Multi-Group Tensor Canonical Correlation Analysis

Zhuoping Zhou
University of Pennsylvania
Philadelphia, USA
zhuopinz@sas.upenn.edu

Boning Tong
University of Pennsylvania
Philadelphia, USA
boningt@seas.upenn.edu

Davoud Ataee Tarzanagh
University of Pennsylvania
Philadelphia, USA
tarzanaq@upenn.edu

Bojian Hou
University of Pennsylvania
Philadelphia, USA
bojian.hou@pennmedicine.upenn.edu

Andrew J. Saykin
Indiana University
Bloomington, USA
asaykin@iupui.edu

Qi Long, Li Shen*
University of Pennsylvania
Philadelphia, USA
[qlong,lishen]@pennmedicine.upenn.edu

Abstract

Tensor Canonical Correlation Analysis (TCCA) is a commonly employed statistical method utilized to examine linear associations between two sets of tensor datasets. However, the existing TCCA models fail to adequately address the heterogeneity present in real-world tensor data, such as brain imaging data collected from diverse groups characterized by factors like sex and race. Consequently, these models may yield biased outcomes. In order to surmount this constraint, we propose a novel approach called Multi-Group TCCA (MG-TCCA), which enables the joint analysis of multiple subgroups. By incorporating a dual sparsity structure and a block coordinate ascent algorithm, our MG-TCCA method effectively addresses heterogeneity and leverages information across different groups to identify consistent signals. This novel approach facilitates the quantification of shared and individual structures, reduces data dimensionality, and enables visual exploration. To empirically validate our approach, we conduct a study focused on investigating correlations between two brain positron emission tomography (PET) modalities (AV-45 and FDG) within an Alzheimer's disease (AD) cohort. Our results demonstrate that MG-TCCA surpasses traditional TCCA in identifying sex-specific cross-modality imaging correlations. This heightened performance of MG-TCCA provides valuable insights for the characterization of multimodal imaging biomarkers in AD.

CCS Concepts: • **Applied computing** → **Imaging**; • **Theory of computation** → *Convex optimization.*

Keywords: Canonical Correlation Analysis, Tensor Decomposition, Neuroimaging, Alzheimer's Disease

ACM Reference Format:
Zhuoping Zhou, Boning Tong, Davoud Ataee Tarzanagh, Bojian Hou, Andrew J. Saykin, and Qi Long, Li Shen. 2023. Multi-Group Tensor Canonical Correlation Analysis. In *14th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '23), September 3–6, 2023, Houston, TX, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3584371.3612962

1 Introduction

Canonical Correlation Analysis (CCA) is a powerful statistical technique widely used to examine the relationship between two sets of variables [10]. It finds applications in diverse fields such as psychology [6, 9], biology [26, 29], neuroscience [1, 27], and medicine [35], for unsupervised or semi-supervised learning. Despite its usefulness, CCA has a few limitations. One limitation of CCA is related to its application on tensor data, which often has a high number of dimensions. This can lead to computational challenges when applying CCA to such data. Additionally, tensor data often exhibit unique structures and interdependencies, such as spatial or temporal associations, that may not be fully captured by traditional CCA. Consequently, interpreting the correlations identified by CCA in the context of the original data structure can be a challenging task.

To overcome these limitations, Tensor Canonical Correlation Analysis (TCCA) has emerged as a valuable alternative to traditional CCA methods [17, 18, 20, 22]. TCCA demonstrates efficacy in handling high-dimensional tensor data by extracting a low-dimensional representation that captures essential information across modalities, which facilitates efficient data representation and visualization. Through the identification of modality-wise weights or coefficients, Tensor CCA yields interpretable results that quantify the significance of each modality in the canonical weights. These

*This work was supported in part by the NIH grants U01 AG066833, RF1 AG063481, U01 AG068057, R01 LM013463, P30 AG073105, and R01 AG071470, and the NSF grant IIS 1837964. Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative database (adni.loni.usc.edu), which was funded by NIH U01 AG024904.

BCB '23, September 3–6, 2023, Houston, TX, USA

ACM ISBN 979-8-4007-0126-9/23/09…$15.00
https://doi.org/10.1145/3584371.3612962

weights enhance the understanding of individual modalities' contributions to the overall correlation structure and offer insights into the relationships between different data modalities.

The motivation for this article lies in the application of TCCA to neuroimaging data in the context of Alzheimer's disease (AD). TCCA has become a valuable tool in AD research for integrating data from various neuroimaging modalities, enabling a comprehensive understanding of brain structure and function changes associated with the disease [5, 23]. However, it is important to acknowledge that TCCA assumes identical brain imaging patterns across all subjects, which may not always hold true due to inter-group differences in brain structure and function. Neuroimaging datasets often include diverse groups, such as individuals of different sexes, education levels, or clinical subgroups [7, 33], posing a challenge for traditional TCCA. This heterogeneity can result in the underrepresentation of minority group conditions in neuroimaging studies, limiting the exploration of important brain imaging biomarkers specific to certain groups. Therefore, a more tailored and group-specific approach is needed to effectively leverage information from different groups while accounting for their unique characteristics.

Methodological Contribution. To bridge this gap, we introduce Multi-Group Tensor TCCA (MG-TCCA), a novel method that enhances traditional TCCA by incorporating dependency information from groups sharing similar characteristics. MG-TCCA enables the identification of consistent signals across diverse groups, preserving individual differences within each group while detecting shared brain imaging patterns; see Figure 1. By quantifying joint variation between data groups, MG-TCCA effectively reduces data dimensionality, facilitating the exploration of common and individual structures. To demonstrate its efficacy, we conducted an empirical study focusing on mapping correlations between two brain PET modalities, AV-45 and FDG, in an AD cohort. Comparing MG-TCCA with traditional TCCA, we showcase MG-TCCA's superior performance in identifying sex-specific cross-modality imaging correlations. This advanced capability of MG-TCCA provides valuable insights for characterizing multimodal imaging biomarkers in AD and holds promise for advancements in personalized medicine.

Innovation from the application perspective. While CCA has been widely applied in brain imaging and genomics studies [27], our MG-TCCA analysis offers several innovative perspectives that differentiate it from existing approaches:
• Existing CCA studies (e.g., [5, 8, 19, 34]) were not specifically designed to analyze 3D image scans as tensor data; this included Du et al [5], where their tensors were formed by regional measures across multiple imaging modalities. In contrast, our MG-TCCA analysis directly handles 3D imaging data, allowing for the natural capture of detailed voxel-level information and spatial relationships using the tensor representation.
• Existing CCA studies (e.g., [5, 8, 19, 34]) did not adequately account for heterogeneity due to pooling from multiple groups, which might lead to biased results. Our MG-TCCA analysis across two sex subgroups is able to identify consistent signals across two groups while adequately accounting for their heterogeneity.
• The MG-TCCA presents a direct and accurate approach for handling the tensor structure and preserving voxel-level information, resulting in estimation bounds that can scale linearly with the sum of factor matrix dimensions, unlike previous studies that use matrix-based variations [28].

2 Preliminary

Notations. Mode i fibres are p_i-dimensional vectors extracted from $\boldsymbol{\mathcal{A}}$ by fixing all the indices $(l_1, \ldots, l_{i-1}, l_{i+1}, \ldots, l_D)$ except l_i. The process of transforming a tensor into a matrix through the reordering of its dimensions is commonly referred to as matricization. The mode i matricization of $\boldsymbol{\mathcal{A}}$ is denoted as $\mathbf{A}_{(i)} \in \mathbb{R}^{p_i \times p_{-i}}$, where $p_{-i} = \prod_{k=1, k\neq i}^{D} p_k$. It rearranges the mode i fibres of $\boldsymbol{\mathcal{A}}$ as the columns of the matrix $\mathbf{A}_{(i)}$. Additionally, the vectorization of mode i matricization of $\boldsymbol{\mathcal{A}}$ is denoted as $\mathbf{a}_{(i)}$. In addition to standard matrix operations, this work utilizes four operations, namely the inner product, the outer product, the Kronecker product, and the Khatri-Rao product, respectively, represented by $< \cdot, \cdot >, \circ, \otimes$, and $\odot$.

The CANDECOMP/PARAFAC (CP) tensor decomposition is a method used to break down a higher-order data array into a sum of rank-one component arrays. Given a rank-R tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{p_1 \times \cdots \times p_D}$, it can be represented as follows:

$$\boldsymbol{\mathcal{X}} = [\![\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_D]\!] = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \circ \mathbf{a}_r^{(2)} \circ \cdots \circ \mathbf{a}_r^{(D)},$$

where

$$\mathbf{A}_i = \begin{pmatrix} \mathbf{a}_1^{(i)} & \mathbf{a}_2^{(i)} & \cdots & \mathbf{a}_R^{(i)} \end{pmatrix} \in \mathbb{R}^{p_i \times R}$$

denotes the mode i factor matrix.

The following representations of variances $\mathrm{Var}(\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{U}} \rangle)$, $\mathrm{Var}(< \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{V}} >)$ and covariance $\mathrm{Cov}(\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{U}} \rangle, \langle \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{V}} \rangle)$ play a crucial role in the design of our algorithm.

Proposition 1 ([22]). *Given datasets $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{p_1 \times p_2 \times \cdots \times p_M}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{p_1 \times q_2 \times \cdots \times q_N}$, let $\boldsymbol{\mathcal{U}} = [\![\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_M]\!]$ and $\boldsymbol{\mathcal{V}} = [\![\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_N]\!]$ be two constant tensors of the same size as $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$ respectively. Define*

$$\mathbf{U}_{(-i)} := [\mathbf{U}_M \odot \cdots \odot \mathbf{U}_{i+1} \odot \mathbf{U}_{i-1} \odot \cdots \odot \mathbf{U}_1] \otimes \mathbf{I}_{p_i}, \quad (1a)$$
$$\mathbf{V}_{(-j)} := [\mathbf{V}_N \odot \cdots \odot \mathbf{V}_{j+1} \odot \mathbf{V}_{j-1} \odot \cdots \odot \mathbf{V}_1] \otimes \mathbf{I}_{q_j}. \quad (1b)$$

Let $\boldsymbol{\Sigma}_{\mathbf{x}_{(i)}}$ denote the covariance of $\mathbf{x}_{(i)} = vec(\mathbf{X}_{(i)})$, where $\mathbf{X}_{(i)}$ is the mode i matricization of $\boldsymbol{\mathcal{X}}$. Define $\boldsymbol{\Sigma}_{\mathbf{y}_{(j)}}, \boldsymbol{\Sigma}_{\mathbf{x}_{(i)}, \mathbf{y}_{(j)}}$

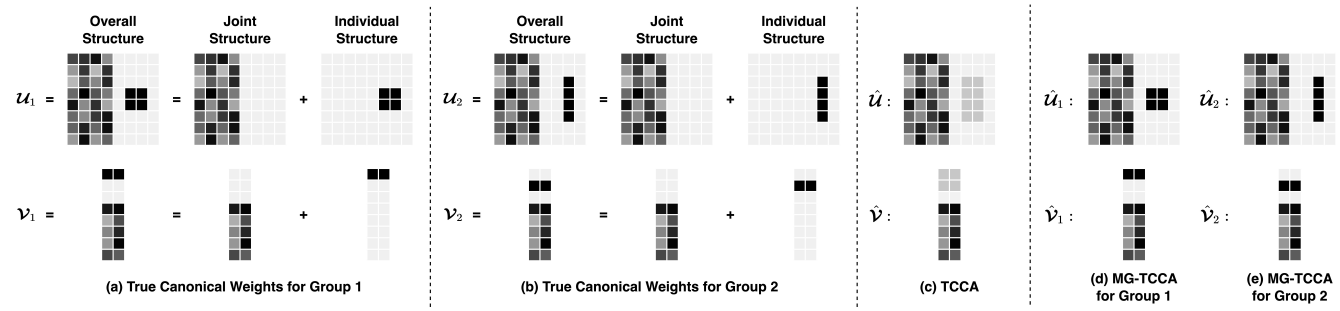


Figure 1. Illustration of TCCA and MG-TCCA. Panels (a) and (b) illustrate the true canonical weights for different groups within the dataset, while Panels (c)-(e) compare the canonical weights generated by TCCA and MG-TCCA. TCCA has limitations in capturing group-specific structures, whereas MG-TCCA is designed to address this drawback by capturing both shared and group-specific structures.

analogously. Then,

$$Var(< \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{U}} >) = \mathbf{u}_i^\top \mathbf{U}_{(-i)}^\top \Sigma_{\mathbf{x}_{(i)}} \mathbf{U}_{(-i)} \mathbf{u}_i, \tag{2a}$$

$$Var(< \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{V}} >) = \mathbf{v}_j^\top \mathbf{V}_{(-j)}^\top \Sigma_{\mathbf{y}_{(j)}} \mathbf{V}_{(-j)} \mathbf{v}_j, \tag{2b}$$

$$Cov(< \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{U}} >, < \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{V}} >) = \mathbf{u}_i^\top \mathbf{U}_{(-i)}^\top \Sigma_{\mathbf{x}_{(i)}, \mathbf{y}_{(j)}} \mathbf{V}_{(-j)} \mathbf{v}_j, \tag{2c}$$

for any $i \in [M]$, $j \in [N]$, *where* $\mathbf{u}_i = vec(\mathbf{U}_i)$, $\mathbf{v}_j = vec(\mathbf{V}_j)$.

2.1 Tensor Canonical Correlation Analysis

Let $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{p_1 \times p_2 \times \cdots \times p_M}$ and $\boldsymbol{\mathcal{Y}}_i \in \mathbb{R}^{q_1 \times q_2 \times \cdots \times q_N}$ be tensorial datasets of order M and N for subject $i = 1, \ldots, n$, respectively. Assume $\boldsymbol{\mathcal{X}}_i$ and $\boldsymbol{\mathcal{Y}}_i$ are centered to zero mean. Let $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{p_1 \times p_2 \times \cdots \times p_M}$ and $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{q_1 \times q_2 \times \cdots \times q_N}$ denote a rank-R tensor weight for $\boldsymbol{\mathcal{X}}$ and a rank-S tensor weight for $\boldsymbol{\mathcal{Y}}$ respectively. TCCA seeks to find tensor weights to maximize the sample Pearson correlation coefficient between two sets of tensorial datasets in the projected space, namely:

$$\begin{aligned} \max_{\boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}} \quad & \frac{1}{n-1} \sum_{i=1}^{n} \langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{U}} \rangle \cdot \langle \boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{V}} \rangle \\ \text{subj. to} \quad & \frac{1}{n-1} \sum_{i=1}^{n} \langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{U}} \rangle^2 \leq 1, \\ & \frac{1}{n-1} \sum_{i=1}^{n} \langle \boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{V}} \rangle^2 \leq 1, \\ & \text{rank}(\boldsymbol{\mathcal{U}}) = R, \ \text{rank}(\boldsymbol{\mathcal{V}}) = S. \end{aligned} \tag{3}$$

Note that when $N = M = 2$, (3) reduces to matrix CCA.

3 Model Framework

The primary objective of MG-TCCA is to facilitate the exploration of association patterns within tensor data consisting of K groups. This method excels at identifying both shared and distinct characteristics among these groups. MG-TCCA employs a dual sparsity structure, incorporating group-wise and individual-wise sparsity. This is accomplished through the use of $L_{2,1}$-norm and L_1-norm.

Let K represent the number of groups under consideration, with each group consisting of n_k samples. Given two tensors $\boldsymbol{\mathcal{X}}_{k,i} \in \mathbb{R}^{p_1 \times \cdots \times p_M}$ and $\boldsymbol{\mathcal{Y}}_{k,i} \in \mathbb{R}^{q_1 \times \cdots \times q_N}$ for group $k = 1, \ldots, K$ and subject $i = 1, \ldots, n_k$, let

$$\boldsymbol{\mathcal{U}}_k = [\![\mathbf{U}_{k,1}, \mathbf{U}_{k,2}, \ldots, \mathbf{U}_{k,M}]\!], \tag{4a}$$

$$\boldsymbol{\mathcal{V}}_k = [\![\mathbf{V}_{k,1}, \mathbf{V}_{k,2}, \ldots, \mathbf{V}_{k,N}]\!] \tag{4b}$$

denote rank-R and rank-S tensor weight for associated with kth diagnostic group respectively. Then, the optimization problem for MG-TCCA is formulated as follows:

$$\begin{aligned} \max_{\{\boldsymbol{\mathcal{U}}_k, \boldsymbol{\mathcal{V}}_k\}_{k=1}^K} \quad & \sum_{k=1}^{K} \frac{1}{n_k - 1} \sum_{i=1}^{n_k} \left\langle \boldsymbol{\mathcal{X}}_{k,i}, \boldsymbol{\mathcal{U}}_k \right\rangle \cdot \left\langle \boldsymbol{\mathcal{Y}}_{k,i}, \boldsymbol{\mathcal{V}}_k \right\rangle \\ \text{subj. to} \quad & \frac{1}{n_k - 1} \sum_{i=1}^{n_k} \left\langle \boldsymbol{\mathcal{X}}_{k,i}, \boldsymbol{\mathcal{U}}_k \right\rangle^2 \leq 1, \ k = 1, \ldots, K, \\ & \frac{1}{n_k - 1} \sum_{i=1}^{n_k} \left\langle \boldsymbol{\mathcal{Y}}_{k,i}, \boldsymbol{\mathcal{V}}_k \right\rangle^2 \leq 1, \ k = 1, \ldots, K, \\ & (1-\alpha) f_1 \left(\{\boldsymbol{\mathcal{U}}_k\}_{k=1}^K \right) + \alpha f_2 \left(\{\boldsymbol{\mathcal{U}}_k\}_{k=1}^K \right) \leq c_1, \\ & (1-\beta) f_1 \left(\{\boldsymbol{\mathcal{V}}_k\}_{k=1}^K \right) + \beta f_2 \left(\{\boldsymbol{\mathcal{V}}_k\}_{k=1}^K \right) \leq c_2. \end{aligned} \tag{5}$$

In (5), different types of norms and regularizations can be used to encourage sparsity and group sparsity. In this paper, we consider the following types of convex regularizations:

$$f_1 \left(\{\boldsymbol{\mathcal{A}}_k\}_{k=1}^K \right) = \sum\nolimits_{k=1}^{K} \|\boldsymbol{\mathcal{A}}_k\|_1, \tag{6a}$$

$$f_2 \left(\{\boldsymbol{\mathcal{A}}_k\}_{k=1}^K \right) = \| \begin{bmatrix} \text{vec}(\boldsymbol{\mathcal{A}}_1) & \ldots & \text{vec}(\boldsymbol{\mathcal{A}}_K) \end{bmatrix} \|_{2,1}. \tag{6b}$$

The parameters $\alpha, \beta \in [0, 1]$ are responsible for regulating the equilibrium between the element-wise and row-wise sparsity levels and $c_1, c_2 > 0$ control the sparsity of tensor

weights. From Prop. 1, Problem (5) is equivalent to:

$$\begin{aligned}\max_{\{\mathcal{U}_k,\mathcal{V}_k\}_{k=1}^K} \quad & \sum_{k=1}^K \mathrm{Cov}\left(\langle \mathcal{X}_k,\mathcal{U}_k\rangle, \langle \mathcal{Y}_k,\mathcal{V}_k\rangle\right)\\ \text{subj. to} \quad & \mathrm{Var}\left(\langle \mathcal{X}_k,\mathcal{U}_k\rangle\right) \le 1,\ k=1,\dots,K,\\ & \mathrm{Var}\left(\langle \mathcal{Y}_k,\mathcal{V}_k\rangle\right) \le 1,\ k=1,\dots,K,\\ & (1-\alpha) f_1\left(\{\mathcal{U}_k\}_{k=1}^K\right) + \alpha f_2\left(\{\mathcal{U}_k\}_{k=1}^K\right) \le c_1,\\ & (1-\beta) f_1\left(\{\mathcal{V}_k\}_{k=1}^K\right) + \beta f_2\left(\{\mathcal{V}_k\}_{k=1}^K\right) \le c_2.\end{aligned} \tag{7}$$

Here, $\{\mathcal{X}_k\}_k$ and $\{\mathcal{Y}_k\}_k$ represent two sets of random tensors aimed at reducing dimensionality.

3.1 A Block Coordinate Ascent Algorithm for MG-TCCA

Next, we propose a block coordinate ascent algorithm for solving (5) (or its equivalent form (7)).

Proposition 1 outlines the block coordinate ascent algorithm, which forms the basis for the development of Algorithm 1. In this algorithm, the factor matrices $(\mathbf{U}_{k,i}, \mathbf{V}_{k,j})$ are updated in pairs, considering various combinations of $k \in [K]$, $i \in [M]$ and $j \in [N]$. By iteratively updating these factor tensors, the Algorithm 1 solves the following series of optimization problems for different $k \in [K]$, $i \in [M]$ and $j \in [N]$, ultimately reaching the desired solution:

$$\begin{aligned}\max_{\mathbf{U}_{k,i},\mathbf{V}_{k,j}} \quad & \mathbf{u}_{k,i}^\top \mathbf{U}_{k,(-i)}^\top \hat{\Sigma}_{\mathbf{x}_{k,(i)},\mathbf{y}_{k,(j)}} \mathbf{V}_{k,(-j)} \mathbf{v}_{k,j}\\ \text{subj. to} \quad & \mathbf{u}_{k,i}^\top \mathbf{U}_{k,(-i)}^\top \hat{\Sigma}_{\mathbf{x}_{k,(i)}} \mathbf{U}_{k,(-i)} \mathbf{u}_{k,i} = 1,\\ & \mathbf{v}_{k,j}^\top \mathbf{V}_{k,(-j)}^\top \hat{\Sigma}_{\mathbf{y}_{k,(j)}} \mathbf{V}_{k,(-j)} \mathbf{v}_{k,j} = 1,\\ & (1-\alpha) f_1\left(\{\mathbf{u}_{k,i}\}_{k=1}^K\right) + \alpha f_2\left(\{\mathbf{u}_{k,i}\}_{k=1}^K\right) \le c_1,\\ & (1-\beta) f_1\left(\{\mathbf{v}_{k,j}\}_{k=1}^K\right) + \beta f_2\left(\{\mathbf{v}_{k,j}\}_{k=1}^K\right) \le c_2.\end{aligned} \tag{8}$$

Here, $\mathbf{u}_{k,i} = \mathrm{vec}(\mathbf{U}_{k,i})$, $\mathbf{v}_{k,j} = \mathrm{vec}(\mathbf{V}_{k,j})$, and $\hat{\Sigma}_{\mathbf{x}_{k,(i)},\mathbf{y}_{k,(j)}}, \hat{\Sigma}_{\mathbf{x}_{k,(i)}}, \hat{\Sigma}_{\mathbf{y}_{k,(j)}}$ are sample estimations of covariance.

Note that $\mathbf{u}_{k,i}$ and $\mathbf{v}_{k,j}$ can be obtained by solving the following classical CCA problem:

$$\max_{\mathbf{u}_{k,i},\mathbf{v}_{k,j}} \rho(\mathbf{u}_{k,i}, \mathbf{v}_{k,j}) := \frac{\mathbf{u}_{k,i}^\top \mathbf{C}_{\mathbf{x}_k,\mathbf{y}_k} \mathbf{v}_{k,j}}{\sqrt{\mathbf{u}_{k,i}^\top \mathbf{C}_{\mathbf{x}_k} \mathbf{u}_{k,i}} \sqrt{\mathbf{v}_{k,j}^\top \mathbf{C}_{\mathbf{y}_k} \mathbf{v}_{k,j}}}, \tag{9}$$

where

$$\mathbf{C}_{\mathbf{x}_k} = \mathbf{U}_{k,(-i)}{}^\top \hat{\Sigma}_{\mathbf{x}_{k(i)}} \mathbf{U}_{k,(-i)}, \tag{10a}$$

$$\mathbf{C}_{\mathbf{y}_k} = \mathbf{V}_{k,(-j)}{}^\top \hat{\Sigma}_{\mathbf{y}_{k(j)}} \mathbf{V}_{k,(-j)}, \tag{10b}$$

$$\mathbf{C}_{\mathbf{x}_k,\mathbf{y}_k} = \mathbf{U}_{k,(-i)}^\top \hat{\Sigma}_{\mathbf{x}_{k(i)},\mathbf{y}_{k(j)}} \mathbf{V}_{k,(-j)}. \tag{10c}$$

By incorporating the numerical method used for computing the generalized eigenvalue problem in classical CCA, we can estimate $\mathbf{u}_{k,i}$ and $\mathbf{v}_{k,j}$ via

$$\begin{pmatrix} \mathbf{0} & \mathbf{C}_{\mathbf{x}_k,\mathbf{y}_k} \\ \mathbf{C}_{\mathbf{y}_k,\mathbf{x}_k} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u}_{k,i} \\ \mathbf{v}_{k.j} \end{pmatrix} = \rho(\mathbf{u}_{k,i}, \mathbf{v}_{k,j}) \begin{pmatrix} \mathbf{C}_{\mathbf{x}_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\mathbf{y}_k} \end{pmatrix} \begin{pmatrix} \mathbf{u}_{k,i} \\ \mathbf{v}_{k,j} \end{pmatrix}. \tag{11}$$

Algorithm 1: A Block Coordinate Ascent Algorithm for MG-TCCA

initialize $\mathbf{U}_{k,i}^{(0)}, \mathbf{V}_{k,j}^{(0)}$ for $(i,j,k) \in [M] \times [N] \times [K]$;
$t \leftarrow 0$;
repeat
  Select $(i,j) \in [M] \times [N]$, Set $\mathbf{s}_1 = 0, \mathbf{s}_2 = 0$;
  for $k = 1, \dots, K$ **do**
    Compute $\mathbf{U}_{k,(-i)}^{(t)}$ using Equation (1a);
    Compute $\mathbf{V}_{k,(-j)}^{(t)}$ using Equation (1b);
    Solve Problem (9) using Equation (11);
    $\mathbf{s}_1 \leftarrow \mathbf{s}_1 + (\mathbf{u}_{k,i}^{(t+1)})^2, \mathbf{s}_2 \leftarrow \mathbf{s}_2 + (\mathbf{v}_{k,j}^{(t+1)})^2$;
    $\rho_k^{(t+1)} \leftarrow \rho(\mathbf{u}_{k,i}, \mathbf{v}_{k,j})$;
  end
  for $k = 1, \dots, K$ **do**
    $\mathbf{u}_{k,i}^{(t+1)} \leftarrow \Theta_\lambda((1-\alpha)\mathbf{u}_{k,i}^{(t+1)}, \alpha\sqrt{\mathbf{s}_1})$;
    $\mathbf{v}_{k,j}^{(t+1)} \leftarrow \Theta_\lambda((1-\beta)\mathbf{v}_{k,j}^{(t+1)}, \beta\sqrt{\mathbf{s}_2})$;
  end
  $t \leftarrow t+1$;
until *mean*$\{\rho_1^{(t+1)}, \rho_2^{(t+1)}, \dots, \rho_K^{(t+1)}\}$ *converges*;
return $\{\hat{\mathcal{U}}_k, \hat{\mathcal{V}}_k\}_{k=1}^K$

Furthermore, to incorporate the $L_{2,1}$-norm and L_1-norm, we utilize the iterative thresholding strategy introduced by [21] for sparse principal component analysis and the hard thresholding strategy introduced by [22] for sparse tensor canonical correlation analysis. Given $\lambda > 0$, the thresholding strategies operate as follows:

$$\Theta_\lambda(\mathbf{w}, \mathbf{a}) = \begin{bmatrix} \theta_\lambda(w_1 + a_1) \\ \vdots \\ \theta_\lambda(w_n + a_n) \end{bmatrix}, \tag{12}$$

where

$$\theta_\lambda(w + a) = \begin{cases} w & |w| + a \ge \lambda \\ 0 & |w| + a < \lambda \end{cases}, \tag{13}$$

where the parameter λ controls the degree of sparsity and is computed based on c_1 (or c_2) in Equation 5 as follows: the vector $\mathbf{w}$ is sorted in descending order, and the value corresponding to index i is designated as λ, such that the cumulative sum of the elements $w_1 + a, w_2 + a, \dots, w_i + a$ is below the threshold c_1 (or c_2). Ultimately, the algorithm can be succinctly summarized as Algorithm 1.

After we obtain the first pair of canonical tensor weights $\{\mathcal{U}, \mathcal{V}\}$, we can employ residuals $\{\mathcal{X}_i - \frac{\langle \mathcal{X}_i, \mathcal{U}\rangle}{\langle \mathcal{U}, \mathcal{U}\rangle}\mathcal{U}, \mathcal{Y}_i - \frac{\langle \mathcal{Y}_i, \mathcal{V}\rangle}{\langle \mathcal{V}, \mathcal{V}\rangle}\mathcal{V}\}$ to fit Model (5) to calculate subsequent pair of canonical tensor weights.

Remark 1. *Algorithm 1 exhibits several advantages in terms of computational and storage costs. In fact, its computational cost is comparable to that of TCCA and Sparse TCCA [21].*

However, the main advantage of MG-TCCA is that its loops can be implemented in parallel, further enhancing its efficiency.

The theorem presented demonstrates the convergence of Algorithm (5) towards a stationary point, which represents a local minimum along each block direction.

Theorem 2. *The estimate* $\{\hat{\boldsymbol{\mathcal{U}}}_k, \hat{\boldsymbol{\mathcal{V}}}_k\}_{k=1}^K$ *obtained from Algorithm 1 corresponds to a stationary point of* (5).

The proof of this theorem is not provided here, as it closely follows the proof of block coordinate maximization for non-differentiable objectives as presented in [3].

4 Experiments

4.1 Synthetic Data

We adopt the probabilistic interpretation of classical CCA proposed by Bach and Jordan [2]. To align the methodology of generating synthetic data for TCCA with classical CCA, we consider vectorized canonical tensors as canonical vectors. This correspondence allows us to employ the following comprehensive procedures. Let $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_d)$ represent a set of d canonical correlations, and let $\mathbf{U} \in \mathbb{R}^{p_1 \times d}$ and $\mathbf{V} \in \mathbb{R}^{q_1 \times d}$ denote two canonical matrices whose columns correspond to the canonical vectors. Based on these given quantities, we formulate the subsequent expression:

$$\Sigma_X = Q_X {R_X}^{-\top} {R_X}^{-1} {Q_X}^\top + T_X(I_{p_1} - Q_X Q_X^\top){T_X}^\top, \quad (14a)$$

$$\Sigma_Y = Q_Y {R_Y}^{-\top} {R_Y}^{-1} {Q_Y}^\top + T_Y(I_{q_1} - Q_Y Q_Y^\top){T_Y}^\top, \quad (14b)$$

$$\Sigma_{XY} = \Sigma_X V \mathrm{diag}(\boldsymbol{\rho}) W^\top \Sigma_Y, \quad (14c)$$

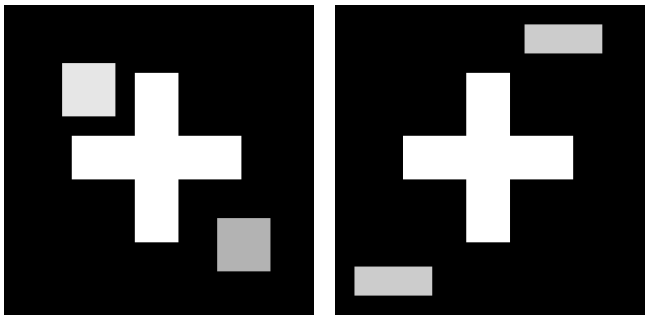
where $\mathbf{U} = Q_X R_X$ and $\mathbf{V} = Q_Y R_Y$ are the QR decomposition of $\mathbf{U}$ and $\mathbf{V}$. In addition, $\mathbf{U}$ and $\mathbf{V}$ satisfy $\mathbf{U}^\top \Sigma_X \mathbf{U} = \mathbf{I}_{p_1}$ and $\mathbf{V}^\top \Sigma_Y \mathbf{V} = \mathbf{I}_{q_1}$ respectively for random $T_X \in \mathbb{R}^{p_1 \times d}$ and $T_Y \in \mathbb{R}^{q_1 \times d}$ in normal distribution. Subsequently, synthetic data is generated through sampling data from the underlying distribution:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{bmatrix}, \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{bmatrix}\right). \quad (15)$$

Within the conducted experimental analysis, it is postulated that the synthetic dataset contains two discrete groups. Each group is associated with closely related yet distinguishable ground truth canonical correlations, denoted as $\boldsymbol{\rho}_1 = 0.95$ and $\boldsymbol{\rho}_2 = 0.97$, along with corresponding canonical tensor weights, namely $\{\boldsymbol{\mathcal{U}}_1, \boldsymbol{\mathcal{V}}_1\}$ and $\{\boldsymbol{\mathcal{U}}_2, \boldsymbol{\mathcal{V}}_2\}$. The true $\boldsymbol{\mathcal{U}}_1$ and $\boldsymbol{\mathcal{U}}_2$ are two 100-dimensional vectors, wherein 96 entries exhibit identical values, while the remaining 4 entries possess distinct values, as illustrated in Figure 2. Furthermore, the true $\boldsymbol{\mathcal{V}}_1$ and $\boldsymbol{\mathcal{V}}_2$ are two 64×64 matrices representing images, as illustrated in Figure 3. In these visual representations, white pixels represent 1, while black pixels represent 0. Drawing upon the aforementioned ground truth information and generative model 15, a synthetic dataset is generated, including a total of 2000 samples. Of these, 1000 samples belong to Group 1, while the remaining 1000 samples belong to Group 2.

(a) $\boldsymbol{\mathcal{U}}_1$ (b) $\boldsymbol{\mathcal{U}}_2$

Figure 2. True canonical tensor weights $\boldsymbol{\mathcal{U}}_1$ for Group 1, and $\boldsymbol{\mathcal{U}}_2$ for Group 2 in numerical experiments



(a) $\boldsymbol{\mathcal{V}}_1$ (b) $\boldsymbol{\mathcal{V}}_2$

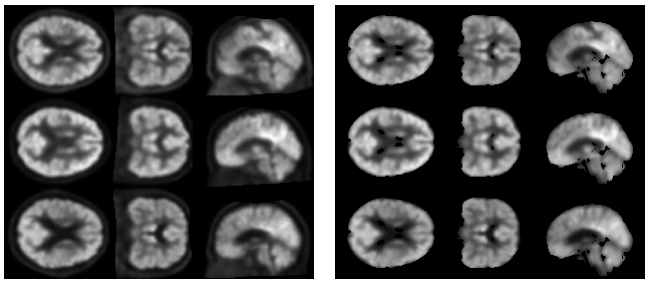
Figure 3. True canonical tensor weights $\boldsymbol{\mathcal{V}}_1$ for Group 1, and $\boldsymbol{\mathcal{V}}_2$ for Group 2 in numerical experiments

4.2 Real Data

Data used in the preparation of this article is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu) [31, 32]. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. All participants provided written informed consent and study protocols were approved by each participating site's Institutional Review Board (IRB). Up-to-date information about the ADNI is available at www.adni-info.org.

We analyze two ADNI imaging modalities: AV-45-PET [12] (measuring amyloid burden) and FDG-PET [12] (measuring glucose metabolism). Both AV-45 and FDG imaging data play a crucial role in the diagnosis and assessment of AD. These imaging techniques facilitate the detection and quantification of beta-amyloid accumulation and glucose metabolism changes, which represent crucial biomarkers of the disease. Furthermore, these biomarkers can be employed to monitor disease progression and evaluate the efficacy of treatments, including pharmaceutical interventions aimed at reducing beta-amyloid accumulation or enhancing glucose metabolism in the brain [4]. In addition, the utilization of AV-45 and FDG datasets in research is also driven by the

aim of examining the association between these two modalities. Previous investigations have furnished proof of such a connection, as illustrated by the findings of Hsiao et al. [11, 30].



(a) Without brain mask **(b)** With brain mask

Figure 4. The efficacy of the brain mask is illustrated by comparing cross-sectional images: the original PET images on the left and the images post-application of the brain mask on the right.

The data processing procedure consists of the following steps. First, the Statistical Parametric Mapping software tool [25] is used to register AV-45-PET scans and FDG-PET scans into the standard brain space defined by the Montreal Neurological Institute (i.e., MNI-space). Second, a brain mask is applied to the original images to minimize the interference of extraneous structures, such as the skull and other sources of noise, as illustrated in Figure 4. Third, to optimize computational efficiency, each image is transformed from dimensions of $(91, 109, 91)$ to $(22, 27, 22)$. In addition, sex is utilized to differentiate groups, and subsequently, the dataset is divided into training and testing sets randomly. The former comprises 242 samples from the female group and 299 samples from the male group, whereas the latter contains 162 samples from the female group and 200 samples from the male group.

4.3 Evaluation Criteria

The outputs of TCCA and MG-TCCA algorithms contain canonical correlation coefficients and canonical tensor weights. Canonical correlation coefficients serve as indicators of the degree of linear association between two sets of variables, while canonical tensor weights denote the respective weighting of each variable in the original variables.

Evaluation Criteria of Synthetic Data. For synthetic data, the assessment of estimation accuracy is conducted through the computation of cosine similarity between the canonical tensor weights $\{\boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}\}$ present in the ground truth and those estimated canonical tensor weights $\{\hat{\boldsymbol{\mathcal{U}}}, \hat{\boldsymbol{\mathcal{V}}}\}$. This evaluation method is defined as follows:

$$\angle\left(\boldsymbol{\mathcal{U}}, \hat{\boldsymbol{\mathcal{U}}}\right) = \frac{\langle \mathbf{u}, \hat{\mathbf{u}} \rangle}{\|\mathbf{u}\|_2 \|\hat{\mathbf{u}}\|_2}, \text{ and } \angle\left(\boldsymbol{\mathcal{V}}, \hat{\boldsymbol{\mathcal{V}}}\right) = \frac{\langle \mathbf{v}, \hat{\mathbf{v}} \rangle}{\|\mathbf{v}\|_2 \|\hat{\mathbf{v}}\|_2}, \quad (16)$$

where $\mathbf{u}$ is the vectorization of $\boldsymbol{\mathcal{U}}$, $\hat{\mathbf{u}}$ is the vectorization of $\hat{\boldsymbol{\mathcal{U}}}$, $\mathbf{v}$ is the vectorization of $\boldsymbol{\mathcal{V}}$, and $\hat{\mathbf{v}}$ is the vectorization of $\hat{\boldsymbol{\mathcal{V}}}$. Additionally, the visualization of the estimated canonical tensor weights is exhibited as a means of facilitating the evaluation and comparison of the recovery performance.

Evaluation Criteria of Real Data. To evaluate the effectiveness of MG-TCCA on ADNI, canonical tensor weights obtained from the training set are utilized on the testing set to derive a testing canonical correlation, which serves as an unbiased estimation of the effectiveness of canonical tensor weights on new data. The evaluation of the testing canonical correlation is carried out using the Pearson correlation coefficient, computed as follows:

$$\rho = \frac{\sum_{i=1}^{n} \langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{U}} \rangle \cdot \langle \boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{V}} \rangle}{\sqrt{\sum_{i=1}^{n} \langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{U}} \rangle^2} \sqrt{\sum_{i=1}^{n} \langle \boldsymbol{\mathcal{Y}}_i, \boldsymbol{\mathcal{V}} \rangle^2}}. \quad (17)$$

Additionally, to demonstrate the benefits of identifying both shared and unique signals across multiple groups, we utilize the automated anatomical labeling (AAL) atlas to partition brain imaging into 116 regions of interest (ROI) and determine the average intensity of canonical tensor weights of each brain region. This approach facilitates a more intuitive differentiation among the canonical tensor weights via visualization.

4.4 Experimental Setup

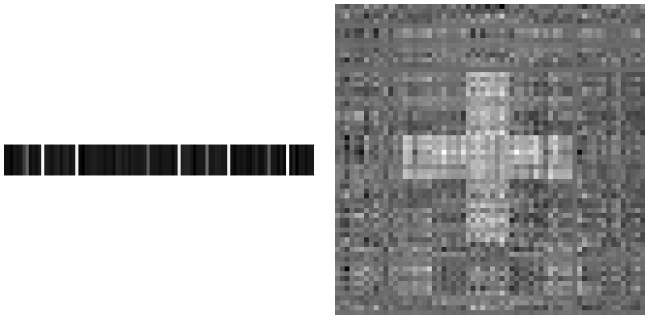
Selection of Tuning Parameter. The hyperparameter λ in 13 controls the sparsity level, similar to the hyperparameters c_1 and c_2 in Problem 5. In numerical experiments, the proportion of zero entries ($p\%$) in the input vectors serves a similar role as λ. For synthetic data, $p\%$ is arbitrarily chosen, while α and β are set to 0.4. With real data, a grid search determines suitable values for $p\%$, α, and β in MG-TCCA. Since both modalities are voxel images, the same $p\%$ is used for both, selected from 0.2, 0.4, 0.6, 0.8. α and β are searched in 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. Various parameter combinations are evaluated on the training set and tested on the testing set, selecting parameters with the lowest overfitting within each group.

Selection of Baseline. The utilization of Classical CCA as the baseline model in this investigation is omitted, given that Min [22] has empirically presented the superior efficacy of TCCA to classical CCA, particularly when applied to tensor data. Consequently, in order to enable subsequent comparisons, a sequence of experiments is conducted in the following manner: initially, the TCCA method is individually employed on each dataset specific to the respective sexes; subsequently, the MG-TCCA method is applied to the entire dataset to derive canonical tensor weights for each distinct group. For synthetic data, in order to conduct further

comparisons, supplementary numerical experiments are conducted using Sparse Tensor Canonical Correlation Analysis (STCCA), as proposed by Min [22].

4.5 Results and Discussions of Synthetic Data

Prior to conducting the comparative experiment, we initially applied TCCA to the entire dataset and visualized its performance. As depicted in Figure 5, the global canonical tensor weights are only able to identify the features shared by both groups, referred to as "cross". However, they are adversely affected by the differences between the two groups in the corners, resulting in an inability to discern the characteristics of individual groups and the presence of significant noise. Therefore, in order to address this issue, it is necessary to propose and validate novel algorithms.



(a) $\hat{\mathcal{U}}$ (b) $\hat{\mathcal{V}}$

Figure 5. Estimated canonical tensor weights $\hat{\mathcal{U}}$ and $\hat{\mathcal{V}}$ of TCCA on the entire synthetic dataset.

Figure 6 displays the visualization of the estimated canonical tensor weights pertaining to Group 1 and Group 2. These estimates are obtained by employing the techniques of TCCA and STCCA on separate groups, as well as MG-TCCA on the complete dataset. Comparisons made with these figures reveal that, despite applying TCCA independently to each group, the TCCA method fails to capture the group-specific features in canonical weights, whereas the STCCA and MG-TCCA methods overcome such limitations, especially for tensor data. The MG-TCCA method demonstrates a high level of effectiveness in discerning distinct signals pertaining to specific groups while concurrently preserving shared signals across all groups. This method exhibits enhanced resilience in the identification of the aforementioned features when compared to STCCA. These observations are consistent with the intended goal of the MG-TCCA approach.

Table 1a and Table 1b compare the estimation accuracy of TCCA, STCCA, and MG-TCCA of 100 replicates. Based on the presented results, the following observations can be made. Firstly, it is evident that the MG-TCCA consistently outperforms the TCCA and STCCA methods in effectively restoring the canonical tensor weights of the ground truth. Secondly, it is worth noting that across different selections

Table 1. Numerical results of synthetic data regarding Cosine Similarity $\angle\left(\mathcal{U},\hat{\mathcal{U}}\right)$ and $\angle\left(\mathcal{V},\hat{\mathcal{V}}\right)$. The best ones in each row are bold and the second best one is underlined. The analysis is carried out by varying the rank of the canonical tensor weights, specifically exploring ranks 1, 2, and 3. In this context, a higher numerical value is indicative of superior model performance.

(a) $\angle\left(\mathcal{U},\hat{\mathcal{U}}\right)$

Group	Rank	TCCA	STCCA	MG-TCCA
Group 1	1	0.9965	0.9965	**0.9967**
	2	0.9969	0.9969	**0.9973**
	3	0.9970	0.9974	**0.9982**
Group 2	1	0.9922	0.9924	**0.9926**
	2	0.9965	0.9967	**0.9969**
	3	0.9983	0.9982	**0.9984**

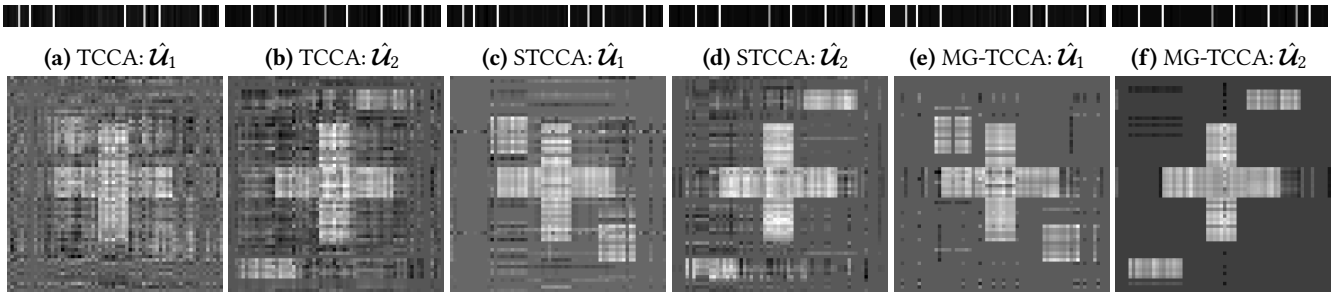
(b) $\angle\left(\mathcal{V},\hat{\mathcal{V}}\right)$

Group	Rank	TCCA	STCCA	MG-TCCA
Group 1	1	0.7958	0.7992	**0.8039**
	2	0.8525	0.8559	**0.8613**
	3	0.8573	0.8829	**0.8986**
Group 2	1	0.7512	0.7539	**0.7545**
	2	0.8620	0.8657	**0.8685**
	3	0.9057	0.9125	**0.9138**

Table 2. Computation time (mean±std) of 100 repeated experiments for different ranks. The computational time of TCCA (STCCA) corresponds to the cumulative sum of the computational time required for applying TCCA (STCCA) to individual group experiments. In contrast, MG-TCCA is applied to the entire dataset. The time is reported in seconds. All experiments are run on Intel(R) Xeon(R) CPU E5-2660.

Rank	TCCA	STCCA	MG-TCCA
1	**7.11 (±0.52)**	12.43 (±0.83)	35.13 (±0.79)
2	**13.64 (±0.13)**	13.66 (±0.19)	30.07 (±0.32)
3	25.44 (±0.24)	20.10 (±0.25)	**16.40 (±0.47)**

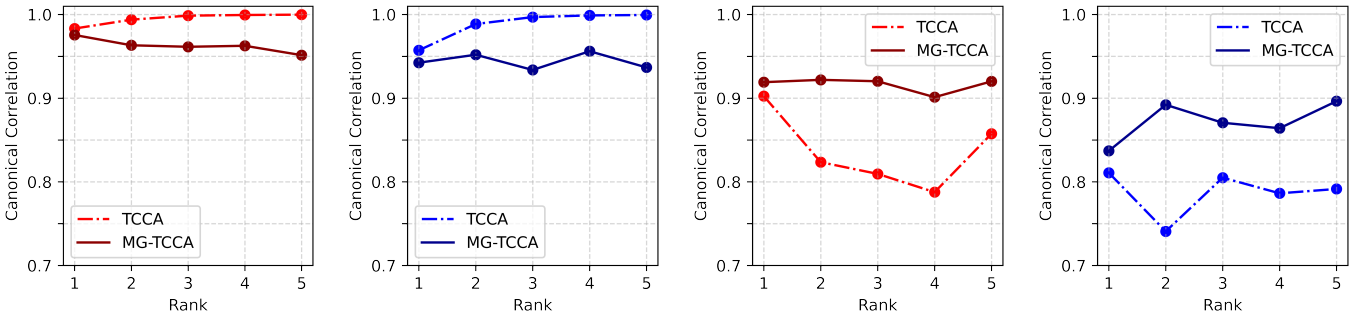
of the rank for canonical tensor weights, MG-TCCA exhibits superior recovery performance compared to the TCCA and STCCA approaches, further accentuating the superiority of the MG-TCCA method. Additionally, a higher rank selection leads to a better performance in estimation accuracy. The computation times for TCCA, STCCA, and MG-TCCA on the dataset are presented in Table 2. The results demonstrate that TCCA exhibits a shorter processing time than STCCA and MG-TCCA when the rank of canonical tensor weights is low. Moreover, it is observed that a greater rank selection for the canonical tensor weights enhances the convergence speed of STCCA and MG-TCCA.



(a) TCCA: $\hat{\mathcal{U}}_1$ **(b)** TCCA: $\hat{\mathcal{U}}_2$ **(c)** STCCA: $\hat{\mathcal{U}}_1$ **(d)** STCCA: $\hat{\mathcal{U}}_2$ **(e)** MG-TCCA: $\hat{\mathcal{U}}_1$ **(f)** MG-TCCA: $\hat{\mathcal{U}}_2$

(g) TCCA: $\hat{\mathcal{V}}_1$ **(h)** TCCA:$\hat{\mathcal{V}}_2$ **(i)** STCCA: $\hat{\mathcal{V}}_1$ **(j)** STCCA: $\hat{\mathcal{V}}_2$ **(k)** MG-TCCA:$\hat{\mathcal{V}}_1$ **(l)** MG-TCCA:$\hat{\mathcal{V}}_2$

Figure 6. The sequence visualizes estimated canonical tensor weights $\hat{\mathcal{V}}_1, \hat{\mathcal{V}}_2$ for synthetic data generated by TCCA, STCCA on distinct groups of the dataset, and MG-TCCA on the entire dataset, arranged from left to right as follows: TCCA on Group 1, TCCA on Group 2, STCCA on Group 1, STCCA on Group 2, MG-TCCA on Group 1, and MG-TCCA on Group 2.



(a) Training Set for Female **(b)** Training Set for Male **(c)** Testing Set for Female **(d)** Testing Set for Male

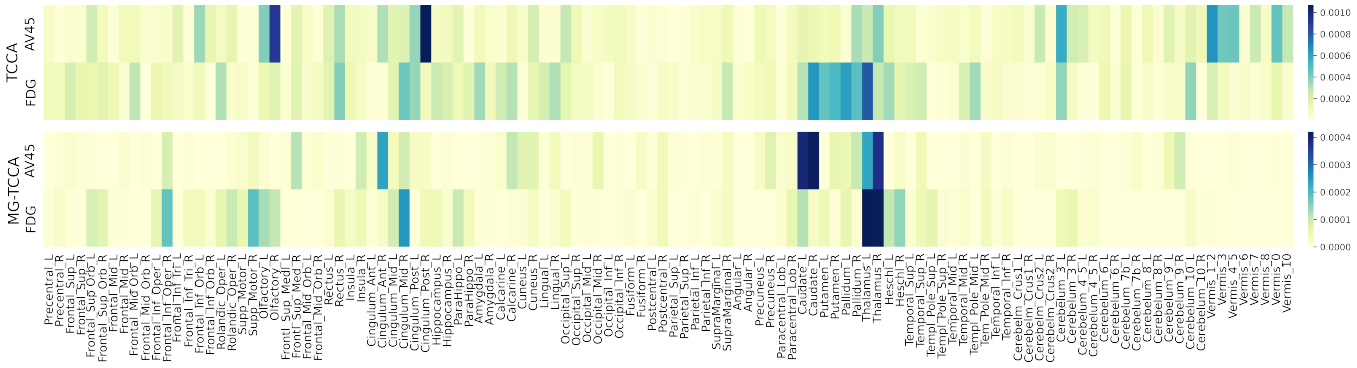
Figure 7. This sequence displays the canonical correlations produced by TCCA on discrete subsets of the data and MG-TCCA on the entire dataset. The subsets are arranged from left to right, consisting of canonical correlations for the female training set, male training set, female testing set, and male testing set.

4.6 Results and Discussions of Real Data

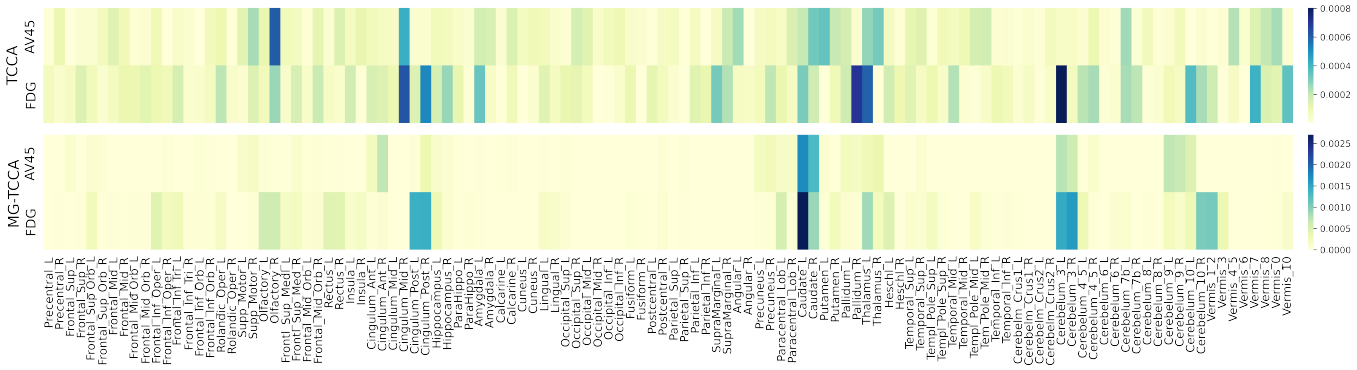
Figure 7 shows the consistent superior performance of the MG-TCCA method over the TCCA method in maximizing canonical correlations between two modalities. This is observed across different choices of the rank of canonical tensor weights, specifically on the testing set within each sex group. Moreover, considering the combined analysis of the training and testing sets, the MG-TCCA method exhibits greater advantages in addressing overfitting than the TCCA method.

Figure 8 displays the absolute value of the average intensity of canonical tensor weights assigned to each brain region, indicating the importance of features between the modalities. The conflicting outcomes between TCCA and MG-TCCA can be attributed to the small sample size in the Male or Female group, leading to overfitting in TCCA, as shown in Figure 7. Further analysis of these figures reveals that TCCA captures numerous superfluous features and exhibits a limited ability to distinguish crucial shared features in the dataset when comparing outcomes between the female and male groups. In contrast, MG-TCCA demonstrates a concentrated signal in the caudate nucleus, with variations observed in the thalamus and cerebellum across the two groups. These findings align with previous studies on abnormal metabolism reduction, accumulation of β-amyloid, and the association between AV-45 and FDG in caudate regions [11, 13, 16], as well as the increased distribution of amyloid in thalamus regions [13, 15]. Additionally, these results support conclusions about sex differences in brain metabolism in the thalamus and cerebellum regions [24].

This experiment primarily contributes to verifying the benefits of MG-TCCA, specifically in terms of enhancing canonical correlations between two sets of variables and detecting sex disparities while preserving shared characteristics in canonical tensor weights. The findings of this study regarding sex-related differences in characteristics associated with AD align with previous research, suggesting the potential for targeted and effective treatments and interventions for these disorders. However, several limitations exist in this study. Firstly, the selection of hyperparameters in the proposed approach is subjective and may not consistently yield optimal outcomes. To address this, employing a bilevel structure **??** or Bayesian optimization [14] is necessary for automatically optimizing these hyperparameters. Secondly, the



(a) Heatmap of feature importance for female



(b) Heatmap of feature importance for male

Figure 8. Comparison of feature importance derived from canonical tensor weights. Each row corresponds to a particular method, either TCCA or MG-TCCA, and contains two modalities, namely AV-45 and FDG. TCCA is utilized to derive tensor weights by separately applying it to distinct Male and Female groups. Conversely, MG-TCCA directly operates on the entire dataset, yielding tensor weights for both Male and Female groups collectively.

efficiency of MG-TCCA is compromised when additional constraints are introduced, particularly when the rank of tensor weights is excessively low. Thirdly, the current experiments only involve mapping two modalities to a one-dimensional subspace, resulting in significant information loss. Future investigations will focus on utilizing higher-rank canonical tensor weights to identify more brain regions associated with the correlation between AV-45 and FDG. Furthermore, this algorithm will allow the exploration of other potentially influential factors, including race, age, educational level, and sex, that may impact Alzheimer's disease.

5 Conclusion

In this study, a novel approach called MG-TCCA is proposed to analyze group-specific correlations and features among two datasets. This method considers the heterogeneity adequately of datasets due to pooling from multiple groups while preserving the advantages of TCCA in identifying consistent signals across groups. The results obtained from synthetic data and real brain imaging data demonstrate the effectiveness of MG-TCCA in identifying group-specific patterns as well as group-shared patterns. This proposed method has the potential to enhance our understanding of group disparities in the brain and their correlations. Additionally, it can be applied to a diverse range of multi-modal brain imaging data, including functional magnetic resonance imaging (fMRI), diffusion tensor imaging (DTI), and PET, to aid in the diagnosis of neurological and psychiatric disorders. It can also be applied to identify correlations between multidimensional data sets at multiple levels, such as genetics, multi-omics, imaging, fluid biomarker, and phenotypic outcome. Further research will focus on validating the method on larger datasets with more imbalanced groups and exploring its potential application in clinical settings.

References

[1] Fares Al-Shargie, Tong Boon Tang, and Masashi Kiguchi. 2017. Assessment of mental stress effects on prefrontal cortical activities using canonical correlation analysis: an fNIRS-EEG study. *Biomedical optics express* 8, 5 (2017), 2583–2598.

[2] Francis R Bach and MI Jordan. 2005. A probabilistic interpretation of canonical correlation analysis. (2005).

[3] Bilian Chen, Simai He, Zhening Li, and Shuzhong Zhang. 2012. Maximum block improvement and polynomial optimization. *SIAM Journal on Optimization* 22, 1 (2012), 87–107.

[4] Jeffrey Cummings. 2019. The role of biomarkers in Alzheimer's disease drug development. *Reviews on Biomarker Studies in Psychiatric and Neurodegenerative Disorders* (2019), 29–61.

[5] Lei Du, Jin Zhang, Fang Liu, Minjianan Zhang, Huiai Wang, Lei Guo, and Junwei Han. 2020. Mining high-order multimodal brain image associations via sparse tensor canonical correlation analysis. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 570–575.

[6] Ralph B Dunham and David J Kravetz. 1975. Canonical correlation analysis in a predictive system. *The Journal of Experimental Education* 43, 4 (1975), 35–42.

[7] Sarah Tomaszewski Farias, DAN Mungas, Bruce Reed, Mary N Haan, and William J Jagust. 2004. Everyday functioning in relation to cognitive functioning and neuroimaging in community-dwelling Hispanic and non-Hispanic older adults. *Journal of the International Neuropsychological Society* 10, 3 (2004), 342–354.

[8] A. Gossmann, P. Zille, V. Calhoun, and Y. P. Wang. 2018. FDR-Corrected Sparse Canonical Correlation Analysis With Applications to Imaging Genomics. *IEEE Trans Med Imaging* 37, 8 (2018), 1761–1774. https://doi.org/10.1109/TMI.2018.2815583

[9] CE Hopkins. 1969. Statistical analysis by canonical correlation: a computer application. *Health services research* 4, 4 (1969), 304.

[10] Harold Hotelling. 1992. Relations between two sets of variates. *Breakthroughs in statistics: methodology and distribution* (1992), 162–190.

[11] Ing-Tsung Hsiao, Chin-Chang Huang, Chia-Ju Hsieh, Wen-Chun Hsu, Shiaw-Pyng Wey, Tzu-Chen Yen, Mei-Ping Kung, and Kun-Ju Lin. 2012. Correlation of early-phase 18 F-florbetapir (AV-45/Amyvid) PET images to FDG images: preliminary studies. *European journal of nuclear medicine and molecular imaging* 39 (2012), 613–620.

[12] William J Jagust, Susan M Landau, Robert A Koeppe, Eric M Reiman, Kewei Chen, Chester A Mathis, Julie C Price, Norman L Foster, and Angela Y Wang. 2015. The Alzheimer's disease neuroimaging initiative 2 PET core: 2015. *Alzheimer's & Dementia* 11, 7 (2015), 757–771.

[13] Jiehui Jiang, Yiwu Sun, Hucheng Zhou, Shaoping Li, Zhemin Huang, Ping Wu, Kuangyu Shi, Chuantao Zuo, Neuroimaging Initiative, et al. 2018. Study of the influence of age in 18F-FDG PET images using a data-driven approach and its evaluation in Alzheimer's disease. *Contrast media & molecular imaging* 2018 (2018).

[14] Kirthevasan Kandasamy, Karun Raju Vysyaraju, Willie Neiswanger, Biswajit Paria, Christopher R Collins, Jeff Schneider, Barnabas Poczos, and Eric P Xing. 2020. Tuning hyperparameters without grad students: Scalable and robust bayesian optimisation with dragonfly. *The Journal of Machine Learning Research* 21, 1 (2020), 3098–3124.

[15] David Keator, Julie C Price, William Charles Kreisl, Michael Phelan, Michael A Yassa, Eric Doran, Christy Hom, Dana Nguyen, Florence Lai, Lisa Taylor, et al. 2019. P4-332: BASELINE AMYLOID (18F-AV-45 PET) DISTRIBUTIONS BY CONSENSUS DIAGNOSIS FROM THE BIOMARKERS OF ALZHEIMER'S DISEASE IN ADULTS WITH DOWN SYNDROME (ADDS) CONSORTIUM. *Alzheimer's & Dementia* 15 (2019), P1423–P1424.

[16] Ji Eun Kim, Dong-Kyun Lee, Ji Hye Hwang, Chan-Mi Kim, Yeji Kim, Jae-Hong Lee, Jong Min Lee, Jee Hoon Roh, et al. 2022. Regional Comparison of Imaging Biomarkers in the Striatum between Early- and Late-onset Alzheimer's Disease. *Experimental Neurobiology* 31, 6 (2022), 401–408.

[17] Tae-Kyun Kim, Shu-Fai Wong, and Roberto Cipolla. 2007. Tensor canonical correlation analysis for action classification. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1–8.

[18] Sun Ho Lee and Seungjin Choi. 2007. Two-dimensional canonical correlation analysis. *IEEE Signal Processing Letters* 14, 10 (2007), 735–738.

[19] D. Lin, V. D. Calhoun, and Y. P. Wang. 2014. Correspondence between fMRI and SNP data by group sparse canonical correlation analysis. *Med Image Anal* 18, 6 (2014), 891–902. https://doi.org/10.1016/j.media.2013.10.010

[20] Yong Luo, Dacheng Tao, Kotagiri Ramamohanarao, Chao Xu, and Yonggang Wen. 2015. Tensor canonical correlation analysis for multi-view dimension reduction. *IEEE transactions on Knowledge and Data Engineering* 27, 11 (2015), 3111–3124.

[21] Zongming Ma. 2013. Sparse principal component analysis and iterative thresholding. (2013).

[22] Eun Jeong Min, Eric C Chi, and Hua Zhou. 2019. Tensor canonical correlation analysis. *Stat* 8, 1 (2019), e253.

[23] Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford R Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. 2005. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's & Dementia* 1, 1 (2005), 55–66.

[24] Alessandra Nicoletti, Roberta Baschi, Calogero Edoardo Cicero, Salvatore Iacono, Vincenzina Lo Re, Antonina Luca, Giuseppe Schirò, Roberto Monastero, et al. 2023. Sex and gender differences in Alzheimer's disease, Parkinson's disease, and Amyotrophic Lateral Sclerosis: a narrative review. *Mechanisms of Ageing and Development* (2023), 111821.

[25] William D Penny, Karl J Friston, John T Ashburner, Stefan J Kiebel, and Thomas E Nichols. 2011. *Statistical parametric mapping: the analysis of functional brain images*. Elsevier.

[26] Harold Pimentel, Zhiyue Hu, and Haiyan Huang. 2018. Biclustering by sparse canonical correlation analysis. *Quantitative Biology* 6, 1 (2018), 56–67.

[27] L. Shen and P. M. Thompson. 2020. Brain Imaging Genomics: Integrated Analysis and Machine Learning. *Proc IEEE Inst Electr Electron Eng* 108, 1 (2020), 125–162.

[28] Hai Shu, Zhe Qu, and Hongtu Zhu. 2022. D-GCCA: Decomposition-based Generalized Canonical Correlation Analysis for Multi-view High-dimensional Data. *Journal of Machine Learning Research* 23, 169 (2022), 1–64.

[29] Michael J Sullivan. 1982. Distribution of Edaphic Diatoms in a Mississippi Salt Marsh: A Canonical Correlation Analysis. *Journal of Phycology* 18, 1 (1982), 130–133.

[30] Matthieu Vanhoutte, Brigitte Landeau, Siya Sherif, Vincent de la Sayette, Sophie Dautricourt, Ahmed Abbas, Alain Manrique, Anne Chocat, and Gaël Chételat. 2021. Evaluation of the early-phase [18F] AV45 PET as an optimal surrogate of [18F] FDG PET in ageing and Alzheimer's clinical syndrome. *NeuroImage: Clinical* 31 (2021), 102750.

[31] Michael W Weiner, Dallas P Veitch, Paul S Aisen, Laurel A Beckett, Nigel J Cairns, Robert C Green, Danielle Harvey, Clifford R Jack, William Jagust, Enchi Liu, et al. 2013. The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's & Dementia* 9, 5 (2013), e111–e194.

[32] Michael W Weiner, Dallas P Veitch, Paul S Aisen, Laurel A Beckett, Nigel J Cairns, Robert C Green, Danielle Harvey, Clifford R Jack Jr, William Jagust, John C Morris, et al. 2017. Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. *Alzheimer's & Dementia* 13, 4 (2017), e1–e85.

[33] Sarah Whittle, Murat Yücel, Marie BH Yap, and Nicholas B Allen. 2011. Sex differences in the neural correlates of emotion: evidence from neuroimaging. *Biological psychology* 87, 3 (2011), 319–333.

[34] J. Yan, L. Du, S. Kim, S. L. Risacher, H. Huang, J. H. Moore, A. J. Saykin, L. Shen, and Initiative Alzheimer's Disease Neuroimaging. 2014. Transcriptome-guided amyloid imaging genetic analysis via a novel structured sparse learning algorithm. *Bioinformatics* 30, 17 (2014), i564–i571. https://doi.org/10.1093/bioinformatics/btu465

[35] Yu Zhang, Guoxu Zhou, Jing Jin, Yangsong Zhang, Xingyu Wang, and Andrzej Cichocki. 2017. Sparse Bayesian multiway canonical correlation analysis for EEG pattern recognition. *Neurocomputing* 225 (2017), 103–110.