RESEARCH

Learning semi-supervised enrichment of longitudinal imaging-genetic data for improved prediction of cognitive decline

Hoon Seo[1], Lodewijk Brand[1], Hua Wang[1*] and for the Alzheimer's Disease Neuroimaging Initiative

Abstract

Background Alzheimer's Disease (AD) is a progressive memory disorder that causes irreversible cognitive decline. Given that there is currently no cure, it is critical to detect AD in its early stage during the disease progression. Recently, many statistical learning methods have been presented to identify cognitive decline with temporal data, but few of these methods integrate heterogeneous phenotype and genetic information together to improve the accuracy of prediction. In addition, many of these models are often unable to handle incomplete temporal data; this often manifests itself in the removal of records to ensure consistency in the number of records across participants.

Results To address these issues, in this work we propose a novel approach to integrate the genetic data and the longitudinal phenotype data to learn a fixed-length "enriched" biomarker representation derived from the temporal heterogeneous neuroimaging records. Armed with this enriched representation, as a fixed-length vector per participant, conventional machine learning models can be used to predict clinical outcomes associated with AD.

Conclusion The proposed method shows improved prediction performance when applied to data derived from Alzheimer's Disease Neruoimaging Initiative cohort. In addition, our approach can be easily interpreted to allow for the identification and validation of biomarkers associated with cognitive decline.

Keywords Alzheimer's disease, Multi-modal, Longitudinal learning, Enrichment

Background

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. The World Health Organization (WHO) predicts that AD will affect 75 million people by 2030 and 132 million people by 2050 [1]. To address this major public health challenge, it is critical to detect AD at an early stage from both the therapeutic and research standpoints. Recent works [2, 3] have analyzed the progression of AD through modeling and predicting clinical assessments. Furthermore, in the last decade [3], rich neuroimaging measurements, such as magnetic resonance imaging (MRI), have been widely used to predict the clinical outcomes associated with AD.

Despite these efforts, many existing approaches [3, 4] suffer from the following limitations. First, because a lot of models routinely carry out the learning tasks at each time

*Correspondence:
Hua Wang
huawangcs@gmail.com
[1] Department of Computer Science, Colorado School of Mines, Golden, Colorado 80401, USA

point of the AD progression separately, they cannot leveage the temporal relationships across the longitudinal records. Given that AD is a progressive neurodegenerative disorder, multiple consecutive records should be analyzed together for keeping track of the disease progression. Therefore, it is ideal that modern statistical learning techniques can study the temporal variations in the records that are consistent with how we expect a progressive disease to behave. Second, temporal records are often missing at certain time points, which results in an inconsistent number of records per participant. This make it difficult to apply traditional statistical methods that work in the setting when the data at all time points provided. Third, current longitudinal methods [4, 5] tend to focus on measurements derived from MRI scans, such as FreeSurfer (FS) and voxel-based morphometry (VBM), rather than genotype information, such as single-nucleotide polymorphisms (SNPs). It is known [6] that genetic factors can be strong predictors of future cognitive decline, therefore it is important to integrate longitudinal phenotype measurements with genetic data that remain constant when AD develops. Finally, the clinical outcomes of participants assessed from cognitive ability tests, such as Ray's Auditory Verbal Learning Test (RAVLT), are often provided in resources such as the Alzheimer's Disease Neuroimaging Initiative (ADNI), which can be used as data labels for better predicting a future AD diagnosis. Therefore it is of great interest to explore how to use such labeled data to learn data representations with improved predictive power.

In an attempt to overcome the first limitation and uncover the temporal structure of brain phenotypes, several longitudinal prediction models [7, 8] have been proposed. However, these models represent the temporal imaging records as a tensor, which inevitably increase the complexity of the prediction problem and require that each participant has the same number of temporal observations. Since each participant must have the same number of observations, the user of these approaches must discard samples that have a number of records below a given threshold, which may potentially lose valuable information of the input data. Other approaches [9, 10] have relied on imputation techniques to estimate the missing records. Yet these imputation methods may incur undesirable artifacts, which may introduce biases into the final predictions of the longitudinal models.

To handle the longitudinal multi-modal prediction problem with incomplete temporal neuroimaging records, in this work we propose a semi-supervised learning method to learn participant-specific projections to enrich the multi-modal phenotypic measurements, which is an extension of our earlier short conference paper [11]. We analyze the consecutive imaging records simultaneously and learn a projection for each participant. To take advantage of temporal and modality relationships, we introduce trace-norm regularization over the concatenation of all participant-specific projections to maintain their global consistency. Furthermore, a structured sparsity-induced norm regularization is applied to learn the group-structured representations of genetic data, which are integrated with the enriched representations for imaging data. Finally, our model factorizes the enriched biomarker representations, available clinical scores, and genetic biomarkers of participants with the common participants representations. The aim of these factorizations is to extract the representations for a participant shared across the different modalities. As a result, the learned projections from imaging data are tightly coupled with the genetic modality and available clinical scores. Provided with the learned projections per-participant, we can transform the multi-modal representations extracted from phenotypes with varied data sizes and the measurements of the genetic biomarkers into an enriched biomarker representation with a fixed length. With a fixed-length vector per participant, we can freely make use of conventional machine learning models to predict clinical scores associated with AD.

Methods

In this section, first we will formalize the problem to learn a fixed-length biomarker representation for each participant. Then we will then gradually develop our learning objective. Finally, an efficient computational algorithm will be derived to solve our proposed objective.

Notations and problem formalization

Throughout this paper, we write matrices as bold uppercase letters and vectors as bold lowercase letters. The i-th row, the j-th column, and the element at i-th row and j-th column of the matrix $\mathbf{M} = \left[m_j^i \right]$ are denoted as $\mathbf{m}^i$, $\mathbf{m}_j$, m_j^i, or $\mathbf{e}_i^T \mathbf{M}$, $\mathbf{M}\mathbf{e}_j$, $\mathbf{e}_i^T \mathbf{M}\mathbf{e}_j$, respectively, where we define $\mathbf{e}_j$ as the j-th column of the identity matrix $\mathbf{I}$. When $p \geq 1$, the ℓ_p-norm of a vector $\mathbf{v} \in \Re^d$ is defined as $\|\mathbf{v}\|_p = \left(\sum_{i=1}^d v_i^p \right)^{\frac{1}{p}}$. For a matrix $\mathbf{M}$, the trace of $\mathbf{M}$ is defined as $\mathrm{tr}(\mathbf{M}) = \sum_i m_i^i$. The Frobenius norm of $\mathbf{M}$ is defined as $\|\mathbf{M}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m |m_j^i|^2}$. The $\ell_{2,1}$-norm of $\mathbf{M}$ is defined as $\|\mathbf{M}\|_{2,1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^m \left| m_j^i \right|^2} = \sum_{i=1}^n \|\mathbf{m}^i\|_2$. 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 $\mathbf{M}$.

Given a neuroimaging dataset, phenotypic measurements are usually described by the 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 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, such

Seo *et al. BMC Medical Informatics and Decision Making* (2024) 24:61

Page 3 of 13

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 last medical record of the i-th participant and use $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$ to summarize these records of all the participants in the studied cohort. Because multiple types of biomarkers, such as VBM and FS markers, can be extracted from the set of brain scans, we concatenate the vector representations of these biomarkers as the phenotypic assessment of a participant. For example, in our study we write $\mathbf{x}_{ij} = [\mathbf{x}_{ij}^{VBM}, \mathbf{x}_{ij}^{FS}]$ and $\mathbf{x}_i = [\mathbf{x}_i^{VBM}, \mathbf{x}_i^{FS}]$, where $1 \leq i \leq n$, $1 \leq j \leq n_i$. Because $\{\mathbf{x}_{ij}\}_{j=1}^{n_i}$, together with $\mathbf{x}_i$, 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, which is also broadly called as the longitudinal measurements in the literature of medical image computing [4, 7, 8, 12, 13]. To make use of $\mathbf{X}_i$ and $\mathbf{x}_i$ together, we can use longitudinal enrichment to learn a fixed-length vector from them [14–17]. Specially, we 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

$\mathcal{W}^T \otimes \mathbf{X} = [\mathbf{W}_1^T \mathbf{x}_1, \mathbf{W}_2^T \mathbf{x}_2, \cdots, \mathbf{W}_n^T \mathbf{x}_n] \in \Re^{r_1 \times n}$ for the entire cohort, *i.e.*, we project $\mathbf{x}_i$ by $\mathbf{W}_i$ for the i-th participant by computing $\mathbf{z}_i = \mathbf{W}_i^T \mathbf{x}_i \in \Re^{r_1}$. A schematic illustration of the projected (enriched) biomarker representations is shown in Fig. 1.

In addition to the phenotypic measurements used in a neuroimaging data set, genotypes of the same cohort may also be available, such as the SNP profiles of the participants, that can be represented by $\mathbf{X}_{SNP} = [\mathbf{x}_1^{SNP}, \cdots, \mathbf{x}_n^{SNP}] \in \Re^{d_{SNP} \times n}$, 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 when AD develops.

Besides the input phenotypic (dynamic) and genotypic (static) data, the outputs of the prediction tasks 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 total number of clinical scores in studied in our work. Here, without losing generality we consider the first l samples as the labeled data for training. Apparently, $\mathbf{Y}_l$ can be used as
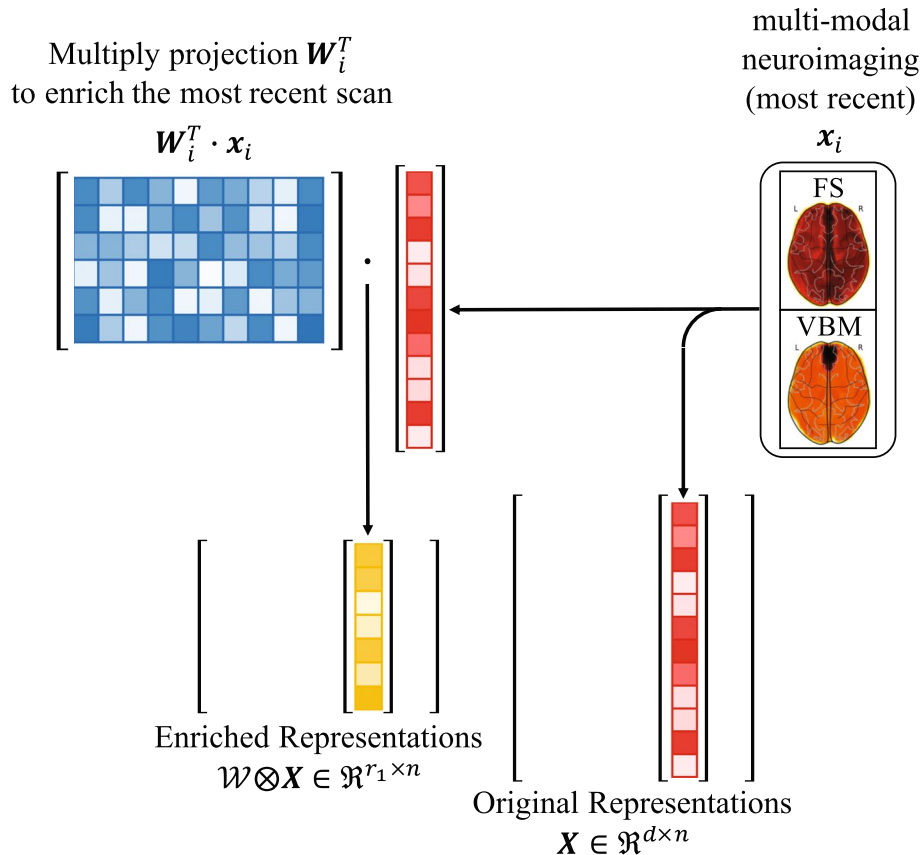

Fig. 1 Illustration of original and enriched biomarker representations. The goal of the enrichment model is to learn the set of projections $\mathcal{W}$ and project the last record. As a result, the dimensionality of enriched representation r_1 is much smaller than the dimensionality of original representation d

the labeled data to enable us to learn the data representation with supervision, which could potentially improve the predictive power of the learned data representations.

In the following subsection, we will develop our learning objective gradually.

Our objective

We start by learning the representations of the static genetic data to utilize the group structures of SNPs [18, 19]. Recent developments in high-throughput genotyping techniques allow new methods to investigate the effect on brain structures and functions of genetic variation. Many previous association studies treated the SNPs as independent units and ignored underlying relationships between the units. However, multiple SNPs from the same gene are naturally related so that such SNPs often jointly perform the genetic functionalities together. To incorporate the group structures associated with SNPs, we propose to learn the representations of the genetic data of a studied cohort by minimizing the following objective:

$$\mathcal{J}_0(\mathbf{H}_0, \mathbf{G}_0) \;=\; \|\mathbf{X}_{SNP} - \mathbf{H}_0\mathbf{G}_0\|_{2,1} + \alpha\|\mathbf{H}_0\|_{G_{2,1}}. \quad (1)$$

In Eq. (1), the first term factorizes $\mathbf{X}_{SNP}$ into $\mathbf{H}_0$ and $\mathbf{G}_0$, where $\mathbf{H}_0$ can be seen as the compressed view of the SNP features [20] and $\mathbf{G}_0$ describes the new representations of the n participants in the subspace spanned by $\mathbf{H}_0$ [21]. To find the group structure of SNPs, we leverage the linkage disequilibrium (LD) [22] which defines the non-random association between alleles at various loci. Then we capture the group-wise sparsity in $\mathbf{H}_0$ by making use of the group $\ell_{2,1}$-norm ($G_{2,1}$-norm) regularization term $\|\mathbf{H}_0\|_{G_{2,1}} = \sum_{k=1}^{K} \left\|\mathbf{H}_0^k\right\|_{2,1}$, where $\mathbf{H}_0 = [\mathbf{H}_0^1; \mathbf{H}_0^2; \cdots; \mathbf{H}_0^K]$ consists of K groups derived from the LD correlations of the SNPs [18, 19]. Here we choose to use the $\ell_{2,1}$-norm distances to improve the robustness of our model against outliers [23–26].

Next we study how to learn a vector representations with fixed length for every participant from their image data in varied sizes. While the genetic profiles of the participants remain constant over time, the functions and structures of the brains of the participants change as AD progresses. Therefore AD progression is characterized by the longitudinal imaging records extracted from the multiple brain scans that change over time. However, the longitudinal imaging records pose a critical challenge to build the predictive models, because different participants may take the brain scans at different time and the number of brain scans of different participants are not same in general. To deal with this difficulty and summarize the brain variations of every participant individually, we propose to learn a vector representation with the fixed

length from the image data of each participant $\{\mathbf{X}_i, \mathbf{x}_i\}$ with the varied size n_i by computing $\mathbf{z}_i = \mathbf{W}_i^T\mathbf{x}_i \in \Re^{r_1}$.

First, to preserve as much dynamic information of $\mathbf{X}_i$ as possible, we propose to learn the projection $\mathbf{W}_i$ for the i-th participant by minimizing the following objective of the principal component analysis (PCA) [27]:

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

Here again we use $\ell_{2,1}$-norm objective in the PCA to enhance the robustness of the learned projection $\mathbf{W}_i$ against outlying samples which is unavoidable in the large dataset [23–26].

Second, besides using the projection learned from each individual participant separately, to maximize the consistency across all the learned projections for the same cohort, we enforce the low-rank consistencies onto the learned projection matrices by introducing two trace-norm regularization terms as following [7, 13, 17]:

$$\mathcal{J}_2(\mathcal{W}) \;=\; \sum_{i=1}^{n} \left\|\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i\right\|_{2,1} + \beta(\left\|\mathbf{W}_{(1)}\right\|_* + \left\|\mathbf{W}_{(2)}\right\|_*),$$
$$s.t.\ \mathbf{W}_i^T\mathbf{W}_i = \mathbf{I},$$
$$(3)$$

where $\mathbf{W}_{(1)} = [\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_n] \in \Re^{d \times (r_1 \times n)}$ and $\mathbf{W}_{(2)} = \left[\mathbf{W}_1^T, \mathbf{W}_2^T, \cdots, \mathbf{W}_n^T\right] \in \Re^{r_1 \times (d \times n)}$ are two unfolded matrices of the local projection tensor $\mathcal{W}$.

Finally, equipped with the learned representations for imaging features in multiple modalities and genetic features, we integrate them together to explore the full potential of an imaging-genetic dataset. First, we write the temporally enriched representations for image data together as $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n] = \mathcal{W}^T \otimes \mathbf{X} \in \Re^{r_1 \times n}$. Following the same idea as before, we factorize $\mathbf{Z}$ and align the factorized data representation with that learned from the static genetic data $\mathbf{G}_0$ by minizing the following objective:

$$\mathcal{J}_3(\mathbf{H}_0, \mathbf{H}_1, \mathbf{G}_0, \mathbf{G}_1, \mathcal{W}) = \gamma_1 \sum_{i=1}^{n} \left\|\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i\right\|_{2,1}$$
$$+ \gamma_2 \left\|\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1\mathbf{G}_1\right\|_{2,1} + \gamma_3 \|\mathbf{X}_{SNP} - \mathbf{H}_0\mathbf{G}_0\|_{2,1} + \gamma_4 \|\mathbf{G}_1 - \mathbf{G}_0\|_{2,1}$$
$$+ \gamma_5 \|\mathbf{H}_0\|_{G_2} + \gamma_6 \left(\left\|\mathbf{W}_{(1)}\right\|_* + \left\|\mathbf{W}_{(2)}\right\|_*\right) + \gamma_7 \|\mathbf{U}\|_1, \quad s.t.\ \mathbf{W}_i^T\mathbf{W}_i = \mathbf{I},$$
$$(4)$$

where $\gamma_1, \gamma_2, \cdots, \gamma_7$ are hyperparameters of our learning model.

Now we can perform the association studies between the clinical scores and the new data representations learned from our model. Suppose that the clinical scores of $\mathbf{Y}_l \in \Re^{c \times l}$ are obtained in c cognitive assessments for the l training samples, we use $\mathbf{F} = [\mathbf{F}_l, \mathbf{F}_u] \in \Re^{c \times n}$ to denote our estimated clinical scores and use the constraint $\mathbf{F}_l = \mathbf{Y}_l$ to make use of the training data $\mathbf{Y}_l$, by which we can conduct the regression analyses by minimizing the following objective:

$$\mathcal{J}(\mathbf{U}, \mathbf{F}, \mathbf{H}_0, \mathbf{H}_1, \mathbf{G}_0, \mathbf{G}_1, \mathcal{W}) = \left\| \mathbf{F} - \mathbf{U}^T \mathbf{G}_1 \right\|_{2,1} + \gamma_1 \sum_{i=1}^{n} \left\| \mathbf{X}_i - \mathbf{W}_i \mathbf{W}_i^T \mathbf{X}_i \right\|_{2,1}$$

$$+ \gamma_2 \left\| \mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1 \mathbf{G}_1 \right\|_{2,1} + \gamma_3 \|\mathbf{X}_{\text{SNP}} - \mathbf{H}_0 \mathbf{G}_0\|_{2,1} + \gamma_4 \|\mathbf{G}_1 - \mathbf{G}_0\|_{2,1} \qquad (5)$$

$$+ \gamma_5 \|\mathbf{H}_0\|_{G_2} + \gamma_6 \left(\left\| \mathbf{W}_{(1)} \right\|_* + \left\| \mathbf{W}_{(2)} \right\|_* \right) + \gamma_7 \|\mathbf{U}\|_1,$$

$$s.t. \quad \mathbf{F}_l = \mathbf{Y}_l, \ \mathbf{W}_i^T \mathbf{W}_i = \mathbf{I}.$$

Our new method is schematically illustrated in Fig. 2.

The solution algorithm

Although our objective in Eq. (5) has clearly motivated, it is difficult to solve in general, because it is non-smooth. Thus in this subsection we drive an efficient solution to optimize our objective. Using the optimization framework presented in the earlier work [28, 29] that proposed the iterative reweighted method to solve non-smooth objectives, we can solve Eq. (5) by an iterative procedure (Algorithm 1 in [28]) in which the key step is to minimize the following objective:

$$\mathcal{J}^R(\mathbf{U}, \mathbf{F}, \mathbf{H}_0, \mathbf{H}_1, \mathbf{G}_0, \mathbf{G}_1, \mathcal{W}) = \text{tr}((\mathbf{U}^T \mathbf{G}_1 - \mathbf{F})^T \mathbf{D}_1 (\mathbf{U}^T \mathbf{G}_1 - \mathbf{F}))$$

$$+ \gamma_1 \sum_{i=1}^{n} \text{tr}((\mathbf{X}_i - \mathbf{W}_i \mathbf{W}_i^T \mathbf{X}_i)^T \mathbf{D}_{2,i}(\mathbf{X}_i - \mathbf{W}_i \mathbf{W}_i^T \mathbf{X}_i))$$

$$+ \gamma_2 \text{tr}((\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1 \mathbf{G}_1)^T \mathbf{D}_3 (\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1 \mathbf{G}_1))$$

$$+ \gamma_3 \text{tr}((\mathbf{X}_{SNP} - \mathbf{H}_0 \mathbf{G}_0)^T \mathbf{D}_4 (\mathbf{X}_{SNP} - \mathbf{H}_0 \mathbf{G}_0))$$

$$+ \gamma_4 \text{tr}((\mathbf{G}_1 - \mathbf{G}_0)^T \mathbf{D}_5 (\mathbf{G}_1 - \mathbf{G}_0)) + \gamma_5 \text{tr}(\mathbf{H}_0^T \mathbf{D}_6 \mathbf{H}_0)$$

$$+ \gamma_6 \text{tr}(\mathbf{W}_{(1)}^T \mathbf{D}_7 \mathbf{W}_{(1)}) + \gamma_6 \text{tr}(\mathbf{W}_{(2)}^T \mathbf{D}_8 \mathbf{W}_{(2)}) + \gamma_7 \sum_{q=1}^{c} (\mathbf{u}_q^T \mathbf{D}_{9,q} \mathbf{u}_q),$$

$$s.t. \quad \mathbf{F}_l = \mathbf{Y}_l, \ \mathbf{W}_i^T \mathbf{W}_i = \mathbf{I}, \qquad (6)$$
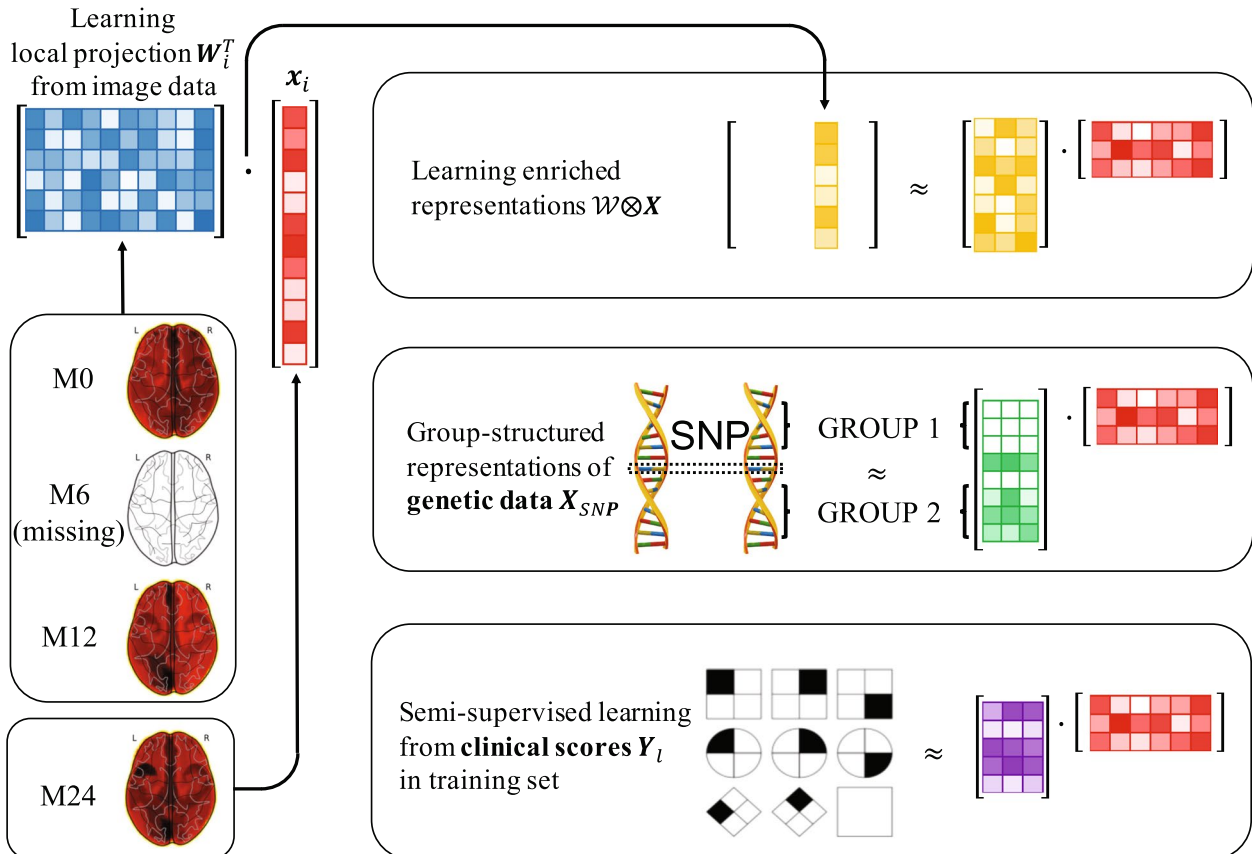

Fig. 2 Overview of proposed semi-supervised learning framework to fully utilize the potential of a longitudinal AD dataset. We use factorization to extract the common representations of participants shared across genetic, image, and clinical scores data. As a result, the genetic and clinical scores data can be reflected in the learned projections $\mathcal{W}$

where $\mathbf{D}_1, \mathbf{D}_{2,i}, \mathbf{D}_3, \mathbf{D}_4, \mathbf{D}_5, \mathbf{D}_{9,q}$ are diagonal matrices whose j-th diagonal element d_*^j of $\mathbf{D}_*$ is as follows:

$$
\begin{aligned}
d_1^j &= \frac{1}{2\sqrt{\left\|\mathbf{e}_j^T(\mathbf{U}^T\mathbf{G}_1 - \mathbf{F})\right\|_2^2 + \delta}}, d_{2,i}^j = \frac{1}{2\sqrt{\left\|\mathbf{e}_j^T(\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i)\right\|_2^2 + \delta}},\\
d_3^j &= \frac{1}{2\sqrt{\left\|\mathbf{e}_j^T(\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1\mathbf{G}_1)\right\|_2^2 + \delta}}, d_4^j = \frac{1}{2\sqrt{\left\|\mathbf{e}_j^T(\mathbf{X}_{SNP} - \mathbf{H}_0\mathbf{G}_0)\right\|_2^2 + \delta}},\\
d_5^j &= \frac{1}{2\sqrt{\left\|\mathbf{e}_j^T(\mathbf{G}_1 - \mathbf{G}_0)\right\|_2^2 + \delta}}, d_{9,q}^j = \frac{1}{2\sqrt{(u_q^j)^2 + \delta}},\\
\mathbf{D}_8 &= \frac{1}{2}(\mathbf{W}_{(2)}\mathbf{W}_{(2)}^T + \delta\mathbf{I})^{-\frac{1}{2}}, \mathbf{D}_7 = \frac{1}{2}(\mathbf{W}_{(1)}\mathbf{W}_{(1)}^T + \delta\mathbf{I})^{-\frac{1}{2}}.
\end{aligned}
\tag{7}
$$

$\mathbf{D}_6$ is a block diagonal matrix, where k-th block (k-th group of SNPs) is $\frac{1}{2}(\left\|\mathbf{H}_0^k\right\|_F^2 + \delta)^{-\frac{1}{2}}\mathbf{I}_k$. $\mathbf{I}_k \in \Re^{d_k \times d_k}$ is a identity matrix, and d_k denotes the number of rows (the number of SNPs) of k-th block of $\mathbf{H}_0 = [\mathbf{H}_0^1; \mathbf{H}_0^2; \cdots; \mathbf{H}_0^K]$, so that $\sum_{k=1}^{K} d_k = d_{SNP}$. The dimensions of the matrices in Eq. (7) are: $\mathbf{D}_1 \in \Re^{c \times c}$, $\mathbf{D}_{2,i} \in \Re^{d \times d}$, $\mathbf{D}_3 \in \Re^{r_1 \times r_1}$, $\mathbf{D}_4 \in \Re^{d_{SNP} \times d_{SNP}}$, $\mathbf{D}_5 \in \Re^{r_2 \times r_2}$, $\mathbf{D}_6 \in \Re^{d_{SNP} \times d_{SNP}}$, $\mathbf{D}_7 \in \Re^{d \times d}$, $\mathbf{D}_8 \in \Re^{r_1 \times r_1}$, $\mathbf{D}_{9,i} \in \Re^{r_2 \times r_2}$.

To minimize the smoothed objective Eq. (6), we use the Alternating Direction Method of Multipliers (ADMM), which is proposed in [30, 31]. 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 $\mathcal{W}$, we rewrite Eq. (6) with the following equivalent objective:

$$
\begin{aligned}
&\mathcal{J}^{ADMM}(\mathbf{U}, \mathbf{F}, \mathcal{W}, \mathbf{H}_0, \mathbf{H}_1, \mathbf{G}_0, \mathbf{G}_1, \mathbf{A}, \mathbf{B}_{(2)}, \Lambda_1, \Lambda_{2,i}, \Lambda_3, \Lambda_{4,i}) =\\
&\mathrm{tr}((\mathbf{A}^T\mathbf{G}_1 - \mathbf{F})^T\mathbf{D}_1(\mathbf{U}^T\mathbf{G}_1 - \mathbf{F}))\\
&+ \gamma_1\sum_{i=1}^{n}\mathrm{tr}((\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i)^T\mathbf{D}_{2,i}(\mathbf{X}_i - \mathbf{B}_i\mathbf{B}_i^T\mathbf{X}_i))\\
&+ \gamma_2\mathrm{tr}((\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1\mathbf{G}_1)^T\mathbf{D}_3(\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1\mathbf{G}_1))\\
&+ \gamma_3\mathrm{tr}((\mathbf{X}_{SNP} - \mathbf{H}_0\mathbf{G}_0)^T\mathbf{D}_4(\mathbf{X}_{SNP} - \mathbf{H}_0\mathbf{G}_0))\\
&+ \gamma_4\mathrm{tr}((\mathbf{G}_1 - \mathbf{G}_0)^T\mathbf{D}_5(\mathbf{G}_1 - \mathbf{G}_0)) + \gamma_5\mathrm{tr}(\mathbf{H}_0^T\mathbf{D}_6\mathbf{H}_0)\\
&+ \gamma_6\mathrm{tr}(\mathbf{W}_{(1)}^T\mathbf{D}_7\mathbf{W}_{(1)}) + \gamma_6\mathrm{tr}(\mathbf{W}_{(2)}^T\mathbf{D}_8\mathbf{B}_{(2)}) + \gamma_7\sum_{q=1}^{c}(\mathbf{a}_q^T\mathbf{D}_{9,q}\mathbf{a}_q)\\
&+ \frac{\mu_1}{2}\left\|\mathbf{F}_l - \mathbf{Y}_l + \frac{1}{\mu_1}\Lambda_1\right\|_F^2 + \sum_{i=1}^{n}\frac{\mu_{2,i}}{2}\left\|\mathbf{W}_i^T\mathbf{B}_i - \mathbf{I} + \frac{1}{\mu_{2,i}}\Lambda_{2,i}\right\|_F^2\\
&+ \frac{\mu_3}{2}\left\|\mathbf{A} - \mathbf{U} + \frac{1}{\mu_3}\Lambda_3\right\|_F^2 + \sum_{i=1}^{n}\frac{\mu_{4,i}}{2}\left\|\mathbf{B}_i - \mathbf{W}_i + \frac{1}{\mu_{4,i}}\Lambda_{4,i}\right\|_F^2,
\end{aligned}
\tag{8}
$$

where $\Lambda_1, \Lambda_{2,i}, \Lambda_3, \Lambda_{4,i}$ 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 $\mathbf{B}_i = \mathbf{W}_i$. The detailed algorithm to minimize Eq. (8) is presented in Algorithm 1. In 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 $\mathbf{PX} + \mathbf{XQ} = \mathbf{R}$. The time complexity of Algorithm 1 is $O(nr_1d^2(d + r_1))$ for each iteration where the step 11 is the most dominant. The detailed derivation of Algorithm 1 is provided in the Appendix.

Algorithm 1 Solve minimization problem in Eq. (8)

```
Input: X, X_SNP, Y_l, X_i (1 ≤ i ≤ n);
Initialization: U, F, W, H_0, H_1, G_0, G_1, A, B_(2), Λ_1, Λ_2,i, Λ_3, Λ_4,i,
    1 < ρ_1, ρ_2,i, ρ_3, ρ_4,i < 2,
    μ_1, μ_2,i, μ_3, μ_4,i, γ_1, γ_2, γ_3, γ_4, γ_5, γ_6, γ_7 > 0, (1 ≤ i ≤ n);
while not converge do
    1. Update D_1, D_2, ⋯, D_9 defined in Eq. (7);
    2. Update H_1 by H_1 = (W^T ⊗ X)G_1^T(G_1G_1^T)^{-1};
    3. Update H_0 by H_0 = sylvester(γ_5/γ_3 D_4^{-1}D_6, G_0G_0^T, X_SNPG_0^T);
    4. Update G_1 by G_1 = (AD_1U^T + UD_1A^T + γ_2H_1^TD_3H_1 + γ_4D_5)^{-1}(½(AD_1 +
        UD_1)F + γ_4D_5G_0 + γ_2H_1^TD_3(W^T ⊗ X));
    5. Update G_0 by G_0 = (γ_3H_0^TD_4H_0 + γ_4D_5)^{-1}(γ_3H_0^TD_4X_SNP + γ_4D_5G_1);
    6. Update B_i (1 ≤ i ≤ n) by
        B_i = (−γ_1(D_2,i(X_i − W_iW_i^TX_i)X_i^T + X_i(X_i − W_iW_i^TX_i)^TD_2,i) + μ_2,iW_iW_i^T +
        μ_4,iI)^{-1}(−γ_6W_iD_8 + (μ_2,i + μ_4,i)W_i − Λ_4,i − W_iΛ_2,i);
    7. Update p-th column of A (1 ≤ p ≤ c) by
        a_p = (2γ_7D_9,p + μ_3I)^{-1}(−G_1(G_1^TU − F^T)D_1 + μ_3U − Λ_3)e_p;
    8. Update p-th column of F (1 ≤ p ≤ l) by
        f_p = (2D_1 + μ_1I)^{-1}(D_1(U^T + A^T)G_1 + μ_1Y_l − Λ_1)e_p;
    9. Update p-th column of F (l + 1 ≤ p ≤ n) by f_p = ½(U^T + A^T)G_1e_p;
    10. Update U by U = −1/μ_3 G_1(G_1^TA − F^T)D_1 + A + 1/μ_3 Λ_3;
    11. Update p-th column of W_i (1 ≤ i ≤ n and 1 ≤ p ≤ r_1) by
        w_i,p = (2γ_1D_7 − γ_1(X_iX_i^TD_2,i − X_iX_i^TB_iB_i^TD_2,i + D_2,iX_iX_i^T −
        D_2,iB_iB_i^TX_iX_i^T) + μ_2,iB_iB_i^T + μ_4,iI + 2γ_2d_3^p x_ix_i^T)^{-1}(B_i((μ_2,i + μ_4,i)I − γ_6D_8 −
        Λ_2,i^T)e_p + 2γ_2d_3^p(e_p^TH_1G_1e_i)x_i + Λ_4,ie_p);
    12. Update Λ_1 by Λ_1 = Λ_1 + μ_1(F_l − Y_l);
    13. Update Λ_2,i (1 ≤ i ≤ n) by Λ_2,i = Λ_2,i + μ_2,i(W_i^TB_i − I);
    14. Update Λ_3 by Λ_3 = Λ_3 + μ_3(A − U);
    15. Update Λ_4,i (1 ≤ i ≤ n) by Λ_4,i = Λ_4,i + μ_4,i(B_i − W_i);
    16. Update μ_1, μ_2,i, μ_3, μ_4,i (1 ≤ i ≤ n) by
        μ_1 = ρ_1μ_1; μ_2,i = ρ_2,iμ_2,i; μ_3 = ρ_3μ_3; μ_4,i = ρ_4,iμ_4,i;
end
Output: W_i (1 ≤ i ≤ n)
```

Results

In this section, we introduce our experimental results about the clinical scores prediction task with the enriched biomarker representation and original biomarker representation to evaluate changes in the prediction performance from enrichment. Then we analyze the AD risk factors identified by the learned projections.

Data preparation

We obtain the data used in our experiments from the ADNI database. We downloaded the MRI scans, SNP

Seo *et al. BMC Medical Informatics and Decision Making* (2024) 24:61

Page 7 of 13

genotypes, and the longitudinal scores of Rey's Auditory Verbal Learning Test (RAVLT) of 821 ADNI-1 participants. We perform voxel-based morphometry (VBM) and FreeSurfer automated parcellation on the MRI data as described by [12] and extract mean modulated gray matter (GM) measures for 90 target regions of interest (ROI). We follow SNP quality control steps discussed in [32]. Among 821 ADNI-1 participants, 412 participants are selected on the basis of existence of MRI records at Month 0/Month 6/Month 12/Month 24. Then we intentionally discard Month 24 scans with 50% probability to evaluate the learning capability of our model from longitudinal data with missing record. Our model learns the enrichment with the neuroimaging records from baseline to the second last visit, and project the last record (Month 12 or Month 24 with 50% probability) to predict the clinical scores at the last time point.

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 of the most recent biomarkers. 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.

We conduct a 5-fold cross-validation to search the set of best hyperparameters for each conventional model. However the naive grid search can be time consuming especially when the combination of many hyperparameters is tuned. Instead of trying all the combinations of hyperparameters, we randomly choose the value from the grid of each hyperparameter. In order to increase the possibility of finding the better hyperparameters in the fewer searches, a randomly selected half of hyperparameters remain the best values found in the previous searches. In the 5-fold cross-validation, we search the best regularization parameter of RR in $\{10^3, 10^2, 10, 1, 10^{-1}, 10^{-2}, 10^{-3}\}$. For SVR, we fine tune the kernel function among sigmoid and radial basis function and box constraints

in $\{10^3, 10^2, 10, 1, 10^{-1}, 10^{-2}, 10^{-3}\}$. We construct a 1-dimensional CNN configured as follows: (1) a convolutional layer with a window size of 5×16 (width $\times$ depth), followed by a rectified linear unit (ReLU) and a max pooling layer with a window size of 1×2; (2) a convolution layer with a window size of 10×32, followed by a ReLU and a max pooling layer with a window size of 1×2; (3) three fully connected layers where the number of nodes and dropout rate for each layer are fine tuned by searching the grid of $\{20, 60, 120\}$ and $\{0.3, 0.5, 0.7\}$ each. The hyperparameters of our enrichment model are tuned as following: $\gamma_1 = 10^{-1}$, $\gamma_2 = 10^{-4}$, $\gamma_3 = 10^{-2}$, $\gamma_4 = 10^{-3}$, $\gamma_5 = 10^{-1}$, $\gamma_6 = 10^{-1}$, $\gamma_7 = 10^{-1}$, $\rho_1 = 1.05$, $\rho_2 = 1.05$, $\rho_3 = 1.15$, $\rho_4 = 1.15$.

Experimental results

Original vs. enriched representation

In the experimental result reported in Fig. 3, 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 result reveals that the prediction from enriched representation are mostly more accurate (9.84% in average) than the predictions supplied from the most recent record of original representation. Interestingly, among the various concatenation of modalities of biomarker measurements, the performance improvements of our enriched representation is larger when the many modalities are given. This indicates that our model fully utilizes the multimodal dynamic data. Especially, the error gap is smallest when only SNPs are given. We suppose that this is because SNPs are static data which do not change along the time, genetic data is not able to provide the enough information about temporal variations of cognitive decline, while our model is designed to learn the temporal variations when dynamic data is given.

Identification of disease-relevant biomarkers

In addition to the cognitive outcomes prediction task, we identify AD relevant biomarkers using the weights of the learned projections. Since p-th feature in the enriched representation $\mathbf{e}_p^T \mathbf{W}_i^T \mathbf{x}_i$ is weighted summation of the original biomarkers measurements, the weights summation in q-th row of the projection $\sum_{i=1}^n \left\| \mathbf{e}_q^T \mathbf{W}_i \right\|_1$ can be interpreted as AD relevance on the q-th biomarker.

Identified Neuroimaging Biomarkers In this aspect, we first identify the AD relevant imaging biomarkers, by plotting the weights of ROIs of VBM and FS in Fig. 4. These brain regions all appear in the medical literature associated with AD-related dimentia. For example,
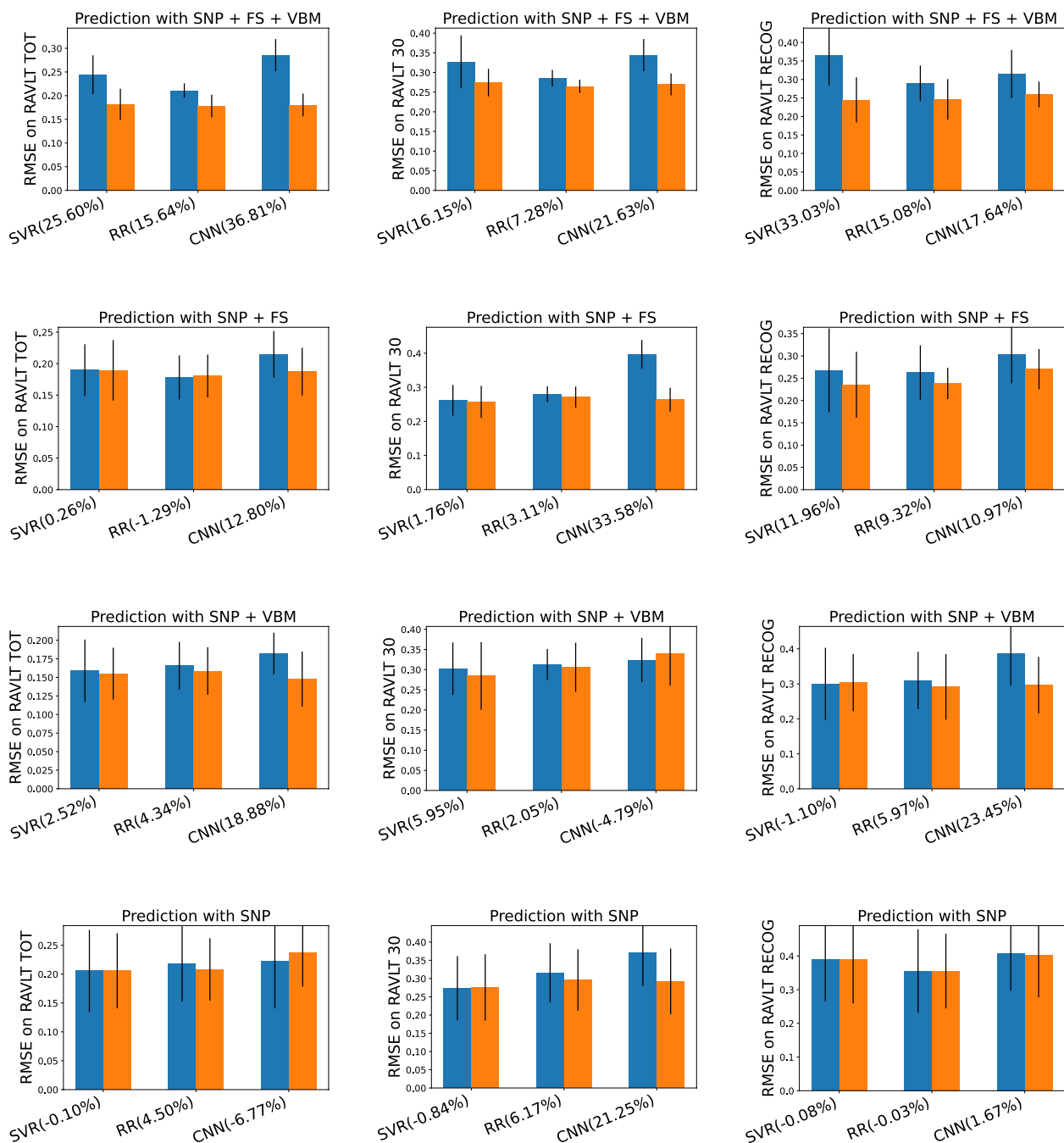
Seo *et al. BMC Medical Informatics and Decision Making*   (2024) 24:61

Page 8 of 13

Fig. 3 Comparison on prediction errors from original (blue) or enriched (orange) representation. The percentage % next to model name indicates the decreased amount in errors of predictions from the enriched representation. We also plot the standard deviation from 5-fold cross-validation at the head of each bar

participants with cognitive decline showed atrophy of the caudate nucleus [33]. The volume of thalamus was significantly reduced [34] in participants diagnosed as probable AD. Furthermore, it has been found that the thalamus region plays an important role in generating attention, and anterior thalamus is in charge of declarative memory functioning [35]. The hippocampus is vulnerable to be damaged from AD [36] and has been shown to affect long-term memory and spatial navigation in participants with AD. Finally, the amygdala

Seo *et al. BMC Medical Informatics and Decision Making* (2024) 24:61

Page 9 of 13
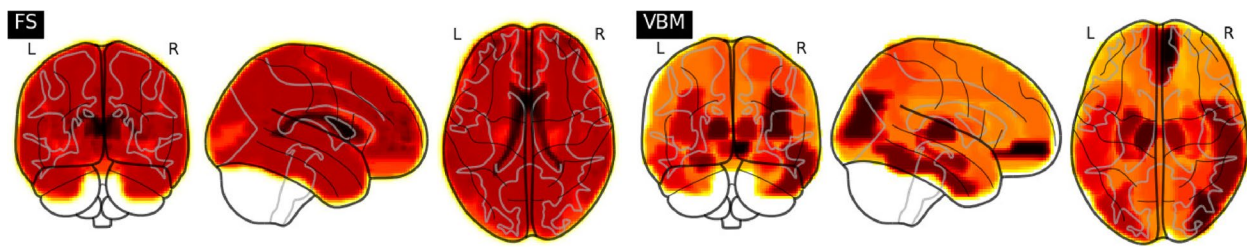
Fig. 4 Visualization of weights distribution over the brain regions. The darker color indicates the larger weight on that region. The top four AD relevant regions are identified in FS: Right Caudate, Brodmann area 24, Left Thalamus, and Left Caudate, in VBM: Left Hippocampus, Left Amygdala, Left Thalamus, and Right Medial Orbito-frontal Cortex

region, also identified by our approach, is also severely affected by AD [37] and is associated with emotional response and decision-making.

Identified Genotypic Biomarkers In addition, we identify AD relevant SNPs by plotting the weights on individual SNPs and AlzGene group of SNPs in Figs. 5 and 6. In Fig. 6, we download and use the AlzGene grouping information constructed by multiple genome-wide association studies listed on (http://www.alzgene.org/) [38]. The standard deviation of weights of SNPs in each AlzGene group is displayed as line length at the head of each bar. The top identified individual SNP, rs10779339, in Fig. 5 has been found to be related to cognitive decline [39]. Among the AlzGene groups in Fig. 6 the SNPs in the ACE (angiotensin-converting enzyme) group have been found to reduce the Amyloid Beta peptide (Aβ) which

is commonly observed in the progression of AD-related cognitive decline [40]. Furthermore, the APOE (apolipo-protein E) gene, also identified by our approach is also involved in Aβ aggregation and clearance [41].

In summary, the complex relationships between cognitive ability and identified biomarkers are clearly identified by our method and are well represented in previous AD research studies. This result supports the utility of our approach as a tool to discover and validate AD risk factors from multi-modal data.

Discussion

Because our enrichment model incorporates the genetic and phenotypic biomarkers, it is possible to learn the enrichment from the different multi-modal data to predict the different target labels. From the experimental
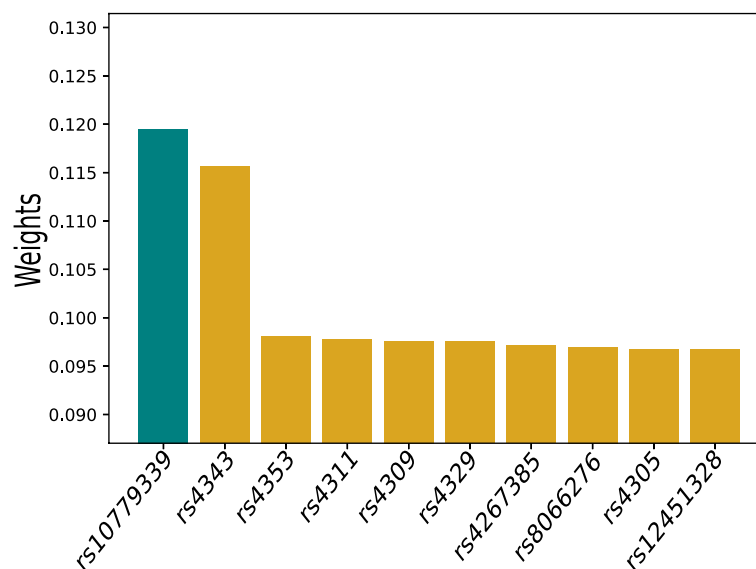

Fig. 5 Weights on individual SNPs. The color of each bar denotes AlzGene group in Fig. 6
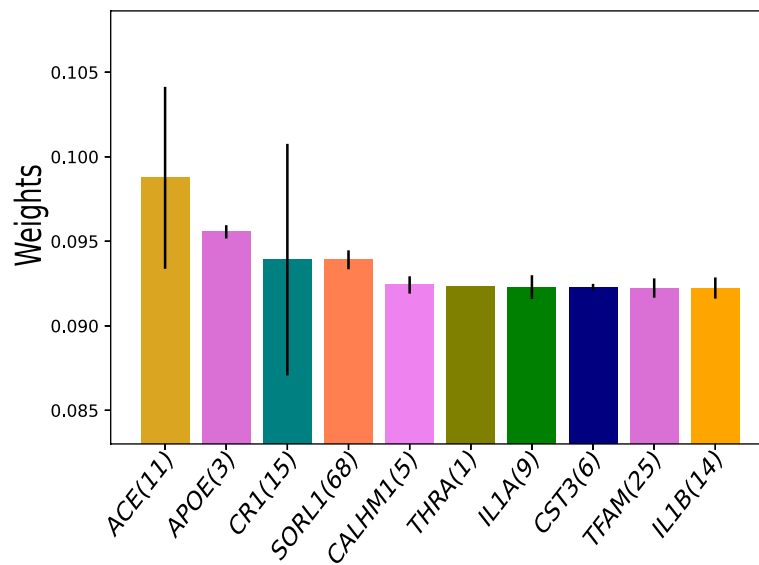
Fig. 6 Weights on AlzGene groups. The number next to the AlzGene group name denotes the number of SNPs in that group

results in Fig. 3, the prediction performance is further improved as the many modalities of data are given. For example, Diffusion Tensor Imaging (DTI) is an effective tool to investigate the white matter organization of the brain and AD progression [42]. Compared to MRI, blood based biomarkers can be measured less costly and intrusive and Aβ levels in blood can aid in the early diagnosis of AD [43]. We can learn the projections for these additional modalities, and concatenate the projections in Eq. (2) and Eq. (5). While considering the complexity of Algorithm 1 proportional to d^3 and the increasing dimensionality d with the many modalities of measurements, our model can be flexibly applied to the diverse datasets and prediction tasks.

Conclusion

Missing data is a major issue in longitudinal multi-modal healthcare datasets. This research aims to devise a novel methodology to learn a consistent 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 records of longitudinal biomarkers on a per-participant basis. Our experiments show that the learned enriched representation outperforms the original measurement in predicting the clinical scores. Finally, the identified AD relevant biomarkers are in nice accordance with existing research findings, indicating the utility of our approach.

Appendix

In this section, we show how Algorithm 1 is derived from our smoothed objective Eq. (6). ADMM solves convex optimization problems by breaking them into smaller pieces that are easier to handle. Specifically, given the following objective with the equality constraint:

$$\min_{x,z} f(x) + g(z), \qquad s.t. \quad h(x,z) = 0, \qquad (9)$$

Algorithm 2 solves the problem in Eq. (9) by decoupling it into subproblems and optimizing each variable while fixing others, where y is the Lagrangian multiplier to the constraint h. We extend Algorithm 2 of two variables to Algorithm 1 of the set of variables $\{\mathbf{U}, \mathbf{F}, \mathcal{W}, \mathbf{H}_0, \mathbf{H}_1, \mathbf{G}_0, \mathbf{G}_1, \mathbf{A}, \mathbf{B}_{(2)}\}$. To be specific, we earn a update equation respect to each variable in Algorithm 1, by taking the derivation over a each variable and set it to zero matrix, while fixing other variables. For the step 2 in Algorithm 1, the derivative is:

$$\frac{\partial \mathcal{J}^{ADMM}}{\partial \mathbf{H}_1} = -2\gamma_2 \mathbf{D}_3 (\mathcal{W} \otimes \mathbf{X} - \mathbf{H}_1 \mathbf{G}_1) \mathbf{G}_1^T. \qquad (10)$$

For the step 3, the derivative is:

$$\frac{\partial \mathcal{J}^{ADMM}}{\partial \mathbf{H}_0} = -2\gamma_3 \mathbf{D}_4(\mathbf{X}_{SNP} - \mathbf{H}_0\mathbf{G}_0)\mathbf{G}_0^T + 2\gamma_5 \mathbf{D}_6\mathbf{H}_0. \quad (11)$$

Algorithm 2 The ADMM algorithm

Set $1 < \rho < 2$ and initialize $\mu > 0$ and y;
while *not converge* **do**
   1. Update x by solving $x^{k+1} = \arg\min_x(f(x) + \frac{\mu}{2}\|h(x,z^k) + \frac{y^k}{\mu}\|^2)$;
   2. Update z by solving $z^{k+1} = \arg\min_z(g(z) + \frac{\mu}{2}\|h(x^{k+1},z) + \frac{y^k}{\mu}\|^2)$;
   3. Update y by $y^{k+1} = y^k + \mu h(x^{k+1}, z^{k+1})$;
   4. Update μ by $\mu = \rho\mu$.

For the step 4, the derivative is:

$$\frac{\partial \mathcal{J}^{ADMM}}{\partial \mathbf{G}_1} = \mathbf{AD}_1(\mathbf{U}^T\mathbf{G}_1 - \mathbf{F}) + \mathbf{UD}_1(\mathbf{A}^T\mathbf{G}_1 - \mathbf{F})$$
$$+ 2\gamma_4\mathbf{D}_5(\mathbf{G}_1 - \mathbf{G}_0) - \gamma_2\mathbf{H}_1^T\mathbf{D}_3(\mathcal{W}^T \otimes \mathbf{X} - \mathbf{H}_1\mathbf{G}_1). \quad (12)$$

For the step 5, the derivative is:

$$\frac{\partial \mathcal{J}^{ADMM}}{\partial \mathbf{G}_0} = -2\gamma_3\mathbf{H}_0^T\mathbf{D}_4(\mathbf{X}_{SNP} - \mathbf{H}_0\mathbf{G}_0) - 2\gamma_4\mathbf{D}_5(\mathbf{G}_1 - \mathbf{G}_0). \quad (13)$$

For the step 6, the derivative is:

$$\frac{\partial \mathcal{J}^{ADMM}}{\partial \mathbf{B}_i} = \gamma_6\mathbf{W}_i\mathbf{D}_8$$
$$- \gamma_1(\mathbf{D}_{2,i}(\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i^T)\mathbf{X}_i^T + (\mathbf{D}_{2,i}(\mathbf{X}_i - \mathbf{W}_i\mathbf{W}_i^T\mathbf{X}_i^T)\mathbf{X}_i^T)^T)\mathbf{B}_i$$
$$+ \mu_{2,i}\mathbf{W}_i\mathbf{W}_i^T\mathbf{B}_i - \mu_{2,i}\mathbf{W}_i + \mathbf{W}_i\Lambda_{2,i} + \mu_{4,i}(\mathbf{B}_i - \mathbf{W}_i + \frac{1}{\mu_{4,i}}\Lambda_{4,i}). \quad (14)$$

For the step 7 ($1 \le p \le c$), the derivative is:

$$\frac{\partial \mathcal{J}^{ADMM}}{\partial \mathbf{a}_p} = (\mathbf{G}_1(\mathbf{G}_1^T\mathbf{U} - \mathbf{F}^T)\mathbf{D}_1 - \mu_3\mathbf{U} + \Lambda_3)\mathbf{e}_p$$
$$+ 2\gamma_7\mathbf{D}_{9,i}\mathbf{a}_p + \mu_3\mathbf{a}_p. \quad (15)$$

For the step 8 ($1 \le p \le l$), the derivative is:

$$\frac{\partial \mathcal{J}^{ADMM}}{\partial \mathbf{f}_p} = \Big(\mathbf{D}_1(\mathbf{U}^T\mathbf{G}_1 - \mathbf{F}) + \mathbf{D}_1(\mathbf{A}^T\mathbf{G}_1 - \mathbf{F})$$
$$+ \mu_1(\mathbf{F} - \mathbf{Y} + \frac{1}{\mu_1}\Lambda_1)\Big)\mathbf{e}_p. \quad (16)$$

For the step 9 ($l + 1 \le p \le n$), the derivative is:

$$\frac{\partial \mathcal{J}^{ADMM}}{\partial \mathbf{f}_p} = (\mathbf{D}_1(\mathbf{U}^T\mathbf{G}_1 - \mathbf{F}) + \mathbf{D}_1(\mathbf{A}^T\mathbf{G}_1 - \mathbf{F}))\mathbf{e}_p. \quad (17)$$

For the step 10, the derivative is:

$$\frac{\partial \mathcal{J}^{ADMM}}{\partial \mathbf{U}} = \mathbf{G}_1(\mathbf{G}_1^T\mathbf{A} - \mathbf{F}^T)\mathbf{D}_1 + \mu_3(\mathbf{A} - \mathbf{U} + \frac{1}{\mu_3}\Lambda_3). \quad (18)$$

For the step 11 ($1 \le p \le r_1$), the derivative is:

$$\frac{\partial \mathcal{J}^{ADMM}}{\partial \mathbf{w}_{i,p}} =$$
$$\big(-\gamma_1(\mathbf{X}_i\mathbf{X}_i^T\mathbf{D}_{2,i} - \mathbf{X}_i\mathbf{X}_i^T\mathbf{B}_i\mathbf{B}_i^T\mathbf{D}_{2,i} + \mathbf{D}_{2,i}\mathbf{X}_i\mathbf{X}_i^T - \mathbf{D}_{2,i}\mathbf{B}_i\mathbf{B}_i^T\mathbf{X}_i\mathbf{X}_i^T)\mathbf{W}_i$$
$$+ \gamma_6\mathbf{B}_i\mathbf{D}_8 + 2\gamma_6\mathbf{D}_7\mathbf{W}_i + \mu_{2,i}\mathbf{B}_i\mathbf{B}_i^T\mathbf{W}_i - \mu_{2,i}\mathbf{B}_i + \mathbf{B}_i\Lambda_{2,i}^T - \mu_{4,i}\mathbf{B}_i$$
$$+ \mu_{4,i}\mathbf{W}_i - \Lambda_{4,i})\mathbf{e}_i + 2\gamma_2 d_3^p\mathbf{x}_i(\mathbf{x}_i^T\mathbf{w}_{i,p} - \mathbf{e}_p^T\mathbf{H}_1\mathbf{G}_1\mathbf{e}_i). \quad (19)$$

By setting the derivative of each step to zero matrix, the update equations in Algorithm 1 can be obtained.

Abbreviations

AD	Alzheimer's disease
MRI	Magnetic resonance imaging
FS	FreeSurfer
VBM	Voxel-based morphometry
SNP	Single-nucleotide polymorphism
RAVLT	Ray's auditory verbal learning test
ADNI	Alzheimer's disease neuroimaging initiative
PCA	principal component analysis
ADMM	Alternating direction method of multipliers
RR	Ridge linear regression
CNN	Convolutional neural network
SVR	Support vector regression
RMSE	Root mean squared error
ReLU	Rectified linear unit

Availability of data and materials
The data used in this work are from public datasets: ADNI (http://adni.loni.usc.edu/). The access to these datasets is managed through secure LONI image and data archive (https://ida.loni.usc.edu/login.jsp) and contingent on adherence to the ADNI data use agreement and the publications' policies. To apply for the access to data please visit: http://adni.loni.usc.edu/data-samples/access-data/.

Declaration

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References

1. World Health Organization. Global action plan on the public health response to dementia 2017-2025. 2017; 2020. https://iris.who.int/bitstream/handle/10665/259615/9789241513487-eng.pdf?sequence=1, https://www.who.int/publications/i/item/global-action-plan-on-the-public-health-response-to-dementia-2017---2025.
2. Karch CM, Goate AM. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. Biol Psychiatry. 2015;77(1):43–51.
3. Stonnington CM, Chu C, Klöppel S, Jack CR Jr, Ashburner J, Frackowiak RS, et al. Predicting clinical scores from magnetic resonance scans in Alzheimer's disease. Neuroimage. 2010;51(4):1405–13.
4. Wang X, Shen D, Huang H. Prediction of memory impairment with MRI Data: a longitudinal study of Alzheimer's disease. In: Medical Image Computing and Computer-Assisted Intervention--MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part I 19. Springer; 2016. p. 273–81.
5. Wang H, Nie F, Huang H. Multi-view clustering and feature learning via structured sparsity. In: International conference on machine learning. PMLR; 2013. p. 352–60.
6. Chi KR. HIT OR MISS? Nature. 2009;461(7265):712.
7. Wang H, Nie F, Huang H, Yan J, Kim S, Risacher S, et al. High-order multi-task feature learning to identify longitudinal phenotypic markers for alzheimer's disease progression prediction. Adv Neural Inf Process Syst. 2012;25.
8. Brand L, Nichols K, Wang H, Huang H, Shen L, Alzheimer's Disease Neuroimaging Initiative. Predicting Longitudinal Outcomes of Alzheimer's Disease via a Tensor-Based Joint Classification and Regression Model. In: Pacific symposium on biocomputing 2020. World Scientific; 2019. p. 7–18.
9. Campos S, Pizarro L, Valle C, Gray KR, Rueckert D, Allende H. Evaluating imputation techniques for missing data in adni: a patient classification study. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 20th Iberoamerican Congress, CIARP 2015, Montevideo, Uruguay, November 9-12, 2015, Proceedings 20; Springer. 2015. p. 3–10.
10. Minhas S, Khanum A, Riaz F, Alvi A, Khan SA, Alzheimer's Disease Neuroimaging Initiative. Early Alzheimer's disease prediction in machine learning setup: empirical analysis with missing value computation. In: Intelligent Data Engineering and Automated Learning--IDEAL 2015: 16th International Conference, Wroclaw, Poland, October 14-16, 2015, Proceedings 16. Springer; 2015. p. 424–32.
11. Seo H, Brand L, Wang H. Learning Semi-Supervised Representation Enrichment Using Longitudinal Imaging-Genetic Data. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE. 2020. p. 1115–118.
12. Risacher SL, Shen L, West JD, Kim S, McDonald BC, Beckett LA, et al. Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. Neurobiol Aging. 2010;31(8):1401–18.
13. Brand L, Nichols K, Wang H, Shen L, Huang H. Joint Multi-Modal Longitudinal Regression and Classification for Alzheimer's Disease Prediction. IEEE Trans Med Imaging. 2019;39(6):1845–55.
14. Lu L, Elbeleidy S, Baker L, Wang H, Huang H, Shen L, ADNI. Improved Prediction of Cognitive Outcomes via Globally Aligned Imaging Biomarker Enrichments over Progressions. In: Medical Image Computing and Computer Assisted Intervention--MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22. Springer; 2019. p. 140–48.
15. Lu L, Elbeleidy S, Baker LZ, Wang H. Learning multi-modal biomarker representations via globally aligned longitudinal enrichment. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, No. 01. 2020. p. 817–824.
16. Lu L, Wang H, Elbeleidy S, Nie F. Predicting Cognitive Declines Using Longitudinally Enriched Representations for Imaging Biomarkers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. p. 4827–836.
17. Seo H, Brand L, Wang H, Nie F. Integrating static and dynamic data for improved prediction of cognitive declines using augmented genotype-phenotype representations. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 1. 2021. p. 522–30.
18. Wang H, Nie F, Huang H, Kim S, Nho K, Risacher SL, et al. Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. Bioinformatics. 2012;28(2):229–37.
19. Yan J, Li T, Wang H, Huang H, Wan J, Nho K, et al. Cortical surface biomarkers for predicting cognitive outcomes using group $\ell_{2,1}$-norm. Neurobiol Aging. 2015;36:S185–93.
20. Wang H, Nie F, Huang H. Large-scale cross-language web page classification via dual knowledge transfer using fast nonnegative matrix tri-factorization. ACM Trans Knowl Disc Data (TKDD). 2015;10(1):1–29.
21. Ding C, He X. K-means clustering via principal component analysis. In: Proceedings of the twenty-first international conference on Machine learning. 2004. p. 29.
22. Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, et al. Powerful SNP-set analysis for case-control genome-wide association studies. Am J Hum Genet. 2010;86(6):929–42.
23. Wang H, Nie F, Huang H. Robust and discriminative distance for multi-instance learning. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE; 2012. p. 2919–924.
24. Nie F, Wang H, Huang H, Ding C. Early active learning via robust representation and structured sparsity. In: Twenty-third international joint conference on artificial intelligence. 2013.
25. Liu K, Wang H, Han F, Zhang H. Visual Place Recognition via Robust ℓ2-Norm Distance Based Holism and Landmark Integration. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. No. 01. 2019. p. 8034–041.
26. Liu K, Brand L, Wang H, Nie F. Learning robust distance metric with side information via ratio minimization of orthogonally constrained l21-norm distances. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. 2019.

27. Jolliffe IT. Principal components in regression analysis. In: Principal component analysis. 1986. pp. 129–55.

28. Liu Y, Guo Y, Wang H, Nie F, Huang H. Semi-supervised classifications via elastic and robust embedding. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 31. No. 1. 2017.

29. Yang H, Liu K, Wang H, Nie F. Learning strictly orthogonal p-order nonnegative laplacian embedding via smoothed iterative reweighted method. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence. 2019. p. 4040–046.

30. Bertsekas DP. Constrained optimization and Lagrange multiplier methods. Academic press; 2014.

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

32. Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, 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. 2010;53(3):1051–63.

33. Apostolova LG, Beyer M, Green AE, Hwang KS, Morra JH, Chou YY, et al. Hippocampal, caudate, and ventricular changes in Parkinson's disease with and without dementia. Mov Disord. 2010;25(6):687–95.

34. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984;34(7):939–939.

35. Van der Werf YD, Witter MP, Uylings HB, Jolles J. Neuropsychology of infarctions in the thalamus: a review. Neuropsychologia. 2000;38(5):613–27.

36. Mu Y, Gage FH. Adult hippocampal neurogenesis and its role in Alzheimer's disease. Mol Neurodegener. 2011;6(1):85.

37. Poulin SP, Dautoff R, Morris JC, Barrett LF, Dickerson BC, Initiative ADN, et al. Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. Psychiatry Res Neuroimaging. 2011;194(1):7–13.

38. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. Nat Genet. 2007;39(1):17–23.

39. Zhu XC, Wang HF, Jiang T, Lu H, Tan MS, Tan CC, et al. Effect of CR1 genetic variants on cerebrospinal fluid and neuroimaging biomarkers in healthy, mild cognitive impairment and Alzheimer's disease cohorts. Mol Neurobiol. 2017;54(1):551–62.

40. Chen GF, Xu TH, Yan Y, Zhou YR, Jiang Y, Melcher K, et al. Amyloid beta: structure, biology and structure-based therapeutic development. Acta Pharmacol Sin. 2017;38(9):1205–35.

41. Kim J, Basak JM, Holtzman DM. The role of apolipoprotein E in Alzheimer's disease. Neuron. 2009;63(3):287–303.

42. Oishi K, Mielke MM, Albert M, Lyketsos CG, Mori S. DTI analyses and clinical applications in Alzheimer's disease. J Alzheimer Dis. 2011;26(s3):287–96.

43. Snyder HM, Carrillo MC, Grodstein F, Henriksen K, Jeromin A, Lovestone S, et al. Developing novel blood-based biomarkers for Alzheimer's disease. Alzheimer Dement. 2014;10(1):109–14.

Publisher's Note