Identifying Alzheimer's Disease Population Structure From Multimodal Imaging Using Deep Multiview Learning Framework

Yixue Feng[1], Mansu Kim[2], Xiaohui Yao[2], Kefei Liu[2], Qi Long[2], Li Shen[2]

[1]School of Engineering and Applied Science University of Pennsylvania, [2]University of Pennsylvania Perelman School of Medicine

Introduction

- **Multimodal neuroimaging** data can provide complementary information about neurodegenerative diseases such as Alzheimer's Disease (AD).
- Multiview learning such as **Canonical Correlation Analysis (CCA)** can learn correlated components from 2 data views.
- We propose to use **Deep Generalized CCA (DGCCA)** to learn correlated components from 3 neuroimaging modalities, and utilize deep neural networks to learn a non-linear projection to the shared latent space.
- We validate the framework by conducting **cluster analysis** and **genetic association** to identify AD population structure.

Data & Materials

- Our data are from the **Alzheimer's Disease Neuroimaging Initiative (ADNI)**.
- We used 3 neuroimaging modalities – structural MRI (processed using voxel based morphometry or **VBM**), Amyloid PET (**AV45**), and **FDG** PET.
- 19 AD candidate SNPs are used in our genetic evaluation.

Diagnosis	Control	Case	P
Number	274	531	-
Gender(M/F)	125/149	282/249	5.25E-02
Age(mean±sd)	74.84±6.35	72.99±8.05	**9.81E-04**
Education(mean±sd)	16.44±2.72	15.99±2.73	**2.71E-02**

P-values were computed using one-way T-test (except for gender using χ^2 test). The bold text denoted p < 0.05.

Table 1: Characteristics of 805 participants from ADNI.

Multiview Learning

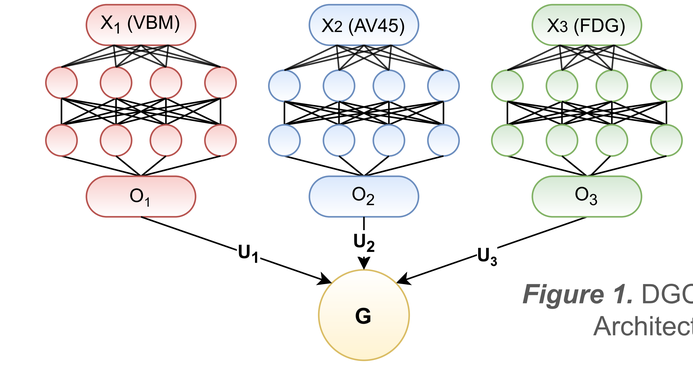

Figure 1. DGCCA Architecture

Experiment Design

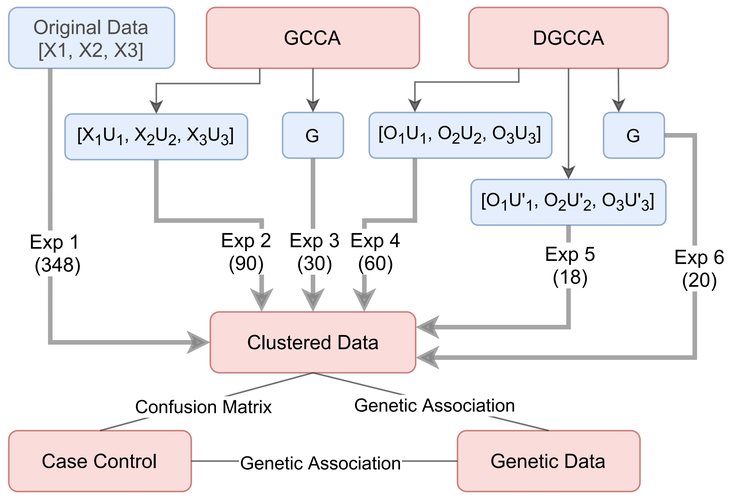

Figure 2. Experiment Design. The number of features is marked under each experiment.

Methods

Multiview Learning

- We extend an existing GCCA implementation to DGCCA using the PyTorch library.

Clustering Analysis

- Six experiments designed above produce different feature sets, each of which are used for clustering
- We use **agglomerative clustering** with Ward's method to generate 2 clusters, similar to the case control partition.
- Evaluate using confusion matrix and **Fisher exact test.**

Genetic Association

- We test for genetic association using **Pearson's Chi-squared test** for each 19 SNPs allele count vs. cluster from 6 experiments and the original case control status

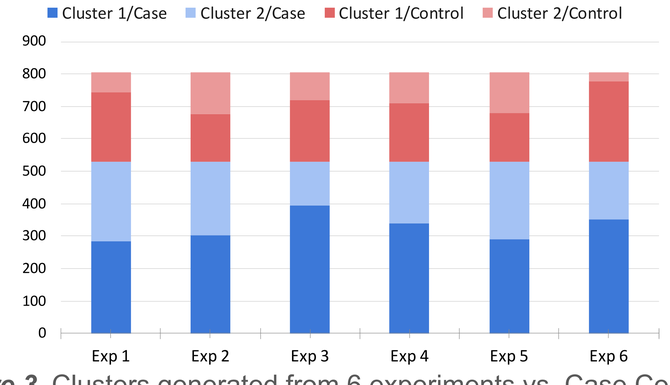

Figure 3. Clusters generated from 6 experiments vs. Case Control

Results

Multiview Learning

- DGCCA features explain more variance in fewer components compared to GCCA.

Clustering Analysis

- Exp 1 (concatenated features, $p = 3.60\times10^{-11}, OR = 0.341$) and Exp 6 (DGCCA, $p = 1.23\times10^{-14}, OR = 4.653$) produced statistically significant result in the Fisher exact test, showing high level of alignment between clusters and case control groups (see ***Figure 3***).

Genetic Association

- The original case control status, features from Exp 1 and Exp 6 produced significant results after FDR correction at $\alpha < 0.05$ while Exp 2-5 did not.
- They all identified APOE, a well known genetic risk factor of AD, and Exp 6 using DGCCA features, also identified rs4247929 from the ABCA7 gene, related to early and late-onset AD.

SNP	Gene	Case Control	Exp 1	Exp 6
rs4247929	ABCA7	7.88E-01	2.18E-01	**2.63E-02**
rs429358	APOE	**1.55E-08**	**3.10E-20**	**9.27E-20**

Conclusion & Discussion

In this study, we apply multiview learning method DGCCA to multimodal neuroimaging data to identify population structure. We found that DGCCA captures more variance in fewer correlated components than its linear counterpart, GCCA, by utilizing non-linear transformation from neural networks. Subsequent clustering analysis shows that it's able to identify case control relationship in an AD cohort. Further genetic association analysis on clustering results demonstrates that DGCCA can yield a population structure with stronger genetic basis than GCCA and single modality features.

Acknowledgements

This work was support in part by National Institute of Health R01 EB022574, R01 LM013463, RF1 AG063481; and National Science Foundation IIS 1837964. Data used in preparation of this article were obtained from the Alzheimer's disease neuroimaging initiative (ADNI) database adni.loni.usc.edu.

Author contact *: Yixue Feng (wendyfyx@seas.upenn.edu)*