

FR-CCA: Fair Multimodal Representation Learning for Equitable Alzheimer's Disease Diagnosis

Bojian Hou*, Zhanliang Wang*, Zhuoping Zhou, Boning Tong, Li Shen[§]

* Equal contribution, Email: bojianh@upenn.edu, zhanliang.wang@penntestmed.upenn.edu

[§] Corresponding author: Email: li.shen@penntestmed.upenn.edu University of Pennsylvania, Philadelphia, PA, USA

1. Introduction

Problem Statement:

- Canonical Correlation Analysis (CCA) is widely used for multimodal data analysis but lacks fairness considerations
- Traditional CCA can amplify societal biases related to sensitive attributes (sex, race, age)
- Previous fair CCA methods focus only on correlation analysis, ignoring downstream classification tasks

Our Contribution:

- Novel Fair Representation CCA (FR-CCA) that ensures independence between learned representations and sensitive attributes
- Jointly optimizes for both fairness and classification performance
- Maintains high correlation while improving fairness in subsequent tasks

2. Illustrations

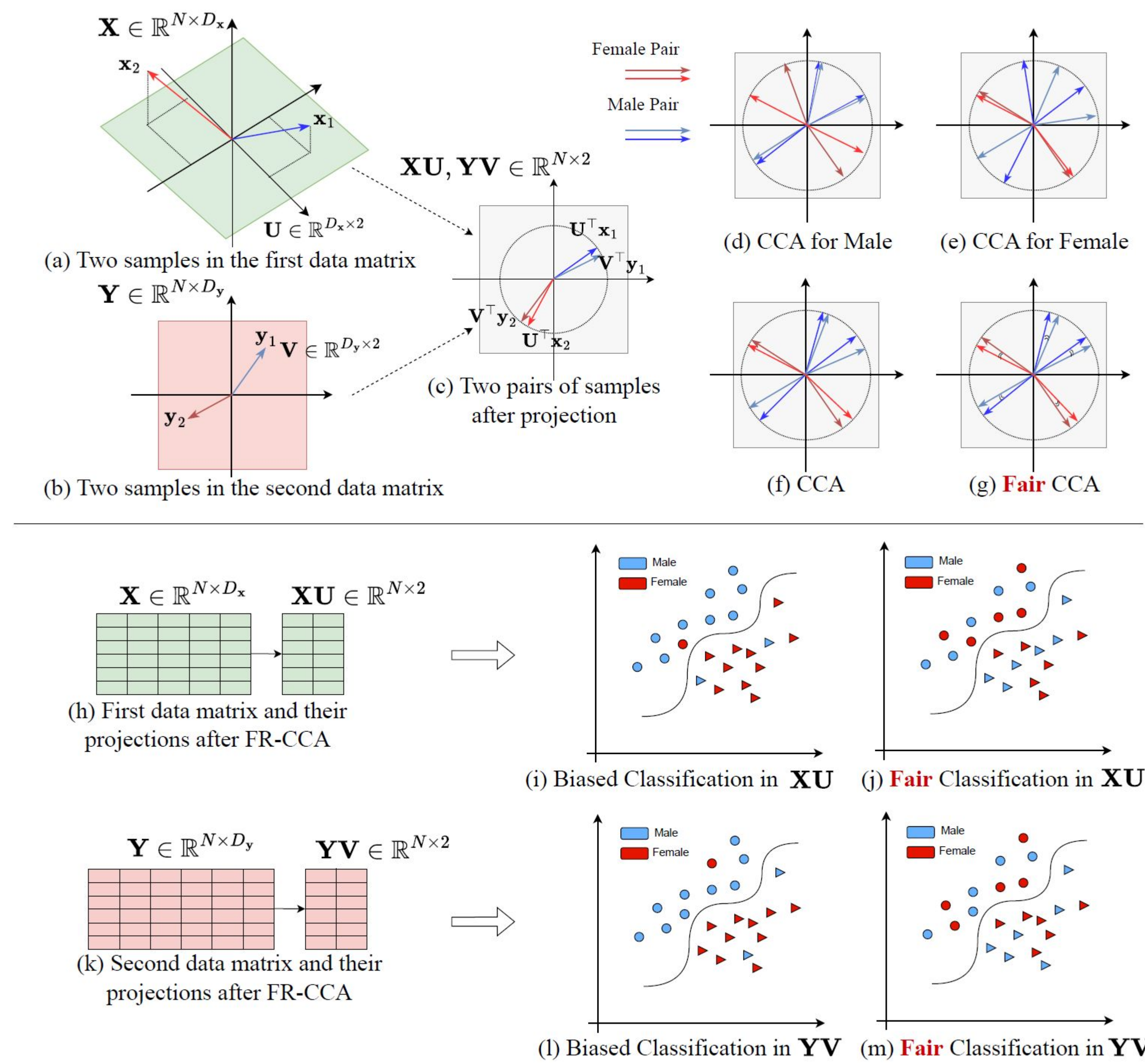


Figure 1: Illustration of FR-CCA with the sensitive attribute sex as an example (female and male). (a)–(c) demonstrate the general framework of CCA, while (d)–(g) provide a comparison of the projected results using various strategies. It is important to note that the correlation between two corresponding samples is inversely associated with the angle formed by their projected vectors. FR-CCA aims to equalize the average angles among different groups. After projection ((h) and (k)), our FR-CCA can lead to fair classification ((j) and (m)), where the classification results will not be affected by the sex information.

3. Methods

FR-CCA Framework:

- **Step 1:** Project data onto nullspace orthogonal to sensitive attributes
- **Step 2:** Apply standard CCA on projected data
- **Step 3:** Train fair classifiers on resulting representations

Key Innovation:

- Enforces statistical independence: $\mathbf{U}^\top \mathbf{X}^\top \hat{\mathbf{z}} = 0$, $\mathbf{V}^\top \mathbf{Y}^\top \hat{\mathbf{z}} = 0$
- Where $\hat{\mathbf{z}}$ is the centered sensitive attribute vector

Optimization Problem:

$$\begin{aligned} \max_{\mathbf{U}, \mathbf{V}} \quad & \text{trace}(\mathbf{U}^\top \mathbf{X}^\top \mathbf{Y} \mathbf{V}) \quad \text{subject to} \quad \mathbf{U} \in \mathcal{U}, \mathbf{V} \in \mathcal{V}, \\ \text{where } \mathcal{U} := & \left\{ \mathbf{U} \in \mathbb{R}^{D_x \times R} \mid \mathbf{U}^\top \mathbf{X}^\top \mathbf{X} \mathbf{U} = \mathbf{I}_R; \quad \forall f: \mathbb{R}^R \rightarrow \mathbb{R}, \right. \\ & \left. f(\mathbf{U}^\top \mathbf{X}_i^\top) \text{ and } z_i \text{ are statistically independent} \right\}, \quad (4) \\ \mathcal{V} := & \left\{ \mathbf{V} \in \mathbb{R}^{D_y \times R} \mid \mathbf{V}^\top \mathbf{Y}^\top \mathbf{Y} \mathbf{V} = \mathbf{I}_R; \quad \forall g: \mathbb{R}^R \rightarrow \mathbb{R}, \right. \\ & \left. g(\mathbf{V}^\top \mathbf{Y}_i^\top) \text{ and } z_i \text{ are statistically independent} \right\}. \end{aligned}$$

$$\begin{aligned} \max_{\mathbf{U}, \mathbf{V}} \quad & \text{trace}(\mathbf{U}^\top \mathbf{X}^\top \mathbf{Y} \mathbf{V}) \quad \text{subject to} \quad \mathbf{U} \in \mathcal{U}', \mathbf{V} \in \mathcal{V}', \\ \text{where } \mathcal{U}' := & \left\{ \mathbf{U} \in \mathbb{R}^{D_x \times R} \mid \mathbf{U}^\top \mathbf{X}^\top \mathbf{X} \mathbf{U} = \mathbf{I}_R; \right. \\ & \left. \forall \mathbf{w}_1 \in \mathbb{R}^R, b_1 \in \mathbb{R}, \text{Cov}(\mathbf{w}_1^\top \mathbf{U}^\top \mathbf{X}_i^\top + b_1, z_i) = 0 \right\}, \quad (6) \\ \mathcal{V}' := & \left\{ \mathbf{V} \in \mathbb{R}^{D_y \times R} \mid \mathbf{V}^\top \mathbf{Y}^\top \mathbf{Y} \mathbf{V} = \mathbf{I}_R; \right. \\ & \left. \forall \mathbf{w}_2 \in \mathbb{R}^R, b_2 \in \mathbb{R}, \text{Cov}(\mathbf{w}_2^\top \mathbf{V}^\top \mathbf{Y}_i^\top + b_2, z_i) = 0 \right\}. \end{aligned}$$

$$\begin{aligned} \forall \mathbf{w}_1 \in \mathbb{R}^R, b_1 \in \mathbb{R} : & \mathbf{w}_1^\top \mathbf{U}^\top \mathbf{X}_i^\top + b_1 \text{ and } z_i \text{ are uncorrelated} \\ \Leftrightarrow \forall \mathbf{w}_1 \in \mathbb{R}^R, b_1 \in \mathbb{R} : & \sum_{i=1}^N (z_i - \bar{z}) \cdot (\mathbf{w}_1^\top \mathbf{U}^\top \mathbf{X}_i^\top + b_1) = 0 \\ \Leftrightarrow \forall \mathbf{w}_1 : & \mathbf{w}_1^\top \mathbf{U}^\top \mathbf{X}^\top \hat{\mathbf{z}} = 0 \Leftrightarrow \mathbf{U}^\top \mathbf{X}^\top \hat{\mathbf{z}} = \mathbf{0}, \end{aligned}$$

Algorithm Overview:

Require: Modalities $\mathbf{X} \in \mathbb{R}^{N \times D_x}$, $\mathbf{Y} \in \mathbb{R}^{N \times D_y}$; Sensitive attribute $\mathbf{z} \in \mathbb{R}^N$; Label $\mathbf{y} \in \mathbb{R}^N$

Ensure: Projected modalities $\mathbf{X}_{\text{fair}}, \mathbf{Y}_{\text{fair}}$; Classification results

Part 1: Learn Fair Representation

- 1: Standardize $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{z}$
- 2: Perform SVD on $\mathbf{X}^\top \mathbf{z}$: $[\mathbf{U}_X, \Sigma_X, \mathbf{V}_X^\top] = \text{SVD}(\mathbf{X}^\top \mathbf{z})$
- 3: Perform SVD on $\mathbf{Y}^\top \mathbf{z}$: $[\mathbf{U}_Y, \Sigma_Y, \mathbf{V}_Y^\top] = \text{SVD}(\mathbf{Y}^\top \mathbf{z})$
- 4: Set $\mathbf{R}_X = \mathbf{U}_X^\top[:, 1:]$ and $\mathbf{R}_Y = \mathbf{U}_Y^\top[:, 1:]$
- 5: Project data to null spaces: $\mathbf{X}_{\text{new}} = \mathbf{X} \mathbf{R}_X$, $\mathbf{Y}_{\text{new}} = \mathbf{Y} \mathbf{R}_Y$
- 6: Fit CCA on projected data: $\Lambda_X, \Lambda_Y = \text{CCA}(\mathbf{X}_{\text{new}}, \mathbf{Y}_{\text{new}})$
- 7: Calculate $\mathbf{U} = \mathbf{R}_X \Lambda_X$, $\mathbf{V} = \mathbf{R}_Y \Lambda_Y$
- 8: Transform data: $\mathbf{X}_{\text{fair}} = \mathbf{X} \mathbf{U}$, $\mathbf{Y}_{\text{fair}} = \mathbf{Y} \mathbf{V}$

Part 2: Subsequent Classification

- 9: Train a classifier (e.g., SVM) on $\mathbf{X}_{\text{fair}}$ with labels $\mathbf{y}$
- 10: Train a classifier (e.g., SVM) on $\mathbf{Y}_{\text{fair}}$ with labels $\mathbf{y}$
- 11: Compute fairness metrics: EOG, GSG, DPG

4. Experimental Setup

Dataset:

- **Synthetic Data:** 500 samples, 55+60 features, simulated sex bias
- **ADNI Real Data:** 375 subjects (182M, 193F)
 - MRI: 66 cortical thickness features
 - AV1451 PET: 68 tau accumulation features
 - Labels: CN (224), MCI (121), AD (30)

Baselines:

- Standard CCA
- Single-objective Fair CCA (SF-CCA)
- Multi-objective Fair CCA (MF-CCA)
- Standard SVM

Fairness Metrics:

- Demographic Parity Gap (DPG)
- Equalized Odds Gap (EOG)
- Group Sufficiency Gap (GSG)

5. Results

Table 1: Numerical results (mean±std) in terms of three fairness metrics of SVM, CCA, SF-CCA, MF-CCA and our FR-CCA (the smaller the better). The best one for each modality in each line among the five methods is in bold.

Modality	Metric	SVM	CCA	SF-CCA	MF-CCA	FR-CCA
Synthetic X	GSG	0.303 ± 0.009	0.278 ± 0.023	0.301 ± 0.015	0.307 ± 0.026	0.242 ± 0.046
	DPG	0.102 ± 0.047	0.077 ± 0.056	0.089 ± 0.022	0.099 ± 0.050	0.026 ± 0.031
	EOG	0.103 ± 0.127	0.064 ± 0.049	0.077 ± 0.070	0.114 ± 0.140	0.012 ± 0.014
	Precision	0.600 ± 0.016	0.591 ± 0.019	0.606 ± 0.045	0.612 ± 0.028	0.635 ± 0.022
	Recall	0.495 ± 0.020	0.457 ± 0.083	0.508 ± 0.084	0.534 ± 0.083	0.599 ± 0.029
Synthetic Y	ROC-AUC Score	0.493 ± 0.017	0.486 ± 0.017	0.479 ± 0.048	0.536 ± 0.049	0.532 ± 0.032
	GSG	0.274 ± 0.038	0.329 ± 0.042	0.286 ± 0.009	0.296 ± 0.004	0.216 ± 0.084
	DPG	0.075 ± 0.069	0.081 ± 0.040	0.100 ± 0.069	0.104 ± 0.057	0.037 ± 0.035
	EOG	0.107 ± 0.131	0.084 ± 0.037	0.077 ± 0.076	0.095 ± 0.117	0.022 ± 0.010
	Precision	0.624 ± 0.026	0.612 ± 0.033	0.632 ± 0.038	0.645 ± 0.029	0.596 ± 0.054
ADNI MRI	Recall	0.523 ± 0.069	0.612 ± 0.033	0.477 ± 0.073	0.512 ± 0.053	0.569 ± 0.017
	ROC-AUC Score	0.519 ± 0.030	0.811 ± 0.016	0.508 ± 0.084	0.537 ± 0.030	0.525 ± 0.037
	GSG	0.329 ± 0.042	0.302 ± 0.230	0.286 ± 0.033	0.296 ± 0.004	0.258 ± 0.038
	DPG	0.081 ± 0.040	0.026 ± 0.032	0.036 ± 0.024	0.104 ± 0.057	0.010 ± 0.011
	EOG	0.084 ± 0.037	0.020 ± 0.018	0.033 ± 0.031	0.095 ± 0.117	0.014 ± 0.011
ADNI AV1451	Precision	0.612 ± 0.033	0.552 ± 0.627	0.619 ± 0.026	0.645 ± 0.029	0.611 ± 0.014
	Recall	0.612 ± 0.033	0.552 ± 0.063	0.619 ± 0.026	0.512 ± 0.053	0.611 ± 0.014
	ROC-AUC Score	0.811 ± 0.016	0.782 ± 0.017	0.800 ± 0.020	0.537 ± 0.030	0.773 ± 0.027
	GSG	0.329 ± 0.042	0.239 ± 0.208	0.286 ± 0.033	0.242 ± 0.042	0.215 ± 0.041
	DPG	0.081 ± 0.040	0.028 ± 0.024	0.036 ± 0.024	0.015 ± 0.013	0.008 ± 0.006
Summary	EOG	0.084 ± 0.037	0.015 ± 0.017	0.033 ± 0.031	0.026 ± 0.022	0.010 ± 0.005
	Precision	0.612 ± 0.033	0.635 ± 0.026	0.619 ± 0.026	0.621 ± 0.015	0.612 ± 0.009
	Recall	0.612 ± 0.033	0.635 ± 0.026	0.619 ± 0.026	0.621 ± 0.015	0.612 ± 0.009
	ROC-AUC Score	0.811 ± 0.016	0.791 ± 0.020	0.800 ± 0.020	0.769 ± 0.024	0.798 ± 0.020
	Win Times	2	4	1	3	14

Key Findings:

- FR-CCA achieves best fairness across all metrics (lowest DPG, EOG, GSG)
- Maintains competitive accuracy while significantly improving fairness
- Consistent performance across both synthetic and real datasets

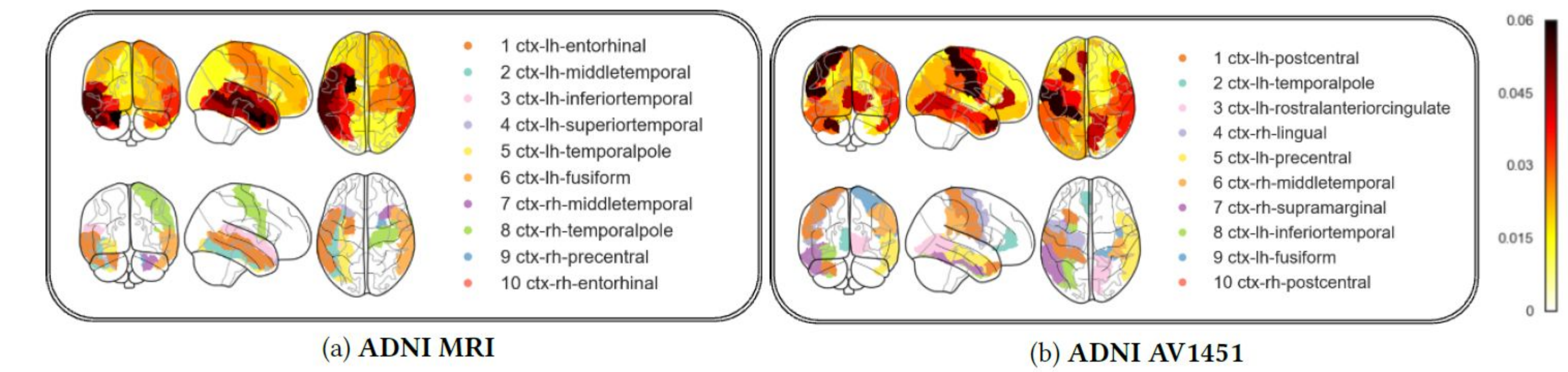


Figure 4: Brain heat map of feature importance for ADNI MRI and ADNI AV1451 using the coefficient of our FR-CCA model. Each modality includes three slices to display the brain region extensively. The heat maps in the first row showcase all the brain regions of interest (ROI), where darker colors represent greater importance for risk prediction. The second row highlights the top ten significant brain regions in each modality, which are annotated in the legends in the right column.

Key Brain Regions (MRI):

- Entorhinal cortex, temporal regions
- Memory and language processing areas

Key Brain Regions (AV1451):

- Postcentral gyrus, temporal pole
- Sensory processing and emotional regulation areas