

Deep Multiview Learning to Identify Population Structure with Multimodal Imaging

Yixue Feng

*School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, USA
wendyfyx@seas.upenn.edu*

Mansu Kim

*Perelman School of Medicine
University of Pennsylvania
Philadelphia, USA
mansu.kim@pennmedicine.upenn.edu*

Xiaohui Yao

*Perelman School of Medicine
University of Pennsylvania
Philadelphia, USA
xiaohui.yao@pennmedicine.upenn.edu*

Kefei Liu

*Perelman School of Medicine
University of Pennsylvania
Philadelphia, USA
kefei.liu@pennmedicine.upenn.edu*

Qi Long

*Perelman School of Medicine
University of Pennsylvania
Philadelphia, USA
qlong@pennmedicine.upenn.edu*

Li Shen

*Perelman School of Medicine
University of Pennsylvania
Philadelphia, USA
li.shen@pennmedicine.upenn.edu*

Abstract—We present an effective deep multiview learning framework to identify population structure using multimodal imaging data. Our approach is based on canonical correlation analysis (CCA). We propose to use deep generalized CCA (DGCCA) to learn a shared latent representation of non-linearly mapped and maximally correlated components from multiple imaging modalities with reduced dimensionality. In our empirical study, this representation is shown to effectively capture more variance in original data than conventional generalized CCA (GCCA) which applies only linear transformation to the multi-view data. Furthermore, subsequent cluster analysis on the new feature set learned from DGCCA is able to identify a promising population structure in an Alzheimer’s disease (AD) cohort. Genetic association analyses of the clustering results demonstrate that the shared representation learned from DGCCA yields a population structure with a stronger genetic basis than several competing feature learning methods.

Keywords—Deep learning, multiview learning, deep generalized canonical correlation analysis, multimodal imaging, image-driven population structure

I. INTRODUCTION

Cluster analysis is a popular machine learning approach used in identifying population structure, and is often applied on brain imaging and genetic data. Clusters can help identify groups of individuals with similar imaging or genetic characteristics [1], and sometimes coupled with feature learning (feature reduction) methods given a large number of imaging features [2]. Multimodal imaging, compared to single imaging modality, are more likely to capture partial but complementary information of population structures from different perspectives [3]. However, many studies typically employed traditional clustering methods on the original features directly. These methods have limited capabilities in automatically learning effective features for the clustering task, in comparison with modern deep learning methods.

To bridge this gap, we propose an effective deep multiview learning framework, and demonstrate its power via

applying it to the multimodal imaging data in an Alzheimer’s disease (AD) cohort for identifying imaging-driven population structure. Our framework is based on an extended version of canonical correlation analysis (CCA), named as *deep generalized CCA (DGCCA)* [4]. CCA is a popular technique to identify linear relationships between two multivariate datasets [5]. Traditional CCA models have two limitations: 1) it cannot be applied to data with more than two modalities, and 2) it cannot capture nonlinear relationships between data modalities.

To overcome the first limitation, CCA can be extended to generalized CCA (GCCA) [6], designed to learn a representation that is able to explain many views of the data, and is a promising strategy to capture meaningful variation shared by multiple imaging modalities. To overcome the second limitation, GCCA can be extended to DGCCA [4], which non-linearly maps the feature space of each imaging modality to a common latent space. To demonstrate the power of DGCCA for effective feature representation learning, we perform an empirical study using the imaging and genetics data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) [7]. DGCCA and several competing feature learning methods, coupled with cluster analysis, are applied to the ADNI multimodal imaging data to identify population structure. Genetic association analyses are subsequently performed on the learned feature representations to evaluate genetic basis for the learned imaging-driven structures. The shared representation learned from DGCCA yields a population structure with a stronger genetic basis than studied competing methods.

II. MATERIALS

To demonstrate the power of DGCCA in learning effective feature representation from multimodal imaging data for detecting population structure, we apply it to the imaging and genetic data in an AD study. This study was approved

Table I
PARTICIPANT CHARACTERISTICS IN OUR EXPERIMENTS. THERE ARE TOTALLY 805 PARTICIPANTS, WHERE HC AND SMC PARTICIPANTS ARE GROUPED AS CONTROLS (N=274), AND EMCI, LMCI AND AD PARTICIPANTS ARE GROUPED AS CASES (N=531).

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 $p < 0.05$.

by institutional review boards of all participating institutions and written informed consent was obtained from all participants or authorized representatives.

Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu) [7]. 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 magnetic resonance imaging (MRI), positron emission tomograph (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. For up-to-date information, see www.adni-info.org.

A. Study Participants

In this work, we analyzed 805 non-Hispanic Caucasian subjects with complete baseline measurements of three studied imaging modalities, genotyping data and visit-matched diagnostic information. Specifically, there are 274 controls (i.e., 196 healthy controls (HC) and 78 normal controls with significant memory concern (SMC)) and 531 cases (i.e., 235 patients with early mild cognitive impairment (EMCI), 162 patients with late mild cognitive impairment (LMCI), and 134 AD patients). Shown in **Table I** are their characteristics.

B. Imaging Data

We focus on analyzing three imaging modalities in ADNI: structural MRI [8] (sMRI, measuring brain morphometry, VBM), amyloid-PET [9] (measuring amyloid burden, AV45), and FDG-PET [10] (measuring glucose metabolism). The multi-modality imaging data were aligned to each participant’s same visit. The sMRI scans were processed with voxel-based morphometry (VBM) using the Statistical Parametric Mapping (SPM) software tool [11]. Generally, all scans were aligned to a T1-weighted template image, segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) maps, normalized to the standard Montreal Neurological Institute (MNI) space as $2 \times 2 \times 2$ mm³ voxels, and were smoothed with an 8mm FWHM

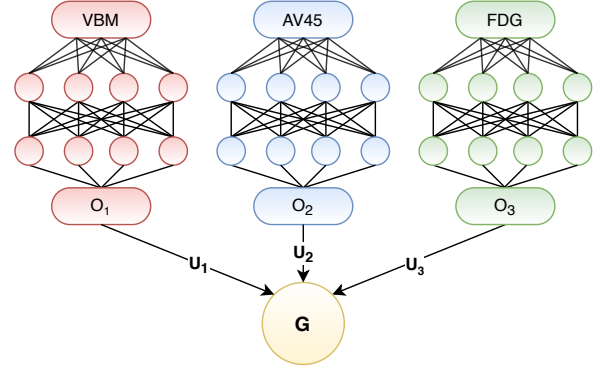


Figure 1. DGCCA architecture. The DGCCA applies a deep neural network to learn non-linear projection from each view to a new representation, and these new representations are then analyzed together via a GCCA model.

kernel. The FDG-PET and AV45-PET scans were also registered into the same MNI space by SPM. The MarsBaR ROI toolbox [12] was used to group voxels into 116 regions-of-interest (ROIs). ROI-level measures were calculated by averaging all the voxel-level measures within each ROI. As mentioned above, participants in this work included 805 non-Hispanic Caucasian subjects with complete baseline ROI-level measurements of three modalities and visit-matched diagnostic information; see **Table I** for their characteristics.

C. Genetic Data

The ADNI genotyping data, acquired on multiple Illumina platforms, have been quality controlled, imputed and combined using the same procedure as described in Yao et al. [13]. To avoid population stratification effect, our analysis was performed on only non-Hispanic Caucasian participants. There were a total of 805 non-Hispanic Caucasian participants (**Table I**) with all imaging, genetic and diagnostic data available. A list of 19 AD candidate SNPs (**Table III**) discovered by a large-scale meta-analytic genome-wide association study (GWAS) [14] was included in the genetic association analysis.

III. MULTIVIEW LEARNING MODELS

Multiview learning refers to the method that learns a single model from multimodal data. In this study, we adopt this approach by learning a latent representation from three brain imaging modalities, VBM, AV45 and FDG. Given limited data and rich feature space, multiview learning can reduce the dimensionality of the data, and learn a shared latent representation [15]. The learned latent representation is expected to capture valuable information fused from all the input views, and has great potential to catch the intrinsic population structure in the studied sample.

A. Generalized CCA (GCCA)

The GCCA is an extended version of CCA to handle more than two views of data [6]. Given M views of data $X_i \in$

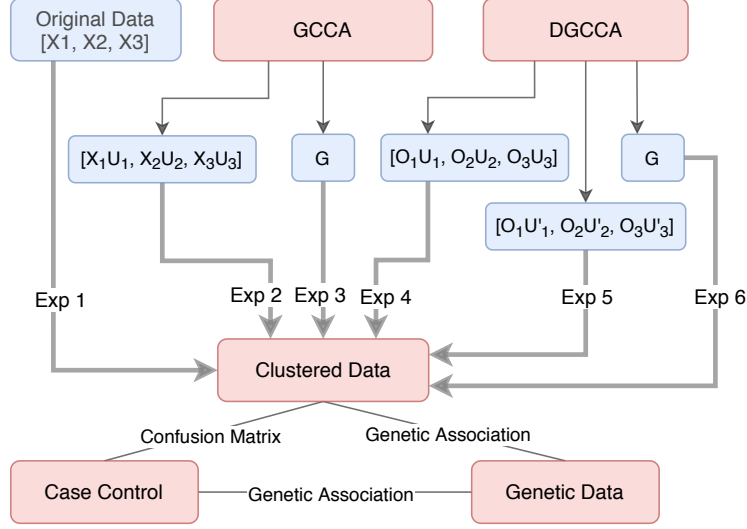


Figure 2. Flowchart of our experiments, which include two major components. The first component is to learn six different feature representations. The second component is to perform clustering on each feature representation to identify a population structure, and then compare the genetic bases of six resulting population structures together with the original case control structure. Experiment 1 (Exp 1), we focused on original data space (i.e., concatenation of VBM, AV45 and FDG) to explain the population structure. Then, in Experiments 2-6 (Exp 2 - Exp 6), we used different latent spaces (i.e., X_iU_i and G for GCCA, and O_iU_i and G for DGCCA) learned from GCCA or DGCCA to better explain the population structure. X_iU_i (and O_iU_i), obtained by applying the learned projection matrices to the original data views, are compared with G , the shared latent feature representation from multiview learning methods.

$R^{N \times p_i}$, where X_i is the i -th view of the data, N is the number of data points, and p_i is the number of features in view i . The goal of GCCA is to learn a shared representation or embedding from all views by optimizing the objective function in Eq. 1.

$$\underset{\{U_i \in R^{p_i \times k}\}_{i=1}^M, G \in R^{N \times k}}{\text{minimize}} \sum_{i=1}^M \|G - X_i U_i\|_F^2 \quad \text{s.t. } G^T G = I_k \quad (1)$$

where G denotes the embedding space, and contains the top k eigenvectors of $\sum_{i=1}^I X_i (X_i^T X_i)^{-1} X_i^T$ as its columns, which we use as the share latent features. U_i denotes projection matrix for the i -th view.

B. Deep GCCA (DGCCA)

The DGCCA applies a deep neural network to learn non-linear projection from each view to a new representation, and these new representations are then analyzed together via a GCCA model. **Fig. 1** shows a schematic design of the DGCCA model [4]. Specifically, DGCCA is defined as follows:

$$\underset{\{U_i \in R^{p_i \times k}\}_{i=1}^M, G \in R^{N \times k}}{\text{minimize}} \sum_{i=1}^M \|G - O_i U_i\|_F^2 \quad \text{s.t. } G^T G = I_k \quad (2)$$

where O_i denotes the output of the final layer in the network for the i -th imaging modality.

IV. EXPERIMENTAL SETUP

A. Experimental Design

We design six experiments to compare six different feature representations for identifying an image-driven population structure. These representations include the direct concatenation of all multimodal imaging features, two latent feature spaces extracted from GCCA, and three latent feature spaces extracted from DGCCA. **Fig. 2** shows the overall flowchart of these experiments, which include two major components. The first component is to learn six different feature representations. The second component is to perform clustering on each feature representation to identify a population structure, and then compare the genetic bases of six resulting population structures together with the original case control structure.

Table II summarizes the feature representations learned from six experiments, which were used for cluster analysis. Details for these feature representations extracted from six experiments are outlined below:

- In Experiment 1 (Exp 1), we concatenate features from all three imaging modalities in the order of VBM, AV45 and FDG. Each modality contains 116 ROI-based features. Thus the total number of features is 348.
- In Experiment 2 (Exp 2), we extract the learned projection matrices for each imaging modality from GCCA U_i , and apply them on the original feature set X_i to get $X_i U_i$ to obtain new feature set for each modality. After that, we concatenate them together across three imaging modalities. For each modality, we keep the first

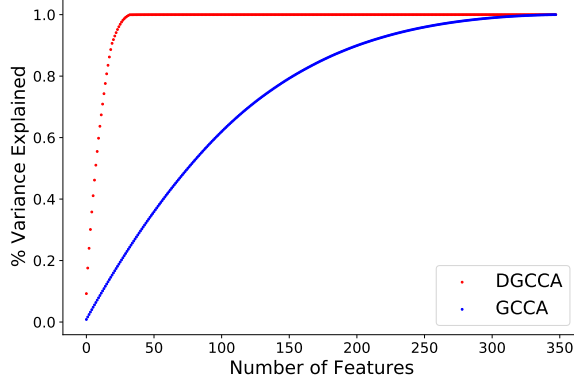


Figure 3. Variance explained against number of features. It is evident that non-linear transformation through neural network implemented in DGCCA can capture more variance in the original data with fewer components than the linear projection implemented in GCCA.

30 components (see **Fig. 3** for the variance captured by these components in our experiments). Thus, the resulting representation contains 90 features.

- In Experiment 3 (Exp 3), we use the shared feature representation G learned directly from GCCA. Similarly to Exp 2, we keep first 30 components to form the new representation.
- In Experiment 4 (Exp 4), we extract the learned projection matrices for each imaging modality from DGCCA U_i , and apply them on the respective output representation from neural network O_i to get $O_i U_i$. After that, we concatenate them together across three imaging modalities. For each modality, we keep the first 20 components (see **Fig. 3** for the variance captured by these components). Thus the resulting representation contains 60 features.
- Experiment 5 (Exp 5) is similar to Exp 4, with the exception that we select top latent features based on correlation matrices in **Fig. 4**. This experiment was designed based on the observation that the first few DGCCA components capture not only most of the data variance, (**Fig. 3**) but also most of the correlations between modalities (**Fig. 4**). Specifically, in our experiments, we chose the first two, eight, eight components for VBM, FDG and AV45 respectively. Thus, this representation consists of 18 features in total.
- In Experiment 6 (Exp 6), we use the shared feature representation G learned directly from DGCCA. Similarly to Exp 4, we keep first 20 components to form the new representation.

Note the difference between Exp 2 vs. Exp 3, and Exp 4 vs. Exp 6, of using concatenation of $X_i U_i$ and G . The objective of GCCA and DGCCA is to directly learn a shared latent space G along with projection matrices U_i . Using $X_i U_i$ allows us to understand the how each modality contributes to the final shared latent space. Concatenation of

Table II
COMPARISON OF FEATURES USED IN CLUSTER ANALYSIS

Experiments	Features used for Clustering	# Features
Exp 1	$[X_1, X_2, X_3]$	348
Exp 2	$[X_1 U_{gcca1}, X_2 U_{gcca2}, X_3 U_{gcca3}]$	90
Exp 3	G_{gcca}	30
Exp 4	$[O_1 U_{dgcca1}, O_2 U_{dgcca2}, O_3 U_{dgcca3}]$	60
Exp 5	$[O_1 U'_{dgcca1}, O_2 U'_{dgcca2}, O_3 U'_{dgcca3}]$	18
Exp 6	G_{dgcca}	20

$X_i U_i$ serves as a comparison with Exp 1 to see if applying GCCA and DGCCA projection matrices on each single modality imaging data can convey as much information as using the full original features space.

B. Clustering and Genetic Association

Clustering along with genetic association can help identify population structures with novel genetic basis. For each experiment, we applied hierarchical agglomerative clustering [16] to all the subjects using the corresponding feature representation, where the Ward’s method was used to minimize the variance when merging clusters. We chose a cluster number of 2, similar to case/control group. We evaluated the clusters by plotting confusion matrices to check for distribution of cases and controls for each cluster. Genetic association is performed by conducting Pearson’s Chi-squared test on genetic data versus clustered data, and the assigned case and control. The goal is to identify genetic markers that are associated with case-control status or the cluster membership identified in each of Exps. 1-6. We used the Benjamini–Hochberg procedure to control the False Discovery Rate (FDR) at $\alpha = 0.05$ [17].

C. Implementation of GCCA and DGCCA

For both GCCA and DGCCA, we applied stratified split on the imaging data 80/20 into training and test set. For GCCA, we used implementation of weighted GCCA in Benton et al. [18], which included added regularization for each view for stabilization. Our DGCCA extends this GCCA implementation to apply non-linear transformations using neural networks. The network for each view is composed of 2 hidden linear layers of 64 nodes with ReLU activation, and trained with Adam optimizer with a learning rate of 0.0008 and weight decay of 0.01. The network outputs have the same dimensions as the inputs with 116 features. Additional regularization was applied using dropout and early stopping. We tuned the model using training loss and evaluated resulting G and U_i by plotting correlation matrices between pairs of modalities.

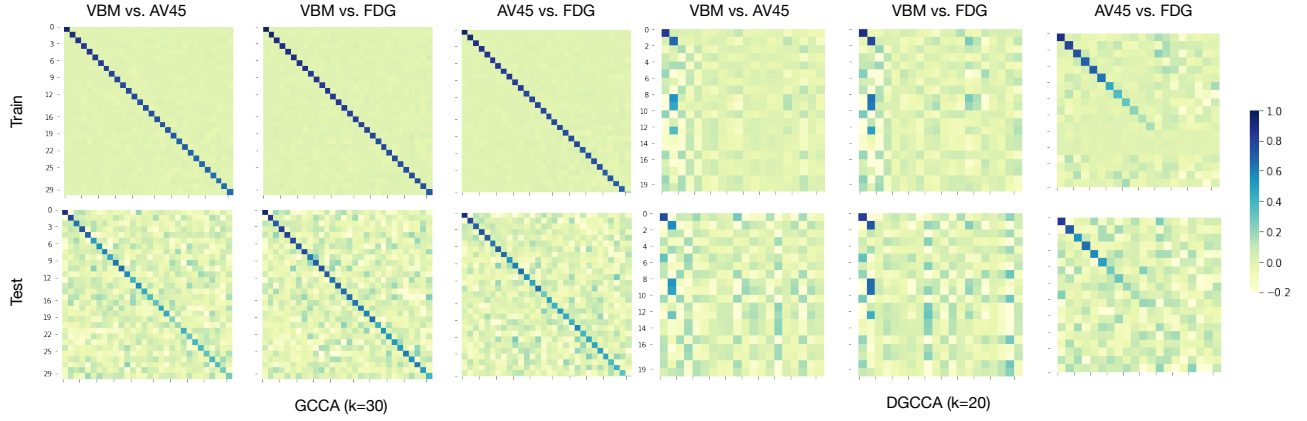


Figure 4. Correlation matrices between pairs of imaging modalities are plotted for extracted GCCA and DGCCA components in the new latent feature spaces. The first row shows the training performances and the second row shows the testing performances. The first three columns show the GCCA results for three pairwise comparisons. The last three columns show the DGCCA results for three pairwise comparisons. While the first 30 canonical components identified by GCCA show the strong correlation between imaging modalities, DGCCA identifies much fewer components (e.g., 2 for VBM, around 8 for AV45 or FDG in the testing results) that are correlated between imaging modalities.

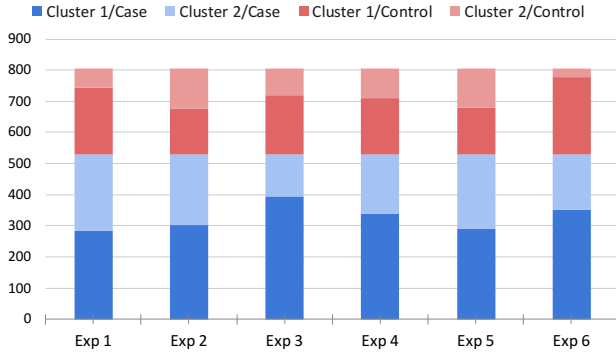


Figure 5. Clusters discovered in each experiment vs. case control status.

V. RESULTS

A. Comparison of GCCA and DGCCA

Table II records the number of features in data extracted for each of the six experiments. Note that the number of latent features k here is different for DGCCA $k = 20$ and GCCA $k = 30$, since we speculate that DGCCA would learn fewer maximally correlated components given the non-linear transformations applied to the data. But we discovered that DGCCA learns maximally correlated components from input views in fewer components compared to GCCA and even 20 components became enough. Therefore, Exp 5 was added where a subset of latent features U'_{dgcca} learned from DGCCA was selected based on diagonal values in correlation matrices (see last three columns in **Fig. 4**).

We compared the performances of GCCA and DGCCA by plotting correlation matrices in the new latent feature space between pairs of imaging modalities (see **Fig. 4**). Since

GCCA and DGCCA maximize correlation between more than two modalities by extracting the top k latent space features (eigenvectors), we evaluate GCCA and DGCCA results by looking at the diagonal values on the correlation matrices. We can see that DGCCA learns much fewer maximally correlated components compared to GCCA, and the chosen $k = 20$, so for Exp 5, we chose the first 2 components for VBM and the first 8 components for FDG and AV45. In addition, DGCCA results show a discrepancy of the maximally correlated components for AV45 and FDG pair compared to the other two pairs, which require future work to be done investigating these differences using CCA, Deep CCA methods [19].

Further comparison between GCCA and DGCCA is done by plotting variance explained by each feature in original feature space ($p = 348$), seen in **Fig. 3**. We can see that non-linear transformation through neural network can capture more variance in the original data with fewer components.

B. Clustering and Genetic Association Analysis

For each of the six experiments, the transformed data is clustered using agglomerative clustering, then confusion matrices are plotted again assigned case and control (HC/SMC vs. EMCI/LMCI/AD), see **Fig. 5**. We can see that in Exp 3 and Exp 6, using learned shared representation G from GCCA and DGCCA representation have low true negative rate but the highest true positive rate. Out of the six experiment, Exp 6 clustered data yield a significant result ($p = 1.23e-14$, OR = 4.653, 95%CI : 3.008, 7.195) in addition to Exp 1 which uses the full feature set, indicating a relatively high level alignment between the data-driven clusters with case control groups.

From the genetic association analyses, p -value after FDR correction from the Pearson's Chi-squared test are recorded

Table III
FDR-CORRECTED p -VALUE FROM CHI-SQUARED TEST IN GENETIC ASSOCIATION ANALYSIS. NO RESULTS FROM EXP 2 - EXP 5 ARE SIGNIFICANT AND THUS NOT SHOWN HERE.

SNP	Gene	Case Control	Exp 1	Exp 6
rs6656401	CR1	6.89e-01	1.11e-01	2.04e-01
rs6733839	BIN1	8.16e-01	6.62e-01	7.68e-01
rs35349669	INPP5D	9.53e-01	9.44e-01	8.68e-01
rs190982	MEF2C	9.53e-01	6.79e-01	9.89e-01
rs10948363	CD2AP	9.53e-01	8.99e-01	8.64e-01
rs2718058	NME8	9.53e-01	6.79e-01	8.68e-01
rs1476679	ZCWPW1	9.53e-01	6.71e-01	8.68e-01
rs11771145	EPHA1	9.53e-01	9.70e-01	4.97e-01
rs28834970	PTK2B	9.53e-01	6.62e-01	8.68e-01
rs9331896	CLU	9.53e-01	5.97e-01	3.29e-01
rs10838725	CELF1	8.16e-01	8.69e-01	9.89e-01
rs983392	MS4A6A	6.89e-01	5.97e-01	8.68e-01
rs10792832	PICALM	9.53e-01	5.97e-01	1.56e-01
rs17125944	FERMT2	8.16e-01	5.97e-01	2.78e-01
rs10498633	SLC24A4	9.52e-01	8.69e-01	9.89e-01
rs4147929	ABCA7	7.88e-01	2.18e-01	2.63e-02
rs429358	APOE	1.55e-08	3.17e-20	9.27e-20
rs3865444	CD33	9.53e-01	5.97e-01	8.68e-01
rs7274581	CASS4	9.53e-01	5.97e-01	9.89e-01

in **Table III**. Our analyses from original case control, Exp 1 and Exp 6 yielded statistically significant results. All three tests produced significant results for APOE, the best known genetic risk factor for AD [20]. In addition, clustered results from Exp 6, using learned shared representation from DGCCA, yielded significant results for SNP rs4147929 from ABCA7 gene ($\chi^2 = 11.777$, FDR-corrected $p < 0.05$). There has been compelling evidence suggesting that ABCA7 is a risk gene for both early and late-onset AD [21]. In Exp 6, our DGCCA method learned the promising feature representation leading to the discovery of a new population structure with a novel genetic basis (i.e., an ABCA7 SNP), which was not detected by the standard case-control status.

VI. CONCLUSION

We have proposed a multi-view representation learning framework using deep generalized CCA (DGCCA), and applied it to multi-modal brain imaging data (VBM, FDG, AV45) for identifying population structure. DGCCA is able to capture original data in much fewer maximally correlated components compared to generalized CCA (GCCA) by applying non-linear transformation to each view. Furthermore, we have shown that the learned shared representation, coupled with cluster analysis, can be utilized to identify promising population structure with a stronger genetic basis.

In the future, we plan to explore the use of our method to identify not only population structure, but also disease subtypes with novel imaging and genetic characteristics.

ACKNOWLEDGMENT

This work was supported 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). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in data analysis or writing of this report. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). A complete listing of ADNI investigators and the complete ADNI Acknowledgement can be found at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

REFERENCES

- [1] R. Filipovych, S. M. Resnick, and C. Davatzikos, "Semi-supervised cluster analysis of imaging data," *Neuroimage*, vol. 54, no. 3, pp. 2185–2197, Feb 2011.
- [2] L. Elliott, K. Sharp, F. Alfaro-Almagro, S. Shi, K. Miller, G. Douaud, J. Marchini, and S. Smith, "Genome-wide association studies of brain imaging phenotypes in uk biobank," *Nature*, vol. 562, 10 2018.
- [3] K. Miller, F. Alfaro-Almagro, N. Bangerter, D. Thomas, E. Yacoub, J. Xu, A. Bartsch, S. Jbabdi, S. Sotiropoulos, J. Andersson, L. Griffanti, G. Douaud, T. Okell, P. Weale, I. Dragonu, S. Garratt, S. Hudson, R. Collins, M. Jenkinson, and S. Smith, "Multimodal population brain imaging in the uk biobank prospective epidemiological study," *Nature neuroscience*, vol. 19, 09 2016.
- [4] A. Benton, H. Khayrallah, B. Gujral, D. A. Reisinger, S. Zhang, and R. Arora, "Deep generalized canonical correlation analysis," in *Proc. of 4th Workshop on Representation Learning for NLP*. Asso. for Computational Linguistics, 2019, pp. 1–6.
- [5] D. M. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, vol. 10, no. 3, pp. 515–34, 2009.
- [6] P. Horst, "Generalized canonical correlations and their applications to experimental data," *Journal of Clinical Psychology*, vol. 17, no. 4, pp. 331–347, 1961.
- [7] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack, J. C. Morris *et al.*, "Recent publications from the Alzheimer's disease neuroimaging initiative: Reviewing progress toward improved ad clinical trials," *Alzheimer's & Dementia*, vol. 13, no. 4, pp. e1–e85, 2017.
- [8] J. Jack, C. R., M. A. Bernstein, B. J. Borowski, J. L. Gunter, N. C. Fox, P. M. Thompson, N. Schuff, G. Krueger, R. J. Killiany, C. S. Decarli, A. M. Dale, O. W. Carmichael, D. Tosun, M. W. Weiner, and Alzheimer's Disease Neuroimaging Initiative, "Update on the magnetic resonance imaging core of the Alzheimer's Disease Neuroimaging Initiative," *Alzheimers Dement*, vol. 6, no. 3, pp. 212–20, 2010.
- [9] W. J. Jagust, S. M. Landau, R. A. Koeppe, E. M. Reiman, K. Chen, C. A. Mathis, J. C. Price, N. L. Foster, and A. Y. Wang, "The Alzheimer's Disease Neuroimaging Initiative 2 PET core: 2015," *Alzheimers Dement*, vol. 11, no. 7, pp. 757–71, 2015.
- [10] W. J. Jagust, D. Bandy, K. Chen, N. L. Foster, S. M. Landau, C. A. Mathis, J. C. Price, E. M. Reiman, D. Skovronsky, R. A. Koeppe, and Alzheimer's Disease Neuroimaging Initiative, "The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core," *Alzheimers Dement*, vol. 6, no. 3, pp. 221–9, 2010.
- [11] J. Ashburner and K. J. Friston, "Voxel-based morphometry—the methods," *Neuroimage*, vol. 11, no. 6, pp. 805–21, 2000.
- [12] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–89, 2002.
- [13] X. Yao, S. L. Risacher, K. Nho, A. J. Saykin, Z. Wang, L. Shen, and I. Alzheimer's Disease Neuroimaging, "Targeted genetic analysis of cerebral blood flow imaging phenotypes implicates the *INPP5D* gene," *Neurobiol Aging*, vol. 81, pp. 213–221, 2019.
- [14] J. C. Lambert, C. A. Ibrahim-Verbaas, D. Harold, A. C. Naj, R. Sims, C. Bellenguez, A. L. DeStafano, J. C. Bis, G. W. Beecham, B. Grenier-Boley, G. Russo, T. A. Thorton-Wells, N. Jones, A. V. Smith, V. Chouraki, C. Thomas *et al.*, "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease," *Nat. Genet.*, vol. 45, no. 12, pp. 1452–1458, 2013.
- [15] A. Serra, P. Galdi, and R. Tagliaferri, "Chapter 13 - multiview learning in biomedical applications," in *Artificial Intelligence in the Age of Neural Networks and Brain Computing*. Academic Press, 2019, pp. 265 – 280.
- [16] L. Rokach and O. Maimon, *Clustering Methods*. Boston, MA: Springer US, 2005, pp. 321–352. [Online]. Available: https://doi.org/10.1007/0-387-25465-X_15
- [17] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995. [Online]. Available: <http://www.jstor.org/stable/2346101>
- [18] A. Benton, R. Arora, and M. Dredze, "Learning multi-view embeddings of twitter users," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 14–19.
- [19] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proceedings of the 30th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 28, no. 3. PMLR, 2013, pp. 1247–1255.
- [20] L. G. Apostolova, S. L. Risacher, T. Duran, E. C. Stage, N. Goukasian, J. D. West, T. M. Do, J. Grotts, H. Wilhalme, K. Nho, M. Phillips, D. Elashoff, and A. J. Saykin, "Associations of the Top 20 Alzheimer Disease Risk Variants With Brain Amyloidosis," *JAMA Neurol*, vol. 75, no. 3, pp. 328–341, 2018.
- [21] A. De Roeck, C. Van Broeckhoven, and K. Sleegers, "The role of ABCA7 in Alzheimer's disease: evidence from genomics, transcriptomics and methylomics," *Acta Neuropathol.*, vol. 138, no. 2, pp. 201–220, 2019.