Integrating Static and Dynamic Data for Improved Prediction of Cognitive Declines Using Augmented Genotype-Phenotype Representations

Hoon Seo[1], **Lodewijk Brand**[1], **Hua Wang**[1*], **Feiping Nie**[2]

[1]Department of Computer Science, Colorado School of Mines, Golden, CO 80401, U.S.A.
[2]School of Computer Science and Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, Shaanxi, P. R. China
seohoon@mymail.mines.edu, lbrand@mymail.mines.edu, huawangcs@gmail.com, feipingnie@gmail.com

Abstract

Alzheimer's Disease (AD) is a chronic neurodegenerative disease that causes severe problems in patients' thinking, memory, and behavior. An early diagnosis is crucial to prevent AD progression; to this end, many algorithmic approaches have recently been proposed to predict cognitive decline. However, these predictive models often fail to integrate heterogeneous genetic and neuroimaging biomarkers and struggle to handle missing data. In this work we propose a novel objective function and an associated optimization algorithm to identify cognitive decline related to AD. Our approach is designed to incorporate dynamic neuroimaging data by way of a participant-specific augmentation combined with multimodal data integration aligned via a regression task. Our approach, in order to incorporate additional side-information, utilizes structured regularization techniques popularized in recent AD literature. Armed with the fixed-length vector representation learned from the multimodal dynamic and static modalities, conventional machine learning methods can be used to predict the clinical outcomes associated with AD. Our experimental results show that the proposed augmentation model improves the prediction performance on cognitive assessment scores for a collection of popular machine learning algorithms. The results of our approach are interpreted to validate existing genetic and neuroimaging biomarkers that have been shown to be predictive of cognitive decline.

Introduction

Alzheimer's Disease (AD) is a serious neurodegenerative condition that destroys brain cells and causes progressive decline in the behavioral and social skills necessary for independent function. A 2017 report (World Health Organization 2017) from the World Health Organization (WHO) estimates that 47 million people suffered from some form of dementia, of which approximately 60-70% is caused by AD. The WHO projects that by 2030, the number of individuals suffering from AD-related dementia will increase to 75 million people. In order to combat this public health crisis it is important that the greater research community develop tools that can (1) predict the cognitive decline in it's early stage to prevent the progression of AD and (2) identify risk factors of AD to assist in the development of treatments that can be used to provide relief for patients suffering from the disease.

Public-private partnerships such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Petersen et al. 2010) have provided a comprehensive dataset consisting of genetic (*e.g.*, *static*) biomarkers, such as Single Nucleotide Polymorphisms (SNPs), and neuroimaging (*e.g.*, *dynamic*) biomarkers derived from brain imaging modalities. Recent algorithmic improvements (Wang et al. 2012d; Brand et al. 2020a) have shown promise in identifying AD from phenotypic changes. Although effective, they have modeled the longitudinal biomarkers as tensors, which inevitably complicates the problem. Other recent works (Lu et al. 2018; Brand et al. 2019), have leveraged longitudinal modalities to identify temporal relationships across the modalities. However, these approaches can only utilize longitudinal data that is consistent across all participants. This reliance on complete data is a major limitation and well-known critique (Caruana et al. 2015) of longitudinal studies. Thus, incorporating multimodal longitudinal data, especially when the participants have incomplete records, is a critical challenge.

In this work, we propose a novel framework to predict the cognitive outcomes of participants within the ADNI cohort. The design of our objective function is motivated by the following four key points. First, we aim to handle the participants within the ADNI that contain a varying number brain scans; we handle this through a per-participant application of principal component analysis (PCA) (Jolliffe 2011) which is used to augment the most recent available data. Second, we employ a multimodal factorization approach to combine the dynamic and static modalities. Third, to guide the multimodal factorization, with any available labeled data, we align the factorization task with a multitask regression to predict the cognitive assessment scores of ADNI participants. Finally, we incorporate structured regularizations to capture the inter-modality relationships across the dynamic modalities and intra-modality relationships within the static modalities. In summary, the primary contributions of this work are:

- A dynamic-static augmentation approach that integrates the most recent neuroimaging data to generate a *fixed-length* representation, which can be readily fed into conventional machine learning models. This allows our

*To whom all correspondence should be addressed.

method to utilize all available dynamic data with an inconsistent number of records.

- A novel objective function based on joint factorization and regression to link the augmented neuroimaging data with genetic data and available cognitive scores.
- We derive an efficient solution based on the smoothed iteratively reweighted method to optimize the proposed objective.
- In our experiments, we apply our augmentation to identify genetic and neuroimaging biomarkers that are predictive of cognitive decline.

Notations and Problem Formalization

Throughout this paper, we write matrices as bold uppercase letters and vectors as bold lowercase letters. The i-th row vector, j-th column vector, and the element at the i-th row and the j-th column of matrix $\mathbf{M}$ are denoted as $\mathbf{m}^i$, $\mathbf{m}_j$, and m_j^i respectively. For a matrix $\mathbf{M} = \left[m_j^i\right]$, its trace is defined as $\text{tr}(\mathbf{M}) = \sum_i m_i^i$ and its Frobenius norm is defined as $\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m}|m_j^i|^2}$. The trace norm of $\mathbf{M}$ is defined as $\|\mathbf{M}\|_* = \sum_{i=1}^{min\{n,m\}}\sigma_i$, where σ_i is the i-th singular value of $\mathbf{M}$.

For a given longitudinal imaging genetic dataset, phenotypic measurements are usually described by biomarkers extracted from brain scans. Mathematically, the medical records of the i-th participant in a studied cohort can be denoted as $\mathcal{X}_i = \{\mathbf{X}_i, \mathbf{x}_i\}$, where $i = 1, 2, \cdots, n$ indicates the index of the participant. Here, $\mathbf{X}_i = [\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in_i}] \in \Re^{d\times n_i}$ collects the available medical records of the i-th participant from the baseline (the first time point) to the second last visit, such that the total number of the medical records of the i-th participant is n_i+1. We note that n_i varies across the dataset due to inconsistent/missing temporal records of the participants. We use $\mathbf{x}_i \in \Re^d$ to denote the medical record of the i-th participant at the last (most recent) time point and use $\mathbf{X}_{MR} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$ to summarize these records of all the participants in the studied cohort. Multiple types of biomarkers can be extracted from the brain scans, such as voxel-based morphometry (VBM) and FreeSurfer (FS) markers. To characterize the data in multiple modalities, we concatenate the vector representations of these biomarkers as the phenotypic vector of a participant, *i.e.*, in our study we write $\mathbf{x}_i = [\mathbf{x}_i^{VBM}, \mathbf{x}_i^{FS}]$ and $\mathbf{x}_{ij} = [\mathbf{x}_{ij}^{VBM}, \mathbf{x}_{ij}^{FS}]$, where $1 \le i \le n$, $1 \le j \le n_i$. Because $\{\mathbf{x}_{ij}\}_{j=1}^{n_i+1}$ describe the temporal changes of the phenotypes of the i-th participant over time, $\mathcal{X}_i$ is a summarization of the ***dynamic*** measurements of the i-th participant. These dynamic measurements are known broadly as *longitudinal* data in the literature of medical image computing and imaging genetic studies.

In addition to the phenotypic measurements of the participants of a neuroimaging data set, genotypes of the same cohort are usually available as well. The SNPs for all participants are represented by $\mathbf{X}_{SNP} = [\mathbf{x}_1^{SNP}, \ldots, \mathbf{x}_n^{SNP}] \in \Re^{d_{SNP}\times n}$, where $\mathbf{x}_i^{SNP}$ is the vector representation of the SNP profile of the i-th participant. Here we note that $\mathbf{X}_{SNP}$ is ***static*** that does not vary over time.

Besides the input phenotypic (dynamic) and genetic (static) data, our approach aims to predict the cognitive status of participants in the ADNI. In a regression on cognitive scores derived from cognitive tests, we use $\mathbf{Y}_l \in \Re^{c\times l}$ to list the clinical scores of the first l participants at their last visit, where c is the number of clinical scores in our study. Here, without loss of generality, we consider the first l ($l \le n$) samples as the labeled data for training. Our task is to learn a projection tensor $\mathcal{W} = \{\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_n\} \in \Re^{d\times r_1\times n}$, by which we can compute the fixed-length biomarker representations $\mathcal{W}^T \otimes \mathbf{X}_{MR} = [\mathbf{W}_1^T\mathbf{x}_1, \mathbf{W}_2^T\mathbf{x}_2, \cdots, \mathbf{W}_n^T\mathbf{x}_n] \in \Re^{r_0\times n}$ by projecting the most recent medical records of $\mathbf{X}_{MR}$ for all participants.

Our Objective

In this section we describe our objective to learn the participant-specific projections $f_i : \Re^d \mapsto \Re^{r_0}$, which can be implemented as a linear projection by computing $\mathbf{z}_i = \mathbf{W}_i^T\mathbf{x}_i$.

Learning Group-Structured Representations for Static Genetic Data

Recent advances in high-throughput genotyping techniques enable new approaches to study the influence of genetic variation on brain structure and function. Traditional association studies typically employ independent and pairwise univariate analysis, which consider SNPs as isolated units and ignores important underlying interacting relationships between the units (Wu et al. 2010). However, certain SNPs are naturally connected via different pathways. For example, multiple SNPs from one gene often carry out genetic functionalities together. Moreover, linkage disequilibrium (LD) (Wu et al. 2010) describes the non-random association between alleles at different loci, through which the SNPs in high LD are linked together in meiosis. Thus, instead of treating SNPs in an isolated manner, it is beneficial to exploit the group structure among SNPs (Wang et al. 2012a; Yan et al. 2015). To utilize the group structure of the SNP data, we propose to learn an inherent representation of the input genetic data by minimizing the following objective:

$$\begin{aligned}\mathcal{J}_0(\mathbf{H}_0, \mathbf{G}_0) &= \|\mathbf{X}_{\text{SNP}} - \mathbf{H}_0\mathbf{G}_0\|_{2,1} + \alpha\|\mathbf{H}_0\|_{G_{2,1}}, \\ s.t. \quad & \mathbf{H}_0^T\mathbf{H}_0 = \mathbf{I}.\end{aligned} \tag{1}$$

The first term in Eq. (1) uses matrix factorization to decouple the input SNP data matrix $\mathbf{X}_{\text{SNP}}$, which is equivalent to perform K-means clustering on $\mathbf{X}_{\text{SNP}}$ (Ding and He 2004; Ding, Li, and Jordan 2008). As a result, $\mathbf{H}_0$ can be viewed as a condensed view of the SNP features and $\mathbf{G}_0$ can be viewed as the new representations of the n participants in the subspace spanned by $\mathbf{H}_0$ (Ding, Li, and Jordan 2008; Wang et al. 2011; Wang, Nie, and Huang 2015). Here we choose to use the $\ell_{2,1}$-norm distances for improved robustness of our model against outliers (Wang, Nie, and Huang 2012; Nie et al. 2013; Liu et al. 2019a,b). Motivated by the earlier works (Wang et al. 2012a; Yan et al. 2015), we can integrate the group structure between SNPs by applying a group $\ell_{2,1}$-norm ($G_{2,1}$-norm) regularization to $\mathbf{H}_0$. Here, the $G_{2,1}$-norm of a matrix $\mathbf{M}$ is defined as

$\|\mathbf{M}\|_{G_{2,1}} = \sum_{k=1}^{K} \left\|\mathbf{M}^k\right\|_{2,1}$ (Wang et al. 2012a), where $\mathbf{M} = [\mathbf{M}^1; \mathbf{M}^2; \cdots; \mathbf{M}^K]$ consists of K groups. These groups can capture LD correlations between SNPs (Wang et al. 2012a) or leverage additional genetic side-information (Yan et al. 2015).

Learning Temporally Augmented Representations for Dynamic Imaging Data

While the SNPs describe the genetic profiles of the participants of a studied cohorts that remain constant over time, the structures and functions of the brains of the participants vary during the development of AD. As a result, the longitudinal imaging records and the measurements of biomarkers extracted from these brain scans are dynamic and change over time. Most importantly, because different participants may take brain scans at different times, the total number of brain scans per participant are not same in general. As a result, it is difficult to directly use the dynamic imaging data to build machine learning models. To address this difficulty, we propose to learn a novel temporally augmented representation with a fixed length for every participant learned from their dynamic imaging data (Lu et al. 2019, 2020a,b).

First, to summarize the brain variations of every participant individually, we learn to map $\mathcal{X}_i = \{\mathbf{X}_i, \mathbf{x}_i\}$ that resides in the high d-dimensional space into a lower r_0-dimensional subspace via a projection by computing $\mathbf{z}_i = \mathbf{W}_i^T \mathbf{x}_i$, where $\mathbf{W}_i$ is learned from $\mathbf{X}_i$ to preserve as much information of $\mathbf{X}_i$ as possible by minimizing the objective of PCA (Jolliffe 2011) as following:

$$\begin{aligned} \mathcal{J}_1(\mathbf{W}_i) &= \left\|\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i\right\|_{2,1}, \\ s.t. \quad & \mathbf{W}_i^T\mathbf{W}_i = \mathbf{I}. \end{aligned} \tag{2}$$

Here again we choose to use the $\ell_{2,1}$-norm in the PCA formulation for its improved robustness against outlying samples that are inevitable when the dataset grows (Wang, Nie, and Huang 2012; Nie et al. 2013; Liu et al. 2019a,b). Now we learn the projections for all the participants altogether to minimize the following objective:

$$\begin{aligned} \mathcal{J}_2(\mathcal{W}) = &\sum_{i=1}^{n} \left\|\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i\right\|_{2,1} \\ &+ \alpha \left\|\mathbf{W}_{(1)}\right\|_* + \beta \left\|\mathbf{W}_{(2)}\right\|_*, \\ s.t. \quad & \mathbf{W}_i^T\mathbf{W}_i = \mathbf{I}, \text{ for } i = 1, 2, \cdots, n, \end{aligned} \tag{3}$$

where $\mathcal{W} = \{\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_n\}$ is the tensor of the n projection matrices, one for each participant. Here, to maximize the consistency across all the learned projection matrices for the same cohort of participants (Wang et al. 2012d,c; Brand et al. 2018, 2019, 2020a), we use the trace norm regularization of $\left\|\mathbf{W}_{(1)}\right\|_*$ and $\left\|\mathbf{W}_{(2)}\right\|_*$ in our objective, where $\mathbf{W}_{(1)} = [\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_n] \in \Re^{d \times r_1 n}$ and $\mathbf{W}_{(2)} = \left[\mathbf{W}_1^T, \mathbf{W}_2^T, \cdots, \mathbf{W}_n^T\right] \in \Re^{r_1 \times dn}$ are the two unfolded matrices of $\mathcal{W}$.

Integrating Static and Multi-Modal Dynamic Data Using Genotype-Phenotype Augmentations

Because multiple types of imaging biomarkers can be extracted from brain scans, the phenotypic measurements can be naturally formulated as multi-modal data (Wang et al. 2012b; Brand et al. 2018, 2019; Lu et al. 2020a; Brand et al. 2020b). Suppose a total of K types of biomarkers are extracted from the brain scans, we have $\mathbf{x}_i = [(\mathbf{x}_i^1)^T, (\mathbf{x}_i^2)^T, \ldots, (\mathbf{x}_i^K)^T]^T \in \Re^d$ and $\mathbf{X}_{MR}^k = [\mathbf{x}_1^k, \mathbf{x}_2^k, \cdots, \mathbf{x}_n^k] \in \Re^{d_k \times n}$, where $\sum_{k=1}^{K} d_k = d$. Using the temporal augmentation, we can write the augmented phenotypic measurements in every modality as:

$$(\mathcal{W}^k)^T \otimes \mathbf{X}_{MR}^k = [(\mathbf{W}_1^k)^T\mathbf{x}_1^k, (\mathbf{W}_2^k)^T\mathbf{x}_2^k, \cdots, (\mathbf{W}_n^k)^T\mathbf{x}_n^k], \tag{4}$$

where $\mathbf{W}_i = [\mathbf{W}_i^1; \mathbf{W}_i^2; \cdots; \mathbf{W}_i^K] \in \Re^{d \times r}$. Then using Eq. (1), we can integrate the static and multi-modal dynamic data by minimizing:

$$\begin{aligned} \mathcal{J}_3(\mathcal{W}, \mathcal{H}, \mathcal{G}) = &\gamma_1 \sum_{i=1}^{n} \left\|\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i\right\|_{2,1} \\ &+ \gamma_2 \sum_{k=1}^{K} \left\|(\mathcal{W}^k)^T \otimes \mathbf{X}_{MR}^k - \mathbf{H}_k\mathbf{G}_k\right\|_{2,1} \\ &+ \gamma_3 \left\|\mathbf{X}_{\text{SNP}} - \mathbf{H}_0\mathbf{G}_0\right\|_{2,1} \\ &+ \gamma_4 \sum_{k=0}^{K} \left\|\mathbf{G} - \mathbf{G}_k\right\|_{2,1} \\ &+ \gamma_6 \left\|\mathbf{H}_{(1)}\right\|_{G_1} + \gamma_7 \left\|\mathbf{H}_0\right\|_{G_{2,1}} \\ &+ \gamma_8 \left\|\mathbf{W}_{(1)}\right\|_* + \gamma_9 \left\|\mathbf{W}_{(2)}\right\|_*, \\ s.t. \quad & \mathbf{W}_i^T\mathbf{W}_i = \mathbf{I}, \ \mathbf{H}_k^T\mathbf{H}_k = \mathbf{I}, \\ & \text{for } i = 1, 2, \cdots, n, \text{ and } k = 0, 1, \cdots, K, \end{aligned} \tag{5}$$

where we write $\mathcal{H} = \{\mathbf{H}_0, \mathbf{H}_1, \cdots, \mathbf{H}_K\}$ and $\mathcal{G} = \{\mathbf{G}, \mathbf{G}_0, \mathbf{G}_1, \cdots, \mathbf{G}_K\}$. In the above objective in Eq. (5), we consider the SNP data as one additional modality beyond the K imaging modalities. As a result, $\mathbf{G}_k$ are the representations of the k-th modality of all the participants in the studied cohort, where $k = 0$ indicates the static genetic modality and $k > 0$ indicates a dynamic phenotypic modality. These new representations in the $K + 1$ modalities are integrated by the group ℓ_1-norm (G_1-norm) (Wang et al. 2012b; Wang, Nie, and Huang 2013; Wang et al. 2013), which is defined as $\|\mathbf{M}\|_{G_1} = \sum_{j=1}^{m} \sum_{k=1}^{K} \left\|\mathbf{m}_j^k\right\|_2$, where $\mathbf{M} = [\mathbf{M}^1; \mathbf{M}^2; \cdots; \mathbf{M}^K]$.

Learning Semi-Supervised Temporal Augmentations

Besides the input data described by the genetic and phenotypic measures, we are often supplied with scores assessed from cognitive tests for clinical diagnoses of a subset of the participants, which can be considered as the labeled data for training a machine learning model. Without loss of generality, we suppose that the clinical scores from c cognitive assessments are known for the first l ($l \leq n$) participants in a studied cohort and we write $\mathbf{G} = [\mathbf{G}_l, \mathbf{G}_u]$. Then we can use a support vector regression (SVR) model to associate the augmented genotype-phenotype representations $\mathbf{G}_l$ to the known cognitive scores of $\mathbf{Y}_l$ by minimizing the fol-

lowing objective:

$$\mathcal{J}(\mathcal{W},\mathcal{H},\mathcal{G},\alpha) = \gamma_1 \sum_{i=1}^{n} \left\|\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i\right\|_{2,1}$$
$$+ \gamma_2 \sum_{k=1}^{K} \left\|(\mathcal{W}^k)^T \otimes \mathbf{X}_{MR}^k - \mathbf{H}_k\mathbf{G}_k\right\|_{2,1}$$
$$+ \gamma_3 \left\|\mathbf{X}_{\text{SNP}} - \mathbf{H}_0\mathbf{G}_0\right\|_{2,1} + \gamma_4 \sum_{k=0}^{K} \left\|\mathbf{G} - \mathbf{G}_k\right\|_{2,1}$$
$$+ \gamma_5 \sum_{o=1}^{c} \text{svr}(\mathbf{y}_l^o, \mathbf{G}_l, \boldsymbol{\alpha}_o - \boldsymbol{\alpha}_o^*)$$
$$+ \gamma_6 \left\|\mathbf{H}\right\|_{G_1} + \gamma_7 \left\|\mathbf{H}_0\right\|_{G_{2,1}} + \gamma_8 \left\|\mathbf{W}_{(1)}\right\|_* + \gamma_9 \left\|\mathbf{W}_{(2)}\right\|_*,$$
$$s.t. \quad \mathbf{W}_i^T\mathbf{W}_i = \mathbf{I},\ \mathbf{H}_k^T\mathbf{H}_k = \mathbf{I},$$
$$\text{for } i = 1, 2, \cdots, n, \text{ and } k = 0, 1, \cdots, K, \tag{6}$$

where $\boldsymbol{\alpha}_o, \boldsymbol{\alpha}_o^* \in \alpha$ are the dual variables of SVR.

The Solution Algorithm

Although our objective in Eq. (6) has been clearly motivated, it is difficult efficiently solve in general, because it is non-smooth. Thus we drive the efficient solution of our objective in this section. Using the optimization framework in the earlier works (Liu et al. 2017; Yang et al. 2019) that introduced the iterative reweighted method to solve non-smooth objectives, we can solve Eq. (6) by an iterative procedure in which the key step is to minimize the following objective:

$$\mathcal{J}^{smooth}(\mathcal{W},\mathcal{H},\mathcal{G},\alpha) =$$
$$\gamma_1 \sum_{i=1}^{n} \text{tr}\left((\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i)^T\mathbf{D}_{1,i}(\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i)\right)$$
$$+ \gamma_2 \sum_{k=1}^{K} \text{tr}\left(((\mathcal{W}^k)^T \otimes \mathbf{X}_{MR}^k - \mathbf{H}_k\mathbf{G}_k)^T\mathbf{D}_{2,k}((\mathcal{W}^k)^T \otimes \mathbf{X}_{MR}^k - \mathbf{H}_k\mathbf{G}_k)\right)$$
$$+ \gamma_3 \text{tr}\left((\mathbf{X}_{\text{SNP}} - \mathbf{H}_0\mathbf{G}_0)^T\mathbf{D}_3(\mathbf{X}_{\text{SNP}} - \mathbf{H}_0\mathbf{G}_0)\right)$$
$$+ \gamma_4 \sum_{k=0}^{K} \text{tr}\left((\mathbf{G} - \mathbf{G}_k)^T\mathbf{D}_{4,k}(\mathbf{G} - \mathbf{G}_k)\right)$$
$$+ \gamma_5 \sum_{o=1}^{c} \text{svr}(\mathbf{y}_l^o, \mathbf{G}_l, \boldsymbol{\alpha}_o - \boldsymbol{\alpha}_o^*)$$
$$+ \gamma_6 \sum_{k=1}^{K} \text{tr}\left(\mathbf{H}_k\mathbf{D}_{6,k}\mathbf{H}_k^T\right) + \gamma_7 \text{tr}(\mathbf{H}_0^T\mathbf{D}_7\mathbf{H}_0)$$
$$+ \gamma_8 \text{tr}(\mathbf{W}_{(1)}^T\mathbf{D}_8\mathbf{W}_{(1)}) + \gamma_9 \text{tr}(\mathbf{W}_{(2)}^T\mathbf{D}_9\mathbf{W}_{(2)})$$
$$s.t. \quad \mathbf{W}_i^T\mathbf{W}_i = \mathbf{I},\ \mathbf{H}_k^T\mathbf{H}_k = \mathbf{I},$$
$$\text{for } i = 1, 2, \cdots, n, \text{ and } k = 0, 1, \cdots, K, \tag{7}$$

where j-th diagonal element of diagonal matrices $\mathbf{D}_{1,i}, \mathbf{D}_{2,k}, \mathbf{D}_3, \mathbf{D}_{4,k}, \mathbf{D}_{6,k}$ is defined as: $\frac{1}{2\sqrt{\left\|\mathbf{x}_i^j - \mathbf{w}_i^j\mathbf{W}_i^T\mathbf{X}_i\right\|_2^2 + \delta}}$, $\frac{1}{2\sqrt{\left\|(\mathcal{W}^{k,j})^T \otimes \mathbf{X}_{MR}^k - \mathbf{h}_k^j\mathbf{G}_k\right\|_2^2 + \delta}}$, $\frac{1}{2\sqrt{\left\|\mathbf{x}_{SNP}^j - \mathbf{h}_0^j\mathbf{G}_0\right\|_2^2 + \delta}}$, $\frac{1}{2\sqrt{\left\|\mathbf{g}^j - \mathbf{g}_g^j\right\|_2^2 + \delta}}$, $\frac{1}{2\sqrt{\left\|\mathbf{h}_{k,j}\right\|_2^2 + \delta}}$, $\mathbf{D}_8 = \frac{1}{2}(\mathbf{W}_{(1)}\mathbf{W}_{(1)}^T + \delta\mathbf{I})^{-\frac{1}{2}}$, $\mathbf{D}_9 = \frac{1}{2}(\mathbf{W}_{(2)}\mathbf{W}_{(2)}^T + \delta\mathbf{I})^{-\frac{1}{2}}$, and $\mathbf{D}_7$ is a block diagonal matrix, whose s-th block is $\frac{1}{2\sqrt{\left\|\mathbf{H}_0^s\right\|_{2,2}^2 + \delta}}\mathbf{I}_s$. Here $\mathbf{I}_s \in \Re^{d_s \times d_s}$ is the identity matrix, and d_s denotes the number of rows of s-th block of $\mathbf{H}_0 = [\mathbf{H}_0^1; \mathbf{H}_0^2; \cdots; \mathbf{H}_0^S]$, for the groups of SNPs. $(\mathcal{W}^{k,j})^T \otimes \mathbf{X}_{MR}^k$ is defined as j-th row of $(\mathcal{W}^k)^T \otimes \mathbf{X}_{MR}^k$. The dimensions of $\mathbf{D}_*$ are: $\mathbf{D}_{1,i} \in \Re^{d\times d}$, $\mathbf{D}_{2,k} \in \Re^{r_0\times r_0}$, $\mathbf{D}_3 \in \Re^{d_{SNP}\times d_{SNP}}$, $\mathbf{D}_{4,k} \in \Re^{r_1\times r_1}$, $\mathbf{D}_{6,k} \in \Re^{r_1\times r_1}$, $\mathbf{D}_7 \in \Re^{d_{SNP}\times d_{SNP}}$, $\mathbf{D}_8 \in \Re^{d\times d}$, $\mathbf{D}_9 \in \Re^{r_0\times r_0}$.

To minimize the smoothed objective in Eq. (7), we use the alternating direction method of multipliers (ADMM) proposed by (Bertsekas 2014; Boyd et al. 2011). The ADMM breaks the complex problem into smaller sub-problems that are easier to solve. Following the ADMM we rewrite the objective in Eq. (7) as an equivalent objective in Eq. (8), by introducing two additional constraints $\mathbf{W}_i = \mathbf{B}_i$ ($\Rightarrow \mathbf{W}_{(1)} = \mathbf{B}_{(1)}$ and $\mathbf{W}_{(2)} = \mathbf{B}_{(2)}$ and $\mathcal{W} = \mathcal{B}$) and $\mathbf{H}_k = \mathbf{A}_k$ ($\Rightarrow \mathcal{H} = \mathcal{A}$) to decouple the $\mathbf{W}_i$ and $\mathbf{H}_k$, for $i = 1, 2, \cdots, n$ and $k = 0, 1, \cdots, K$:

$$\mathcal{J}^{ADMM}(\mathcal{W},\mathcal{H},\mathcal{G},\alpha,\mathcal{B},\mathcal{A}) =$$
$$\gamma_1 \sum_{i=1}^{n} \text{tr}\left((\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i)^T\mathbf{D}_{1,i}(\mathbf{X}_i - \mathbf{B}_i\mathbf{B}_i^T\mathbf{X}_i)\right)$$
$$+ \gamma_2 \sum_{k=1}^{K} \text{tr}\left(((\mathcal{W}^k)^T \otimes \mathbf{X}_{MR}^k - \mathbf{H}_k\mathbf{G}_k)^T\mathbf{D}_{2,k}((\mathcal{W}^k)^T \otimes \mathbf{X}_{MR}^k - \mathbf{A}_k\mathbf{G}_k)\right)$$
$$+ \gamma_3 \text{tr}\left((\mathbf{X}_{\text{SNP}} - \mathbf{H}_0\mathbf{G}_0)^T\mathbf{D}_3(\mathbf{X}_{\text{SNP}} - \mathbf{A}_0\mathbf{G}_0)\right)$$
$$+ \gamma_4 \sum_{k=0}^{K} \text{tr}\left((\mathbf{G} - \mathbf{G}_k)^T\mathbf{D}_{4,k}(\mathbf{G} - \mathbf{G}_k)\right)$$
$$+ \gamma_5 \sum_{o=1}^{c} \text{svr}(\mathbf{y}_l^o, \mathbf{G}_l, \boldsymbol{\alpha}_o - \boldsymbol{\alpha}_o^*)$$
$$+ \gamma_6 \sum_{k=1}^{K} \text{tr}\left(\mathbf{H}_k\mathbf{D}_{6,k}\mathbf{A}_k^T\right) + \gamma_7 \text{tr}\left(\mathbf{H}_0^T\mathbf{D}_7\mathbf{A}_0\right)$$
$$+ \gamma_8 \text{tr}(\mathbf{W}_{(1)}^T\mathbf{D}_8\mathbf{B}_{(1)}) + \gamma_9 \text{tr}(\mathbf{W}_{(2)}^T\mathbf{D}_9\mathbf{B}_{(2)})$$
$$+ \sum_{i=1}^{n} \frac{\mu_{1,i}}{2} \left\|\mathbf{W}_i^T\mathbf{B}_i - \mathbf{I} + \frac{1}{\mu_{1,i}}\boldsymbol{\Lambda}_{1,i}\right\|_F^2$$
$$+ \sum_{i=1}^{n} \frac{\mu_{2,i}}{2} \left\|\mathbf{B}_i - \mathbf{W}_i + \frac{1}{\mu_{2,i}}\boldsymbol{\Lambda}_{2,i}\right\|_F^2$$
$$+ \sum_{k=0}^{K} \frac{\mu_{3,k}}{2} \left\|\mathbf{H}_k^T\mathbf{A}_k - \mathbf{I} + \frac{1}{\mu_{3,k}}\boldsymbol{\Lambda}_{3,k}\right\|_F^2$$
$$+ \sum_{k=0}^{K} \frac{\mu_{4,k}}{2} \left\|\mathbf{A}_k - \mathbf{H}_k + \frac{1}{\mu_{4,k}}\boldsymbol{\Lambda}_{4,k}\right\|_F^2, \tag{8}$$

where $\boldsymbol{\Lambda}_{1,i}, \boldsymbol{\Lambda}_{2,i}, \boldsymbol{\Lambda}_{3,k}, \boldsymbol{\Lambda}_{4,k}$ are the Lagrange multipliers

Algorithm 1: Solve the minimization problem in Eq. (8)

Input: $\mathbf{X}_{MR},\ \mathbf{X}_i,\ \mathbf{X}_{SNP},\ \mathbf{Y}_l,$ for $1 \le i \le n$;
Initialization: $\mathcal{W},\ \mathbf{B}_i,\ \mathbf{H}_k,\ \mathbf{A}_k,\ \mathbf{G}_k,\ \boldsymbol{\alpha}_o - \boldsymbol{\alpha}_o^*,\ \mathbf{\Lambda}_{1,i},\ \mathbf{\Lambda}_{2,i},\ \mathbf{\Lambda}_{3,k},\ \mathbf{\Lambda}_{4,k}$
$1 < \rho_{1,i},\ \rho_{2,i},\ \rho_{3,k},\ \rho_{4,k} < 2$,
$\mu_{1,i},\ \mu_{2,i},\ \mu_{3,k},\ \mu_{4,k} > 0$,
$\gamma_1,\ \gamma_2,\ \gamma_3,\ \gamma_4,\ \gamma_5,\ \gamma_6,\ \gamma_7,\ \gamma_8,\ \gamma_9 > 0$,
for $1 \le i \le n$ and $0 \le k \le K$ and $1 \le o \le c$;
while *not converge* **do**

1. Update $\mathbf{D}_{1,i}, \mathbf{D}_{2,k}, \cdots, \mathbf{D}_9$ as defined in Eq. (7);
2. Update $\mathbf{H}_k$ $(1 \le k \le K)$ by $\mathbf{H}_k =$ $(\mu_{3,k}\mathbf{A}_k\mathbf{A}_k^T + \mu_{4,k}\mathbf{I})^{-1}(\gamma_2\mathbf{D}_{2,k}((\mathcal{W}^k)^T \otimes \mathbf{X}_{MR} - \mathbf{A}_k\mathbf{G}_k)\mathbf{G}_k^T - \mathbf{A}_k(\gamma_6\mathbf{D}_{6,k} + \mathbf{\Lambda}_{3,k}^T) + (\mu_{3,k} + \mu_{4,k})\mathbf{A}_k + \mathbf{\Lambda}_{4,k})$;
3. Update $\mathbf{H}_0$ by $\mathbf{H}_0 = (\mu_{3,0}\mathbf{A}_0\mathbf{A}_0^T + \mu_{4,0}\mathbf{I})^{-1}(\gamma_3\mathbf{D}_3(\mathbf{X}_{SNP} - \mathbf{A}_0\mathbf{G}_0)\mathbf{G}_0^T + ((\mu_{4,0} + \mu_{3,0})\mathbf{I} - \gamma_7\mathbf{D}_7)\mathbf{A}_0 - \mathbf{A}_0\mathbf{\Lambda}_{3,0}^T + \mathbf{\Lambda}_{4,0})$;
4. Update $\mathbf{A}_k$ $(1 \le k \le K)$ by $\mathbf{A}_k =$ $(\mu_{3,k}\mathbf{H}_k\mathbf{H}_k^T + \mu_{4,k}\mathbf{I})^{-1}(\gamma_2\mathbf{D}_{2,k}((\mathcal{W}^k)^T \otimes \mathbf{X}_{MR} - \mathbf{H}_k\mathbf{G}_k)\mathbf{G}_k^T - \gamma_6\mathbf{H}_k\mathbf{D}_{6,k} + (\mu_{3,k} + \mu_{4,k})\mathbf{H}_k - \mathbf{H}_k\mathbf{\Lambda}_{3,k} - \mathbf{\Lambda}_{4,k})$;
5. Update $\mathbf{A}_0$ by $\mathbf{A}_0 = (\mu_{3,0}\mathbf{H}_0\mathbf{H}_0^T + \mu_{4,0}\mathbf{I})^{-1}(\gamma_3\mathbf{D}_3(\mathbf{X}_{SNP} - \mathbf{H}_0\mathbf{G}_0)\mathbf{G}_0^T - \gamma_7\mathbf{D}_7\mathbf{H}_0 + (\mu_{4,0} + \mu_{3,0})\mathbf{H}_0 - \mathbf{H}_0\mathbf{\Lambda}_{3,0} - \mathbf{\Lambda}_{4,0})$;
6. Update $\mathbf{G}_k$ $(1 \le k \le K)$ by $\mathbf{G}_k = (\frac{\gamma_2}{2}(\mathbf{H}_k^T\mathbf{D}_{2,k}\mathbf{A}_k + \mathbf{A}_k^T\mathbf{D}_{2,k}\mathbf{H}_k) + \gamma_4\mathbf{D}_{4,k})^{-1}(\gamma_4\mathbf{D}_{4,k}\mathbf{G} + \frac{\gamma_2}{2}(\mathbf{H}_k^T + \mathbf{A}_k^T)\mathbf{D}_{2,k}((\mathcal{W}^k)^T \otimes \mathbf{X}_{MR}^k))$;
7. Update $\mathbf{G}_0$ by $\mathbf{G}_0 = (\frac{\gamma_3}{2}(\mathbf{H}_0^T\mathbf{D}_3\mathbf{A}_0 + \mathbf{A}_0^T\mathbf{D}_3\mathbf{H}_0) + \gamma_4\mathbf{D}_{4,0})^{-1}(\frac{\gamma_3}{2}(\mathbf{H}_0^T + \mathbf{A}_0^T)\mathbf{D}_3\mathbf{X}_{SNP} + \gamma_4\mathbf{D}_{4,0}\mathbf{G})$;
8. Update $\boldsymbol{\alpha}_o - \boldsymbol{\alpha}_o^*$ $(1 \le o \le c)$ by solving Support Vector Regression problem;
9. Update $\mathbf{G}_l$ by the root of equation: $2\gamma_4\sum_{k=0}^K \mathbf{D}_{4,k}(\mathbf{G}_l - \mathbf{G}_{k,l}) + 4\gamma_5\sum_{o=1}^c \Big(\mathbf{G}_l\Big(k'\big(s(\mathbf{G}_l)\big) \circ \big((\boldsymbol{\alpha}_o - \boldsymbol{\alpha}_o^*)(\boldsymbol{\alpha}_o - \boldsymbol{\alpha}_o^*)^T\big)\Big) - \mathbf{G}_l\mathbf{D}_o^\alpha\Big) = \mathbf{0}$;
For $\mathbf{G}_{k,r}$, the right submatrix from l-th column of $\mathbf{G}_k$, update $\mathbf{G}_r$ by $\mathbf{G}_r = (\sum_{k=0}^K \mathbf{D}_{4,k})^{-1}\sum_{k=0}^K(\mathbf{D}_{4,k}\mathbf{G}_{k,r})$;
10. Update $\mathbf{B}_i$ $(1 \le i \le n)$ by $\mathbf{B}_i = (-\gamma_1(\mathbf{D}_{1,i}(\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i)\mathbf{X}_i^T + (\mathbf{D}_{1,i}(\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i)\mathbf{X}_i^T)^T) + \mu_{1,i}\mathbf{W}_i\mathbf{W}_i^T + \mu_{2,i}\mathbf{I})^{-1}(-\gamma_8\mathbf{D}_8\mathbf{W}_i - \gamma_9\mathbf{W}_i\mathbf{D}_9 + (\mu_{1,i} + \mu_{2,i})\mathbf{W}_i - \mathbf{W}_i\mathbf{\Lambda}_{1,i} - \mathbf{\Lambda}_{2,i})$;
11. Update $\mathbf{w}_{i,q}$ $(1 \le i \le n$ and $1 \le q \le r_0)$ by $\mathbf{w}_{i,q} = (-\gamma_1\big((\mathbf{X}_i\mathbf{X}_i^T(\mathbf{I} - \mathbf{B}_i\mathbf{B}_i^T)\mathbf{D}_{1,i}) + (\mathbf{X}_i\mathbf{X}_i^T(\mathbf{I} - \mathbf{B}_i\mathbf{B}_i^T)\mathbf{D}_{1,i})^T\big) + 2\gamma_2\mathbf{D}_{i,q}^w + \mu_{1,i}\mathbf{B}_i\mathbf{B}_i^T + \mu_{2,i}\mathbf{I})^{-1}(\gamma_2\mathbf{x}_{i,q}^w - \gamma_8\mathbf{D}_8\mathbf{b}_{i,q} - \gamma_9\mathbf{B}_i\mathbf{d}_{9,q} + \mu_{1,i}\mathbf{b}_{i,q} - \mathbf{B}_i(\boldsymbol{\lambda}_{1,i}^q)^T + \mu_{2,i}\mathbf{b}_{i,q} + \boldsymbol{\lambda}_{2,i,q})$;
12. Update $\mathbf{\Lambda}_{1,i}$ $(1 \le i \le n)$ by $\mathbf{\Lambda}_{1,i} = \mathbf{\Lambda}_{1,i} + \mu_{1,i}(\mathbf{W}_i^T\mathbf{B}_i - \mathbf{I})$;
13. Update $\mathbf{\Lambda}_{2,i}$ $(1 \le i \le n)$ by $\mathbf{\Lambda}_{2,i} = \mathbf{\Lambda}_{2,i} + \mu_{2,i}(\mathbf{B}_i - \mathbf{W}_i)$;
14. Update $\mathbf{\Lambda}_{3,k}$ $(0 \le k \le K)$ by $\mathbf{\Lambda}_{3,k} = \mathbf{\Lambda}_{3,k} + \mu_{3,k}(\mathbf{H}_k^T\mathbf{A}_k - \mathbf{I})$;
15. Update $\mathbf{\Lambda}_{4,k}$ $(0 \le k \le K)$ by $\mathbf{\Lambda}_{4,k} = \mathbf{\Lambda}_{4,k} + \mu_{4,k}(\mathbf{A}_k - \mathbf{H}_k)$;
16. Update $\mu_{1,i}, \mu_{2,i}, \mu_{3,k}, \mu_{4,k}$ $(1 \le i \le n$ and $0 \le k \le K)$ by $\mu_{1,i} = \rho_{1,i}\mu_{1,i};\ \mu_{2,i} = \rho_{2,i}\mu_{2,i};\ \mu_{3,k} = \rho_{3,k}\mu_{3,k};\ \mu_{4,k} = \rho_{4,k}\mu_{4,k}$;

end
Output:$\mathbf{W}_i$ $(1 \le i \le n)$

for the constraints $\mathbf{W}_i^T\mathbf{W}_i = \mathbf{I},\ \mathbf{W}_i = \mathbf{B}_i,\ \mathbf{H}_k^T\mathbf{H}_k = \mathbf{I},\ \mathbf{H}_k = \mathbf{A}_k$. The detailed algorithm to minimize the smoothed ADMM objective in Eq. (8) is presented in Algorithm 1. In step 9 of Algorithm 1, $\mathbf{D}_o^\alpha \in \Re^{l\times l}$ is defined as a diagonal matrix whose q-th diagonal element is $\sum_{p=1}^{r_1}\mathbf{g}_l^p\mathbf{s}_{l,q}^o$, where $\mathbf{s}_{l,q}^o$ is q-th column of $k'\left(s(\mathbf{G}_l)\right) \circ \big((\boldsymbol{\alpha}_o - \boldsymbol{\alpha}_o^*)(\boldsymbol{\alpha}_o - \boldsymbol{\alpha}_o^*)^T\big)$. In step 11, $\mathbf{D}_{i,q}^w \in \Re^{d\times d}$ is defined as block-diagonal matrix whose k-th block is $d_{2,k,q}^q\mathbf{x}_i^k(\mathbf{x}_i^k)^T$ and all other elements are zeros, and $\mathbf{x}_{i,q}^w \in \Re^d$ is defined as $\mathbf{x}_{i,q}^w = [(\mathbf{d}_{2,1}^q(\mathbf{H}_1 + \mathbf{A}_1)\mathbf{g}_{1,i})\mathbf{x}_i^1;\ (\mathbf{d}_{2,2}^q(\mathbf{H}_2 + \mathbf{A}_2)\mathbf{g}_{2,i})\mathbf{x}_i^2; \cdots; (\mathbf{d}_{2,K}^q(\mathbf{H}_K + \mathbf{A}_K)\mathbf{g}_{K,i})\mathbf{x}_i^K]$. The time complexity of Algorithm 1 is $O(nr_0d^2(n+d))$ for each iteration, where step 11 of Algorithm 1 is the most dominant. Due to the space limitation, the derivation details for Algorithm 1 will be supplied in the extended journal version of this paper.

Experiments

In this section, we compare the regression performance on the cognitive scores prediction task between the augmented and original representations; this is intended to showcase how the proposed augmentation method improves the final predictive capacity. In addition, we analyze the AD-relevant biomarkers identified by the proposed augmentation.

Dataset Description The data used in the following experiments are obtained from the ADNI database (adni.loni.usc.edu). We download the magnetic resonance imaging (MRI) scans, SNP genotypes, and demographic information of 821 ADNI-1 participants. We perform VBM and FS automated parcellation on the MRI data following (Risacher et al. 2010) and extracted mean modulated

gray matter (GM) measures for 90 target regions of interest (ROI), and follow the SNP quality control steps discussed in Shen et al. 2010. We build the dataset $\mathbf{X}_i$ and $\mathbf{X}_{MR}$ from 342 participants who have complete records at Month 0/Month 6/Month 12/Month 24. In order to test the robustness of our augmentation to the missing records, we intentionally discard Month 24 scans with 50% probability. To construct the target label matrix $\mathbf{Y}$, we use the longitudinal cognitive scores of Rey's Auditory Verbal Learning Test (RAVLT TOT, RAVLT 30, RAVLT RECOG), Fluency Test (FLU VEG, FLU ANIM), and Alzheimer's Disease Assessment Scale (ADAS) at the most recent time point. The input features and cognitive scores in the resulting dataset are normalized and are randomly shuffled at the beginning of each experiment to ensure representative results.

Experiment Settings We split the dataset into a training and test set with a proportion of 80% and 20%, such that the number of participants is $l = 274$ in the training set and $n - l = 68$ in the testing set. During the data augmentation step, our model learns a projection $\mathbf{W}_i$ for each participant $i = 1, 2, \cdots, n$ from the cognitive scores in the training set, and the available SNPs and MRI images in both the training and testing sets. Once the $\mathbf{W}_i$'s have been learned, we predict the cognitive scores in the test set with the original representation $\mathbf{x}_i$ and the augmented representation $\mathbf{W}_i^T\mathbf{x}_i$, and compare root mean squared errors (RMSE) of the two predictions. After augmentation, we use the following prediction models: Least Absolute Shrinkage and Selection Operator (LASSO), Ridge linear Regression (RR), Convolutional Neural Network (CNN), and Support Vector Regression (SVR). We use a 5-fold cross-validation within the training set to identify the best hyperparameters. The validation set in this inner cross-validation is chosen to be 20% of the training set. Due to space limitation, the report for fine tuning the hyperparameters will be supplied in the extended journal version of this paper.

Original *vs* Augmented, Prediction Error

In Table 1, we report the RMSE values of predictions with the original representation versus augmented representation. Following the results in Table 1 with SNPs + FS + VBM modalities, we can see that the prediction error of the augmented representation is decreased by 9.68% on average, when compared to the original representation.

Brain Imaging Identification

In addition to the regression analysis, we identify the AD relevant biomarkers by analyzing the weights of the learned projections $\mathbf{W}_i$ over each biomarker. Conceptually, because each feature of the augmented representation is an weighted summation over the biomarkers of the original representation, higher weights indicate that those biomarkers are emphasized by the proposed augmentation. In order to investigate the emphasized neuroimaging biomarkers, we plot the absolute values of the weights over the FS and VBM ROIs averaged across all participants. In Figure 1 we report the brain biomarkers identified by our method; some of these have been shown in the literature to be related to AD. For example, studies (Carmichael et al. 2007; Jack et al. 2004) have shown that ventricular volume, and its rate of change, is related to the vulnerability to cognitive decline and dementia. Specifically, Carmichael et al. have identified that larger ventricles in healthy participants increase the risk of a dementia-related disease in the future.

SNPs Identification

We also analyze the weights over the AlzGene groups of SNPs to identify the AD relevant groups of SNPs. The AlzGene groups have been constructed by the multiple genome-wide association studies listed on the AlzGene website (http://www.alzgene.org/). In Fig. 2, we identify the two AlzGene groups; encoding angiotensin I converting enzyme (ACE), and apolipoprotein E (APOE). These AlzGene groups have been shown to be related with cognitive decline. From previous AD studies (Miners et al. 2009), ACE genes are able to reduce the amyloid Beta peptide or Aβ in brain, and the accumulation of Aβ is commonly observed in the progress of AD (Chen et al. 2017). The amyloid hypothesis suggests a reasonable mechanism for how the accumulation of Aβ can result neuronal malfunction (Hardy and Selkoe 2002). On the other hand, APOE genes are strongly related to Aβ (Näslund et al. 1995) this aligned with previous findings that indicate that APOE isoforms are involved with Aβ aggregation and clearance (Kim, Basak, and Holtzman 2009). An experiment (Bales et al. 1997) reveals that the mice with APOE removed shows the decreased cerebral Aβ level in their descendants. The ϵ4 allele of the APOE gene, also identified by our approach, has been associated with early onset AD (Corder et al. 1993).

Conclusion

In this work, we propose an augmentation model to learn a fixed length vector summarizing the measurements of each participant from both static and dynamic modalities. We consider the inter and intra-modality group structures, tasks correlations between different cognitive measures, and consistency across the temporal records with the varying number of records. Our experiment shows that the proposed augmentation improves the accuracy of predictions on cognitive scores from the longitudinal data of different combination of modalities. We also observe that our approach successfully identifies AD-relevant biomarkers among the brain regions and SNPs, further verifying the correctness and utility of our approach.

Acknowledgments

This work was supported in part by the National Science Foundation (NSF) under the grants of IIS 1652943, IIS 1849359, CNS 1932482 and CCF 2029543.

We thank Dr. Heng Huang at the Department of Electrical and Computer Engineering, University of Pittsburgh and Dr. Li Shen at the Department of Biostatistics, Epidemiology and Informatics Perelman School of Medicine, University of Pennsylvania for providing the data used in all our experiments.

Modalities	Target Label	Representation	LASSO	SVR	RR	CNN
SNP + FS + VBM	RAVLT TOT	Original	0.1832	0.1756	0.1821	0.1705
		Augmented	**0.1732**	**0.1543**	**0.1328**	**0.1632**
	RAVLT 30	Original	0.2725	0.2525	0.2452	0.3440
		Augmented	**0.2609**	**0.2242**	**0.2204**	**0.3069**
	RAVLT RECOG	Original	0.2815	0.2562	0.2458	0.2856
		Augmented	**0.2560**	**0.2443**	**0.2223**	**0.2531**
	ADAS	Original	0.3258	0.3314	0.3056	0.3282
		Augmented	**0.3158**	**0.3150**	**0.2703**	**0.2660**
	FLU ANIM	Original	0.2568	0.2547	0.2568	0.2435
		Augmented	**0.2388**	**0.2369**	**0.2272**	**0.2097**
	FLU VEG	Original	0.2568	0.2258	0.2147	0.2058
		Augmented	**0.2310**	**0.2100**	**0.1832**	**0.1948**

Table 1: RMSE of predictions with augmented or original representation. The most accurate prediction is denoted as bold font.

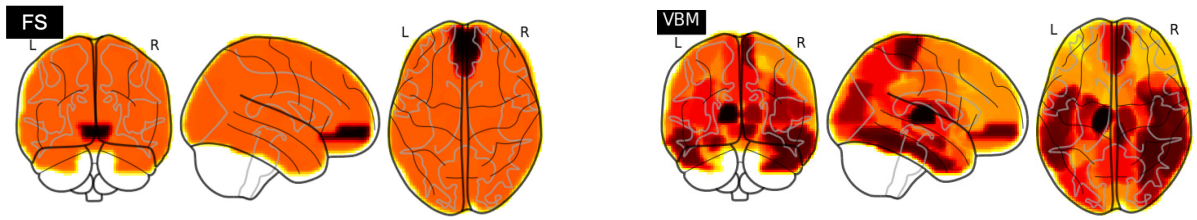

Figure 1: We plot the projections weights over the brain regions. The top five important regions are identified in **FS:** Right Caudate, Left Lateral Ventricle, Right Lateral Ventricle, Left Caudate, and Left Thalamus. **VBM:** Left Thalamus, Left Hippocampus, Right Occipital Mid, Left Amygdala, and Right Temporal Inf.

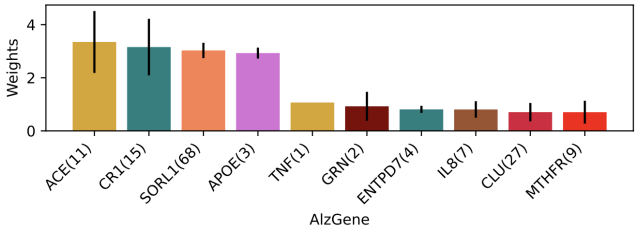

Figure 2: The averaged weights of SNPs on each AlzGene group. We denote the number of SNPs in each group as the number next to AlzGene group name, and the standard deviation as the line at the head of each bar.

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). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

Bales, K. R.; Verina, T.; Dodel, R. C.; Du, Y.; Altstiel, L.; Bender, M.; Hyslop, P.; Johnstone, E. M.; Little, S. P.; Cummins, D. J.; et al. 1997. Lack of apolipoprotein E dramatically reduces amyloid β-peptide deposition. *Nature genetics* 17(3): 263–264.

Bertsekas, D. P. 2014. *Constrained optimization and Lagrange multiplier methods*. Academic press.

Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J.; et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning* 3(1): 1–122.

Brand, L.; Nichols, K.; Wang, H.; Huang, H.; and Shen, L. 2020a. Predicting Longitudinal Outcomes of Alzheimer's Disease via a Tensor-Based Joint Classification and Regression Model. In *Pac Symp Biocomput*, 7–18. World Scientific.

Brand, L.; Nichols, K.; Wang, H.; Shen, L.; and Huang, H. 2019. Joint Multi-Modal Longitudinal Regression and Classification for Alzheimer's Disease Prediction. *IEEE Transactions on Medical Imaging* .

Brand, L.; O'Callaghan, B.; Sun, A.; and Wang, H. 2020b. Task Balanced Multimodal Feature Selection to Predict the Progression of Alzheimer's Disease. In *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, 196–203. IEEE.

Brand, L.; Wang, H.; Huang, H.; Risacher, S.; Saykin, A.; Shen, L.; et al. 2018. Joint high-order multi-task feature learning to predict the progression of alzheimer's disease. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 555–562. Springer.

Carmichael, O. T.; Kuller, L. H.; Lopez, O. L.; Thompson, P. M.; Dutton, R. A.; Lu, A.; Lee, S. E.; Lee, J. Y.; Aizenstein, H. J.; Meltzer, C. C.; et al. 2007. Ventricular volume and dementia progression in the Cardiovascular Health Study. *Neurobiology of aging* 28(3): 389–397.

Caruana, E. J.; Roman, M.; Hernández-Sánchez, J.; and Solli, P. 2015. Longitudinal studies. *Journal of thoracic disease* 7(11): E537.

Chen, G.-f.; Xu, T.-h.; Yan, Y.; Zhou, Y.-r.; Jiang, Y.; Melcher, K.; and Xu, H. E. 2017. Amyloid beta: structure, biology and structure-based therapeutic development. *Acta Pharmacologica Sinica* 38(9): 1205–1235.

Corder, E. H.; Saunders, A. M.; Strittmatter, W. J.; Schmechel, D. E.; Gaskell, P. C.; Small, G.; Roses, A.; Haines, J.; and Pericak-Vance, M. A. 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261(5123): 921–923.

Ding, C.; and He, X. 2004. K-means clustering via principal component analysis. In *Proceedings of the twenty-first international conference on Machine learning*, 29.

Ding, C. H.; Li, T.; and Jordan, M. I. 2008. Convex and semi-nonnegative matrix factorizations. *IEEE transactions on pattern analysis and machine intelligence* 32(1): 45–55.

Hardy, J.; and Selkoe, D. J. 2002. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *science* 297(5580): 353–356.

Jack, C.; Shiung, M.; Gunter, J.; O'brien, P.; Weigand, S.; Knopman, D. S.; Boeve, B. F.; Ivnik, R.; Smith, G.; Cha, R.; et al. 2004. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. *Neurology* 62(4): 591–600.

Jolliffe, I. 2011. Principal component analysis. In *International encyclopedia of statistical science*, 1094–1096. Springer.

Kim, J.; Basak, J. M.; and Holtzman, D. M. 2009. The role of apolipoprotein E in Alzheimer's disease. *Neuron* 63(3): 287–303.

Liu, K.; Brand, L.; Wang, H.; and Nie, F. 2019a. Learning robust distance metric with side information via ratio minimization of orthogonally constrained '21-norm distances. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*.

Liu, K.; Wang, H.; Han, F.; and Zhang, H. 2019b. Visual Place Recognition via Robust L2-Norm Distance Based Holism and Landmark Integration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 8034–8041.

Liu, Y.; Guo, Y.; Wang, H.; Nie, F.; and Huang, H. 2017. Semi-supervised classifications via elastic and robust embedding. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Lu, L.; Elbeleidy, S.; Baker, L.; Wang, H.; Huang, H.; Shen, L.; et al. 2019. Improved Prediction of Cognitive Outcomes via Globally Aligned Imaging Biomarker Enrichments over Progressions. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 140–148. Springer.

Lu, L.; Elbeleidy, S.; Baker, L. Z.; and Wang, H. 2020a. Learning Multi-Modal Biomarker Representations via Globally Aligned Longitudinal Enrichments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 817–824.

Lu, L.; Wang, H.; Elbeleidy, S.; and Nie, F. 2020b. Predicting cognitive declines using longitudinally enriched representations for imaging biomarkers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4827–4836.

Lu, L.; Wang, H.; Yao, X.; Risacher, S.; Saykin, A.; and Shen, L. 2018. Predicting progressions of cognitive outcomes via high-order multi-modal multi-task feature learning. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 545–548. IEEE.

Miners, S.; Ashby, E.; Baig, S.; Harrison, R.; Tayler, H.; Speedy, E.; Prince, J. A.; Love, S.; and Kehoe, P. G. 2009. Angiotensin-converting enzyme levels and activity in Alzheimer's disease: differences in brain and CSF ACE and association with ACE1 genotypes. *American journal of translational research* 1(2): 163.

Näslund, J.; Thyberg, J.; Tjernberg, L. O.; Wernstedt, C.; Karlström, A. R.; Bogdanovic, N.; Gandy, S. E.; Lannfelt, L.; Terenius, L.; and Nordstedt, C. 1995. Characterization of stable complexes involving apolipoprotein E and the amyloid β peptide in Alzheimer's disease brain. *Neuron* 15(1): 219–228.

Nie, F.; Wang, H.; Huang, H.; and Ding, C. H. 2013. Early Active Learning via Robust Representation and Structured Sparsity. In *IJCAI*, 1572–1578.

Petersen, R. C.; Aisen, P.; Beckett, L. A.; Donohue, M.; Gamst, A.; Harvey, D. J.; Jack, C.; Jagust, W.; Shaw, L.; Toga, A.; et al. 2010. Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. *Neurology* 74(3): 201–209.

Risacher, S. L.; Shen, L.; West, J. D.; Kim, S.; McDonald, B. C.; Beckett, L. A.; Harvey, D. J.; Jack Jr, C. R.; Weiner, M. W.; Saykin, A. J.; et al. 2010. Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. *Neurobiology of aging* 31(8): 1401–1418.

Shen, L.; Kim, S.; Risacher, S. L.; Nho, K.; Swaminathan, S.; West, J. D.; Foroud, T.; Pankratz, N.; Moore, J. H.; Sloan, C. D.; et al. 2010. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage* 53(3): 1051–1063.

Wang, H.; Nie, F.; and Huang, H. 2012. Robust and discriminative distance for multi-instance learning. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2919–2924. IEEE.

Wang, H.; Nie, F.; and Huang, H. 2013. Multi-view clustering and feature learning via structured sparsity. In *International conference on machine learning*, 352–360.

Wang, H.; Nie, F.; and Huang, H. 2015. Large-scale cross-language web page classification via dual knowledge transfer using fast nonnegative matrix trifactorization. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 10(1): 1–29.

Wang, H.; Nie, F.; Huang, H.; and Ding, C. 2011. Nonnegative matrix tri-factorization based high-order co-clustering and its fast implementation. In *2011 IEEE 11th international conference on data mining*, 774–783. IEEE.

Wang, H.; Nie, F.; Huang, H.; and Ding, C. 2013. Heterogeneous visual features fusion via sparse multimodal machine. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3097–3102.

Wang, H.; Nie, F.; Huang, H.; Kim, S.; Nho, K.; Risacher, S. L.; Saykin, A. J.; Shen, L.; and Initiative, A. D. N. 2012a. Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. *Bioinformatics* 28(2): 229–237.

Wang, H.; Nie, F.; Huang, H.; Risacher, S. L.; Saykin, A. J.; Shen, L.; and Initiative, A. D. N. 2012b. Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. *Bioinformatics* 28(12): i127–i136.

Wang, H.; Nie, F.; Huang, H.; Yan, J.; Kim, S.; Nho, K.; Risacher, S. L.; Saykin, A. J.; Shen, L.; and Initiative, A. D. N. 2012c. From phenotype to genotype: an association study of longitudinal phenotypic markers to Alzheimer's disease relevant SNPs. *Bioinformatics* 28(18): i619–i625.

Wang, H.; Nie, F.; Huang, H.; Yan, J.; Kim, S.; Risacher, S.; Saykin, A.; and Shen, L. 2012d. High-order multi-task feature learning to identify longitudinal phenotypic markers for alzheimer's disease progression prediction. In *Advances in neural information processing systems*, 1277–1285.

World Health Organization. 2017. Global action plan on the public health response to dementia 2017–2025 .

Wu, M. C.; Kraft, P.; Epstein, M. P.; Taylor, D. M.; Chanock, S. J.; Hunter, D. J.; and Lin, X. 2010. Powerful SNP-set analysis for case-control genome-wide association studies. *The American Journal of Human Genetics* 86(6): 929–942.

Yan, J.; Li, T.; Wang, H.; Huang, H.; Wan, J.; Nho, K.; Kim, S.; Risacher, S. L.; Saykin, A. J.; Shen, L.; et al. 2015. Cortical surface biomarkers for predicting cognitive outcomes using group L2,1-norm. *Neurobiology of aging* 36: S185–S193.

Yang, H.; Liu, K.; Wang, H.; and Nie, F. 2019. Learning Strictly Orthogonal p-Order Nonnegative Laplacian Embedding via Smoothed Iterative Reweighted Method. In *IJCAI*, volume 2019, 4040–4046.