | Multiset CCA with reference | Supervised fusion technique to link common patterns with a prior known variable  |   |

Abbreviation: CCA, Canonical correlation analysis.

multivariately correlated without determining the direction of the linear relationship.

## 5.2 | Constrained CCA

### 5.2.1 | Advantages

Incorporating constraints in CCA can in general avoid overfitted and unstable results in CCA. More specifically, different constraints can benefit neuroscience research in various ways.

Sparse CCA incorporates the  $L_1$ -norm penalty on the canonical coefficients  $u_k$ ,  $k = 1, 2$  such that noninformative features are automatically removed by suppressing their weights. Thus, sparse CCA is suitable for high-dimensional co-linear data, such as whole-brain voxel-wise activities or genetic data. In practice, the within-modality covariance matrices  $\Sigma_{kk}$ ,  $k = 1, 2$  are replaced with the identity matrix  $I$  in sparse CCA, since estimating  $\Sigma_{kk}$  from the high-dimensional collinear data are both memory and time consuming. This replacement saves both computation time and physical resources, and is widely adopted in the neuroscience field.

Structure and discriminant sparse CCA removes noninformative features and incorporates prior information about the data in the algorithms simultaneously. Prior knowledge about feature structure or group assignment of each observation are required, respectively, for these two techniques. In neuroscience applications, information implanted in features can improve the performance and effectiveness of sparse CCA (Du, Liu, Zhang, et al., 2017) and guide the algorithm to produce more biologically meaningful results (Du, Huang, et al., 2016a; Liu et al., 2017). Alternatively, with group assignments implanted in each observation, discriminant sparse CCA is able to discover group discriminant features, which can later improve the performance of supervised classification (Wang et al., 2019).

Other constraints are also beneficial in neuroscience research. For instance, the  $L_2$ -norm penalty on canonical coefficients retains all features in the model with regularized weights, and therefore most of the variance can be maintained in a stable model (Dashtestani et al., 2019). In addition, when applied to task fMRI activation detection, locally constrained CCA penalizes weights on the neighboring voxels to guarantee the dominance of the central voxel and therefore, is able to reduce false positives (Cordes et al., 2012b; Zhuang et al., 2017).

### 5.2.2 | Limitations

One major limitation of constrained CCA is the requirement of expertise in optimization techniques. By having additional penalty terms on canonical coefficients or covariance matrices, analytical solutions of constrained CCA no longer exist, and, instead, iterative optimization methods are required to solve the constrained CCA problems efficiently.

The predefined constraint itself also requires prior knowledge about the data. For structure and discriminant sparse CCA, prior

information about the observation domain or the feature domain is required. Furthermore, in neuroscience application, the constraint itself is usually data specific. For instance, when applying local constrained CCA to task fMRI activation detection, the predefined constraint should be strong enough to penalize neighboring voxels, but loose enough to guarantee the multivariate contribution of neighboring voxels to the central voxel. This constraint can only be selected through simulating a series of synthetic data that mimic real fMRI signals, which requires prior knowledge of the data and is time-consuming.

## 5.3 | Nonlinear CCA

### 5.3.1 | Advantages

By definition, nonlinear CCA is able to uncover multivariate nonlinear relationships between two modalities, which commonly exist in neuroscience variables. For instance, during an fMRI task, collected fMRI signals are nonlinearly correlated with the task design due to the unknown hemodynamic response function; and kernel CCA can extract this multivariate nonlinear relationship and produce a localized brain activation map (Hardoon et al., 2007).

In general, kernel CCA first implicitly transforms the original feature space into a kernel space with a predefined kernel function. With this transform, nonlinear relationship between two modalities can be discovered. Furthermore, in the new kernel space, kernel CCA can be solved efficiently with a closed-form analytical solution.

Temporal kernel CCA shares similar advantages with kernel CCA, with additional benefits from considering temporal delays between modalities when applied to simultaneously collected data. In neuroscience research, simultaneously collected EEG/fMRI data are a typical candidate for temporal kernel CCA, as neural activities collected by fMRI data, which are the blood oxygenated level-dependent signals, contain temporal delays caused by the hemodynamic response function (Ogawa, Lee, Kay, & Tank, 1990), as compared to the simultaneously collected EEG signals.

Deep CCA, a purely data-driven technique, can reveal unknown nonlinear relationships between variables without assuming any predefined nonlinear intermodality relationship. It has the potential to be applied to neuroscience data that contains enough samples for training a deep learning schema.

### 5.3.2 | Limitations

For kernel CCA, a predefined kernel function needs to be selected and this selection will affect final results. This choice of kernel functions requires additional knowledge about data and the kernel function. Another major limitation of both kernel CCA and temporal kernel CCA is that it is difficult to project the kernel space ( $H_1$  and  $H_2$ ) back to the original feature space ( $Y_1$  and  $Y_2$ ), leading to additional difficulties in interpreting results (Hardoon et al., 2007). For instance, when applying

kernel CCA to link fMRI task stimuli and collected BOLD signals for activation detection, the obtained high-dimensional features cannot be mapped backwards to an individual voxel in order to assign the activation value because the feature embedded for commonly used nonlinear kernels (e.g., Gaussian kernel and power kernel) have information from multiple voxels. Therefore, kernel CCA with a general nonlinear kernel remains unsolved for fMRI activation analysis, and only linear kernels were used for constructing activation maps in fMRI.

Unlike kernel CCA, deep CCA does not require a predefined function and learns the nonlinear feature mapping from the data itself. However, in deep CCA, the number of unknown parameters significantly increases with the number of layers, which requires much more samples in the training data. In neuroscience data, it is usually difficult to have enough number of subjects as training samples for deep CCA. Furthermore, deep learning expertise is also required for selecting the appropriate deep learning structures for nonlinear feature mapping.

## 5.4 | Multiset CCA

### 5.4.1 | Advantages

In neuroscience research, more than two variables are commonly collected for the same set of subjects. Multiset CCA uncovers multivariate joint relationships among multiple variables, which is well defined to link all collected data in this case. Furthermore, if data from one subject are treated as one modality (or variable), multiset CCA will also discover the common patterns across subjects, which becomes a powerful data-driven group analysis method.

Sparse multiset CCA combines more than two modalities and suppresses noninformative features simultaneously, and therefore shares the advantages and limitations with both multiset CCA and sparse CCA.

Multiset CCA with reference is the only supervised CCA technique and is proposed specifically for neuroscience applications. It discovers joint multivariate relationships among variables in response to a specific reference variable. For instance, using this method, common brain changes from structural, fMRI and diffusion MRI with respect to a specific neuropsychological measurement can be discovered.

### 5.4.2 | Limitations

There are five possible objective functions for multiset CCA optimization, and different objective functions will lead to various results. The closed-form analytical solution only exists for SUMCOR and SSQCOR objective functions. Optimization expertise are required to solve multiset CCA with other objective functions, and with constraints as well. Another major limitation of multiset CCA is that the number of final canonical components output from the algorithm does not represent the intersected common patterns across all modalities, or subjects. Instead, multiset CCA discovers the unified similarities among every modality pair (Levin-Schwartz, Song, Schreier, Calhoun, & Adali, 2016).

## 5.5 | Abstract

To summarize, conventional CCA uncovers joint multivariate linear relationships between two modalities and can be quickly and easily applied. In neuroscience research, due to the existing multiple modalities and nonlinear intermodality relationships, multiset CCA and nonlinear CCA have their own advantages when applied accordingly to appropriate variables. Constraints can be applied in these three methods to stabilize results, remove noninformative features, and produce supervised meaningful results. However, optimization expertise and prior knowledge about the data are required to select the appropriate constraints.

## 6 | CHOOSING THE APPROPRIATE CCA TECHNIQUE

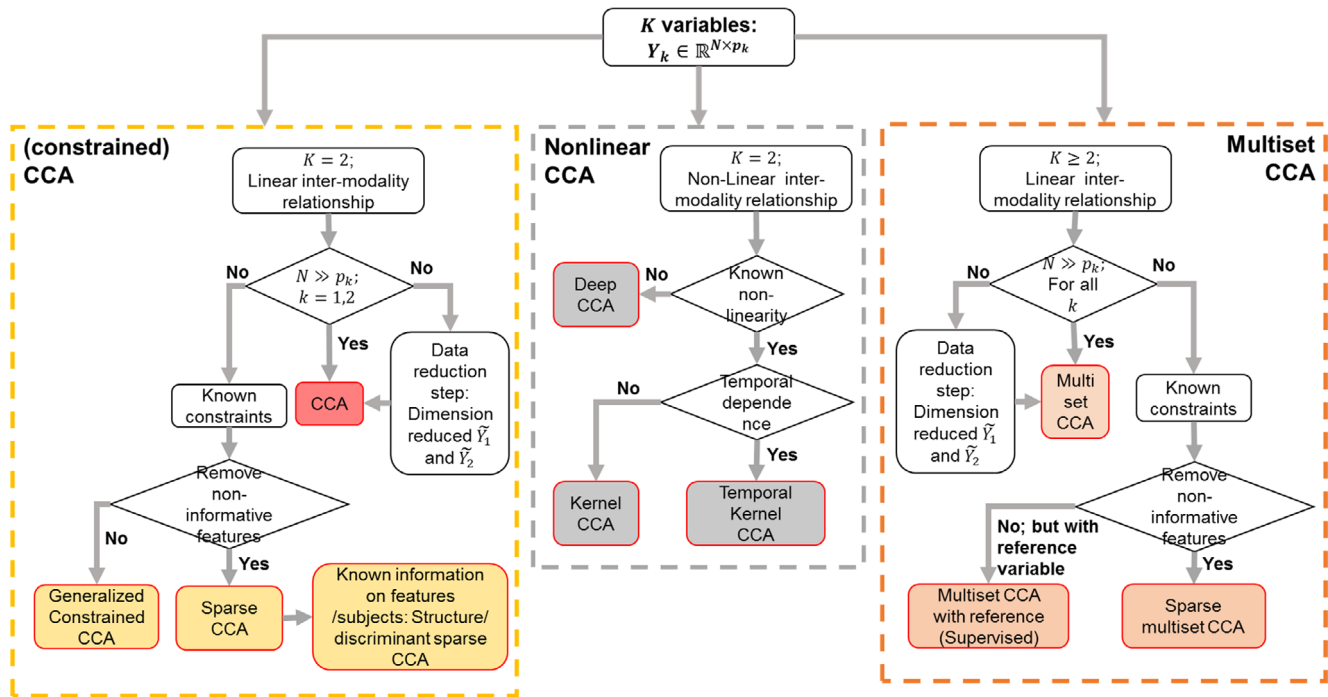
The first step in selecting a CCA technique is to decide what type of neuroscience application is of interest. Based on the types of combined modalities, CCA applications can be summarized into four categories (a–d): (a) finding relationship among multiple measurements; (b) detecting brain activations in response to task stimuli; (c) uncovering common patterns among multiple subjects; and (d) denoising the raw data. Table 7 summarizes current and potential techniques that can be applied for each application.

After determining the application of interest, the flowchart in Figure 4 provides a detailed guidance in selecting an appropriate CCA technique. Based on the number of variables ( $K$ ) and linear or nonlinear intermodality relationships, three major applications are mostly common in neuroscience research: uncover linear relationship

**TABLE 7** Current applied and potential CCA techniques for each application

| Applications                                       | Currently applied   | Potential techniques  |
|--|---|---|
| Link two modalities                                | <ul style="list-style-type: none"> <li>CCA</li> <li>Sparse CCA</li> <li>Structure/discriminant sparse CCA</li> <li>Kernel CCA</li> <li>Temporal kernel CCA</li> </ul> | <ul style="list-style-type: none"> <li>Deep CCA</li> </ul>  |
| Detect task fMRI activation                        | <ul style="list-style-type: none"> <li>CCA</li> <li>Constrained CCA</li> <li>Kernel CCA</li> </ul>  | <ul style="list-style-type: none"> <li>Deep CCA</li> <li>Sparse CCA</li> </ul>                          |
| Uncover common patterns across multiple modalities | <ul style="list-style-type: none"> <li>Multiset CCA</li> <li>Sparse multiset CCA</li> </ul>   | <ul style="list-style-type: none"> <li>Multiset constrained CCA</li> <li>Deep CCA</li> </ul>            |
| Denoise raw data                                   | <ul style="list-style-type: none"> <li>CCA</li> </ul>   | <ul style="list-style-type: none"> <li>Constrained CCA</li> <li>Kernel CCA</li> <li>Deep CCA</li> </ul> |

Abbreviations: CCA, canonical correlation analysis; fMRI, functional magnetic resonance imaging.



**FIGURE 4** Selecting a canonical correlation analysis (CCA)-technique that suits your application. Three scenarios are most commonly encountered in neuroscience applications: CCA with and without constraints (dashed yellow box); nonlinear CCA (dashed gray box) and multiset CCA (dashed orange box)

between two variables (dashed yellow box); find nonlinear relationship between two variables (dashed gray box) and discover covariated patterns among more than two variables (dashed orange box). Detailed choices are further made based on the number of observations and number of features within each variable, known prior knowledge about the variable, such as feature structures, and specific questions of interest for research studies.

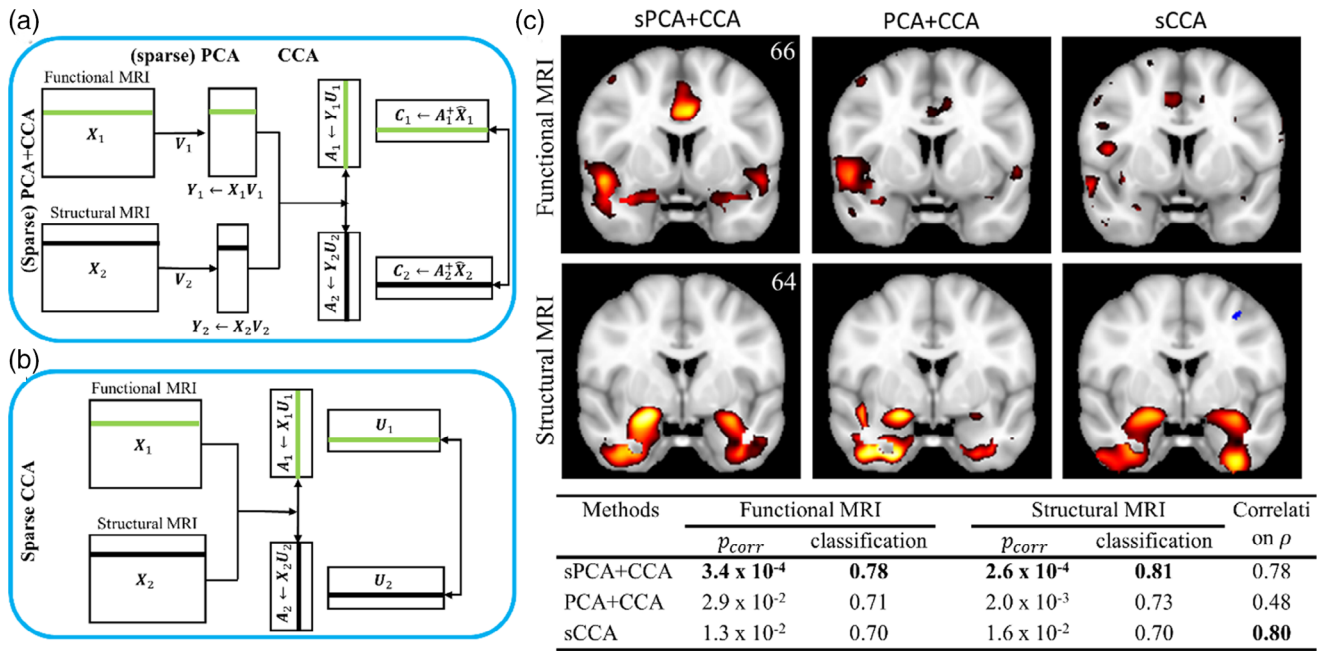
Furthermore, here, we give an experimental example of CCA applications in neuroscience research.

Among many neuroscience applications, CCA is commonly used as a data fusion technique to uncover the association between two datasets. In the following, we demonstrate how to follow the guidance in Figure 4 to link disease-related pathology using fMRI and structural MRI data from cognitive normal subjects and subjects with mild cognitive impairment (MCI). As a prodromal stage of Alzheimer's disease, both functional and structural pathology are expected in MCI subjects. Yang, Zhuang, Bird, et al. (2019) used CCA to examine the disease-related links between voxel-wise functional information (e.g., eigenvector centrality mapping from fMRI data,  $\mathbf{X}_1 \in \mathbb{R}^{N \times p_1}$ ) and voxel-wise structural information (e.g., voxel-based morphometry from T1 structural MRI data,  $\mathbf{X}_2 \in \mathbb{R}^{N \times p_2}$ ), where  $N$  is the number of subjects, and  $p_1$  and  $p_2$  are the number of voxel-wise features for fMRI and structural MRI data, respectively. Since there are only two imaging modalities in the analysis, multiset CCA is not an option for this case. Considering that deep CCA requires a large number of samples but  $N \ll p_1$  or  $p_2$ , and kernel CCA has the difficulty to project coefficients back to original voxel-wise feature space as mentioned in Section 5.3, a linear relationship

between these two imaging modalities is considered. There are two approaches for the scenario that the number of samples is much less than the number of features.

The first approach is to perform dimension reduction before feeding data into conventional CCA as shown in Figure 5a. Yang, Zhuang, Bird, et al. (2019) used PCA or sPCA (Witten et al., 2009) for dimension reduction and then fed CCA with dimension-reduced data  $\mathbf{Y}_1$  and  $\mathbf{Y}_2$ . CCA found a set of canonical coefficients  $\mathbf{U}_k$ ,  $k = 1, 2$  and the corresponding canonical variables  $\mathbf{A}_k$ . The voxel-wise weight coefficient can be obtained with a pseudo inverse operation. The other approach is to implement constrained CCA as shown in Figure 5b. With the assumption that a proportion of voxels in the brain is not informative for finding the association between fMRI and structural MRI data, sparse CCA was applied with  $\mathbf{X}_1$  and  $\mathbf{X}_2$  directly without dimension reduction step (Yang, Zhuang, Bird, et al., 2019). The canonical coefficients  $\mathbf{U}_k$ ,  $k = 1, 2$  are in the voxel-wise feature space, thus no operation is required to calculate voxel-wise weight coefficients.

The voxel-wise weight coefficients play a role in uncovering which brain regions are most relevant for finding the association between datasets. The voxel-wise weight maps for the most significant disease-related component in  $\mathbf{A}_k$  for (s)PCA + CCA and sparse CCA is shown in Figure 5c. A nonparametric permutation test is applied to test the significance of the association between fMRI and structural MRI data with  $p$  values shown at the bottom of Figure 5c. In this study, the canonical variables  $\mathbf{A}_k$  computed from sPCA + CCA have the highest classification accuracy for both fMRI and structural MRI data.



**FIGURE 5** Example of choosing canonical correlation analysis (CCA) variants by following the guideline. Voxel-wise functional and structural MRI information from cognitive normal subjects and subjects with mild cognitive impairment were used for data fusion analysis. (a) Schematic diagram of (sparse) principal component analysis (PCA) + CCA. The abbreviation sPCA stands for sparse PCA. (b) Schematic diagram of sparse CCA (sCCA). (c) Top panel shows the most disease-discriminant functional and structural component and the bottom panel shows the correlation between datasets ( $\rho$ ), the significance of the correlation derived from nonparametric permutation test ( $p_{corr}$ ) and the classification accuracy for each method

## 7 | FUTURE DIRECTION OF CCA IN NEUROSCIENCE APPLICATIONS

Currently, when applying CCA to data with a smaller number of observations than features, either a data reduction or feature selection step is performed as a preprocessing step, or an  $L_1$  norm penalty is added as a constraint to remove noninformative features. Future efforts should be made toward incorporating prior information on feature structures of input variables that are more reasonable or more biologically meaningful, and canonical correlation values should be computed in a one step process that includes prior information. Furthermore, applying CCA and its variant techniques to uncover joint multivariate relationships between two modalities has dominated the current CCA applications in the neuroscience field. In these applications, various techniques have been proposed to incorporate prior information within variables to boost the model performance, such as considering group-discriminant features to strengthen group separation. However, much less effort was put to incorporate these prior information within the variables in multiset CCA. In neuroscience research, collecting multiple modalities of a single subject has become a commonplace, and with more than two variables, multiset CCA should be considered for this multimodal data-fusion more often. Future efforts toward incorporating prior information within each variable to further improve the performance of multiset CCA could shed new lights in neuroscience research. For instance, we suggest incorporating group information in multiset CCA to extract common group-discriminant patterns among

multiple measurements derived from fMRI, or to uncover correlated group-discriminant feature among brain imaging data and behavioral or clinical measurements. Furthermore, nonlinear relationships among multiple modalities have not been explored within multiset CCA in neuroscience research. It might be of interest to incorporate kernels in multiset CCA to uncover covariates nonlinear patterns among multiple brain imaging data, or to input each variable through multiple layers to generate “deep” features before applying multiset CCA.

In addition, future efforts are also required to statistically interpret CCA results. Currently, a parametric statistical significance of CCA model is only well defined for conventional CCA. Statistical significances of CCA variants are usually determined nonparametrically through permutation tests, which are time-consuming and methods dependent. Furthermore, even using permutation tests, statistical significance can only be determined for each canonical correlation value, instead of canonical coefficients. Therefore, we cannot determine the statistical significance of a specific feature in the model. Identifying important features as potential biomarkers is usually an end goal in neuroscience. Therefore, developing test statistics to interpret CCA results by determining statistically important features would also benefit neuroscience research.

## 8 | CONCLUSION

Uncovering multivariate relationships between modalities of the same subjects have gained significant attentions in neuroscience research.



CCA is a powerful tool to investigate these joint associations and has been widely applied. Multiple CCA-variant techniques have been proposed to fulfill specific analysis requirements. In this study, we reviewed CCA and its variant techniques from a technical perspective, with summarized applications in neuroscience research. Of each CCA-related technique, detailed formulation and solution, relationship with other techniques, current applications, advantages, and limitations are provided. Selecting the most appropriate CCA-related technique to take full advantages of available information embedded in every variable in joint multimodal research might shed new lights in our understandings of normal development, aging, and disease processes.

## 9 | CODE AVAILABILITY

Python-based CCA toolbox (Bilenko & Gallant, 2016) is available on github: <http://github.com/gallantlab/pyrcca>; CCA package in R can be found in González, Déjean, Martin, and Baccini (2008). Codes for applying CCA and kernel CCA to detect task-fMRI activations are available on github (Yang, Zhuang, et al., 2018; Zhuang et al., 2017): [https://github.com/pipiyang/CCA\\_GUI](https://github.com/pipiyang/CCA_GUI). Bayesian CCA with group-wise ARD prior and the relevant techniques are implemented in R CCAGFA package (<https://cran.r-project.org/web/packages/CCAGFA/index.html>).

## ACKNOWLEDGMENTS

The study is supported by the National Institute of Health (grants 1R01EB014284); Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health, Grant/Award Number: 5P20GM109025; The Keep Memory Alive Foundation Young Scientist Award; A private grant from the Peter and Angela Dal Pezzo funds; A private grant from Lynn and William Weidner; A private grant from Stacie and Chuck Matthewson.

## DATA AVAILABILITY STATEMENT

There is no data or code involved in this review article.

## ORCID

Xiaowei Zhuang  <https://orcid.org/0000-0002-0581-6530>

Zhengshi Yang  <https://orcid.org/0000-0002-9796-9680>

Dietmar Cordes  <https://orcid.org/0000-0001-6574-5546>

## REFERENCES

- Abrol, A., Rashid, B., Rachakonda, S., Damaraju, E., & Calhoun, V. D. (2017). Schizophrenia shows disrupted links between brain volume and dynamic functional connectivity. *Frontiers in Neuroscience*, 11(624). <https://doi.org/10.3389/fnins.2017.00624>
- Abraham, H. D., & Duffy, F. H. (1996). Stable quantitative EEG difference in post-LSD visual disorder by split-half analysis: evidence for disinhibition. *Psychiatry Research*, 67, 173–187. [https://doi.org/10.1016/0925-4927\(96\)02833-8](https://doi.org/10.1016/0925-4927(96)02833-8)
- Adhikari, B. M., Hong, L. E., Sampath, H., Chiappelli, J., Jahanshad, N., Thompson, P. M., ... Kochunov, P. (2019). Functional network connectivity impairments and core cognitive deficits in schizophrenia. *Human Brain Mapping*, 40, 4593–4605. <https://doi.org/10.1002/hbm.24723>
- Afshin-Pour, B., Grady, C., & Strother, S. (2014). Evaluation of spatio-temporal decomposition techniques for group analysis of fMRI resting state data sets. *NeuroImage*, 87, 363–382. <https://doi.org/10.1016/j.neuroimage.2013.10.062>
- Afshin-Pour, B., Hossein-Zadeh, G.-A., Strother, S. C., & Soltanian-Zadeh, H. (2012). Enhancing reproducibility of fMRI statistical maps using generalized canonical correlation analysis in NPAIRS framework. *NeuroImage*, 60, 1970–1981. <https://doi.org/10.1016/j.neuroimage.2012.01.137>
- Andrew, G., Arora, R., Bilmes, J., & Livescu, K. (2013). Deep canonical correlation analysis. In *International conference on machine learning* (pp. 1247–1255).
- Ashad Alam, M., Komori, O., Deng, H.-W., Calhoun, V. D., & Wang, Y.-P. (2019). Robust kernel canonical correlation analysis to detect gene-gene co-associations: A case study in genetics. *Journal of Bioinformatics and Computational Biology*, 17, 1950028. <https://doi.org/10.1142/S0219720019500288>
- Ashrafulla, S., Haldar, J. P., Joshi, A. A., & Leahy, R. M. (2013). Canonical Granger causality between regions of interest. *NeuroImage*, 83, 189–199. <https://doi.org/10.1016/j.neuroimage.2013.06.056>
- Avants, B. B., Cook, P. A., Ungar, L., Gee, J. C., & Grossman, M. (2010). Dementia induces correlated reductions in white matter integrity and cortical thickness: A multivariate neuroimaging study with sparse canonical correlation analysis. *NeuroImage*, 50, 1004–1016. <https://doi.org/10.1016/j.neuroimage.2010.01.041>
- Badea, A., Delpratt, N. A., Anderson, R. J., Dibb, R., Qi, Y., Wei, H., ... Colton, C. (2019). Multivariate MR biomarkers better predict cognitive dysfunction in mouse models of Alzheimer's disease. *Magnetic Resonance Imaging*, 60, 52–67. <https://doi.org/10.1016/j.mri.2019.03.022>
- Bai, Y., Zille, P., Hu, W., Calhoun, V. D., & Wang, Y.-P. (2019). Biomarker identification through integrating fMRI and epigenetics. *IEEE Transactions on Biomedical Engineering*, 67, 1186–1196. <https://doi.org/10.1109/TBME.2019.2932895>
- Bartlett, M. S. (1939). A note on tests of significance in multivariate analysis. *Mathematical Proceedings of the Cambridge Philosophical Society*, 35, 180–185.
- Baumeister, T. R., Lin, S.-J. J., Vavasour, I., Kolind, S., Kosaka, B., Li, D. K. B., ... McKeown, M. J. (2019). Data fusion detects consistent relations between non-lesional white matter myelin, executive function, and clinical characteristics in multiple sclerosis. *NeuroImage: Clinical*, 24, 101926. <https://doi.org/10.1016/j.nicl.2019.101926>
- Baxter, L. C., Sparks, D. L., Johnson, S. C., Lenoski, B., Lopez, J. E., Connor, D. J., & Sabbagh, M. N. (2006). Relationship of cognitive measures and gray and white matter in Alzheimer's disease. *Journal of Alzheimer's Disease*, 9, 253–260. <https://doi.org/10.3233/JAD-2006-9304>
- Bedi, G., Carrillo, F., Cecchi, G. A., Slezak, D. F., Sigman, M., Mota, N. B., et al. (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. *NPJ Schizophrenia*, 1, 15030.
- Bilenko, N. Y., & Gallant, J. L. (2016). Pyrcca: Regularized kernel canonical correlation analysis in Python and its applications to neuroimaging. *Frontiers in Neuroinformatics*, 10(49). <https://doi.org/10.3389/fninf.2016.00049>
- Bologna, M., Guerra, A., Paparella, G., Giordo, L., Fegatelli, D. A., Vestri, A. R., ... Berardelli, A. (2018). Neurophysiological correlates of bradykinesia in Parkinson's disease. *Brain*, 141, 2432–2444. <https://doi.org/10.1093/brain/awy155>
- Brookes, M. J., O'Neill, G. C., Hall, E. L., Woolrich, M. W., Baker, A., Palazzo Corner, S., et al. (2014). Measuring temporal, spectral and spatial changes in electrophysiological brain network connectivity. *NeuroImage*, 91, 282–299. <https://doi.org/10.1016/j.neuroimage.2013.12.066>
- Breakspear, M., Brammer, M. J., Bullmore, E. T., Das, P., & Williams, L. M. (2004). Spatiotemporal wavelet resampling for functional neuroimaging

- data. *Human Brain Mapping*, 23, 1–25. <https://doi.org/10.1002/hbm.20045>
- Brier, M. R., McCarthy, J. E., Benzinger, T. L. S., Stern, A., Su, Y., Friedrichsen, K. A., ... Vlassenko, A. G. (2016). Local and distributed PIB accumulation associated with development of preclinical Alzheimer's disease. *Neurobiology of Aging*, 38, 104–111. <https://doi.org/10.1016/j.neurobiolaging.2015.10.025>
- Brooke, A., Kendrick, D., Meeraus, A., & Rama, R. (1998). GAMS: A user's guide (p. 1998). Washington, DC: GAMS Development Corp.
- Browne, M. W. (1979). The maximum-likelihood solution in inter-battery factor analysis. *The British Journal of Mathematical and Statistical Psychology*, 32, 75–86.
- Chen, Y.-L., Kolar, M., & Tsay, R. S. (2019). Tensor canonical correlation analysis. *arXiv. Prepr arXiv190605358*.
- Chenausky, K., Kernbach, J., Norton, A., & Schlaug, G. (2017). White matter integrity and treatment-based change in speech performance in minimally verbal children with autism spectrum disorder. *Frontiers in Human Neuroscience*, 11(175). <https://doi.org/10.3389/fnhum.2017.00175>
- Churchill, N. W., Yourganov, G., Spring, R., Rasmussen, P. M., Lee, W., Ween, J. E., & Strother, S. C. (2012). PHYCAA: Data-driven measurement and removal of physiological noise in BOLD fMRI. *NeuroImage*, 59, 1299–1314. <https://doi.org/10.1016/j.neuroimage.2011.08.021>
- Cordes, D., Jin, M., Curran, T., & Nandy, R. (2012a). The smoothing artifact of spatially constrained canonical correlation analysis in functional MRI. *International Journal of Biomedical Imaging*, 2012, 1–11. <https://doi.org/10.1155/2012/738283>
- Cordes, D., Jin, M., Curran, T., & Nandy, R. (2012b). Optimizing the performance of local canonical correlation analysis in fMRI using spatial constraints. *Human Brain Mapping*, 33, 2611–2626. <https://doi.org/10.1002/hbm.21388>
- Correa, N. M., Adali, T., Li, Y., & Calhoun, V. D. (2010). Canonical correlation analysis for data fusion and group inferences. *IEEE Signal Processing Magazine*, 27, 39–50. <https://doi.org/10.1109/MSP.2010.936725>. Canonical
- Correa, N. M., Eichele, T., Adali, T., Li, Y.-O., & Calhoun, V. D. (2010). Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI. *NeuroImage*, 50, 1438–1445. <https://doi.org/10.1016/j.neuroimage.2010.01.062>
- Cremers, H. R., Wager, T. D., & Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PLoS One*, 12, 1–20. <https://doi.org/10.1371/journal.pone.0184923>
- Dashtestani, H., Zaragoza, R., Pirsivash, H., Knutson, K. M., Kermanian, R., Cui, J., ... Gandjbakhche, A. (2019). Canonical correlation analysis of brain prefrontal activity measured by functional near infra-red spectroscopy (fNIRS) during a moral judgment task. *Behavioural Brain Research*, 359, 73–80. <https://doi.org/10.1016/j.bbr.2018.10.022>
- de Cheveigne, A., Wong, D. D. E., Di Liberto, G. M., Hjortkjaer, J., Slaney, M., & Lalor, E. (2018). Decoding the auditory brain with canonical component analysis. *NeuroImage*, 172, 206–216. <https://doi.org/10.1016/j.neuroimage.2018.01.033>
- Deleus, F., & Van Hulle, M. M. (2011). Functional connectivity analysis of fMRI data based on regularized multiset canonical correlation analysis. *Journal of Neuroscience Methods*, 197, 143–157.
- Deligianni, F., Carmichael, D. W., Zhang, G. H., Clark, C. A., & Clayden, J. D. (2016). NODDI and tensor-based microstructural indices as predictors of functional connectivity. *PLoS One*, 11, 1–17. <https://doi.org/10.1371/journal.pone.0153404>
- Deligianni, F., Centeno, M., Carmichael, D. W., & Clayden, J. D. (2014). Relating resting-state fMRI and EEG whole-brain connectomes across frequency bands. *Frontiers in Neuroscience*, 8(258). <https://doi.org/10.3389/fnins.2014.00258>
- Dell'Osso, L., Rugani, F., Maremmani, A. G. I., Bertoni, S., Pani, P. P., & Maremmani, I. (2014). Towards a unitary perspective between post-traumatic stress disorder and substance use disorder. Heroin use disorder as case study. *Comprehensive Psychiatry*, 55, 1244–1251. <https://doi.org/10.1016/j.comppsy.2014.03.012>
- Dinga, R., Schmaal, L., Penninx, B. W. J. H., van Tol, M. J., Veltman, D. J., van Velzen, L., ... Marquand, A. F. (2019). Evaluating the evidence for biotypes of depression: Methodological replication and extension of. *NeuroImage: Clinical*, 22, 101796. <https://doi.org/10.1016/j.nicl.2019.101796>
- Dmochowski, J. P., Ki, J. J., DeGuzman, P., Sajda, P., & Parra, L. C. (2018). Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity. *NeuroImage*, 180, 134–146. <https://doi.org/10.1016/j.neuroimage.2017.05.037>
- Dong, L., Zhang, Y., Zhang, R., Zhang, X., Gong, D., Valdes-Sosa, P. A., ... Yao, D. (2015). Characterizing nonlinear relationships in functional imaging data using eigenspace maximal information canonical correlation analysis (emiCCA). *NeuroImage*, 109, 388–401. <https://doi.org/10.1016/j.neuroimage.2015.01.006>
- Drud, A. (1985). CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems. *Mathematical Programming*, 31, 153–191.
- Du, L., Huang, H., Yan, J., Kim, S., Risacher, S., Inlow, M., ... Shen, L. (2016a). Structured sparse CCA for brain imaging genetics via graph OSCAR. *BMC Systems Biology*, 10(Suppl 3), 68. <https://doi.org/10.1186/s12918-016-0312-1>
- Du, L., Huang, H., Yan, J., Kim, S., Risacher, S. L., Inlow, M., ... Shen, L. (2016b). Structured sparse canonical correlation analysis for brain imaging genetics: An improved GraphNet method. *Bioinformatics*, 32, 1544–1551. <https://doi.org/10.1093/bioinformatics/btw033>
- Du, L., Jingwen, Y., Kim, S., Risacher, S. L., Huang, H., Inlow, M., ... Shen, L. (2014). A novel structure-aware sparse learning algorithm for brain imaging genetics. *Medical Image Computing and Computer-Assisted Intervention*, 17, 329–336. [https://doi.org/10.1007/978-3-319-10443-0\\_42](https://doi.org/10.1007/978-3-319-10443-0_42)
- Du, L., Liu, K., Yao, X., Risacher, S. L., Guo, L., Saykin, A. J., & Shen, L. (2019). Diagnosis status guided brain imaging genetics via integrated regression and sparse canonical correlation analysis. *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 2019, 356–359. <https://doi.org/10.1109/ISBI.2019.8759489>
- Du, L., Liu, K., Yao, X., Yan, J., Risacher, S. L., Han, J., ... Shen, L. (2017). Pattern discovery in brain imaging genetics via SCCA modeling with a generic non-convex penalty. *Scientific Reports*, 7, 14052. <https://doi.org/10.1038/s41598-017-13930-y>
- Du, L., Liu, K., Zhang, T., Yao, X., Yan, J., Risacher, S. L., ... Shen, L. (2017). A novel SCCA approach via truncated l1-norm and truncated group Lasso for brain imaging genetics. *Bioinformatics*, 34, 278–285. <https://doi.org/10.1093/bioinformatics/btx594>
- Du, L., Liu, K., Zhu, L., Yao, X., Risacher, S. L., Guo, L., ... Shen, L. (2019). Identifying progressive imaging genetic patterns via multi-task sparse canonical correlation analysis: A longitudinal study of the ADNI cohort. *Bioinformatics*, 35, i474–i483. <https://doi.org/10.1093/bioinformatics/btz320>
- Du, L., Yan, J., Kim, S., Risacher, S. L., Huang, H., Inlow, M., ... Shen, L. (2015). GN-SCCA: GraphNet based sparse canonical correlation analysis for brain imaging genetics. *Brain Informatics Heal 8th Int Conf BIH 2015*, London, UK, August 30–September 2, 2015 proceedings BIH (8th 2015 London, England) 9250, 275–284.
- Du, L., Zhang, T., Liu, K., Yao, X., Yan, J., Risacher, S. L., ... Shen, L. (2016). Sparse canonical correlation analysis via truncated l1-norm with application to brain imaging genetics. *Proceedings IEEE International Conference on Bioinformatics and Biomedicine*, 2016, 707–711. <https://doi.org/10.1109/BIBM.2016.7822605>
- Duda, J. T., Detre, J. A., Kim, J., Gee, J. C., & Avants, B. B. (2013). Fusing functional signals by sparse canonical correlation analysis improves network reproducibility. *Medical Image Computing and Computer-Assisted Intervention*, 16, 635–642. [https://doi.org/10.1007/978-3-642-40760-4\\_79](https://doi.org/10.1007/978-3-642-40760-4_79)
- Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., et al. (2017). Resting-state connectivity biomarkers define

- neurophysiological subtypes of depression. *Nature Medicine*, 23, 28–38. <https://doi.org/10.1038/nm.4246>
- El-Shabrawy, N., Mohamed, A. S., Youssef, A.-B. M., & Kadah, Y. M. (2007). Activation detection in functional MRI using model-free technique based on CCA-ICA analysis. In 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, 2007 (pp. 3430–3433). <https://doi.org/10.1109/IEMBS.2007.4353068>
- Fang, J., Lin, D., Schulz, S. C., Xu, Z., Calhoun, V. D., & Wang, Y.-P. (2016). Joint sparse canonical correlation analysis for detecting differential imaging genetics modules. *Bioinformatics*, 32, 3480–3488. <https://doi.org/10.1093/bioinformatics/btw485>
- Friman, O., Borga, M., Lundberg, P., & Knutsson, H. (2003). Adaptive analysis of fMRI data. *NeuroImage*, 19, 837–845. [https://doi.org/10.1016/S1053-8119\(03\)00077-6](https://doi.org/10.1016/S1053-8119(03)00077-6)
- Friman, O., Cedefamn, J., Lundberg, P., Borga, M., & Knutsson, H. (2001). Detection of neural activity in functional MRI using canonical correlation analysis. *Magnetic Resonance in Medicine*, 45, 323–330. [https://doi.org/10.1002/1522-2594\(200102\)45:2<323::AID-MRM1041>3.0.CO;2-#](https://doi.org/10.1002/1522-2594(200102)45:2<323::AID-MRM1041>3.0.CO;2-#)
- Fujiwara, Y., Miyawaki, Y., & Kamitani, Y. (2013). Modular encoding and decoding models derived from Bayesian canonical correlation analysis. *Neural Computation*, 25, 979–1005. [https://doi.org/10.1162/NECO\\_a\\_00423](https://doi.org/10.1162/NECO_a_00423)
- Gaebler, M., Biessmann, F., Lamke, J.-P., Muller, K.-R., Walter, H., & Hetzer, S. (2014). Stereoscopic depth increases intersubject correlations of brain networks. *NeuroImage*, 100, 427–434. <https://doi.org/10.1016/j.neuroimage.2014.06.008>
- González, I., Déjean, S., Martin, P. G. P., & Baccini, A. (2008). CCA: An R package to extend canonical correlation analysis. *Journal of Statistical Software*, 23, 1–14. <https://doi.org/10.18637/jss.v023.i12>
- Gossmann, A., Zille, P., Calhoun, V., & Wang, Y.-P. (2018). FDR-corrected sparse canonical correlation analysis with applications to imaging genomics. *IEEE Transactions on Medical Imaging*, 37, 1761–1774. <https://doi.org/10.1109/TMI.2018.2815583>
- Graa, O., & Rekik, I. (2019). Multi-view learning-based data proliferator for boosting classification using highly imbalanced classes. *Journal of Neuroscience Methods*, 327, 108344. <https://doi.org/10.1016/j.jneumeth.2019.108344>
- Grellmann, C., Bitzer, S., Neumann, J., Westlye, L. T., Andreassen, O. A., Villringer, A., & Horstmann, A. (2015). Comparison of variants of canonical correlation analysis and partial least squares for combined analysis of MRI and genetic data. *NeuroImage*, 107, 289–310. <https://doi.org/10.1016/j.neuroimage.2014.12.025>
- Grosenick, L., Shi, T. C., Gunning, F. M., Dubin, M. J., Downar, J., & Liston, C. (2019). Functional and Optogenetic approaches to discovering stable subtype-specific circuit mechanisms in depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4, 554–566. <https://doi.org/10.1016/j.bpsc.2019.04.013>
- Gulin, S. L., Perrin, P. B., Stevens, L. F., Villaseñor-Cabrera, T. J., Jimenez-Maldonado, M., Martinez-Cortes, M. L., & Arango-Lasprilla, J. C. (2014). Health-related quality of life and mental health outcomes in Mexican TBI caregivers. *Families, Systems & Health*, 32, 53–66. <https://doi.org/10.1037/a0032623>
- Hackmack, K., Weygandt, M., Wuerfel, J., Pfueller, C. F., Bellmann-Strobl, J., Paul, F., & Haynes, J.-D. (2012). Can we overcome the “clinico-radiological paradox” in multiple sclerosis? *Journal of Neurology*, 259, 2151–2160. <https://doi.org/10.1007/s00415-012-6475-9>
- Hallez, H., de Vos, M., Vanrumste, B., van Hese, P., Asseconci, S., van Laere, K., ... Lemahieu, I. (2009). Removing muscle and eye artifacts using blind source separation techniques in ictal EEG source imaging. *Clinical Neurophysiology*, 120, 1262–1272. <https://doi.org/10.1016/j.clinph.2009.05.010>
- Hao, X., Li, C., Du, L., Yao, X., Yan, J., Risacher, S. L., ... Zhang, D. (2017). Mining outcome-relevant brain imaging genetic associations via three-way sparse canonical correlation analysis in Alzheimer's disease. *Scientific Reports*, 7, 44272. <https://doi.org/10.1038/srep44272>
- Hao, X., Li, C., Yan, J., Yao, X., Risacher, S. L., Saykin, A. J., ... Zhang, D. (2017). Identification of associations between genotypes and longitudinal phenotypes via temporally-constrained group sparse canonical correlation analysis. *Bioinformatics*, 33, i341–i349. <https://doi.org/10.1093/bioinformatics/btx245>
- Hardoon, D. R., Mourão-Miranda, J., Brammer, M., & Shawe-Taylor, J. (2007). Unsupervised analysis of fMRI data using kernel canonical correlation. *NeuroImage*, 37, 1250–1259. <https://doi.org/10.1016/j.neuroimage.2007.06.017>
- Hardoon, D. R., Szedmak, S., & Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 16, 2639–2664. <https://doi.org/10.1162/0899766042321814>
- Hirjak, D., Rashidi, M., Fritze, S., Bertolino, A. L., Geiger, L. S., Zang, Z., et al. (2019). Patterns of co-altered brain structure and function underlying neurological soft signs in schizophrenia spectrum disorders. *Human Brain Mapping*, 40, 5029–5041.
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28, 321–377.
- Hu, W., Lin, D., Calhoun, V. D., & Wang, Y.-P. (2016). Integration of SNPs-FMRI-methylation data with sparse multi-CCA for schizophrenia study. Conf Proc. Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf. 2016, 3310–3313.
- Hu, W., Lin, D., Cao, S., Liu, J., Chen, J., Calhoun, V. D., & Wang, Y.-P. (2018). Adaptive sparse multiple canonical correlation analysis with application to imaging (epi)genomics study of schizophrenia. *IEEE Transactions on Biomedical Engineering*, 65, 390–399. <https://doi.org/10.1109/TBME.2017.2771483>
- Irimia, A., & van Horn, J. D. (2013). The structural, connectomic and network covariance of the human brain. *NeuroImage*, 66, 489–499. <https://doi.org/10.1016/j.neuroimage.2012.10.066>
- Janani, A. S., Grummett, T. S., Bakhshayesh, H., Lewis, T. W., DeLosAngeles, D., Whitham, E. M., ... Pope, K. J. (2020). Fast and effective removal of contamination from scalp electrical recordings. *Clinical Neurophysiology*, 131, 6–24. <https://doi.org/10.1016/j.clinph.2019.09.016>
- Jang, H., Kwon, H., Yang, J.-J., Hong, J., Kim, Y., Kim, K. W., ... Lee, J.-M. (2017). Correlations between gray matter and White matter degeneration in pure Alzheimer's disease, pure subcortical vascular dementia, and mixed dementia. *Scientific Reports*, 7, 9541. <https://doi.org/10.1038/s41598-017-10074-x>
- Ji, J., Porjesz, B., Begleiter, H., & Chorlian, D. (1999). P300: the similarities and differences in the scalp distribution of visual and auditory modality. *Brain Topography*, 11, 315–327. <https://doi.org/10.1023/a:1022262721343>
- John, M., Lencz, T., Ferbinteanu, J., Gallego, J. A., & Robinson, D. G. (2017). Applications of temporal kernel canonical correlation analysis in adherence studies. *Statistical Methods in Medical Research*, 26, 2437–2454.
- Kang, K., Kwak, K., Yoon, U., & Lee, J.-M. M. (2018). Lateral ventricle enlargement and cortical thinning in idiopathic normal-pressure hydrocephalus patients. *Scientific Reports*, 8, 1–9. <https://doi.org/10.1038/s41598-018-31399-1>
- Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, 58, 433–451. <https://doi.org/10.1093/biomet/58.3.433>
- Kim, M., Won, J. H., Youn, J., & Park, H. (2019). Joint-connectivity-based sparse canonical correlation analysis of imaging genetics for detecting biomarkers of Parkinson's disease. *IEEE Transactions on Medical Imaging*, 39, 23–34. <https://doi.org/10.1109/TMI.2019.2918839>
- Klami, A., & Kaski, S. (2007). Local dependent components. Proceedings of the 24th International Conference on Machine Learning. 425–432.
- Klami, A., Virtanen, S., & Kaski, S. (2013). Bayesian canonical correlation analysis. *Journal of Machine Learning Research*, 14, 965–1003.



- Koskinen, M., & Seppa, M. (2014). Uncovering cortical MEG responses to listened audiobook stories. *NeuroImage*, 100, 263–270. <https://doi.org/10.1016/j.neuroimage.2014.06.018>
- Kottaram, A., Johnston, L. A., Cocchi, L., Ganella, E. P., Everall, I., Pantelis, C., ... Zalesky, A. (2019). Brain network dynamics in schizophrenia: Reduced dynamism of the default mode network. *Human Brain Mapping*, 40, 2212–2228. <https://doi.org/10.1002/hbm.24519>
- Kroonenberg, P. M., & de Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45, 69–97.
- Kucukboyaci, N. E., Girard, H. M., Hagler, D. J. J., Kuperman, J., Tecoma, E. S., Iragui, V. J., ... McDonald, C. R. (2012). Role of frontotemporal fiber tract integrity in task-switching performance of healthy controls and patients with temporal lobe epilepsy. *Journal of the International Neuropsychological Society*, 18, 57–67. <https://doi.org/10.1017/S1355617711001391>
- Kuo, Y.-L. L., Kutch, J. J., & Fisher, B. E. (2019). Relationship between inter-hemispheric inhibition and dexterous hand performance in musicians and non-musicians. *Scientific Reports*, 9, 1–10. <https://doi.org/10.1038/s41598-019-47959-y>
- Langers, D. R. M., Krumbholz, K., Bowtell, R. W., & Hall, D. A. (2014). Neuroimaging paradigms for tonotopic mapping (I): The influence of sound stimulus type. *NeuroImage*, 100, 650–662. <https://doi.org/10.1016/j.neuroimage.2014.07.044>
- Lankinen, K., Saari, J., Hari, R., & Koskinen, M. (2014). Intersubject consistency of cortical MEG signals during movie viewing. *NeuroImage*, 92, 217–224. <https://doi.org/10.1016/j.neuroimage.2014.02.004>
- Lankinen, K., Saari, J., Hlushchuk, Y., Tikka, P., Parkkonen, L., Hari, R., & Koskinen, M. (2018). Consistency and similarity of MEG- and fMRI-signal time courses during movie viewing. *NeuroImage*, 173, 361–369. <https://doi.org/10.1016/j.neuroimage.2018.02.045>
- Lankinen, K., Smeds, E., Tikka, P., Pihko, E., Hari, R., & Koskinen, M. (2016). Haptic contents of a movie dynamically engage the spectator's sensorimotor cortex. *Human Brain Mapping*, 37, 4061–4068. <https://doi.org/10.1002/hbm.23295>
- Laskaris, L., Zalesky, A., Weickert, C. S., di Biase, M. A., Chana, G., Baune, B. T., ... Cropley, V. (2019). Investigation of peripheral complement factors across stages of psychosis. *Schizophrenia Research*, 204, 30–37. <https://doi.org/10.1016/j.schres.2018.11.035>
- Lee, S. H., & Choi, S. (2007). Two-dimensional canonical correlation analysis. *IEEE Signal Processing Letters*, 14(10), 735–738.
- Lee, W. H., Moser, D. A., Ing, A., Doucet, G. E., & Frangou, S. (2019). Behavioral and health correlates of resting-state metastability in the human connectome project. *Brain Topography*, 32, 80–86. <https://doi.org/10.1007/s10548-018-0672-5>
- Leibach, G. G., Stern, M., Arelis, A. A., Islas, M. A. M., & Barajas, B. V. R. (2016). Mental health and health-related quality of life in multiple sclerosis caregivers in Mexico. *International Journal of MS Care*, 18, 19–26. <https://doi.org/10.7224/1537-2073.2014-094>
- Leonenko, G., Di Florio, A., Allardyce, J., Forty, L., Knott, S., Jones, L., et al. (2018). A data-driven investigation of relationships between bipolar psychotic symptoms and schizophrenia genome-wide significant genetic loci. *American Journal of Medical Genetics*, 177, 468–475. <https://doi.org/10.1002/ajmg.b.32635>
- Lerman-Sinkoff, D. B., Kandala, S., Calhoun, V. D., Barch, D. M., & Mamah, D. T. (2019). Transdiagnostic multimodal neuroimaging in psychosis: Structural, resting-state, and task magnetic resonance imaging correlates of cognitive control. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4, 870–880. <https://doi.org/10.1016/j.bpsc.2019.05.004>
- Lerman-Sinkoff, D. B., Sui, J., Rachakonda, S., Kandala, S., Calhoun, V. D., & Barch, D. M. (2017). Multimodal neural correlates of cognitive control in the human connectome project. *NeuroImage*, 163, 41–54. <https://doi.org/10.1016/j.neuroimage.2017.08.081>
- Levin-Schwartz, Y., Song, Y., Schreier, P. J., Calhoun, V. D., & Adali, T. (2016). Sample-poor estimation of order and common signal subspace with application to fusion of medical imaging data. *NeuroImage*, 134, 486–493. <https://doi.org/10.1016/j.neuroimage.2016.03.058>
- Li, J., Bolt, T., Bzdok, D., Nomi, J. S., Yeo, B. T. T., Spreng, R. N., & Uddin, L. Q. (2019). Topography and behavioral relevance of the global signal in the human brain. *Scientific Reports*, 9, 1–10. <https://doi.org/10.1038/s41598-019-50750-8>
- Li, J., Chen, Y., Taya, F., Lim, J., Wong, K., Sun, Y., & Bezerianos, A. (2017). A unified canonical correlation analysis-based framework for removing gradient artifact in concurrent EEG/fMRI recording and motion artifact in walking recording from EEG signal. *Medical & Biological Engineering & Computing*, 55, 1669–1681. <https://doi.org/10.1007/s11517-017-1620-3>
- Liao, J., Zhu, Y., Zhang, M., Yuan, H., Su, M.-Y., Yu, X., & Wang, H. (2010). Microstructural white matter abnormalities independent of white matter lesion burden in amnesic mild cognitive impairment and early Alzheimer disease among Han Chinese elderly. *Alzheimer Disease and Associated Disorders*, 24, 317–324. <https://doi.org/10.1097/WAD.0b013e3181df1c7b>
- Lin, S. J., Baumeister, T. R., Garg, S., & McKeown, M. J. (2018). Cognitive profiles and hub vulnerability in Parkinson's disease. *Frontiers in Neurology*, 9, 482.
- Lin, D., Calhoun, V. D., & Wang, Y.-P. (2014). Correspondence between fMRI and SNP data by group sparse canonical correlation analysis. *Medical Image Analysis*, 18, 891–902. <https://doi.org/10.1016/j.media.2013.10.010>
- Lin, S.-J., Lam, J., Beveridge, S., Vavasour, I., Traboulsee, A., Li, D. K. B., ... Kosaka, B. (2017). Cognitive performance in subjects with multiple sclerosis is robustly influenced by gender in canonical-correlation analysis. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 29, 119–127. <https://doi.org/10.1176/appi.neuropsych.16040083>
- Lin, S.-J. J., Vavasour, I., Kosaka, B., Li, D. K. B. B., Traboulsee, A., MacKay, A., & McKeown, M. J. (2018). Education, and the balance between dynamic and stationary functional connectivity jointly support executive functions in relapsing-remitting multiple sclerosis. *Human Brain Mapping*, 39, 5039–5049. <https://doi.org/10.1002/hbm.24343>
- Lisowska, A., & Reikik, I. (2019). Joint pairing and structured mapping of convolutional brain morphological multiplexes for early dementia diagnosis. *Brain Connectivity*, 9, 22–36. <https://doi.org/10.1089/brain.2018.0578>
- Liu, J., & Calhoun, V. D. (2014). A review of multivariate analyses in imaging genetics. *Frontiers in Neuroinformatics*, 8(29). <https://doi.org/10.3389/fninf.2014.00029>
- Liu, K., Yao, X., Yan, J., Chasioti, D., Risacher, S., Nho, K., ... Shen, L. (2017). Transcriptome-guided imaging genetic analysis via a novel sparse CCA algorithm. Graphs in Biomedical Image Analysis, Computational Anatomy and Imaging Genetics First International Workshop, GRAIL 2017, 6th International Workshop, MFCA 2017, and Third International Workshop, MICGen 2017, Held in Conjunction with MICCAI 2017, Québec City, Canada, September 10–14, 2017, Proceedings 10551, 220–229.
- Liu, L., Wang, Q., Adeli, E., Zhang, L., Zhang, H., & Shen, D. (2018). Exploring diagnosis and imaging biomarkers of Parkinson's disease via iterative canonical correlation analysis based feature selection. *Computerized Medical Imaging and Graphics*, 67, 21–29. <https://doi.org/10.1016/j.compmedimag.2018.04.002>
- Liu, Y., & Ayaz, H. (2018). Speech recognition via fNIRS based brain signals. *Frontiers in Neuroscience*, 12(695). <https://doi.org/10.3389/fnins.2018.00695>
- Lopez, E., Steiner, A. J., Smith, K., Thaler, N. S., Hardy, D. J., Levine, A. J., et al. (2017). Diagnostic utility of the HIV dementia scale and the international HIV dementia scale in screening for HIV-associated neurocognitive disorders among Spanish-speaking adults. *Applied Neuropsychology Adult*, 24, 512–521.
- Lottman, K. K., White, D. M., Kraguljac, N. V., Reid, M. A., Calhoun, V. D., Catao, F., & Lahti, A. C. (2018). Four-way multimodal fusion of 7 T

- imaging data using an mCCA+jICA model in first-episode schizophrenia. *Human Brain Mapping*, 39, 1–14. <https://doi.org/10.1002/hbm.23906>
- Luo, Y., Tao, D., Ramamohanarao, K., Xu, C., & Wen, Y. (2015). Tensor canonical correlation analysis for multi-view dimension reduction. *IEEE Transactions on Knowledge and Data Engineering*, 27, 3111–3124.
- McCrary, S. J., & Ford, I. (1991). Multivariate analysis of spect images with illustrations in Alzheimer's disease. *Statistics in Medicine*, 10, 1711–1718. <https://doi.org/10.1002/sim.4780101109>
- McMillan, C. T., Toledo, J. B., Avants, B. B., Cook, P. A., Wood, E. M., Suh, E., ... Grossman, M. (2014). Genetic and neuroanatomic associations in sporadic frontotemporal lobar degeneration. *Neurobiology of Aging*, 35, 1473–1482. <https://doi.org/10.1016/j.neurobiolaging.2013.11.029>
- Mihalik, A., Ferreira, F. S., Rosa, M. J., Moutoussis, M., Ziegler, G., Monteiro, J. M., ... Mourao-Miranda, J. (2019). Brain-behaviour modes of covariation in healthy and clinically depressed young people. *Scientific Reports*, 9, 11536. <https://doi.org/10.1038/s41598-019-47277-3>
- Mirza, M. B., Adams, R. A., Mathys, C., & Friston, K. J. (2018). Human visual exploration reduces uncertainty about the sensed world. *PLoS One*, 13, e0190429. <https://doi.org/10.1371/journal.pone.0190429>
- Mishra, V. R., Zhuang, X., Sreenivasan, K. R., Banks, S. J. S. J., Yang, Z., Bernick, C., & Cordes, D. (2017). Multimodal MR imaging signatures of cognitive impairment in active professional fighters. *Radiology*, 285, 555–567. <https://doi.org/10.1148/radiol.2017162403>
- Moser, D. A., Doucet, G. E., Lee, W. H., Rasgon, A., Krinsky, H., Leib, E., ... Frangou, S. (2018). Multivariate associations among behavioral, clinical, and multimodal imaging phenotypes in patients with psychosis. *JAMA Psychiatry*, 75, 386–395.
- Mohammadi-Nejad, A.-R., Hossein-Zadeh, G.-A., & Soltanian-Zadeh, H. (2017). Structured and sparse canonical correlation analysis as a brain-wide multi-modal data fusion approach. *IEEE Transactions on Medical Imaging*, 36, 1438–1448. <https://doi.org/10.1109/TMI.2017.2681966>
- Murayama, Y., Biessmann, F., Meinecke, F. C., Muller, K.-R., Augath, M., Oeltermann, A., & Logothetis, N. K. (2010). Relationship between neural and hemodynamic signals during spontaneous activity studied with temporal kernel CCA. *Magnetic Resonance Imaging*, 28, 1095–1103. <https://doi.org/10.1016/j.mri.2009.12.016>
- Nandy, R., & Cordes, D. (2004). Improving the spatial specificity of canonical correlation analysis in fMRI. *Magnetic Resonance in Medicine*, 52, 947–952. <https://doi.org/10.1002/mrm.20234>
- Nandy, R. R., & Cordes, D. (2003). Novel nonparametric approach to canonical correlation analysis with applications to low CNR functional MRI data. *Magnetic Resonance in Medicine*, 50, 354–365. <https://doi.org/10.1002/mrm.10537>
- Neal, R. M. (2012). *Bayesian learning for neural networks*, Berlin, Germany: Springer Science & Business Media.
- Neumann, J., von Cramon, D. Y., Forstmann, B. U., Zysset, S., & Lohmann, G. (2006). The parcellation of cortical areas using replicator dynamics in fMRI. *Neuroimage*, 32, 208–219. <https://doi.org/10.1016/j.neuroimage.2006.02.039>
- Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, 87, 9868–9872. <https://doi.org/10.1073/pnas.87.24.9868>
- Ouyang, X., Chen, K., Yao, L., Hu, B., Wu, X., Ye, Q., & Guo, X. (2015). Simultaneous changes in gray matter volume and white matter fractional anisotropy in Alzheimer's disease revealed by multimodal CCA and joint ICA. *Neuroscience*, 301, 553–562. <https://doi.org/10.1016/j.neuroscience.2015.06.031>
- Palaniyappan, L., Mota, N. B., Oowise, S., Balain, V., Copelli, M., Ribeiro, S., & Liddle, P. F. (2019). Speech structure links the neural and socio-behavioural correlates of psychotic disorders. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 88, 112–120. <https://doi.org/10.1016/j.pnpbp.2018.07.007>
- Peng, Y., Zhang, D., & Zhang, J. (2010). A new canonical correlation analysis algorithm with local discrimination. *Neural Processing Letters*, 31, 1–15. <https://doi.org/10.1007/s11063-009-9123-3>
- Pustina, D., Avants, B., Faseyitan, O. K., Medaglia, J. D., & Coslett, H. B. (2018). Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. *Neuropsychologia*, 115, 154–166. <https://doi.org/10.1016/j.neuropsychologia.2017.08.027>
- Qi, S., Abbott, C. C., Narr, K. L., Jiang, R., Upston, J., McClintock, S. M., ... Calhoun, V. D. (2020). Electroconvulsive therapy treatment responsive multimodal brain networks. *Human Brain Mapping*, 41, 1775–1785. <https://doi.org/10.1002/hbm.24910>
- Qi, S., Calhoun, V. D., van Erp, T. G. M., Bustillo, J., Damaraju, E., Turner, J. A., ... Sui, J. (2018). Multimodal fusion with reference: Searching for joint neuromarkers of working memory deficits in schizophrenia. *IEEE Transactions on Medical Imaging*, 37, 93–105. <https://doi.org/10.1109/TMI.2017.2725306>
- Qi, S., Yang, X., Zhao, L., Calhoun, V. D., Perrone-Bizzozero, N., Liu, S., ... Ma, X. (2018). MicroRNA132 associated multimodal neuroimaging patterns in unmedicated major depressive disorder. *Brain*, 141, 916–926. <https://doi.org/10.1093/brain/awx366>
- Rodrigue, A. L., McDowell, J. E., Tandon, N., Keshavan, M. S., Tamminga, C. A., Pearson, G. D., ... Clementz, B. A. (2018). Multivariate relationships between cognition and brain anatomy across the psychosis Spectrum. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3, 992–1002. <https://doi.org/10.1016/j.bpsc.2018.03.012>
- Rodu, J., Klein, N., Brincat, S. L., Miller, E. K., & Kass, R. E. (2018). Detecting multivariate cross-correlation between brain regions. *Journal of Neurophysiology*, 120, 1962–1972. <https://doi.org/10.1152/jn.00869.2017>
- Rosa, M. J., Mehta, M. A., Pich, E. M., Risterucci, C., Zelaya, F., Reinders, A. A. T. S., ... Marquand, A. F. (2015). Estimating multivariate similarity between neuroimaging datasets with sparse canonical correlation analysis: An application to perfusion imaging. *Frontiers in Neuroscience*, 9(366). <https://doi.org/10.3389/fnins.2015.00366>
- Rydel, J., Knutsson, H., & Borga, M. (2006). On rotational invariance in adaptive spatial filtering of fMRI data. *NeuroImage*, 30, 144–150. <https://doi.org/10.1016/j.neuroimage.2005.09.002>
- Sato, J. R., Fujita, A., Cardoso, E. F., Thomaz, C. E., Brammer, M. J., & Amaro, E. J. (2010). Analyzing the connectivity between regions of interest: An approach based on cluster granger causality for fMRI data analysis. *NeuroImage*, 52, 1444–1455. <https://doi.org/10.1016/j.neuroimage.2010.05.022>
- Shams, S.M., Hossein-Zadeh, G.A., & Soltanian-Zadeh, H. (2006). Multi-subject activation detection in fMRI by testing correlation of data with a signal subspace. *Magnetic Resonance Imaging*, 24, 775–784. <https://doi.org/10.1016/j.mri.2006.03.008>
- Shen, H., Chau, D. K. P., Su, J., Zeng, L.L., Jiang, W., He, J., ... Hu, D. (2016). Brain responses to facial attractiveness induced by facial proportions: Evidence from an fMRI study. *Scientific Reports*, 6, 35905. <https://doi.org/10.1038/srep35905>
- Sheng, J., Kim, S., Yan, J., Moore, J., Saykin, A., & Shen, L. (2014). Data synthesis and method evaluation for brain imaging genetics. *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 2014, 1202–1205. <https://doi.org/10.1109/ISBI.2014.6868091>
- Sintini, I., Schwarz, C. G., Martin, P. R., Graff-Radford, J., Machulda, M. M., Senjem, M. L., ... Whitwell, J. L. (2019). Regional multimodal relationships between tau, hypometabolism, atrophy, and fractional anisotropy in atypical Alzheimer's disease. *Human Brain Mapping*, 40, 1618–1631. <https://doi.org/10.1002/hbm.24473>
- Sintini, I., Schwarz, C. G., Senjem, M. L., Reid, R. I., Botha, H., Ali, F., ... Whitwell, J. L. (2019). Multimodal neuroimaging relationships in progressive supranuclear palsy. *Parkinsonism & Related Disorders*, 66, 56–61. <https://doi.org/10.1016/j.parkreldis.2019.07.001>
- Smirnov, D., Lachat, F., Peltola, T., Lahnakoski, J. M., Koistinen, O.-P., Glerean, E., ... Nummenmaa, L. (2017). Brain-to-brain hyperclassification

- reveals action-specific motor mapping of observed actions in humans. *PLoS One*, 12, e0189508. <https://doi.org/10.1371/journal.pone.0189508>
- Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E. J., Glasser, M. F., ... Miller, K. L. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, 18, 1565–1567. <https://doi.org/10.1038/nn.4125>
- Somers, B., & Bertrand, A. (2016). Removal of eye blink artifacts in wireless EEG sensor networks using reduced-bandwidth canonical correlation analysis. *Journal of Neural Engineering*, 13, 66008. <https://doi.org/10.1088/1741-2560/13/6/066008>
- Soto, J. L. P., Lachaux, J.-P., Baillet, S., & Jerbi, K. (2016). A multivariate method for estimating cross-frequency neuronal interactions and correcting linear mixing in MEG data, using canonical correlations. *Journal of Neuroscience Methods*, 271, 169–181. <https://doi.org/10.1016/j.jneumeth.2016.07.017>
- Stout, D. M., Buchsbaum, M. S., Spadoni, A. D., Risbrough, V. B., Strigo, I. A., Matthews, S. C., & Simmons, A. N. (2018). Multimodal canonical correlation reveals converging neural circuitry across trauma-related disorders of affect and cognition. *Neurobiology of Stress*, 9, 241–250. <https://doi.org/10.1016/j.ynstr.2018.09.006>
- Sui, J., Adali, T. T., Pearlson, G., Yang, H., Sponheim, S. R., White, T., ... Calhoun, V. D. (2010). A CCA+ICA based model for multi-task brain imaging data fusion and its application to schizophrenia. *NeuroImage*, 51, 123–134. <https://doi.org/10.1016/j.neuroimage.2010.01.069>
- Sui, J., Adali, T. T., Yu, Q., Chen, J., & Calhoun, V. D. (2012). A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of Neuroscience Methods*, 204, 68–81. <https://doi.org/10.1016/j.jneumeth.2011.10.031>
- Sui, J., He, H., Pearlson, G. D., Adali, T., Kiehl, K. A., Yu, Q., ... Calhoun, V. D. (2013). Three-way (N-way) fusion of brain imaging data based on mCCA+jICA and its application to discriminating schizophrenia. *NeuroImage*, 66, 119–132. <https://doi.org/10.1016/j.neuroimage.2012.10.051>
- Sui, J., Pearlson, G., Caprihan, A., Adali, T., Kiehl, K. A., Liu, J., ... Calhoun, V. D. (2011). Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model. *NeuroImage*, 57, 839–855. <https://doi.org/10.1016/j.neuroimage.2011.05.055>
- Sui, J., Pearlson, G. D., Du, Y., Yu, Q., Jones, T. R., Chen, J., ... Calhoun, V. D. (2015). In search of multimodal neuroimaging biomarkers of cognitive deficits in schizophrenia. *Biological Psychiatry*, 78, 794–804. <https://doi.org/10.1016/j.biopsych.2015.02.017>
- Sui, J., Qi, S., van Erp, T. G. M., Bustillo, J., Jiang, R., Lin, D., ... Calhoun, V. D. (2018). Multimodal neuromarkers in schizophrenia via cognition-guided MRI fusion. *Nature Communications*, 9, 3028. <https://doi.org/10.1038/s41467-018-05432-w>
- Szefer, E., Lu, D., Nathoo, F., Beg, M. F., & Graham, J. (2017). Multivariate association between single-nucleotide polymorphisms in Alzgene linkage regions and structural changes in the brain: Discovery, refinement and validation. *Statistical Applications in Genetics and Molecular Biology*, 16, 349–365. <https://doi.org/10.1515/sagmb-2016-0077>
- Thye, M., & Mirman, D. (2018). Relative contributions of lesion location and lesion size to predictions of varied language deficits in post-stroke aphasia. *NeuroImage: Clinical*, 20, 1129–1138. <https://doi.org/10.1016/j.nicl.2018.10.017>
- Tian, Y., Zalesky, A., Bousman, C., Everall, I., & Pantelis, C. (2019). Insula functional connectivity in schizophrenia: Subregions, gradients, and symptoms. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4, 399–408. <https://doi.org/10.1016/j.bpsc.2018.12.003>
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso Robert Tibshirani. *Journal of the Royal Statistical Society, Series B*, 58, 267–288. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>
- Tsvetanov, K. A., Henson, R. N. A., Tyler, L. K., Razi, A., Geerligs, L., Ham, T. E., & Rowe, J. B. (2016). Extrinsic and intrinsic brain network connectivity maintains cognition across the lifespan despite accelerated decay of regional brain activation. *The Journal of Neuroscience*, 36, 3115–3126. <https://doi.org/10.1523/JNEUROSCI.2733-15.2016>
- Valakos, D., Karantinos, T., Evdokimidis, I., Stefanis, N. C., Avramopoulos, D., & Smyrnis, N. (2018). Shared variance of oculomotor phenotypes in a large sample of healthy young men. *Experimental Brain Research*, 236, 2399–2410. <https://doi.org/10.1007/s00221-018-5312-5>
- Varoquaux, G., Sadaghiani, S., Pinel, P., Kleinschmidt, A., Poline, J. B., & Thirion, B. (2010). A group model for stable multi-subject ICA on fMRI datasets. *NeuroImage*, 51, 288–299. <https://doi.org/10.1016/j.neuroimage.2010.02.010>
- Vatansever, D., Bzdok, D., Wang, H.-T., Mollo, G., Sormaz, M., Murphy, C., ... Jefferies, E. (2017). Varieties of semantic cognition revealed through simultaneous decomposition of intrinsic brain connectivity and behaviour. *NeuroImage*, 158, 1–11. <https://doi.org/10.1016/j.neuroimage.2017.06.067>
- Vergult, A., de Clercq, W., Palmini, A., Vanrumste, B., Dupont, P., van Huffel, S., & van Paesschen, W. (2007). Improving the interpretation of ictal scalp EEG: BSS-CCA algorithm for muscle artifact removal. *Epilepsia*, 48, 950–958. <https://doi.org/10.1111/j.1528-1167.2007.01031.x>
- Viviano, J. D., Buchanan, R. W., Calarco, N., Gold, J. M., Foussias, G., Bhagwat, N., ... Green, M. (2018). Resting-state connectivity biomarkers of cognitive performance and social function in individuals with schizophrenia spectrum disorder and healthy control subjects. *Biological Psychiatry*, 84, 665–674. <https://doi.org/10.1016/j.biopsych.2018.03.013>
- von Luhmann, A., Boukouvalas, Z., Muller, K.-R., & Adali, T. (2019). A new blind source separation framework for signal analysis and artifact rejection in functional near-infrared spectroscopy. *NeuroImage*, 200, 72–88. <https://doi.org/10.1016/j.neuroimage.2019.06.021>
- Wan, J., Kim, S., Inlow, M., Nho, K., Swaminathan, S., Risacher, S. L., ... Shen, L. (2011). Hippocampal surface mapping of genetic risk factors in AD via sparse learning models. *Medical Image Computing and Computer-Assisted Intervention*, 14, 376–383. [https://doi.org/10.1007/978-3-642-23629-7\\_46](https://doi.org/10.1007/978-3-642-23629-7_46)
- Wang, C. (2007). Variational Bayesian approach to canonical correlation analysis. *IEEE Transactions on Neural Networks*, 18, 905–910. <https://doi.org/10.1109/TNN.2007.891186>
- Wang, H. T., Poerio, G., Murphy, C., Bzdok, D., Jefferies, E., & Smallwood, J. (2018). Dimensions of experience: Exploring the heterogeneity of the wandering mind. *Psychological Science*, 29, 56–71. <https://doi.org/10.1177/0956797617728727>
- Wang, M., Shao, W., Hao, X., Shen, L., & Zhang, D. (2019). Identify consistent cross-modality imaging genetic patterns via discriminant sparse canonical correlation analysis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1. <https://doi.org/10.1109/TCBB.2019.2944825>
- Wee, C.Y., Tuan, T. A., Broekman, B. F. P., Ong, M. Y., Chong, Y.S., Kwek, K., et al. (2017). Neonatal neural networks predict children behavioral profiles later in life. *Human Brain Mapping*, 38, 1362–1373. <https://doi.org/10.1002/hbm.23459>
- Will, G.J. J., Rutledge, R. B., Moutoussis, M., & Dolan, R. J. (2017). Neural and computational processes underlying dynamic changes in self-esteem. *Elife*, 6, 1–21. <https://doi.org/10.7554/eLife.28098>
- Witten, D. M., Tibshirani, R., & Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10, 515–534. <https://doi.org/10.1093/biostatistics/kxp008>
- Witten, D. M., & Tibshirani, R. J. (2009). Extensions of sparse canonical correlation analysis with applications to genomic data. *Statistical Applications in Genetics and Molecular Biology*, 8, 1–27. <https://doi.org/10.2202/1544-6115.1470>
- Xia, C. H., Ma, Z., Ciric, R., Gu, S., Betzel, R. F., Kaczkurkin, A. N., ... Satterthwaite, T. D. (2018). Linked dimensions of psychopathology



- and connectivity in functional brain networks. *Nature Communications*, 9, 1–14. <https://doi.org/10.1038/s41467-018-05317-y>
- Yan, J., Du, L., Kim, S., Risacher, S. L., Huang, H., Moore, J. H., ... Shen, L. (2014). Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm. *Bioinformatics*, 30, i564–i571. <https://doi.org/10.1093/bioinformatics/btu465>
- Yan, J., Risacher, S. L., Nho, K., Saykin, A. J., & Shen, L. I. (2017). Identification of discriminative imaging proteomics associations in Alzheimer's disease via a novel sparse correlation model. *Pacific Symposium on Bio-computing*, 22, 94–104. [https://doi.org/10.1142/9789813207813\\_0010](https://doi.org/10.1142/9789813207813_0010)
- Yang, B., Cao, J., Zhou, T., Dong, L., Zou, L., & Xiang, J. (2018). Exploration of neural activity under cognitive reappraisal using simultaneous EEG-fMRI data and kernel canonical correlation analysis. *Computational and Mathematical Methods in Medicine*, 2018, 3018356. <https://doi.org/10.1155/2018/3018356>
- Yang, Z., Zhuang, X., Bird, C., Sreenivasan, K., Mishra, V., Banks, S., & Cordes, D. (2019). Performing sparse regularization and dimension reduction simultaneously in multimodal data fusion. *Frontiers in Neuroscience*, 13(878). <https://doi.org/10.3389/fnins.2019.00642>
- Yang, Z., Zhuang, X., Sreenivasan, K., & Mishra, V. (2019). Robust Motion regression of resting-state data using a convolutional neural network model. *Frontiers in Neuroscience*, 13, 1–14. <https://doi.org/10.3389/fnins.2019.00169>
- Yang, Z., Zhuang, X., Sreenivasan, K., Mishra, V., Curran, T., Byrd, R., ... Cordes, D. (2018). 3D spatially-adaptive canonical correlation analysis: Local and global methods. *NeuroImage*, 169, 240–255. <https://doi.org/10.1016/j.neuroimage.2017.12.025>
- Yang, Z., Zhuang, X., Sreenivasan, K., Mishra, V., Curran, T., & Cordes, D. (2020). A robust deep neural network for denoising task-based fMRI data: An application to working memory and episodic memory. *Medical Image Analysis*, 60, 101622. <https://doi.org/10.1016/j.media.2019.101622>
- Yu, Q., Erhardt, E. B., Sui, J., Du, Y., He, H., Hjelm, D., ... Calhoun, V. D. (2015). Assessing dynamic brain graphs of time-varying connectivity in fMRI data: Application to healthy controls and patients with schizophrenia. *NeuroImage*, 107, 345–355. <https://doi.org/10.1016/j.neuroimage.2014.12.020>
- Zarnani, K., Nichols, T. E., Alfaro-Almagro, F., Fagerlund, B., Lauritzen, M., Rostrup, E., & Smith, S. M. (2019). Discovering markers of healthy aging: A prospective study in a Danish male birth cohort. *Aging (Albany NY)*, 11, 5943–5974. <https://doi.org/10.18632/aging.102151>
- Zhang, Q., Borst, J. P., Kass, R. E., & Anderson, J. R. (2017). Inter-subject alignment of MEG datasets in a common representational space. *Human Brain Mapping*, 38, 4287–4301. <https://doi.org/10.1002/hbm.23689>
- Zhao, F., Qiao, L., Shi, F., Yap, P.-T., & Shen, D. (2017). Feature fusion via hierarchical supervised local CCA for diagnosis of autism spectrum disorder. *Brain Imaging and Behavior*, 11, 1050–1060. <https://doi.org/10.1007/s11682-016-9587-5>
- Zhu, X., Suk, H.-I., Lee, S.-W., & Shen, D. (2016). Canonical feature selection for joint regression and multi-class identification in Alzheimer's disease diagnosis. *Brain Imaging and Behavior*, 10, 818–828. <https://doi.org/10.1007/s11682-015-9430-4>
- Zhuang, X., Walsh, R. R., Sreenivasan, K., Yang, Z., Mishra, V., & Cordes, D. (2018). Incorporating spatial constraint in co-activation pattern analysis to explore the dynamics of resting-state networks: An application to Parkinson's disease. *NeuroImage*, 172, 64–84. <https://doi.org/10.1016/j.neuroimage.2018.01.019>
- Zhuang, X., Yang, Z., Curran, T., Byrd, R., Nandy, R., & Cordes, D. (2017). A family of locally constrained CCA models for detecting activation patterns in fMRI. *NeuroImage*, 149, 63–84. <https://doi.org/10.1016/j.neuroimage.2016.12.081>
- Zhuang, X., Yang, Z., Sreenivasan, K. R., Mishra, V. R., Curran, T., Nandy, R., & Cordes, D. (2019). Multivariate group-level analysis for task fMRI data with canonical correlation analysis. *NeuroImage*, 194, 25–41. <https://doi.org/10.1016/j.neuroimage.2019.03.030>
- Zille, P., Calhoun, V. D., & Wang, Y.-P. (2018). Enforcing co-expression within a brain-imaging genomics regression framework. *IEEE Transactions on Medical Imaging*, 37, 2561–2571. <https://doi.org/10.1109/TMI.2017.2721301>

**How to cite this article:** Zhuang X, Yang Z, Cordes D. A technical review of canonical correlation analysis for neuroscience applications. *Hum Brain Mapp*. 2020;1–27. <https://doi.org/10.1002/hbm.25090>