Learning Semi-Supervised Representation Enrichment Using Longitudinal Imaging-Genetic Data

Hoon Seo
Department of Computer Science
Colorado School of Mines
Golden, CO, USA
seohoon@mymail.mines.edu

Lodewijk Brand
Department of Computer Science
Colorado School of Mines
Golden, CO, USA
lbrand@mymail.mines.edu

Hua Wang
Department of Computer Science
Colorado School of Mines
Golden, CO, USA
huawangcs@gmail.com

for the Alzheimer's Disease Neuroimaging Initiative

***Abstract*—Alzheimer's Disease (AD) is a progressive memory disorder that causes irreversible cognitive decline. Recently, many statistical learning methods have been presented to predict cognitive declines by using longitudinal imaging data. However, missing records that broadly exist in the longitudinal neuroimaging data have posed a critical challenge for effectively using these data in machine learning models. To tackle this difficulty, in this paper we propose a novel approach to integrate longitudinal (dynamic) phenotypic data and static genetic data to learn a fixed-length biomarker representation using the enrichment learned from the temporal data in multiple imaging modalities. Armed with this enriched biomarker representation, as a fixed-length vector per participant, conventional machine learning models can be used to predict clinical outcomes associated with AD. We have applied our new method on the Alzheimer's Disease Neruoimaging Initiative (ADNI) cohort and achieved promising experimental results that validate its effectiveness.**

***Index Terms*—Alzheimer's Disease; Multi-Modal; Longitudinal Imaging Data; Enrichment**

I. Introduction

Alzheimer's disease (AD) is a neurodegernative condition in which people suffer from the progressive deterioration of cognitive functions such as memory, language, and judgment. To address this major public health crisis, it is critical to identify signs of AD at an early stage from both the therapeutic and research standpoints. Recent works [1]–[3] have analyzed the progression of AD through the modeling and prediction of clinical assessments. Furthermore, in the last decade [4], rich neuroimaging measurements, such as magnetic resonance imaging (MRI) scans, have been widely used to predict the clinical outcomes associated with AD.

Corresponding author: Hua Wang (huawangcs@gmail.com). 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.

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 full list of funding sources for ADNI is provided in the document "Alzheimers Disease Neuroimaging Initiative (ADNI) Data Sharing and Publication Policy" available through adni.loni.usc.edu/.

Incomplete temporal neuroimaging records pose a challenge to effectively use longitudinal multi-modal imaging data. To tackle this difficulty, we propose a semi-supervised learning method to learn a participant-specific projection to enrich the multi-modal phenotypic and genotypic measurements. We uncover the temporal relationships across the imaging biomarkers by learning a projection for each participant from the available neuroimaging records. Since the number of the records in a studied cohort varies between participants, we introduce the trace norm regularizations over the concatenation of all projections matrices. Furthermore, our model factorizes the enriched biomarker representations, available cognitive scores, and genetic biomarkers of the participants, where a latent participant representation shared across the different modalities can be learned. Finally, a structured sparsity-induced norm regularization is applied to the genotypic biomarker data. The purpose of this group-wise regularization is to strengthen the weights of the learned projections related with the subgroups of SNPs that are more predictive of a given clinical outcome. Previous studies [5], [6] have identified that groups of SNPs operate together. Provided with the learned projections per-participant, we can transform the multi-modal phenotypic and genetic biomarker representations with varied lengths into an enriched representation in the format of a fixed length vector. Using this learned fixed-length vector, one can easily make use of conventional machine learning models to predict clinical scores associated with AD.

II. Our Method

A. Notations and problem formulation

Throughout this paper, we write matrices as bold uppercase letters and vectors as bold lowercase letters. For a matrix $\mathbf{M} = [m_j^i]$, 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 group ℓ_2-norm [5], [6] is defined as $\|\mathbf{M}\|_{G_2} = \sum_{m=1}^{k}\|\mathbf{M}^m\|_2$, where $\mathbf{M} = [\mathbf{M}^1; \mathbf{M}^2; \cdots; \mathbf{M}^k]$ consists of k blocks. The

978-1-7281-6215-7/20/$31.00 ©2020 IEEE

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 dataset, phenotypes are usually described by the biomarkers extracted from brain scans that vary over time. 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 (first time point) to the second last visit, where $n_i + 1$ denotes the total number of the medical records of the i-th participant. 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 last medical record of the i-th participant and use $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$ to summarize these recordss of all the participants. Because multiple types of biomarkers can be extracted from the brain scans, such as the voxel-based morphometry (VBM) markers and FreeSurfer (FS) markers extracted from the MRI images. We thereby concatenate the vector representations of these biomarkers as the phenotypic assessment of a participant. For example, 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 $1 \leq i \leq n,\ 1 \leq j \leq n_i$.

In addition to the phenotypic measurements of the participants of a neuroimaging data set, genotypes of the same cohort may also be available, such as the single nucleotide polymorphisms (SNPs) that can be represented by $\mathbf{X}_{SNP} = [\mathbf{x}_1^{SNP}, \ldots, \mathbf{x}_n^{SNP}]$, 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 constant that does not vary over time.

Besides the input phenotypic and genotypic data, the outputs of the task of predicting clinical outcomes are cognitive status of the participants, which are usually assessed by the clinical scores of a set of cognitive tests. We use $\mathbf{Y}_l \in \Re^{c \times l}$ to list the clinical scores of the first l participants at her or his last visit, where c is the number of clinical scores. Here, without losing generality we consider the first l samples as the labeled data for training. Then 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 $\mathcal{W}^T \otimes \mathbf{X} = [\mathbf{W}_1^T \mathbf{x}_1, \mathbf{W}_2^T \mathbf{x}_2, \cdots, \mathbf{W}_n^T \mathbf{x}_n] \in \Re^{r_1 \times n}$ by projecting the medical records of $\mathbf{X}$ for all the participants.

B. Our Objective

In this section we describe our objective to learn the projections $f : \Re^d \mapsto \Re^{r_1}$, which can be formulated as a linear projection by computing $\mathbf{z}_i = \mathbf{W}_i^T \mathbf{x}_i$.

First, to eliminate the redundant information in the biomarker measurements over time [7], we choose to use the Principal Component Analysis (PCA) to preserve as much information of every participant as possible that minimize the following objective:

$$\mathcal{J}(\mathcal{W}) = \sum_{i=1}^{n} \left\|\mathbf{X}_i - \mathbf{W}_i \mathbf{W}_i^T \mathbf{X}_i\right\|_2, \ s.t. \quad \mathbf{W}_i^T \mathbf{W}_i = \mathbf{I}. \quad (1)$$

Second, to capture the longitudinal patterns when AD develops, we consider the following two types of tasks correlations [8], [9]. First, for an individual cognitive measure, although its association to the imaging features at different stages of the disease could be different, its associations patterns at two consecutive time points tend to be similar. Second, we know that during the AD progression, different cognitive measures are interrelated to each other. Mathematically, we can describe these two types of correlations by minimizing the ranks of the coefficient matrices unfolded from the coefficient tensor along different modes as following [9]:

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

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 unfolded matrices of the projection tensor $\mathcal{W}$ along the first and second modes, respectively.

Third, given that cognitive status of the first l participants $\mathbf{Y}_l$ are available as the training data, our model aims to learn the relationship between the output clinical scores and input biomarker measurements. We introduce the variable of $\mathbf{F} = [\mathbf{F}_l, \mathbf{F}_u] \in \Re^{c \times n}$ as our estimated predictive outcomes and set the constraint of $\mathbf{F}_l = \mathbf{Y}_l$ for the training data. We factorize $\mathbf{F}$ into two sets of latent factors, *i.e.*, $\mathbf{F} \approx \mathbf{U}^T \mathbf{G}_1$, where $\mathbf{G}_1 \in \Re^{r_2 \times n}$ is the representations of the imaging biomarkers of the participants in the studied cohort in the latent subspace of $\mathbf{U}^T \in \Re^{c \times r_2}$. Likewise, we factorize the enriched biomarker representations $\mathcal{W}^T \otimes \mathbf{X} \approx \mathbf{H}_1 \mathbf{G}_1$, where $\mathbf{G}_1 \in \Re^{r_2 \times n}$ again is vector representation of the participants in the latent subspace of $\mathbf{H}_1 \in \Re^{r_1 \times r_2}$. Here the operation of $\mathcal{W}^T \otimes \mathbf{X}$ indeed enriches the baseline biomarker representations using the longitudinal phenotypic measurements over time [7], [10], [11]. By sharing the data representations by the same factor matrix $\mathbf{G}_1$ in the above two factorizations, the longitudinally enriched biomarker representations for each participant are connected with the clinical scores. To summarize, we develop Eq. (2) as follows:

$$\begin{aligned} \mathcal{J}(\mathbf{U}, \mathbf{F}, \mathbf{H}_1, \mathbf{G}_1, \mathcal{W}) \ = &\sum_{i=1}^{n} \left\|\mathbf{X}_i - \mathbf{W}_i \mathbf{W}_i^T \mathbf{X}_i\right\|_2 \\ &+ \left(\left\|\mathbf{W}_{(1)}\right\|_* + \left\|\mathbf{W}_{(2)}\right\|_*\right) + \left\|\mathbf{F} - \mathbf{U}^T \mathbf{G}_1\right\|_2 + \|\mathbf{U}\|_1 \\ &+ \left\|\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1 \mathbf{G}_1\right\|_2, \\ s.t. \quad &\mathbf{W}_i^T \mathbf{W}_i = \mathbf{I}, \ \mathbf{F}_l = \mathbf{Y}_l, \end{aligned} \quad (3)$$

where $\|\mathbf{U}\|_1$ is the regularization term to avoid overfitting our learning model.

Finally, apart from making use of the information conveyed by the longitudinal phenotypic measurements of the imaging biomarkers and the cognitive status of the participants, we can also take advantage of the genetic information encoded by the SNP profiles. Similarly, we factorize the SNPs data of all the participants $\mathbf{X}_{SNP} = \mathbf{H}_2 \mathbf{G}_2$, where $\mathbf{H}_2 \in \Re^{d_{SNP} \times r_2}$ and $\mathbf{G}_2 \in \Re^{r_2 \times n}$. To ensure the consistency between the

learned phenotypic and genotypic representations, we use a soft constraint $\|\mathbf{G}_1 - \mathbf{G}_2\|_2$ and develop our objective as follows:

$$\begin{aligned}
&\mathcal{J}(\mathbf{U}, \mathbf{F}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{G}_1, \mathbf{G}_2, \mathcal{W}) = \\
&\left\|\mathbf{F} - \mathbf{U}^T\mathbf{G}_1\right\|_2 + \gamma_1 \sum_{i=1}^{n} \left\|\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i\right\|_2 \\
&+ \gamma_2 \left\|\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1\mathbf{G}_1\right\|_2 + \gamma_3 \left\|\mathbf{X}_{\text{SNP}} - \mathbf{H}_2\mathbf{G}_2\right\|_2 \qquad (4) \\
&+ \gamma_4 \left\|\mathbf{G}_1 - \mathbf{G}_2\right\|_2 + \gamma_5 \left\|\mathbf{H}_2\right\|_{G_2} + \gamma_6 \left(\left\|\mathbf{W}_{(1)}\right\|_* + \left\|\mathbf{W}_{(2)}\right\|_*\right) \\
&+ \gamma_7 \left\|\mathbf{U}\right\|_1, \quad s.t. \quad \mathbf{F}_l = \mathbf{Y}_l,\ \mathbf{W}_i^T\mathbf{W}_i = \mathbf{I},
\end{aligned}$$

where $\gamma_1, \gamma_2, \cdots, \gamma_7$ are hyperparameters to adjust the impact of each term. In Eq. (4), we use $\|\mathbf{H}_2\|_{G_2}$ to leverage the group structures of the SNPs data [5], [6].

C. The Solution Algorithm

While our objective in Eq. (4) is clearly motivated, it is difficult to solve in general, because it is a non-smooth optimization objective [12]. Thus we drive an efficient algorithm to solve our objective. Using the optimization framework in the earlier works [13], [14] that proposed the iterative reweighted method to solve non-smooth objectives, we can solve Eq. (4) by an iterative procedure [13, Algorithm 1] in which the key step is to minimize the following objective:

$$\begin{aligned}
&\mathcal{J}^R(\mathbf{U}, \mathbf{F}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{G}_1, \mathbf{G}_2, \mathcal{W}) = \\
&\operatorname{tr}((\mathbf{U}^T\mathbf{G}_1 - \mathbf{F})^T\mathbf{D}_1(\mathbf{U}^T\mathbf{G}_1 - \mathbf{F})) \\
&+ \gamma_1 \sum_{i=1}^{n} \operatorname{tr}((\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i)^T\mathbf{D}_{2,i}(\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i)) \\
&+ \gamma_2 \operatorname{tr}((\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1\mathbf{G}_1)^T\mathbf{D}_3(\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1\mathbf{G}_1)) \\
&+ \gamma_3 \operatorname{tr}((\mathbf{X}_{SNP} - \mathbf{H}_2\mathbf{G}_2)^T\mathbf{D}_4(\mathbf{X}_{SNP} - \mathbf{H}_2\mathbf{G}_2)) \\
&+ \gamma_4 \operatorname{tr}((\mathbf{G}_1 - \mathbf{G}_2)^T\mathbf{D}_5(\mathbf{G}_1 - \mathbf{G}_2)) + \gamma_5 \operatorname{tr}(\mathbf{H}_2^T\mathbf{D}_6\mathbf{H}_2) \\
&+ \gamma_6 \operatorname{tr}(\mathbf{W}_{(1)}^T\mathbf{D}_7\mathbf{W}_{(1)}) + \gamma_6 \operatorname{tr}(\mathbf{W}_{(2)}^T\mathbf{D}_8\mathbf{W}_{(2)}) \\
&+ \gamma_7 \sum_{p=1}^{c}((\mathbf{u}_p)^T\mathbf{D}_{9,p}\mathbf{u}_p), \quad s.t. \quad \mathbf{F}_l = \mathbf{Y}_l,\ \mathbf{W}_i^T\mathbf{W}_i = \mathbf{I},
\end{aligned} \qquad (5)$$

where $[\mathbf{D}_1]_j^j, [\mathbf{D}_{2,i}]_j^j, [\mathbf{D}_3]_j^j, [\mathbf{D}_{9,p}]_j^j, [\mathbf{D}_4]_j^j, [\mathbf{D}_5]_j^j$ are defined in the following diagonal matrices respectively:

$$\begin{aligned}
&\frac{1}{2\sqrt{\|[\mathbf{U}^T\mathbf{G}_1 - \mathbf{F}]^j\|_2^2 + \delta}}, \ \frac{1}{2\sqrt{\left\|[\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i]^j\right\|_2^2 + \delta}} \\
&\frac{1}{2\sqrt{\|[\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1\mathbf{G}_1]^j\|_2^2 + \delta}}, \ \frac{1}{2\sqrt{(u_p^j)^2 + \delta}}, \\
&\frac{1}{2\sqrt{\|[\mathbf{X}_{SNP} - \mathbf{H}_2\mathbf{G}_2]^j\|_2^2 + \delta}}, \ \frac{1}{2\sqrt{\|[\mathbf{G}_1 - \mathbf{G}_2]^j\|_2^2 + \delta}}, \\
&\mathbf{D}_7 = \frac{1}{2}(\mathbf{W}_{(1)}\mathbf{W}_{(1)}^T + \delta\mathbf{I})^{-\frac{1}{2}}, \ \mathbf{D}_8 = \frac{1}{2}(\mathbf{W}_{(2)}\mathbf{W}_{(2)}^T + \delta\mathbf{I})^{-\frac{1}{2}}.
\end{aligned} \qquad (6)$$

$\mathbf{D}_6$ is a block diagonal matrix, where j-th block is $\frac{1}{2\sqrt{\|\mathbf{H}_2^j\|_F^2 + \delta}}\mathbf{I}_j$. $\mathbf{I}_j \in \Re^{d_j \times d_j}$ is an identity matrix, and d_j denotes the number of rows of j-th block of $\mathbf{H}_2 = [\mathbf{H}_2^1; \mathbf{H}_2^2; \cdots; \mathbf{H}_2^k]$, so that $\sum_{j=1}^{k} d_j = d_{SNP}$. The dimensions of the matrices in Eq. (6) are: $\mathbf{D}_1 \in \Re^{c \times c}$, $\mathbf{D}_{2,i} \in \Re^{d \times d}$, $\mathbf{D}_3 \in \Re^{r_1 \times r_1}$, $\mathbf{D}_4 \in \Re^{d_{SNP} \times d_{SNP}}$, $\mathbf{D}_5 \in \Re^{r_2 \times r_2}$, $\mathbf{D}_6 \in \Re^{d_{SNP} \times d_{SNP}}$, $\mathbf{D}_7 \in \Re^{d \times d}$, $\mathbf{D}_8 \in \Re^{r_1 \times r_1}$, $\mathbf{D}_{9,p} \in \Re^{r_2 \times r_2}$.

To minimize the smoothed objective Eq. (5), we use the Alternating Direction Method of Multipliers (ADMM) proposed by [15]. By introducing two more constraints $\mathbf{A} = \mathbf{U}$ and $\mathbf{B}_{(2)} = \mathbf{W}_{(2)} \Leftrightarrow \mathbf{B}_i = \mathbf{W}_i$ $(i = 1, 2, \cdots, n)$ to decouple the $\mathbf{U}$ and $\mathcal{W}$, the smoothed objective in Eq. (5) can be minimized by Algorithm 1. Due to space limit, the detailed derivations of Algorithm 1 will be provided in the extended journal version of this paper.

Algorithm 1: Solve minimization problem in Eq. (5)

Input: $\mathbf{X}$, $\mathbf{X}_i$, $\mathbf{Y}_l$ $(1 \le i \le n)$
Initialization:
$\mathbf{U}, \mathbf{F}, \mathcal{W}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{G}_1, \mathbf{G}_2, \mathbf{A}, \mathbf{B}_{(2)}, \mathbf{\Lambda}_1, \mathbf{\Lambda}_{2,i}, \mathbf{\Lambda}_3, \mathbf{\Lambda}_{4,i}$,
$1 < \rho_1, \rho_{2,i}, \rho_3, \rho_{4,i} < 2$,
$\mu_1, \mu_{2,i}, \mu_3, \mu_{4,i}, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7 > 0$;
while *not converge* **do**
1. Update $\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_9$ defined in Eq. (6);
2. Update $\mathbf{H}_1$ by $\mathbf{H}_1 = (\mathcal{W}^T \otimes \mathbf{X})\mathbf{G}_1^T(\mathbf{G}_1\mathbf{G}_1^T)^{-1}$;
3. Update $\mathbf{H}_2$ by
$\mathbf{H}_2 = sylvester(\frac{\gamma_5}{\gamma_3}\mathbf{D}_4^{-1}\mathbf{D}_6, \mathbf{G}_2\mathbf{G}_2^T, \mathbf{X}_{SNP}\mathbf{G}_2^T)$;
4. Update $\mathbf{G}_1$ by
$\mathbf{G}_1 = (\frac{1}{2}(\mathbf{A}\mathbf{D}_1\mathbf{U}^T + \mathbf{U}\mathbf{D}_1\mathbf{A}^T) + \gamma_2\mathbf{H}_1^T\mathbf{D}_3\mathbf{H}_1 +$
$\gamma_4\mathbf{D}_5)^{-1}(\frac{1}{2}(\mathbf{A}\mathbf{D}_1 + \mathbf{U}\mathbf{D}_1)\mathbf{F} + \gamma_4\mathbf{D}_5\mathbf{G}_2 + \gamma_2\mathbf{H}_1^T\mathbf{D}_3(\mathcal{W}^T \otimes \mathbf{X}))$;
5. Update $\mathbf{G}_2$ by
$\mathbf{G}_2 = (\gamma_3\mathbf{H}_2^T\mathbf{D}_4\mathbf{H}_2 + \gamma_4\mathbf{D}_5)^{-1}(\gamma_3\mathbf{H}_2^T\mathbf{D}_4\mathbf{X}_{SNP} + \gamma_4\mathbf{D}_5\mathbf{G}_1)$;
6. Update $\mathbf{B}_i$ $(1 \le i \le n)$ by
$\mathbf{B}_i = (-\gamma_1(\mathbf{D}_{2,i}(\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i)\mathbf{X}_i^T + \mathbf{X}_i(\mathbf{X}_i -$
$\mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i)^T\mathbf{D}_{2,i}) + \mu_{2,i}\mathbf{W}_i\mathbf{W}_i^T + \mu_{4,i}\mathbf{I})^{-1}(-\gamma_6\mathbf{W}_i\mathbf{D}_8 +$
$\mu_{4,i}\mathbf{W}_i - \mathbf{\Lambda}_{4,i} + \mu_{2,i}\mathbf{W}_i - \mathbf{W}_i\mathbf{\Lambda}_{2,i})$;
7. Update $[\mathbf{A}]_p$ $(1 \le p \le c)$ by $[\mathbf{A}]_p =$
$(2\gamma_7\mathbf{D}_{9,p} + \mu_3\mathbf{I})^{-1}[-\mathbf{G}_1(\mathbf{G}_1^T\mathbf{U} - \mathbf{F}^T)\mathbf{D}_1 + \mu_3\mathbf{U} - \mathbf{\Lambda}_3]_p$;
8. Update $[\mathbf{F}]_p$ $(1 \le p \le l)$ by $[\mathbf{F}]_p =$
$(2\mathbf{D}_1 + \mu_1\mathbf{I})^{-1}([\mathbf{D}_1(\mathbf{U}^T + \mathbf{A}^T)\mathbf{G}_1]_p + \mu_1[\mathbf{Y}_l]_p - [\mathbf{\Lambda}_1]_p)$;
9. Update $[\mathbf{F}]_p$ $(l+1 \le p \le n)$ by $[\mathbf{F}]_p = \frac{1}{2}[(\mathbf{U}^T + \mathbf{A}^T)\mathbf{G}_1]_p$;
10. Update $\mathbf{U}$ by $\mathbf{U} = -\frac{1}{\mu_3}\mathbf{G}_1(\mathbf{G}_1^T\mathbf{A} - \mathbf{F}^T)\mathbf{D}_1 + \mathbf{A} + \frac{1}{\mu_3}\mathbf{\Lambda}_3$;
11. Update $[\mathbf{W}_i]_p$ $(1 \le i \le n)$ and $(1 \le p \le r_1)$ by
$[\mathbf{W}_i]_p = (-\gamma_1(\mathbf{X}_i\mathbf{X}_i^T\mathbf{D}_{2,i} - \mathbf{X}_i\mathbf{X}_i^T\mathbf{B}_i\mathbf{B}_i^T\mathbf{D}_{2,i} +$
$\mathbf{D}_{2,i}\mathbf{X}_i\mathbf{X}_i^T - \mathbf{D}_{2,i}\mathbf{B}_i\mathbf{B}_i^T\mathbf{X}_i\mathbf{X}_i^T) + 2\gamma_6\mathbf{D}_7 + \mu_{2,i}\mathbf{B}_i\mathbf{B}_i^T +$
$\mu_{4,i}\mathbf{I} + 2\gamma_2[\mathbf{D}_3]_p^p[\mathbf{X}]_i[\mathbf{X}^T]^i)^{-1}(\mathbf{B}_i(-\gamma_6[\mathbf{D}_8]_p + (\mu_{2,i} +$
$\mu_{4,i})[\mathbf{I}]_p - [\mathbf{\Lambda}_{2,i}^T]_p) + 2\gamma_2[\mathbf{D}_3]_p^p[\mathbf{H}_1\mathbf{G}_1]_i^p[\mathbf{X}]_i + [\mathbf{\Lambda}_{4,i}]_p)$;
12. Update $\mathbf{\Lambda}_1$ by $\mathbf{\Lambda}_1 = \mathbf{\Lambda}_1 + \mu_1(\mathbf{F}_l - \mathbf{Y}_l)$;
13. Update $\mathbf{\Lambda}_{2,i}$ $(1 \le i \le n)$ by $\mathbf{\Lambda}_{2,i} = \mathbf{\Lambda}_{2,i} + \mu_{2,i}(\mathbf{W}_i^T\mathbf{B}_i - \mathbf{I})$;
14. Update $\mathbf{\Lambda}_3$ by $\mathbf{\Lambda}_3 = \mathbf{\Lambda}_3 + \mu_3(\mathbf{A} - \mathbf{U})$;
15. Update $\mathbf{\Lambda}_{4,i}$ $(1 \le i \le n)$ by $\mathbf{\Lambda}_{4,i} = \mathbf{\Lambda}_{4,i} + \mu_{4,i}(\mathbf{B}_i - \mathbf{W}_i)$;
16. Update $\mu_1, \mu_{2,i}, \mu_3, \mu_{4,i}$ $(1 \le i \le n)$ by
$\mu_1 = \rho_1\mu_1;\ \mu_{2,i} = \rho_{2,i}\mu_{2,i};\ \mu_3 = \rho_3\mu_3;\ \mu_{4,i} = \rho_{4,i}\mu_{4,i}$;
end
Output: $\mathbf{W}_i$ $(1 \le i \le n)$

In Algorithm 1, $\mathbf{\Lambda}_1, \mathbf{\Lambda}_{2,i}, \mathbf{\Lambda}_3, \mathbf{\Lambda}_{4,i}$ are the Lagrangian multipliers for the constraints $\mathbf{F}_l = \mathbf{Y}_l$, $\mathbf{W}_i^T\mathbf{W}_i = \mathbf{I}$, $\mathbf{A} = \mathbf{U}$, and $\mathbf{B}_i = \mathbf{W}_i$ $(1 \le i \le n)$. In step 3 of Algorithm 1, we use solution of Sylvester equation, such that $sylvester(\mathbf{P}, \mathbf{Q}, \mathbf{R})$ gives an unique and exact solution for $\mathbf{X}$ of equation $\mathbf{PX} + \mathbf{XQ} = \mathbf{R}$ if it exists. The time complexity of Algorithm 1 is $O(nr_1d^2(d + r_1))$ for each iteration where the step 11 of Algorithm 1 is the most dominant.

III. Experiments

Experimental data. We experiment with the data obtained from the ADNI database (http://www.adni.loni.usc.edu). We

TABLE I
RMSE OF PREDICTIONS ON CLINICAL SCORES. THE SMALLER RMSE IS DENOTED AS BOLD FONT.

Target class	Representation	SVR	RR	CNN
RAVLT TOT	Original	0.2441	0.2110	0.2855
	Enriched	**0.1816**	**0.1780**	**0.1804**
RAVLT 30	Original	0.3270	0.2856	0.3440
	Enriched	**0.2742**	**0.2648**	**0.2696**
RAVLT RECOG	Original	0.3657	0.2897	0.3147
	Enriched	**0.2449**	**0.2460**	**0.2592**

downloaded the MRI scans, SNP genotypes, and the longitudinal scores of Rey's Auditory Verbal Learning Test (RAVLT) of 821 ADNI-1 participants. We perform VBM and FreeSurfer automated parcellation on the MRI data as described by [16] and extract mean modulated gray matter (GM) measures for 90 target regions of interest (ROI). We follow SNP quality control steps discussed in [17]. Among 821 ADNI-1 participants, 412 participants are selected on the basis of existence of MRI records at Month 0 (baseline)/Month 6/Month 12/Month 24. We discard Month 24 scans with 50% probability to evaluate the learning capability of our model from longitudinal data with missing records.

Experimental settings. In our experiments we aim to predict RAVLT clinical scores in the test set using two types of the inputs — the learned enriched representation and original representation at the baseline time point. We use the different concatenations of SNPs, FS, and VBM modalities to assess the prediction performance of our model with diverse modalities. We split the dataset into a training and test set with a proportion of 80% and 20% each, therefore the number of participants is $l = 323$ in the training set and $n - l = 89$ in the test set. The SNPs and MRI images of all n participants and clinical scores of only the l participants in training set are provided for our model to learn enriched representation. To predict the $n-l$ clinical scores in test set, we use the following conventional prediction models: Ridge linear Regression (RR), Convolutional Neural Network (CNN), and Support Vector Regression (SVR) which is the regression version of Support Vector Machine. We conduct a 5-fold cross-validation to search the set of best hyperparameters for each conventional model, and to control the effects of hyperparameters.

Experimental results. In the experimental result reported in Table I, we compute the Root Mean Squared Error (RMSE) between the ground truth clinical scores and predicted clinical scores from both the original and enriched representations. The results show that the prediction performance by the enriched representations are consistently more accurate than the predictions using the original baseline representation, which firmly demonstrate the effectiveness of our new method.

IV. CONCLUSION

Missing data is one of the most challenging issue on using longitudinal multi-modal healthcare datasets. This research aims to devise a novel method to learn a fixed-length representation for all the participants in the ADNI dataset. The learned biomarker representation summarizes the genetic biomarkers and their group structure, known clinical scores, and all the available measurements of longitudinal biomarkers on a per-participant basis. Our experiments show that the learned enriched representation outperforms the baseline measurement in predicting the clinical scores.

REFERENCES

[1] C. M. Karch and A. M. Goate, "Alzheimer's disease risk genes and mechanisms of disease pathogenesis," *Biological psychiatry*, vol. 77, no. 1, pp. 43–51, 2015.

[2] L. Brand, H. Wang, H. Huang, S. Risacher, A. Saykin, L. Shen *et al.*, "Joint high-order multi-task feature learning to predict the progression of alzheimer's disease," in *MICCAI*, 2018, pp. 555–562.

[3] L. Brand, K. Nichols, H. Wang, L. Shen, and H. Huang, "Joint multi-modal longitudinal regression and classification for alzheimer's disease prediction," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1845–1855, 2019.

[4] C. M. Stonnington, C. Chu, S. Klöppel, C. R. Jack Jr, J. Ashburner, R. S. Frackowiak, A. D. N. Initiative *et al.*, "Predicting clinical scores from magnetic resonance scans in alzheimer's disease," *Neuroimage*, vol. 51, no. 4, pp. 1405–1413, 2010.

[5] H. Wang, F. Nie, H. Huang, S. Kim, K. Nho, S. L. Risacher, A. J. Saykin, L. Shen, and A. D. N. Initiative, "Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the adni cohort," *Bioinformatics*, vol. 28, no. 2, pp. 229–237, 2012.

[6] J. Yan, T. Li, H. Wang, H. Huang, J. Wan, K. Nho, S. Kim, S. L. Risacher, A. J. Saykin, L. Shen *et al.*, "Cortical surface biomarkers for predicting cognitive outcomes using group $\ell_{2,1}$-norm," *Neurobiology of aging*, vol. 36, pp. S185–S193, 2015.

[7] L. Lu, S. Elbeleidy, L. Baker, H. Wang, H. Huang, L. Shen *et al.*, "Improved prediction of cognitive outcomes via globally aligned imaging biomarker enrichments over progressions," in *MICCAI*, 2019, pp. 140–148.

[8] H. Wang, F. Nie, H. Huang, J. Yan, S. Kim, K. Nho, S. L. Risacher, A. J. Saykin, L. Shen, and A. D. N. Initiative, "From phenotype to genotype: an association study of longitudinal phenotypic markers to alzheimer's disease relevant snps," *Bioinformatics*, vol. 28, no. 18, pp. i619–i625, 2012.

[9] H. Wang, F. Nie, H. Huang, J. Yan, S. Kim, S. Risacher, A. Saykin, and L. Shen, "High-order multi-task feature learning to identify longitudinal phenotypic markers for alzheimer's disease progression prediction," in *NIPS*, 2012, pp. 1277–1285.

[10] L. Lu, S. Elbeleidy, L. Z. Baker, and H. Wang, "Learning multi-modal biomarker representations via globally aligned longitudinal enrichments," in *AAAI*, vol. 34, no. 01, 2020, pp. 817–824.

[11] L. Lu, H. Wang, S. Elbeleidy, and F. Nie, "Predicting cognitive declines using longitudinally enriched representations for imaging biomarkers," in *IEEE/CVF CVPR*, 2020, pp. 4827–4836.

[12] F. Nie, H. Huang, C. Ding, D. Luo, and H. Wang, "Robust principal component analysis with non-greedy l1-norm maximization," in *IJCAI*, vol. 22, no. 1, 2011, p. 1433.

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

[14] H. Yang, K. Liu, H. Wang, and F. Nie, "Learning strictly orthogonal *p*-order nonnegative laplacian embedding via smoothed iterative reweighted method." in *IJCAI*, 2019, pp. 4040–4046.

[15] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[16] S. L. Risacher, L. Shen, J. D. West, S. Kim, B. C. McDonald, L. A. Beckett, D. J. Harvey, C. R. Jack Jr, M. W. Weiner, A. J. Saykin *et al.*, "Longitudinal mri atrophy biomarkers: relationship to conversion in the adni cohort," *Neurobiology of aging*, vol. 31, no. 8, pp. 1401–1418, 2010.

[17] L. Shen, S. Kim, S. L. Risacher, K. Nho, S. Swaminathan, J. D. West, T. Foroud, N. Pankratz, J. H. Moore, C. D. Sloan *et al.*, "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*, vol. 53, no. 3, pp. 1051–1063, 2010.